European Journal of Economic and Social Systems
Vol. 15, N°3 (2001) 131–152

IMPLICATIONS OF ROUTINE-BASED LEARNING FOR DECISION MAKING

Thomas Brenner[1]

Abstract. Various mathematical models have been proposed in the economic literature for the modelling of boundedly rational learning. Each of these models considers one type of learning, like imitation, satisficing, trial-and-error or reinforcement learning. This paper combines several types of boundedly rational learning and develops a more general learning model. Then, the characteristics of this learning model are examined for a quite general decision situation involving risk. It is studied under which conditions learning leads to an expected utility maximising behaviour. Furthermore, it is analysed what kind of deviations from the utility maximising behaviour are caused by the different aspects of learning.

Classification Codes. C44, D83.

1. Introduction

For many decades economists have modelled decision making according to the rational choice paradigm which assumes that people always choose those actions that give rise to the highest utility. Usually it was even assumed that they have complete knowledge about each of the possible actions and are therefore able to choose the best one. It was acknowledged that in some situations they might lack the necessary information and have to collect information or learn. This means, according to the view of most economists, that individuals might not choose the best actions right from the beginning in such situations. However, learning would surely lead them to finally choose the best actions, it is usually argued. Only in recent years this belief has become contested. A huge number of experiments in which individuals have even after a quite long period not chosen the optimal action

Keywords and phrases: Decision making, learning, mathematical modelling.

[1] Max-Planck-Institute for Research into Economic Systems, Evolutionary Economics Unit, Kahlaische Str. 10, 07745 Jena, Germany.

has triggered a debate about the rationality of human behaviour. There are still different opinions about how these deviations from the expected behaviour have to be interpreted (see for example Harrison, 1994; Börgers, 1996; Slembeck, 1999; Weibull, 2000). However, there is strong evidence that people learn and behave in many situations according to simple rules (see for example the evidence collected on frugal rules in Gigerenzer and Goldstein, 1996 and the experimental evidence for boundedly rational rules in Slonim, 1999 and reinforcement learning in Roth and Erev, 1995 and Erev and Roth, 1998).

The theoretical literature reacted to this debate by an intensive study of the convergence processes of different learning models. Most of these studies, however, have been motivated by the desire to prove that learning processes lead to utility maximising behaviour. Already in the 50's, shortly after the concept of Nash equilibria had been established (*cf.* Nash, 1950), a learning process was proposed that leads to behaviour in correspondence with the Nash equilibrium in some games (see Brown, 1951). This learning process is called fictitious play. It was established to solve the problem that players might not know what their opponents will do when they play a game for the first time. To show that the Nash equilibrium is nevertheless a valuable concept, it was proved that learning, in the form of fictitious play, makes the behaviour of players converge to the Nash equilibrium. In the meantime many works have been conducted that examine whether certain learning processes converge to utility maximising behaviour or not (see Day, 1967 for satisficing; Marcet and Sargent, 1989 for least squares learning; Marimon, 1993 for adaptive learning; Börgers and Sarin, 1997 for reinforcement learning; Mailath, 1998 for a general discussion of evolutionary game theory, and Schlag, 1998 for imitation). Some of these approaches show that such learning processes indeed lead to utility maximising behaviour, while others identify the circumstances that are necessary for learning to cause utility maximising behaviour. It has been shown that utility maximising behaviour is not the general result of all learning processes in all situations.

The approach that is proposed here also aims to study whether learning leads to utility maximising behaviour. However, it takes two further steps that have not been taken in the literature or have been restricted to behaviour in games. First, it uses a learning model that combines many of the different aspects of learning that are usually studied separately. Second, it aims to make some statements about the structure of the differences between the behaviour that results from learning and utility maximisation. Furthermore, it does not analyse behaviour in a game, as most papers on learning nowadays do. It focuses on individual decision making in multi-arm bandit situations.

Although rarely mentioned in literature, each of the existing models includes in general only one aspect of learning and neglects other aspects (some models include two aspects or even three aspects, see, *e.g.*, Camerer and Ho, 1999 and Levine and Pesendorfer, 2000). Furthermore, different learning processes occur in different situations. Thus, only a study of the implications of all kinds of learning processes will offer a complete answer to the question of whether learning leads to utility maximising behaviour. The list of learning processes that are frequently

used in economics is long (see Brenner, 1999, Chapt. 3 for a detailed list of learning models). However, learning models can be classified into three categories according to the situations in which they occur and according to their structure (see Brenner, 1999, Chapt. 2 and 3 for a detailed discussion): 1) Non-cognitive learning: The main process of non-cognitive learning is called reinforcement learning and was first investigated by I. P. Pawlov (see Pawlov, 1953 for a description on all major initial finding on this learning process). Reinforcement learning is an innate process that leads to a more frequent occurrence of behaviours with positive consequences and to a less frequent occurrence of behaviours with negative consequences. Learning is dominated by this process whenever people are not aware of their own learning. However, it has recently been proved to describe learning processes quite well under other circumstances (see Roth and Erev, 1995 and Erev and Roth, 1998). Two types of models have been proposed for such learning processes: the Bush-Mosteller model (see Bush and Mosteller, 1955) and the model of melioration learning (see Herrnstein and Prelec, 1991). The implications of both models have been compared with the predictions of utility maximisations in the literature (see Börgers and Sarin, 1997 and Brenner and Witt, 1997). 2) Routine-based learning: This kind of learning is mainly used in situations in which the individuals are aware of the situation they face and have a fixed mental model with respect to the situation. A typical example is a situation in which the alternatives are known but not the utilities which they give rise to. However, other situations with a fixed structure and a clear aim, like the evolutionary process described by Nelson and Winter (1982) also belongs to this group of situations. In such situations the individuals rely on their own experience and the information from others. They usually behave according to certain routines in such situations. These routines determine how they collect information and use it. Most of the learning models in the literature, like fictitious play, satisficing, and imitation, belong to this type of learning. 3) Learning by cognitive association: If the situation is less clear or the behaviour of others is not well-known, people have to develop an understanding of the situation and the behaviour of others. They develop so-called mental models that help them to understand the situation and find the adequate behaviour. This type of learning is described in cognitive learning models in psychology. However, it is seldom used in economics and a commonly used mathematical formulation is missing.

This paper studies the second type of learning. While the first type has been analyse comprehensively in the literature (see Börgers and Sarin, 1997 and Brenner and Witt, 1997) and the third type seems to be too complex to study it in a general way, the second type has been analysed to some extent in the literature. The second type of learning comprises many different aspects. Some of them have been studied separately. Here a more general approach is taken, that includes many, if not all of these aspects. The considered aspects are the occasional trial of new behaviours, satisficing, imitation, and the information gathering using own experiences and communication with others. Such a model has been first used in Brenner (1997) and has been further developed in Brenner (1999). It is used here because it combines many of the aspects that are used in learning models in the

literature and allows to study their implications for decision making simultaneously.

The paper proceeds as follows. The basic concept of the learning model and the specific situation for which its implications are studied are discussed in the next section. In Section 3 the model is described in detail. This model is analysed mathematically in Section 4. The results of this analysis are compared to the predictions of traditional decision theory in Section 5. Section 6 concludes.

2. Basic concept and situation

Individual learning processes are studied here. It is assumed that N individuals exist (each labeled by a natural number i with $i \in \{1, 2, ..., N\}$). However, these individuals do not interact such that the action of one individual influences the utility obtained by another individual. The only way in which the individuals interact is communication and imitation.

The learning model that is used here is designed to include most aspects that seem relevant for routine-based learning. Let me restate that routine-based learning occurs whenever individuals have a fixed mental model of the situation they face but lack sufficient knowledge about the usefulness of different actions. A fixed mental model means that they have a clear notion of the possible actions they might take and of how their actions interact in principle with the surrounding and/or the actions of other people. This notion might be correct or not. In this approach it is just assumed that the mental model does not change during the learning process that is considered here. If this assumption is not given learning would not follow a fixed routine, but would involve changes of mental models and therefore changes of routines. This implies two conditions for routine-based learning: The individual has to face repeatedly exactly the same situation and all possible actions have to be known to the individual.

Many of the learning models in the literature describe routines that might be applied in such a situation. Here a model is proposed that is a combination of many of these routines. Namely, four aspects are included in the model that is used here: the occasional trial of alternative actions (often called exploration or innovation in the literature), satisficing, the collection of information about the average outcomes of actions, and imitation. Each of these aspects has been modelled separately and some have been modelled in combination in the literature.

In many learning models in the literature stochastic elements are included. The authors usually interpret these stochastic elements as the results of errors or innovations. They assume either that unintended changes in behaviour occur from time to time (see *e.g.* Binmore and Samuelson, 1994; Samuelson, 1994; Wu and Axelrod, 1995) or that individuals explore alternative actions randomly from time to time (see *e.g.* Young, 1993 and Levine and Pesendorfer, 2000). One might also argue that novelty constitutes a basic desire of human beings as it is argued in psychology (*cf.* Scitovsky, 1992) and that therefore people change their behaviour due to the satisfaction that such a change offers. It might be argued that changes

caused by the desire for the new are directed (see Witt, 1987 for a discussion). The model that is used here follows the latter interpretation. This means that, although the action of changing is stochastic, the choice of the new action depends on former experience and information from others. Errors are included in the model anyway due to its stochastic formulation.

The concept of satisficing is based on the psychological finding that individuals are tempted to change their behaviour as long as they are not satisfied by the outcome of their actions (a detailed description can be found in Simon 1987). This concept has been used in several learning models (see *e.g.* Day, 1967; Witt, 1986; Binmore and Samuelson, 1994; Börger and Sarin, 1996; Dawid, 1997 and Posch, 1999). Working with the concept of satisficing implies that an aspiration level has to be defined. It has been shown in experiments that the aspiration level adapts to the outcomes that have been experienced in the past (see *e.g.* Festinger, 1942). Nevertheless, for simplicity of the model and analogous to most learning models in the literature that include the concept of satisficing, it is assumed here that the aspiration level remains constant. The analysis of this paper concentrates on the behaviour that learning converges to. After such a convergence the aspiration level can be assumed to have converged to a fairly stable value as well so that the assumption of a constant aspiration level has no significant impact on the results.

Imitation has been intensively studied from a psychological perspective in the 60's, 70's and 80's (the major results and concepts can be found in Bandura, 1979 and Latané, 1981). Quite a number of learning models that are used in economics are based on this mechanism (see *e.g.* Sinclair, 1990; Kandori, Mailath and Rob, 1993; Witt, 1996; Schlag, 1998 and Levine and Pesendorfer, 2000). The mechanism is the same in all these models except the one by Levine and Pesendorfer: the better the results of a decision maker are the more probable her/his action will be imitated. This way of modelling does neither exactly correspond to the concept of observational learning (see Bandura, 1979) nor to the concept of social impact (see Latané, 1981). The model that is used here will be more precise in this respect.

Most learning models in the economic literature are based on the aspect that individuals collect information about the outcomes of different actions, build expectations and take the action that leads to highest expected utility (see *e.g.* Brown, 1951; Ellison, 1993; Young, 1993; Samuelson, 1994; Crawford, 1995; Sarin and Vahid, 1999 and Sarin, 2000). This aspect is also included in the model that is proposed here (in a way that is similar to the approach of Sarin and Vahid, 1999), although the individuals are not assumed to choose always the action that leads to the highest expected utility.

The structure of the model that is used here is depicted in Figure 1. It is assumed that an individual faces repeatedly the same situation. The action that is taken by the individual leads only to one piece of new information each time: the utility which the action has given rise to. Furthermore, the individual communicates with others and receives information about their choice and the utility they obtained. The number of possible actions and the context of the situation remain constant, so that the individual can be assumed to have complete knowledge about

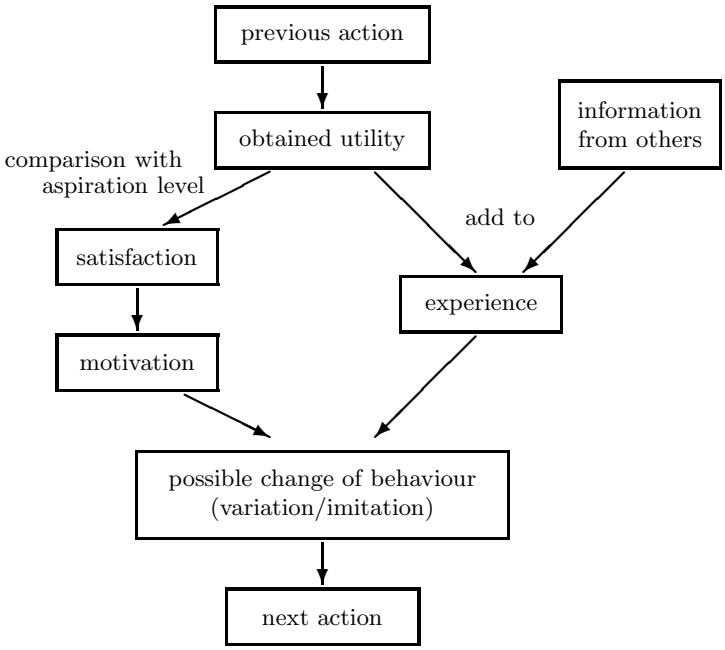

FIGURE 1. Structure of the learning model.

the situation, except the utility that each action gives rise to, and does not change the mental model about the situation.

The information about the utility that has been obtained after the previous action is used in two ways as depicted in Figure 1. First, a comparison of the utility obtained with the aspiration level determines the actual satisfaction of the individual. Second, the obtained utility and the information from others is added to the experience that has been already made in the past. The satisfaction influences the decision whether to take the same action again or to change behaviour. As long as the individual is satisfied with the utility obtained there is no reason to change behaviour. The lower the satisfaction, however, the more is the individual motivated to choose another action next time. Therefore, the degree of satisfaction leads to a respective degree of motivation. If the individual decides to change behaviour, there are two possibilities to do so: (s)he might try another action to explore the respective outcomes or (s)he might imitate another individual. The experience that has been collected in the past determines which action is explored next or who is imitated, respectively. Thus, neither exploration nor imitation is blind nor does it depend only on the utility that others obtain. The information that is gathered in the past is used to guide these processes, while the satisfaction determines whether the behaviour is changed at all. This learning processes is repeated.

Due to the requirements for routine-based learning to occur, the situation that is analysed here is restricted to a repeated situation with a fixed set of possible actions. It is assumed that there are A possible actions which are labeled by $a = 1, ..., A$. Each action results in a utility $u(a, t)$ each time t (time is discrete: $t \in \mathbb{N}$). The utility $u(a, t)$ is defined such that is reflects the evaluation of the outcome of an action. It is not a von Neumann-Morgenstern utility since a transformation of the values changes the learning process.

The utility that each action a gives rise to is assumed not to be constant. It varies with time stochastically. Nevertheless, the probability of each possible value $u(a, t)$ is constant so that the action itself stays the same. Consequently, a fixed probability distribution exists for the utility that can be derived from each of the actions. This probability distribution is denoted by $Q(u|a)$ which means that $Q(u|a)$ represents the probability that an individual choosing action a will receive a utility of u (one might think of a lottery and the individuals have to choose between different lotteries without knowing their characteristics). According to this probability distribution the actual utility is determined stochastically each time. The individuals are not aware of this probability distribution. They are aware of the existence of the alternative actions and each time they have chosen one of them they recognise the utility this alternative has given rise to.

3. Learning model

In this section a brief outline of the mathematical formulation of the learning model is given. More details about the reasons for certain aspects of the model can be found in Brenner (1999, Chapt. 3).

Each time t each individual i has to take one action $a_i(t)$. The utility $u_i(a_i(t), t)$ derived from this action is determined stochastically according to the probability distribution $Q(u|a)$. $u_i(a_i(t), t)$ is the only information that individuals receives about the outcome of their choice.

When individuals first face the repeated situation they have no information about the alternative actions but step by step they gain experience. There is also the possibility for the individuals to exchange information. The collected information about an action is called experience here. The experience of individual i about action a is denoted by $\eta_i(a, t)$ at time t.

$\eta_i(a, t)$ represents the utility that the individual remembers as the average utility that action a gives rise to. It adapts towards each new information without neglecting former experiences. Therefore, the new value of $\eta_i(a, t+1)$ is defined as the weighted average of the old value $\eta_i(a, t)$ and the new experience $u_i(a_i(t), t)$. The weight of the new experience is defined for each individual i by κ_{ii} $(0 < \kappa_{ii} < 1)$. In addition, people are able to exchange their experiences. This information exchange might be different for different pairs of individuals. Therefore, a parameter $\frac{\kappa_{ij}}{\kappa_{ii}}$ is defined which denotes the amount of information that individual j transfers to individual i (this exchange of information, if it is due to observation, corresponds to the concept of observational learning; see Bandura, 1979). If experience

is fully communicated, the fraction $\frac{\kappa_{ij}}{\kappa_{ii}}$ equals one. In general $\kappa_{ij} \leq \kappa_{ii}$ holds and the κs have to be restricted such that $\sum_{j=1}^{I} \kappa_{ij} \leq 1$. The dynamics of $\eta_i(a,t)$ are given by

$$\eta_i(a,t+1) = \Big(1 - \sum_{j=1}^{I} \delta(a = a_j(t))\kappa_{ij}\Big) \cdot \eta_i(a,t) + \sum_{j=1}^{I} \delta(a = a_j(t)) \cdot \kappa_{ij} \cdot u_j(a_j(t),t) \tag{3.1}$$

with

$$\delta(a = a_j(t)) = \begin{cases} 1 \text{ for } a = a_j(t) \\ 0 \text{ for } a \neq a_j(t) \end{cases} . \tag{3.2}$$

The function $\delta(a = a_j(t))$ has to be used because only for those actions that are chosen, new information is obtained.

Besides the collection of experience, the utility that is obtained also determines the satisfaction of an individual. It is compared to the individual's aspiration level z_i. Utilities that are above the aspiration level satisfy the individual while utilities below the aspiration level disappoint them. In the approach that is used here the aspiration level is assumed to be constant in time.

Individuals are assumed to compare the previous outcome with their aspiration level. A variable $s_i(t)$, named satisfaction, is defined to consider this fact. This satisfaction is defined by

$$s_i(t) = u_i(a_i(t),t) - z_i \, . \tag{3.3}$$

As long as the satisfaction $s_i(t)$ is greater than zero, individuals feel no need to change behaviour. If $s_i(t)$ falls below zero, instead, individual i is not satisfied and has the desire to change something to obtain higher utilities. However, as it is argued above, even if individuals are satisfied with their current situation, there is a certain probability that they change behaviour. Therefore, the probability for a change of behaviour is given by the sum of a certain constant value m_0 and a value that depends on the actual satisfaction. This probability or desire to change behaviour is called motivation $m_i(t)$ here and is defined by

$$m_i(t) = \begin{cases} m_0 - s_i(t) & \text{for } s_i < 0 \\ m_0 & \text{for } s_i \geq 0 \end{cases} . \tag{3.4}$$

Once an individual has decided to change behaviour, there are two ways in which such a change can be conducted. First, the individual might explore an alternative action that has not been used so far or abandoned in the past. This kind of change is named variation here. Second, the individual might look for actions that are successfully applied by other individuals. This kind of change is named imitation here. Whether an individual rather uses variation or imitation is a personal characteristic and depends on the situation. In this approach two

parameters ν_V and ν_I are used to represent the respective probabilities. They have to be defined such that the probabilities given in (3.5) and (3.6) are always smaller than one.

Therefore, whether an individual changes behaviour according to the process of variation depends on her/his motivation $m_i(t)$ and the parameter ν_V. Furthermore, variation is assumed to be directed. The experience $\eta_i(a,t)$ that has been collected in the past is used to determine which action is chosen. The higher the value of $\eta_i(a,t)$ the more likely action a is chosen. The probability $p_V(i,\tilde{a},t)$ for individual i to change at time t from the previous choice to action $\tilde{a}$ due to variation is assumed to be

$$p_V(i,\tilde{a},t) = \nu_V \cdot m_i(t) \cdot \exp\left[\zeta \eta_i(\tilde{a},t)\right]. \tag{3.5}$$

ζ is a parameter that determines the influence of the experience on the choice.

Whether an individual changes behaviour according to the process of imitation depends also on her/his motivation $m_i(t)$ and on the parameter ν_I. Furthermore, only actions are imitated that are observed to lead to a utility higher than the own aspiration level (see Brenner, 1999 for a discussion about what can be observed and different ways of modelling the imitation process). The higher the observed utility is, the more likely the respective action is imitated. Finally, the own experience again plays a role. The probability $p_I(i,j,t)$ for individual i to imitate individual j at time t is given by

$$p_I(i,j,t) = \frac{\nu_I}{N} \cdot m_i(t) \cdot \exp\left[\zeta \cdot \eta_i(a_j(t),t)\right] \cdot \Theta(u_j(a_j(t),t)) \tag{3.6}$$

where

$$\Theta(x) = \begin{cases} 0 & \text{for } x \leq z_i \\ x - z_i & \text{for } x > z_i. \end{cases} \tag{3.7}$$

This definition of imitation includes the aspects of observational learning, since the utility that is obtained by the other individuals influences its probability, and the aspects of social impact, since the imitation of a certain behaviour becomes more likely the more other individuals show this behaviour.

4. Mathematical analysis

For each individual the process that is set up above is a stochastic process with a fixed number of possible actions. The processes of variation and imitation determine the probabilities for switches between these actions. Therefore, the learning process constitutes a Markov chain. To analyse such a system a probability distribution for all possible states is defined for each time. This probability distribution is denoted by $P_i(a,t)$ where $P_i(a,t)$ denotes the probability of individual i to chose action a at time t.

The use of the probabilities $P_i(a,t)$ does not mean that the individuals choose their actions randomly, like it is for example assumed for reinforcement learning (see Bush and Mosteller, 1955 for a detailed discussion). At each point in time every individual chooses a clearly determined action. However, the changes in behaviour are random. Therefore, predictions of behaviour can only be made in the form of probabilities. $P_i(a,t)$ means that the model predicts that individual i uses action a at time t with probability $P_i(a,t)$. The formulation of such predictions is necessary to deduce statements about the behaviour of individuals in the long run, given that they learn according to the learning model that is used.

Therefore, the mathematical aim of this approach is to deduce statements about the dynamics of the probability distribution $P_i(a,t)$. To this end, the dynamics of the probability distribution have to be combined with the probabilities for the changes of behaviour that have been defined above. Let us denote the probability that individual i changes from action a to action $\tilde{a}$ at time t by $r_i(a \to \tilde{a}, t)$. Then, the dynamics of the probability distribution $P_i(a,t)$ are given by

$$P_i(a,t+1) - P_i(a,t) = \sum_{\tilde{a}=1}^{A} r_i(\tilde{a} \to a, t) \cdot P_i(\tilde{a}, t) - \sum_{\tilde{a}=1}^{A} r_i(a \to \tilde{a}, t) \cdot P_i(a,t) \; . \tag{4.1}$$

To apply this equation, the transition probabilities $r_i(a \to \tilde{a}, t)$ have to be deduced from the learning model.

In the above learning model two kinds of behaviour changes appear: variation and imitation. For variations the probabilities for switches between actions are directly given by equation (3.5). In the case of imitations the probabilities that are given by equation (3.6) have to be summed up for all other individuals who have chosen the respective action. Thus, the transition probabilities are given by

$$r_i(a \to \tilde{a}, t) = p_V(i, \tilde{a}, t) + \sum_{j=1}^{I} \delta(a_j(t) = \tilde{a}) \cdot p_I(i,j,t) \; . \tag{4.2}$$

Inserting equations (3.5) and (3.6) in this equation results in

$$r_i(a \to \tilde{a}, t) = m_i(t) \cdot \exp\Big[\zeta \eta_i(\tilde{a}, t)\Big] \times \left[\nu_V + \frac{\nu_I}{N} \cdot \sum_{j=1}^{N} \Big[\delta(a_j(t) = \tilde{a}) \cdot \Theta(u_j(a_j(t), t))\Big] \right] . \tag{4.3}$$

Equation (4.3) does not satisfy the requirements for a Markov chain and the respective solution methods. Markov chains require that the transition probabilities depend only on the actual state of the system. The probability distribution has to be defined for all possible states. Above the probability distribution is defined for all possible actions. This implies that the action which is chosen by individual i is

assumed to be the only variable that describes the state of the individual. However, the transition probabilities also depend on the motivation and the experience of the individual and the utility obtained by all other individuals. Thus, although equation (4.1) has the mathematical form that is used for Markov chains, it does not describe such a process.

One of the problems is that the transition probabilities for individual i depend on the behaviour of the other individuals. Therefore, the behaviour of an individual cannot be analysed separately. A possibility to solve this problem, is to analyse the behaviour on the population level. Two assumptions are necessary to make such an analysis feasible without using simulations. First, a homogeneous population has to be assumed. This means that all parameters are identical for all individuals (the index i is omitted below). The behaviour, however, might be different for the individuals due to stochastic elements in the learning process. Second, the number of individuals has to be sufficiently high, so that assuming $N \to \infty$ leads to an adequate approximation of the real situation.

These assumptions imply that the share $x(a,t)$ of individuals in the population who choose action a at time t equals the probability $P_i(a,t)$. Thus, the dynamics on the population level are given in correspondence with equation (4.1) by

$$x(a,t+1) - x(a,t) = \sum_{\tilde{a}=1}^{A} r(\tilde{a} \to a,t) \cdot x(\tilde{a},t) - \sum_{\tilde{a}=1}^{A} r(a \to \tilde{a},t) \cdot x(a,t)\,. \qquad (4.4)$$

Equation (4.4) describes the dynamics of the population. As a consequence, the transition probabilities $r(a \to \tilde{a},t)$ have also to be defined on the population level, *i.e.*, dependent on the shares $x(a,t)$. Since the population is assumed to be infinitely large, this means that the average transition probability for an individual to switch from action a to action $\tilde{a}$ at time t has to be calculated. To this end, the probability for each state of an individual, including motivation and experience, that takes action a at time t and the possible actions and utilities of all other individuals has to be calculated. Then, the average transition probability is given as the sum of the individual transition probability $r_i(a \to \tilde{a},t)$ for each state of the individual multiplied by the respective probability.

The utilities that are obtained by other individuals are independent from the individuals action, motivation and experiences. Therefore, the term $\frac{1}{N} \sum_{j=1}^{N} \Big[\delta(a_j(t) = \tilde{a}) \cdot \Theta(u(a_j(t),t))\Big]$ can be analysed separately. Due to the assumption of an infinitesimally large population, an infinitesimally large number $x(\tilde{a},t) \cdot N$ of individuals choose action $\tilde{a}$ at time t. Each of them obtains a different utility according to $Q(u|\tilde{a})$. Hence,

$$\frac{1}{N} \sum_{j=1}^{N} \Big[\delta(a_j(t) = \tilde{a}) \cdot \Theta(u(a_j(t),t))\Big] = x(\tilde{a},t) \cdot \int_z^{\infty} Q(u|\tilde{a}) \cdot u \,\mathrm{d}u\,. \qquad (4.5)$$

Below the shortening

$$\theta_{\tilde{a}} = \int_z^\infty Q(u|\tilde{a}) \cdot u \, \mathrm{d}u \tag{4.6}$$

is used.

The experience $\eta_i(\tilde{a}, t)$ is also independent of the action $a_i(t)$ currently taken and the motivation $m_i(t)$. Therefore, it is sufficient to calculate the probability distribution for $\eta_i(\tilde{a}, t)$ separately. Instead of the dynamic definition of $\eta_i(\tilde{a}, t)$ that is given in equation (3.1), $\eta_i(\tilde{a}, t)$ can also be written in the form of an exponentially weighted sum:

$$\begin{aligned} \eta_i(\tilde{a}, t) = \sum_{\tau=0}^{t} \Big[\Big(& \delta(a_i(\tau) = \tilde{a}) \cdot \kappa_{ii} \cdot u_i(\tilde{a}, \tau) \\ & + \sum_{\substack{j=0 \\ j \neq i}}^{N} \kappa_{ij} \cdot \delta(a_j(\tau) = \tilde{a}) \cdot u_j(\tilde{a}, \tau) \Big) \\ & \cdot \prod_{\tilde{t}=\tau}^{t} \Big(1 - \delta(a_i(\tilde{t}) = \tilde{a}) \cdot \kappa_{ii} - \sum_{\substack{j=0 \\ j \neq i}}^{N} \kappa_{ij} \cdot \delta(a_j(\tilde{t}) = \tilde{a}) \Big) \Big] \, . \end{aligned} \tag{4.7}$$

Above the assumptions have been introduced that all individuals are identical (implying that κ_{ij} is the same for all pairs of individuals), and that there is an infinitesimally large number of individuals. This implies that equation (4.7) results in

$$\begin{aligned} \eta_i(\tilde{a}, t) = \sum_{\tau=0}^{t} \Big[\Big(& \delta(a_i(\tau) = \tilde{a}) \cdot \kappa_{ii} \cdot u_i(\tilde{a}, \tau) \\ & + \kappa_{ij} \cdot x(\tilde{a}, \tau) \cdot \bar{u}(\tilde{a}) \Big) \\ & \cdot \prod_{\tilde{t}=\tau}^{t} \Big(1 - \delta(a_i(\tilde{t}) = \tilde{a}) \cdot \kappa_{ii} - \kappa_{ij} \cdot x(\tilde{a}, \tilde{t}) \Big) \Big] \end{aligned} \tag{4.8}$$

where $\bar{u}(\tilde{a})$ denotes the average utility that action $\tilde{a}$ gives rise to, defined by

$$\bar{u}(a) = \int_{-\infty}^{\infty} u \cdot Q(u|a) \mathrm{d}u \, . \tag{4.9}$$

According to equation (4.8) $\eta_i(\tilde{a}, t)$ is a weighted sum of the stochastic values of $u_i(\tilde{a}, \tau)$ and the value of $\bar{u}(\tilde{a})$. If individual i has not chosen action $\tilde{a}$ for a long time, the value of $\eta_i(\tilde{a}, t)$ converges to $\bar{u}(\tilde{a})$, given that $\kappa_{ij} > 0$. If, instead, individual i has chosen action $\tilde{a}$ quite often in the past, $\eta_i(\tilde{a}, t)$ is the weighted sum of many independent stochastic values and $\bar{u}(\tilde{a})$. The probability distribution of a sum of many independent stochastic values can be approximated by a normal distribution. The average of such a sum is given by the sum of the averages of the stochastic values. In the case of equation (4.9) the average is $\bar{u}(\tilde{a})$. The variance can also be calculated on the basis of the variances of the summed stochastic values. However, it depends on the history of the choices by the individual in the present case. In general it can be stated that it decreases with κ_{ij} and increase

with κ_{ii} and is proportional to the variance

$$\sigma_u(\tilde{a}) = \int_{-\infty}^{\infty} (u - \bar{u}(\tilde{a}))^2 \cdot Q(u|\tilde{a}) \mathrm{d}u \ . \tag{4.10}$$

A new parameter κ is defined such that the variance of $\eta_i(\tilde{a})$ is given by $\sigma_\eta(\tilde{a}) = \kappa \cdot \sigma_u(\tilde{a})$. Then, the probability distribution $W(\eta|\tilde{a})$ for the values of $\eta_i(\tilde{a}, t)$ can be approximated by a normal distribution $N(\bar{u}(\tilde{a}), \sigma_\eta(\tilde{a}))$.

Finally, the motivation $m_i(t)$ has to be analysed. Its value depends directly on the utility $u_i(a, t)$ that is obtained by individual i. The respective probability distribution is given by

$$R(m|a) = \begin{cases} Q(z - m|a) & \text{if } m > 0 \\ \int_z^\infty Q(u|a) \mathrm{d}u & \text{if } m = 0 \\ 0 & \text{if } m < 0 \end{cases} \ . \tag{4.11}$$

The average transition probability $r(a \to \tilde{a}, t)$ is given by

$$r(a \to \tilde{a}, t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} m \cdot \exp\Big[\zeta\eta\Big] \cdot \Big[\ \nu_V + \nu_I \cdot \theta_{\tilde{a}} \cdot x(\tilde{a}, t)\Big] \cdot W(\eta|\tilde{a}) \cdot R(m|a) \ \mathrm{d}\eta \ \mathrm{d}m \ . \tag{4.12}$$

The following two abbreviations are introduced:

$$\mu_a = \int_{-\infty}^{z} Q(u|a) \cdot (z - u) \ \mathrm{d}u \tag{4.13}$$

and

$$\begin{aligned} n_{\tilde{a}} &= \int_{-\infty}^{\infty} \exp\left[\zeta\eta\right] \cdot \frac{1}{\sqrt{\pi \cdot \sigma_u(\tilde{a})}} \cdot \exp\left[-\frac{(\eta - \bar{u}(\tilde{a}))^2}{\kappa \cdot \sigma_u(\tilde{a})}\right] \ \mathrm{d}\eta \\ &= \exp\left[\zeta \cdot \bar{u}(\tilde{a}) + \frac{1}{2}\kappa \cdot \zeta^2 \cdot \sigma_u(\tilde{a})\right] \end{aligned} \tag{4.14}$$

Using these abbreviations the transition probabilities are given by

$$r(a \to \tilde{a}, t) = \mu_a \cdot n_{\tilde{a}} \cdot \Big\{\nu_V + \nu_I \cdot x(\tilde{a}, t) \cdot \theta_{\tilde{a}}\Big\} \ . \tag{4.15}$$

Inserting (4.15) into equation (4.4) results in

$$\begin{aligned} x(a, t+1) - x(a, t) = \sum_{\tilde{a}=1}^{A} \Big[&\mu_{\tilde{a}} \cdot n_a \cdot \Big[\nu_V + \nu_I \cdot x(a, t) \cdot \theta_a\Big] \cdot x(\tilde{a}, t) \\ &- \mu_a \cdot n_{\tilde{a}} \cdot \Big[\nu_V + \nu_I \cdot x(\tilde{a}, t) \cdot \theta_{\tilde{a}}\Big] \cdot x(a, t)\Big] \ . \end{aligned} \tag{4.16}$$

Equation (4.16) now represents a closed system of equations that describes the dynamics of behaviour on the population level. The aim of this mathematical analysis is to identify the behaviour to which routine-based learning converges. This means that the attractors of the dynamic system (4.16) have to be identified. This results in the following:

Theorem 1. *Let $\nu_V > 0$ and $\mu_a > 0$ for all a. Then, the dynamical system that is defined by equation (4.16) has one unique stable state to which it converges independent of the initial state. This stable state is given by*

$$x_{st}(a) = \frac{\nu_V \cdot n_a}{\mu_a \cdot \left(C - \frac{\nu_I \cdot n_a \cdot \theta_a}{\mu_a}\right)} \tag{4.17}$$

where C is determined by $\sum_{a=1}^{A} x_{st}(a) = 1$.

The proof of Theorem 1 is given in the appendix.

5. Implication of routine-based learning for decision making

With the help of Theorem 1 it is possible to address the basic question of this paper: whether routine-based learning leads to utility maximising behaviour and if not, how it deviates from it.

To this end, it has to be first discussed what utility maximising behaviour means in the situation that is studied here. For each outcome a cardinal utility value has been defined. However, outcomes are random events and choosing one action means facing a probability distribution over outcomes and therefore utility values. Utility maximisation implies that all individuals choose always the same action, the action with the highest expected utility (the case of a few actions with exactly the same expected utility is neglected here). This action is the same for all individuals because they are assumed to be identical here. Therefore, utility maximisation implies in general that $x_{st}(a)$ should be one for one action and zero for all other actions.

The stable state (4.17) of the learning process contradicts this implication if $\nu_V > 0$ and $\mu_a > 0$ for all a. According to equation (4.17) each of the actions is taken with a certain probability within the population. There is no elimination of behaviours as long as not either $\nu_V = 0$ or $\mu_a = 0$ for at least one action a. Therefore, routine-based learning, as it is modelled here, does not converge to utility maximising behaviour as long as $\nu_V > 0$ and $m_0 > 0$.

$\nu_V = 0$ means that changes of behaviour appear only due to imitation processes. Such an assumption seems not to be realistic because it excludes any kind of variation, including variation that is motivated by the dissatisfaction. This would imply that behaviour is only changed due to imitation processes. However, μ_a might equal zero for some actions if $m_0 = 0$ holds. $m_0 = 0$ excludes the exploration of alternative actions. Some models in the literature assume a basic rate

of exploration that vanishes with time. In favour of such an approach it can be argued that the more experience individuals have collected about their possible actions, the more they tend to exploit this knowledge instead of investigating in further exploration. In favour of a constant value of m_0 it can be argued that we live in a world that changes permanently, so that a basic rate of exploration is an innate feature of human beings.

If, however, the basic rate of exploration m_0 vanishes, behaviours might disappear. One additional condition is required for such a disappearance of behaviours: There have to exist actions that always lead to satisficing outcomes, meaning that $\exists a : Q(u|a) = 0 \; \forall \; u < z$. If at least one such action exists, $\mu_a = 0$ holds for this action. As a consequence, equation (4.17) does not hold for this action. Instead, a situation in which all individuals take this action is an absorbing state, meaning that once such a state is reached, there will be no further changes of behaviour. Any state in which all individuals choose the same action and this action leads always to satisficing outcomes, constitutes such an absorbing state. No other absorbing states exist, so that the learning process ends in one of these states. However, the learning process might converge to any behaviour that always leads to satisficing outcomes. Such lock-ins might, therefore, hinder the learning process from converging to the utility maximising behaviour. It is only guaranteed that all actions that cause at least sometimes dissatisfying outcomes are eliminated. However, the utility maximising action might be eliminated as well.

The results can be summed up as follows:

Result 1. *Routine-based learning, as it is modelled here, does in general not converge to utility maximising behaviour. Variation leads to stochastic behaviour and saves actions from being eliminated, while pure imitation or imitation combined with satisficing might by chance eliminate the utility maximising behaviour.*

Remark. Variation in general causes an eventual choice of alternatives that have become extinct or have not been chosen before. Hence, it prevents actions from being eliminated. However, if variation is restricted to those actions that have led to satisfactory outcomes in the past or not been taken before, actions might be eliminated. In contrast, imitation causes individuals to behave conformly. Without the process of variation, no new actions are chosen, while eventually actions are eliminated from the repertoire of the population. In an infinitely large population this elimination might lead to utility maximising behaviour because elimination occurs according to the laws of imitation (see *e.g.* the analysis in Schlag, 1998). If the population is finite and the imitation process is stochastic, actions are eliminated by chance and the utility maximising behaviour might disappear. Satisficing causes individuals to stop changing their behaviour. Hence, utility maximising might not be reached.

This leads to the question of whether the utility maximising behaviour is at least the most frequent action. The frequency of an action a depends on the values μ_a, n_a and θ_a. While n_a and θ_a increase this frequency, μ_a decreases it. Equations (4.6), (4.13) and (4.14) reveal the following characteristics of the values:

First, the following definition is helpful to compare actions.

Definition 1. An action a is said to dominate action $\tilde{a}$ if and only if

$$\int_v^\infty Q(u|a)\mathrm{d}u \geq \int_v^\infty Q(u|\tilde{a})\mathrm{d}u \qquad \forall v \tag{5.1}$$

and $\exists v: \int_v^\infty Q(u|a)\mathrm{d}u > \int_v^\infty Q(u|\tilde{a})\mathrm{d}u$.

Using Definition 1, the characteristics of the functions μ_a, n_a and θ_a imply that an action that dominates another action is chosen more frequently in the long run. Thus, routine-based learning converges to a state in which dominating actions are chosen more frequently.

Second, an action that leads on average to a higher utility does not necessarily imply higher values n_a and θ_a and a lower value μ_a. The value of n_a depends on the average $\bar{u}(a)$ and the variance $\sigma_u(a)$ of the utility that action a gives rise to. θ_a depends on the likelihood of the utility to be above the aspiration level z and the distribution of these values. Finally, μ_a depends on the likelihood of the utility to be below the aspiration level z and the distribution of these values.
Nevertheless, the values of μ_a, n_a and θ_a are completely determined by the probability distribution $Q(u|a)$. Therefore, we might ask whether an imaginary utility function $\tilde{u}(a)$ could be defined such that the learning model is consistent with optimising behaviour given this utility function. Such an imaginary utility function cannot be found, because the introduction of new alternative actions or the elimination of possible actions might change the relative impact of the variation and the imitation processes (due to a change of C) and might therefore change the ordering of actions. This does not happen if one action causes higher values of n_a and θ_a and a lower value of μ_a, but if variation and imitation processes lead to different rankings, the ordering might depend on the number and characteristics of alternative actions.

Result 2. *According to routine-based learning, dominating actions are chosen more frequently in the long run. However, actions with a higher expected utility are not necessarily chosen more frequently and no utility function can be formulated – dependent on the probability distribution $Q(u|a)$ – that would allow to deduce the frequency of each action from this function.*

Remark. The action with the highest expected utility is not necessarily chosen most often because the imitation and the satisficing processes depend not only on the expected utility but also on the distribution of the utility that is caused by an action. Schlag (1998) already found that imitation leads to utility maximisation only if it has a certain structure (probability of imitation linear in the outcome). For other forms of imitation, like the one used here, it deviates from utility maximising behaviour. Satisficing deviates in principle from utility maximising behaviour because it is only important whether the outcome is above or below the aspiration level. Especially how far an outcome is above the aspiration level does not have any impact on satisficing. Hence, the likelihood to choose an action depends on the distribution of the utility that it gives rise to and not only on the expected utility. However, one might expect that a new utility function

could be defined dependent on this distribution such that individuals choose the action with the highest expected new utility most often. This is not possible because the likelihood to choose an action depends on the alternatives that exist. This is caused by the mixture of the processes variation, imitation and satisficing, that all lead to different relations between actions.

However, if each action leads always to the same utility $u(a)$, their frequencies are given by

$$x_{st}(a) = \begin{cases} \dfrac{\nu_V \cdot \exp[\zeta \cdot u(a)]}{m_0 \cdot C - \nu_I \cdot (u(a) - z) \cdot \exp[\zeta \cdot u(a)]} & \text{if } u(a) > z \\ \dfrac{\nu_V \cdot \exp[\zeta \cdot u(a)]}{C \cdot (z - u(a) + m_0)} & \text{if } u(a) \leq z \end{cases} . \tag{5.2}$$

In this case the action that gives rise to the highest utility is chosen most frequently.

Besides the question of which action is chosen most frequently, the question of how much the frequency varies between the actions is of interest. Does the routine-based learning process lead to a nearly equally frequent choice of all actions or is one action chosen with a probability of almost one? The relation between the frequencies of different actions is called the concentration of choice here. Of course, this concentration depends on the difference in the outcomes of the actions, meaning here the values of μ_a, n_a and θ_a. However, it also depends on the parameters ζ, ν_V, ν_I and m_0. This latter dependence is studied here. To this end, the frequencies of two actions a and $\tilde{a}$ are compared:

$$\frac{x_{st}(a)}{x_{st}(\tilde{a})} = \frac{n_a \cdot [\mu_{\tilde{a}} \cdot C - \nu_I \cdot n_{\tilde{a}} \cdot \theta_{\tilde{a}}]}{n_{\tilde{a}} \cdot [\mu_a \cdot C - \nu_I \cdot n_a \cdot \theta_a]} . \tag{5.3}$$

It results that a change of ν_I which determines the likelihood of imitations does not influence the concentration of choice. As long as ν_I does not become zero it does not change $x_{st}(a)$ since each change of ν_I leads to a respective change of C (in order to satisfy the normation condition).

Although ν_V does not appear in equation (5.3), a change of ν_V influences the concentration of choice. An increase of ν_V leads to a respective increase of C due to the normation condition. As a consequence, the influence of μ_a on the frequency of action a is increased. In general it can be assumed that actions that are chosen more frequently are characterised by a higher value of μ_a, because the values μ_a, θ_a and n_a are strongly correlated (compare their definitions in equations (4.6), (4.13) and (4.14)). Thus, an increase of ν_V leads in general to an increase of the concentration of choice.

The basic rate of variation m_0 which is due to the basic desire for novelty and exploration, has a similar impact, although in the opposite direction. If m_0 increases, C has to decrease to satisfy the normation condition. Hence, an increase of m_0 lowers in general the concentration of choice. However, both effects are not very strong.

A much stronger effect is caused by a variation of ζ. An increase of ζ has an exponential effect on the values of n_a. Again, in general those actions that are more frequently chosen are the actions with a higher value of n_a. Thus, an increase of ζ will increase the concentration of choice strongly. In addition, the impact of n_a, meaning the collected experience, on the decision making increases. If ζ is very high, the action with the highest value of n_a is chosen with a probability of almost one. However, it is interesting that at the same time n_a is more and more determined by the variance $\sigma_u(a)$ of the utility that action a gives rise to. In general it seems that an increase of the influence of past experience should move decision making towards utility maximisation. If the utility of an action does not vary, meaning $\sigma_u(a) = 0$, this is the case. For $\zeta \to \infty$ the action with the highest average utility $\bar{u}(a)$ is chosen with probability one, if $\sigma_u(a) = 0$ holds for all actions a. However, if $\sigma_u(a) > 0$ for several actions a, the individuals tend to choose only those actions that lead occasionally to the highest utility. This effect is somewhat eliminated by a decrease of κ which is reached by either increasing the amount of information that is exchanged in the population or by larger memories of the individuals (smaller value of κ_{ii}). This can be summed up as follows.

Result 3. *The likelihood of imitation ν_I has no impact on the concentration of choice, while the likelihood of variation ν_V increases this concentration in favour of those actions that are less dissatisfying. A high basic rate of variation m_0 decreases in general the concentration of choice, while a high influence of experience ζ leads to a nearly exclusive choice of the action with the highest value of n_a. In the case of a strong exchange of information in the population, a huge memory of the individuals, or diminishing variations in the utility of actions this almost leads to utility maximisation.*

Remark. The impacts of ν_I and ν_V seem to depend on the specific choice of the model. In contrary, the effect that a high basic rate of variation decreases the concentration of choice is obvious. Similarly, decision making according to the average experience in the past generally leads to myopic utility maximising, as it has been already found for reinforcement learning (see Börgers and Sarin, 1997) and melioration learning (see Brenner and Witt, 1997).

6. Conclusions

In the recent economic literature the question of whether learning processes converge to utility maximisation has been frequently addressed. Various learning models have been studied. Each of the models is characterised by one or two fundamental mechanisms of learning. The model that has been proposed and studied above combines four of these mechanisms: variation, satisficing, imitation and the collection of experiences. Therefore, it is able to give a more comprehensive picture of the implications of learning for decision making. Three main results are obtained.

First, a learning process that includes variation does not converge to a choice of one action by all individuals. If, instead, only satisficing, imitation and the

collection of experiences play a role, the behaviour might lock-in with all individuals showing the same behaviour that might be not optimal. A certain amount of variation seems therefore to be helpful.

Second, the distribution of behaviour in the population that results from learning cannot be described by a utility function. Even if stochastic decision making is assumed, like in the logit model (see Mc Fadden, 1984), so that the utility assigned to each action has only to represent its frequency in comparison with the frequency of other action, no utility function can be found that remains constant if the set of possible actions changes. Nevertheless, dominating actions are always chosen more frequently.

Third, the concentration of choice increases in general with an increase of ν_V and ζ and a decrease of m_0. If ζ becomes very high, the variation in decision making vanishes. However, the action that is almost surely chosen in this case is not necessarily the one that gives rise to the highest expected utility. Occasional high utilities become very important in this case if the individuals are not able to consider a large amount of information for their decision making. Therefore, expected utility maximisation seems to be a good approximation for routine-based learning only if learning is dominated by the collection of experiences and either the individuals are able to collect a huge amount of information due to communication or a large memory, or the variances in the utilities of each action are very small.

Hence, some conditions for the application of the concept of utility maximisation have been given here and the directions of the deviation from this concept have been discussed for the case that these conditions are not given. The study has been restricted to routine-based learning. Non-cognitive learning has been studied in the literature. What remains to be done to complete the picture is to analyse the implications of associative learning. However, before this can be done, more knowledge about the mechanisms that lead to the change of mental models has to be acquired and the modelling of these mechanisms has to be improved.

Appendix

Proof of Theorem 1. The proof proceeds as follows. First, it is proved that every stationary state has to satisfy equation (4.17). Second, it is proved that this stationary state is unique. Finally, its stability is proved.

Inserting the condition for the stationarity of a state, $x(a, t+1) = x(a, t)$ into equation (4.16) this equation can be transformed into

$$\frac{\nu_V \cdot n_a + \nu_I \cdot n_a \cdot \theta_a \cdot x_{st}(a)}{\mu_a \cdot x_{st}(a)} = \frac{\nu_V \cdot n_{\tilde{a}} + \nu_I \cdot \sum_{\tilde{a}=1}^{A} [n_{\tilde{a}} \cdot \theta_{\tilde{a}} \cdot x_{st}(\tilde{a})]}{\sum_{\tilde{a}=1}^{A} [\mu_{\tilde{a}} \cdot x_{st}(\tilde{a})]} \,. \tag{A.1}$$

The right-hand side of this equation is the same for all actions a. Thus, it can be replaced by a constant C which is determined by the normation condition. Then, equation (A.1) and equation (4.17) state the same relation. Therefore, all stationary states have to satisfy equation (4.17).

For C chosen nearly equal to but larger than $\max_a(\frac{\nu_I \cdot n_a \cdot \theta_a}{\mu_a})$ the sum $\sum_{a=1}^{A} x_{st}(a)$ becomes infinitely large. For C infinitely large the sum equals zero. Furthermore, $x_{st}(a)$ in (4.17) is a continuous function in C for $C > \frac{\nu_I \cdot n_a \cdot \theta_a}{\mu_a}$. Therefore, $\sum_{a=1}^{A} x_{st}(a) = 1$ for one and only one value of C which is larger than $\max_a(\frac{\nu_I \cdot n_a \cdot \theta_a}{\mu_a})$. Thus, the solution is unique.

To investigate the stability of this solution, the Jabobi-matrix is calculated. This results in

$$\mathbf{J}_{ab} = \begin{cases} -\sum\limits_{\tilde{a}=1(\neq a)}^{A} g_{\tilde{a}a} & \text{for } b = a \\ g_{ab} & \text{for } b \neq a \end{cases} \tag{A.2}$$

where

$$g_{ab} = \nu_V \cdot n_b \cdot \mu_a \cdot \frac{x_{st}(a)}{x_{st}(b)} > 0 \; . \tag{A.3}$$

The proof that the eigenvalues of $\mathbf{J}$ are all negative except of one which is zero, is done be contradiction. Let me assume that one eigenvalue λ_1 is positive. For this eigenvalue the following equation has to be satisfied by some eigenvector $\mathbf{v}_1$:

$$\mathbf{L} \cdot \mathbf{v}_1 = 0 \; . \tag{A.4}$$

The matrix $\mathbf{L}$ is given by

$$\mathbf{L}_{ab} = \begin{cases} -\sum\limits_{\tilde{a}=1(\neq a)}^{A} g_{\tilde{a}a} - \lambda_1 & \text{for } b = a \\ g_{ab} & \text{for } b \neq a \end{cases} \; . \tag{A.5}$$

This matrix has the following characteristics. All diagonal elements are negative, while all non-diagonal elements are positive. In each column the absolute value of the negative diagonal element is larger than the sum of all positive non-diagonal elements. Equation (A.4) has a solution if the rows in matrix $\mathbf{L}$ are linearly dependent. This means that factors f_i ($f_i \neq 0$ for at least one of them) can be found such that $\sum_{i=1}^{A} f_i \cdot L_{ij} = 0$ for all j. Let me assumed that such set of factors has been found. If none of the factors is greater than zero, they are all multiplied by -1 (they still satisfy the condition). Then, the highest value f_h is identified. $f_h > 0$ holds. It results that $\sum_{i=1}^{A} f_i \cdot L_{ih} < 0$ since $-L_{hh} > \sum_{i=1,\neq j}^{A} L_{ih}$ and $f_h \geq f_i$ for all i. This contradicts $\sum_{i=1}^{A} f_i \cdot L_{ij} = 0$. Thus, no set f_i is found to satisfy $\sum_{i=1}^{A} f_i \cdot L_{ij} = 0$ and, therefore, the rows of $\mathbf{L}$ are linearly independent. Thus, no eigenvalue with $\lambda > 0$ exists. □

Acknowledgements. I thank Drew Fudenberg, Ulrich Witt, Armin Haas, and Helmar Abele for stimulating discussions and helpful comments.

References

Bandura A. (1979) *Sozial-kognitive Lerntheorie*. Stuttgart: Klett-Cotta.

Binmore K. and Samuelson L. (1994) *Muddling Through: Noisy Equilibrium Selection*. Discussion Paper No. B-275, Sonderforschungsbereich 303, Bonn.

Börgers T. (1996) On the Relevance of Learning and Evolution to Economic Theory. *The Economic Journal* **106**, 1374–1385.

Börgers T. and Sarin R. (1996) *Naive Reinforcement Learning With Endogenous Aspirations*. mimeo, University College London.

Börgers T. and Sarin R. (1997) Learning Through Reinforcement and Replicator Dynamics. *J. Econ. Theory* **77**, 1–16.

Brenner T. (1997) Decision Making and the Exchange of Information. In: F. Schweitzer, ed., *Self-Organization of Complex Structures: From Individual to Collective Dynamics, Vol. II*, pp. 379–392. London: Gordon and Breach.

Brenner T. (1999) *Modelling Learning in Economics*. Cheltenham: Edgar Elgar.

Brenner T. and Witt U. (1997) *Frequency-Dependent Pay-offs, Replicator Dynamics, and Learning Under the Matching Law*. Papers on Economics and Evolution #9706, Max-Planck-Institute, Jena.

Brown G.W. (1951) Iterative Solution of Games by Fictitious Play. In: *Activity Analysis of Production and Allocation*, pp. 374–376. New York: John Wiley and Sons.

Bush R.R. and Mosteller F. (1955) *Stochastic Models for Learning*. New York: John Wiley and Sons.

Camerer C. and Ho T.-H. (1999) Experience-Weighted Attraction Learning in Normal Form Games. *Econometrica* **67**, 827–874.

Crawford V.P. (1995) Adaptive Dynamics in Coordination Games. *Econometrica* **63**, 103–143.

Dawid H. (1997) Learning of Equilibria by a Population with Minimal Information. *J. Econ. Behavior and Organization* **32**, 1–18.

Day R.H. (1967) Profits, Learning and the Convergence of Satisficing to Marginalism. *Quarte. J. Econ.* **81**, 302–311.

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

Erev I. and Roth A.E. (1998) Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *Am. Econ. Rev.* **88**, 848–881.

Festinger L. (1942) A Theoretical Interpretation of Shifts in Level of Aspiration. *Psycholo. Rev.* **49**, 235–250.

Gigerenzer G. and Goldstein D.G. (1996) Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psycholo. Rev.* **103**, 650–669.

Harrison G.W. (1994) Expected Utility Theory and the Experimentalists. *Emp. Econ.* **19**, 223–253.

Herrnstein R.J. and Prelec D. (1991) Melioration: A Theory of Distributed Choice. *J. Econ. Perspectives* **5**, 137–156.

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

Latané B. (1981) The Psychology of Social Impact. *Am. Psychol.* **36**, 343–356.

Levine D.K. and Pesendorfer W. (2000) *Evolution Through Imitation in a Single Population*. mimeo, UCLA and Princeton.

Mailath G.J. (1998) Do People Play Nash Equilibrium? Lessons From Evolutionary Game Theory. *J. Econ. Lit.* **36**, 1347–1374.

Marcet A. and Sargent T.J. (1989) Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models. *J. Econ. Theory* **48**, 337–368.

Marimon R. (1993) Adaptive Learning, Evolutionary Dynamics and Equilibrium Selection in Games. *Eur. Econ. Rev.* **37**, 603–611.

Mc Fadden D.L. (1984) Econometric Analysis of Quantitative Response Models. In: Z. Griliches and M. D. Intriligator, Eds., *Handbook of Econometrics, Vol. II*, pp. 1395–1457. Amsterdam: Elsevier Science Publisher.

Nash J.F. (1950) Equilibrium Points in n-Person Games. *Proceedings of the National Academy of Science (US)* **36**, 48–49.

Nelson R.R. and Winter S.G. (1982) *An Evolutionary Theory of Economic Change*. Cambridge: The Belknap Press.

Pawlov I.P. (1953) *Sämtliche Werke, Bd. IV*. Berlin: Akademie-Verlag.

Posch M. (1999) Win-Stay, Lose-Shift Strategies for Repeated Games – Memory Length, Aspiration Levels and Noise. *J. Theoret. Biol.* **198**, 183–195.

Roth A.E. and Erev I. (1995) Learning in Extensive Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Run. *Games and Economic Behavior* **6**, 164–212.

Samuelson L. (1994) Stochastic Stability in Games with Alternative Best Replies. *J. Econ. Theory* **64**, 35–65.

Sarin R. (2000) Decision Rules with Bounded Memory. *J. Econ. Theory* **90**, 151–160.

Sarin R. and Vahid F. (1999) Payoff Assessments without Probabilities: A Simple Dynamic Model of Choice. *Games and Economic Behavior* **28**, 294–309.

Schlag K.H. (1998) Why Imitate, and If So, How? A Bounded Rational Approach to Multi-armed Bandits. *J. Econ. Theory* **78**, 130–156.

Scitovsky T. (1992) *The Joyless Economy, revised edition*. New York: Oxford University Press.

Simon H.A. (1987) *Satisficing*. The New Palgrave Dictionary of Economics, Vol. 4, Macmillan Press, London, 243–245.

Sinclair P.J. (1990) The Economics of Imitation. *Scottish J. Political Econ.* **37**, 113–144.

Slembeck T. (1999) *Learning in Economics: Where Do We Stand? A Behavioral View on Learning in Theory, Practice and Experiments*. Discussion paper no. 9907, University of St. Gallen, Department of Economics.

Slonim R.L. (1999) Learning Rules of Thumb or Learning more Rational Rules. *J. Econ. Behavior and Organization* **38**, 217–236.

Weibull J.W. (2000) *Testing Game Theory*. mimeo, Stockholm.

Witt U. (1986) How Can Complex Economic Behavior Be Investigated? The Example of the Ignorant Monopolist Revisited. *Behav. Sci.* **31**, 173–188.

Witt U. (1987) How Transaction Rights Are Shaped to Channel Innovativeness. *J. Institutional and Theoretical Economics* **143**, 180–195.

Witt U. (1996) Bounded Rationality, Social Learning, and Viable Moral Conduct in a Prisoners' Dilemma. In: M. Perlman and E. Helmstädter, eds., *Behavioral Norms, Technological Progress and Economic Dynamics: Studies in Schumpeterian Economics*, pp. 33–49. Ann Arbor: Michigan University Press.

Wu J. and Axelrod R. (1995) How to Cope with Noise in the Iterated Prisoner's Dilemma. *J. Conflict Resolution* **39**, 183–189.

Young P. (1993) The Evolution of Conventions. *Econometrica* **61**, 57–84.