



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2002

---

## **A myopic adjustment process leading to best-reply matching**

Kosfeld, M ; Droste, E ; Voorneveld, M

DOI: [https://doi.org/10.1016/S0899-8256\(02\)00007-6](https://doi.org/10.1016/S0899-8256(02)00007-6)

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-2248>

Journal Article

Originally published at:

Kosfeld, M; Droste, E; Voorneveld, M (2002). A myopic adjustment process leading to best-reply matching. *Games and Economic Behavior*, 40(2):270-290.

DOI: [https://doi.org/10.1016/S0899-8256\(02\)00007-6](https://doi.org/10.1016/S0899-8256(02)00007-6)

# A Myopic Adjustment Process Leading to Best-Reply Matching\*

Michael Kosfeld<sup>†</sup>

*Institute for Empirical Research in Economics  
University of Zurich  
Blümlisalpstrasse 10  
CH-8006 Zurich  
Switzerland*

Edward Droste

*KPMG  
Economic Consulting  
P.O. Box 559  
NL-2130 AN Hoofddorp  
The Netherlands*

Mark Voorneveld<sup>‡</sup>

*Department of Economics  
Stockholm School of Economics  
Box 6501  
SE-113 83 Stockholm  
Sweden*

---

\*Published in: *Games and Economic Behavior* 40, 2002, 270-298. We would like to thank Eric van Damme, Urs Fischbacher, Bill Sandholm, Dolf Talman, and an anonymous referee for useful comments and suggestions. We are particularly grateful to Josef Hofbauer for many helpful comments and for pointing us to a general convergence result in two-player games. The result is included in the appendix of the paper.

<sup>†</sup>This author's research has been financially supported by the European Commission through a Marie Curie Fellowship at CentER for Economic Research, Tilburg University.

<sup>‡</sup>Part of this author's research was financially supported by the Dutch Foundation for Mathematical Research (SWON) through project 613-04-051.

*Proposed Running Head:*  
Adjustment to Best-Reply Matching

*Corresponding author:*  
Michael Kosfeld  
Institute for Empirical Research in Economics  
University of Zurich  
Blümlisalpstrasse 10  
CH-8006 Zurich  
Switzerland  
phone: +41 1 634 3761  
fax: +41 1 634 4907  
email: kosfeld@iew.unizh.ch

**Abstract**

We analyze a myopic strategy adjustment process in strategic-form games. It is shown that the steady states of the continuous time limit, which is constructed assuming frequent play and slow adjustment of strategies, are exactly the best-reply matching equilibria, as discussed in Droste *et al.* (2000). In a best-reply matching equilibrium every player ‘matches’ the probability of playing a pure strategy to the probability that this pure strategy is a best reply to the pure strategy profile played by his opponents. We derive stability results for the steady states of the continuous time limit in  $2 \times 2$  bimatrix games and coordination games. Analyzing the asymptotic behavior of the stochastic adjustment process in discrete time shows convergence to minimal curb sets of the game. Moreover, absorbing states of the process correspond to best-reply matching equilibria of the game. *Journal of Economic Literature* Classification Numbers: C72, D83.

# 1 Introduction

In this paper, we introduce an adjustment process that models the evolution of mixed strategies in finite normal-form games. Two interpretations are available for the adjustment process. The first is based on a multipopulation model of social learning, as commonly studied in evolutionary game theory. In this set-up mixed strategies represent shares of pure strategies in the populations corresponding to different player positions in the underlying game. The second interpretation of the process focuses on individual decision-making and is more closely related to the psychological literature on individual learning. Here, mixed strategies capture probabilistic choice behavior of individual players.

Independent of the interpretation, we take the best-reply structure of the normal-form game as the main driving force behind the adjustment process. The idea we pursue is the following: any time the game is played, pure strategies that are best replies to the realized strategy profile are reinforced, by increasing the probability they are chosen in the future. On an individual level this assumption is motivated by the experience of regret in case a player has not played a best reply to the profile of his opponents (or, similarly, rejoice in case he has played a best reply). On an aggregate, or social, level the idea is that every time a constant share of strategies dies and is replaced by new strategies that are best replies to the observed outcome.

Models of reinforcement learning have attracted much interest recently, e.g. to explain experimental data (Roth and Erev, 1995; Erev and Roth, 1998; and Camerer and Ho, 1999), or for building microfoundations to aggregate dynamics in evolutionary game theory (Börgers and Sarin, 1997). While this literature shares many fundamental principles with our model, we shall emphasize a main difference that we believe to be of crucial importance. Contrary to most other models the reinforcement of strategies in our set-up is not based on actually earned payoffs but on the experience of having played a best reply. Since the attribute of being a best reply is a pure ordinal concept, this consequently leads to an ordinal model of reinforcement learning.

In the course of analyzing the adjustment process, which is a discrete-time Markov process with infinite state space, we find it necessary to distinguish between behavior in finite time and asymptotic behavior. Assuming frequent interaction and slow adjustment, behavior in finite time can be approximated by a dynamical system of deterministic differential equations. Frequent interaction and slow adjustment mean that each discrete-time period contains ‘many’ repetitions of the game, and that the players’ adjustments of their mixed strategies between two successive repetitions are ‘small’, respectively. The approximation is based on results from Norman (1972) that recently have also been used in Börgers and

Sarin (1997). For a general discussion of approximation of stochastic processes by dynamical systems, including also a stronger approximation result see, e.g., Benaïm and Weibull (1999) and Sandholm (1999).<sup>1</sup> In our set-up analyzing the steady states of the deterministic approximation and their stability properties provides us with information about the players' behavior in finite time.

A main feature of the deterministic approximation is that its steady states predict matching behavior. In fact, they predict so-called best-reply matching behavior, meaning that every player plays a pure strategy with a probability which is generically equal to the probability that this pure strategy is a best reply to the pure-strategy profile of his opponents. Matching behavior arising in a repeated interaction scenario suggests that a player eventually 'learns' the probabilities with which the pure-strategy profiles of his opponents occur. However, once these probabilities are 'known', he does not reply by playing the pure strategy that maximizes expected utility but instead matches the probabilities. Matching behavior is often associated with the decision-theoretic setting where a player repeatedly faces a two-armed bandit problem. A two-armed bandit is a slot machine with two arms, each operating with a fixed probability. At each point in time, the player's task is to predict which arm will be operating. In such a setting matching behavior is supported by the experiments of, e.g., Edwards (1956), Siegel and Goldstein (1959), and Suppes and Atkinson (1960). In this paper, we propose a model which predicts that matching behavior may also occur in interactive situations.

The asymptotic behavior of the stochastic adjustment process can typically be quite different from the asymptotic behavior of the deterministic approximation. More precisely, we show that asymptotic behavior of the adjustment process is characterized by a set of mixed-strategy profiles whose supports are contained in a minimal curb set. Minimal curb sets are introduced by Basu and Weibull (1991). A nonempty product set of pure strategies is said to be closed under best replies if all best replies against all these pure strategies are contained in the set. Minimal sets with this property are called minimal curb sets. Still, it may happen that asymptotic behavior of the adjustment process and the deterministic approximation coincide. A necessary condition for this to happen is the existence of an absorbing state of the adjustment process. Namely, as we will show, the absorbing states of the adjustment process reflect best-reply matching behavior.

As mentioned above, on the individual level of players the adjustment process is closely related to the literature on reinforcement learning as pioneered by Estes (1950) and Bush and Mosteller (1955), see also Cross (1983). Reinforcement learning assumes that players behave

---

<sup>1</sup>Contrary to our model, however, these articles consider Markov processes with a finite state space.

in a stimulus-response fashion where pure strategies are positively or negatively reinforced through their payoffs. Since these models do not require players to develop complicated cognitive processes in order to form beliefs about their environment, they are very attractive from the players' rationality or knowledge point of view. On this point, the adjustment process considered in this paper is closely related to the reinforcement learning models. Namely, we assume players to myopically adjust their mixed strategies in response to the experiences they have when playing the game. Players negatively reinforce pure strategies which are not a best reply to the realized outcome of the game. This also indicates the main difference between the adjustment process and the reinforcement learning models: the adjustment process is based on an ordinal concept of best replies, while models of reinforcement learning are based on cardinal concepts of payoffs.

The learning models analyzed by Börgers and Sarin (1995, 1997) are of special relevance to the adjustment process considered in this paper. Not only are the mathematical techniques used in the work of Börgers and Sarin to some extent similar to the ones we use here, but also, under certain conditions Börgers and Sarin (1995) predict matching behavior in interactive situations. The learning process analyzed in Börgers and Sarin (1995) differs, however, significantly from the present adjustment process as it explicitly models an endogenously evolving aspiration level.

Hurkens (1995) considers a learning process in finite normal-form games which is represented by a discrete-time Markov process with a finite state space. The players are assumed to play best replies against their beliefs, which in turn are formed on the basis of the pure-strategy profiles used in the recent past. It is shown that the learning process leads the players to play pure-strategy profiles from a minimal curb set.

The adjustment process is also related to the adaptive procedure of Hart and Mas-Colell (2000). Although Hart and Mas-Colell focus on a process leading to correlated equilibria, the intuition underlying their procedure coincides to a certain extent with the regret considerations mentioned above. In fact, they assume that players adjust the probabilities assigned to pure strategies by measuring the average regret of not having played that strategy in the past. Notice that we assume players to focus on recent history only. Furthermore, contrary to the ordinal approach presented in this paper, Hart and Mas-Colell measure regret by actual payoff differences.

The paper is organized as follows. In Section 2 we introduce the class of normal-form games considered in this paper and we state the adjustment process describing how the players' behavior evolves over time. Two alternative interpretations of the adjustment process are discussed in Section 3. Sections 4 and 5 deal with behavior in finite time. The main result in Section 4 identifies the dynamical system which approximates the adjustment process in

finite time. Section 5 deals with the stability properties of the steady states of the dynamical system. (We include a further result in the appendix.) The asymptotic long-run behavior of the adjustment process is analyzed in Section 6. Section 7 concludes.

## 2 Model

The finite normal-form game that we consider in this paper is represented by a tuple  $G = \langle N, (S_i)_{i \in N}, (\succeq_i)_{i \in N} \rangle$ , where  $N = \{1, \dots, n\}$  is a finite set of players. Every player  $i \in N$  has a finite set  $S_i$  of pure strategies and a binary relation  $\succeq_i$  over the joint pure-strategy set  $S = \prod_{i \in N} S_i$ , which reflects his preferences over the outcomes of the game  $G$ . The binary relation  $\succeq_i$  is assumed to be reflexive and its asymmetric part  $\succ_i$ , defined for all  $s, \tilde{s} \in S$  by

$$s \succ_i \tilde{s} \Leftrightarrow [s \succeq_i \tilde{s} \text{ and } \tilde{s} \not\succeq_i s],$$

is assumed to be acyclic. In the following, we also consider the more stringent case in which the preference relations  $\succeq_i$  induce von Neumann-Morgenstern utility functions  $u_i : S \rightarrow \mathbb{R}$ . The corresponding game is denoted by  $G = \langle N, (S_i)_{i \in N}, (u_i)_{i \in N} \rangle$ . We write  $S_{-i} = \prod_{j \in N \setminus \{i\}} S_j$  for notational convenience. For a pure-strategy tuple  $s = (s_1, \dots, s_n) \in S$  and a player  $i \in N$ , we let  $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n) \in S_{-i}$  and, with a slight abuse of notation,  $s = (s_i, s_{-i}) \in S$ .

The set of mixed strategies of a player  $i \in N$  is denoted by

$$\Delta_i := \{ \sigma_i : S_i \rightarrow \mathbb{R} \mid \forall s_i \in S_i : \sigma_i(s_i) \geq 0, \sum_{s_i \in S_i} \sigma_i(s_i) = 1 \}$$

and the set of mixed strategies of a player  $i \in N$ , such that a positive probability is assigned to every pure strategy, is given by

$$\text{int}(\Delta_i) := \{ \sigma_i : S_i \rightarrow \mathbb{R} \mid \forall s_i \in S_i : \sigma_i(s_i) > 0, \sum_{s_i \in S_i} \sigma_i(s_i) = 1 \}.$$

Analogous to the pure-strategy case, we use notations  $\Delta$ ,  $\Delta_{-i}$ ,  $\sigma_{-i}$ , and  $\sigma = (\sigma_i, \sigma_{-i})$ . Furthermore, for a mixed-strategy profile  $\sigma \in \Delta$  and a player  $i \in N$ , we let  $\sigma_{-i}(s_{-i}) := \prod_{j \in N \setminus \{i\}} \sigma_j(s_j)$  denote the probability that the opponents of player  $i$  play the pure-strategy profile  $s_{-i} \in S_{-i}$ . Similarly, for a mixed-strategy profile  $\sigma \in \Delta$ ,  $\sigma(s) = \prod_{i \in N} \sigma_i(s_i)$  gives the probability that the pure-strategy profile  $s \in S$  is realized.

Finally, for every player  $i \in N$  and every pure-strategy profile of his opponents  $s_{-i} \in S_{-i}$ , we denote the set of best replies, i.e., the set of pure strategies that player  $i$  cannot improve upon, by

$$B_i(s_{-i}) := \{ s_i \in S_i \mid \nexists \tilde{s}_i \in S_i : (\tilde{s}_i, s_{-i}) \succ_i (s_i, s_{-i}) \}.$$

Since  $S_i$  is finite and  $\succ_i$  is acyclic, we know that  $B_i(s_{-i})$  is nonempty. We assume that every player  $i \in N$  has knowledge of the set  $B_i(s_{-i})$  for every pure-strategy profile of his opponents  $s_{-i} \in S_{-i}$ .

The game  $G$  introduced above is played repeatedly by the  $n$  players. The iterations of the game are indexed by  $k \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$ . We take a probabilistic choice approach, meaning that at each time  $k \in \mathbb{N}_0$  every player  $i \in N$  plays a pure strategy  $s_i \in S_i$ , but this pure strategy is drawn from the probability distribution over his set of pure strategies as induced by his mixed strategy  $\sigma_i^k \in \Delta_i$ . Hence, at each time  $k \in \mathbb{N}_0$ , every player  $i \in N$  is completely characterized by his mixed strategy  $\sigma_i^k \in \Delta_i$ . The state of the game, which characterizes every player  $i \in N$ , at time  $k \in \mathbb{N}_0$  is therefore given by a mixed-strategy profile  $\sigma^k = (\sigma_1^k, \dots, \sigma_n^k) \in \Delta$ . By  $\sigma_i^k(s_i)$ , we denote the probability that player  $i$ , at time  $k$ , plays the pure strategy  $s_i \in S_i$ . Furthermore, for a mixed-strategy profile  $\sigma^k \in \Delta$  and a player  $i \in N$ , we let  $\sigma_{-i}^k(s_{-i}) := \prod_{j \in N \setminus \{i\}} \sigma_j^k(s_j)$  denote the probability that player  $i$ 's opponents, at time  $k$ , play the pure-strategy profile  $s_{-i} \in S_{-i}$ . Similarly, for a mixed-strategy profile  $\sigma^k \in \Delta$ ,  $\sigma^k(s) = \prod_{i \in N} \sigma_i^k(s_i)$  gives the probability that the pure-strategy profile  $s \in S$  is realized at time  $k$ . The initial state of the game  $\sigma^0 = (\sigma_1^0, \dots, \sigma_n^0) \in \Delta$  is exogenously given.

Now, we describe the adjustment rule that specifies how the players' mixed strategies, as represented by the state of the game  $\sigma^k \in \Delta$ , evolve over time. Consider a fixed period  $k \in \mathbb{N}_0$  and let the pure-strategy profile  $s \in S$  be realized in this period. We assume that after the game  $G$  is played every player  $i \in N$  observes the realized pure-strategy profile of his opponents  $s_{-i} \in S_{-i}$  and then adjusts his mixed strategy as follows

$$\sigma_i^{k+1}(s_i) = \begin{cases} (1 - \theta) \sigma_i^k(s_i) + \frac{\theta}{|B_i(s_{-i})|} & \text{if } s_i \in B_i(s_{-i}), \\ (1 - \theta) \sigma_i^k(s_i) & \text{otherwise,} \end{cases} \quad (1)$$

where  $0 < \theta < 1$  is exogenously given. For a given initial mixed-strategy profile  $\sigma^0 \in \Delta$  and parameter  $\theta$ , the adjustment rule (1) determines a discrete-time Markov process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  with infinite state space  $\Delta$ . An extensive discussion of the adjustment rule (1) is contained in the next section.

### 3 Two Alternative Interpretations

In this section, we discuss two interpretations of the model, and of the adjustment rule (1) in particular, as introduced in the previous section. The first interpretation we offer is based on a multipopulation model of social learning. The second interpretation relies on principles of individual learning which originated in the psychological literature.



### 3.1. A Model of Social Learning

Suppose there exist  $n$  large (technically infinite) populations of agents, each corresponding to a particular player position in the underlying game. Every agent in population  $i \in N$  is programmed to a pure strategy  $s_i \in S_i$ . The fraction of agents in population  $i$  programmed to pure strategy  $s_i$  in period  $k \in \mathbb{N}_0$  is given by  $\sigma_i^k(s_i)$ . We examine the idea that the multipopulation model describes a process of social learning. In a model of social learning, agents typically ask around and learn from other agents in the populations. Consider a social-learning process such that in each period  $k$ , some fraction  $\theta$  of the agents in every population  $i \in N$  is randomly drawn to play the game  $G$ . The agents who are called to play the game are randomly matched in  $n$ -tuples such that each agent is matched with exactly one agent from every other population. After all  $n$ -tuples of agents have played the game, the participating agents leave the system and are replaced by new agents who learn something about the prevailing states of the  $n$  populations. More precisely, suppose that the new agents sample exactly one of the outcomes of the games, say the pure-strategy profiles  $s \in S$ . Note that the probability of sampling a pure-strategy profile thus equals its share in the current populations.

Now, consider a new agent who replaces an agent exiting population  $i$ . This new agent makes a once-and-for-all choice of pure strategy, which he does by adopting the best reply to the observed pure-strategy profile  $s_{-i} \in S_{-i}$ . In the case of multiple best replies, we assume that a new agent adopts each of the best replies with equal probability. Obviously, the fraction of agents in population  $i$  programmed to pure strategy  $s_i$  in period  $k + 1$ , i.e.,  $\sigma_i^{k+1}(s_i)$ , is then equal to  $(1 - \theta) \sigma_i^k(s_i) + \frac{\theta}{|B_i(s_{-i})|}$  if  $s_i \in B_i(s_{-i})$  and equal to  $(1 - \theta) \sigma_i^k(s_i)$  otherwise. This is exactly the adjustment rule (1).

The social-learning model states that agents are being periodically replaced and that only new agents make choices. This causes the state of the adaptive system, i.e., the distribution of strategies currently played, to be only a function of last period's state and not of the entire history. However, instead of replacing agents, it is also possible to assume that the agents who are called to play the game (and then adjust their pure strategy) do not remember their own past experience.

### 3.2. Principles of Individual Learning

Alternatively, the adjustment rule (1) can be interpreted as a model of individual learning that is based on common principles from the psychological literature. Adjustment rule (1) states that a player  $i \in N$  in period  $k \in \mathbb{N}_0$ , after the game  $G$  is played, myopically adjusts his mixed strategy by first proportionally decreasing all probabilities currently assigned to

the pure strategies by a fraction  $0 < \theta < 1$ . The parameter  $\theta$  reflects the so-called recency or forgetting effect, meaning that a player's recent experience plays a larger role than past experience in determining his behavior.

Having proportionally decreased all probabilities then leaves the player with a probability  $\theta$  that is reallocated over all pure strategies which are best replies to the pure-strategy profile of his opponents that is realized in period  $k$ . Reallocating the probability  $\theta$  over all best replies corresponds to the so-called generalization or experimentation effect, stating that players will not quickly become locked into one alternative in exclusion of all others. The probability is equally distributed over all best replies, since as far as the recent experience is concerned no best reply is better than the others. Therefore all alternatives are treated similarly.

Examination of the psychological literature on individual learning led Roth and Erev (1995) and Erev and Roth (1998) to the conclusion that the recency effect and the generalization effect are robust phenomena observed in a wide variety of experimental settings. Both effects should therefore be incorporated in a model studying learning in games.

Note that the pure strategy currently played by player  $i \in N$  does not directly influence how he himself adjusts his mixed strategy since the adjustment rule (1) is completely determined by the current pure-strategy profile of his opponents. The pure strategy currently played by player  $i$  does, however, influence the adjustment rule of all of his opponents. Consequently, the pure strategy currently played by player  $i$  influences the future mixed-strategy profiles of his opponents and therefore it will influence his own future behavior indirectly.

According to Erev and Roth, the law of effect is another principle of human learning in strategic environments. The law of effect, as first introduced by Edward L. Thorndike (cf. Erev and Roth, 1998, p.859), states that choices which have led to good outcomes in the past are more likely to be repeated in the future. The adjustment rule (1) reflects the law of effect even though the rule does not necessarily imply that the probability assigned to a player's pure strategy increases if it is a best reply to the realized pure-strategy profile of his opponents. Indeed, the probability assigned to a best reply may decrease only if this strategy is one of multiple best replies to the opponents' strategy profile and if this strategy currently receives more probability than the alternative best replies. In this case the law of effect interferes with the generalization effect.

## 4 Behavior in Finite Time

We find it necessary to distinguish between behavior predicted by the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  in finite time, on the one hand, and asymptotic behavior, on the other hand. This

and the next section deal with the first kind of behavior.

Before we analyze behavior in finite time it is useful to describe the expected movement of the state of the game under the adjustment rule (1). Consider the  $k$ -th repetition of the game  $G$  and let the current state of the game  $\sigma^k = \sigma$ . Given that  $\sigma^k = \sigma$ , next period's state of the game  $\sigma^{k+1}$  is a random variable. In fact, let  $E \left[ \sigma_i^{k+1}(s_i) - \sigma_i^k(s_i) \mid \sigma^k = \sigma \right]$  denote the expected value of  $\sigma_i^{k+1}(s_i) - \sigma_i^k(s_i)$  conditional on  $\sigma^k = \sigma$ , then it holds that

$$\begin{aligned}
& E \left[ \sigma_i^{k+1}(s_i) - \sigma_i^k(s_i) \mid \sigma^k = \sigma \right] \\
&= \sum_{\{s_{-i} \in S_{-i} \mid s_i \in B_i(s_{-i})\}} \left( (1 - \theta) \sigma_i(s_i) + \frac{\theta}{|B_i(s_{-i})|} - \sigma_i(s_i) \right) \sigma_{-i}(s_{-i}) \\
&\quad + ((1 - \theta) \sigma_i(s_i) - \sigma_i(s_i)) \left( 1 - \sum_{\{s_{-i} \in S_{-i} \mid s_i \in B_i(s_{-i})\}} \sigma_{-i}(s_{-i}) \right) \\
&= \sum_{\{s_{-i} \in S_{-i} \mid s_i \in B_i(s_{-i})\}} \left( \frac{\theta}{|B_i(s_{-i})|} - \theta \sigma_i(s_i) \right) \sigma_{-i}(s_{-i}) \\
&\quad - \theta \sigma_i(s_i) \left( 1 - \sum_{\{s_{-i} \in S_{-i} \mid s_i \in B_i(s_{-i})\}} \sigma_{-i}(s_{-i}) \right) \\
&= \theta \left( \left( \sum_{\{s_{-i} \in S_{-i} \mid s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \sigma_{-i}(s_{-i}) \right) - \sigma_i(s_i) \right)
\end{aligned} \tag{2}$$

for all  $k \in \mathbb{N}_0$ ,  $\sigma \in \Delta$ ,  $i \in N$ , and  $s_i \in S_i$ . As we will see below, the expected movement (2) of the state of the game is indispensable for describing the behavior of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  in finite time.

To be able to analyze the behavior of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  in finite time, we assume that each discrete-time period  $k \in \mathbb{N}_0$  contains ‘many’ repetitions of the game  $G$ , and that the adjustments of their behavior which the players make between two repetitions of the game are ‘small’.

Formally, let the time that passes between two repetitions of the game be denoted by  $0 < \eta \leq 1$ , which is equivalent to saying that each discrete-time period  $k$  contains  $\frac{1}{\eta}$  repetitions of the game. After each repetition of the game, the players adjust their behavior by  $\eta$  times what is prescribed by the adjustment rule (1), meaning that the players’ adjustment of their behavior ‘slows down’ at the same rate at which the time distance between two repetitions shrinks. The above considerations give rise to a slight modification of the adjustment rule (1). In fact, they lead us to consider the adjustment rule given by

$$\sigma_i^{\eta, k+1}(s_i) = \begin{cases} (1 - \eta\theta) \sigma_i^{\eta, k}(s_i) + \frac{\eta\theta}{|B_i(s_{-i})|} & \text{if } s_i \in B_i(s_{-i}), \\ (1 - \eta\theta) \sigma_i^{\eta, k}(s_i) & \text{otherwise,} \end{cases} \tag{3}$$

for all  $i \in N$  and  $k \in \mathbb{N}_0$ . Again, for a given parameter  $\theta$  with  $0 < \theta < 1$ , we are left with a discrete-time Markov process  $\{\sigma^{\eta,k}\}_{k \in \mathbb{N}_0}$  with infinite state space  $\Delta$ , provided that we specify the initial random variable  $\sigma^{\eta,0}$ . Since we let the time interval between two repetitions of the game be equal to  $\eta$ , the random variable  $\sigma^{\eta,k}$  gives the state of the game at ‘real’ time  $\eta k$ .

We are interested in the limit of the state of the game  $\sigma^{\eta,k}$  as  $\eta \rightarrow 0$ . As will be shown below, taking this limit not only enables us to work in continuous time, but also to approximate the adjustment process  $\{\sigma^{\eta,k}\}_{k \in \mathbb{N}_0}$  by a linear dynamical system of deterministic differential equations. To describe the dynamical system, we introduce a function  $\hat{\sigma}^t : \mathbb{R}_+ \rightarrow \Delta$  which is differentiable with respect to  $t \in \mathbb{R}_+$ . The derivative of  $\hat{\sigma}^t$  is exactly the expected movement (2) of the state of the game and is therefore given by

$$\frac{d\hat{\sigma}_i^t(s_i)}{dt} = \theta \left( \left( \sum_{\{s_{-i} \in S_{-i} | s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \hat{\sigma}_{-i}^t(s_{-i}) \right) - \hat{\sigma}_i^t(s_i) \right) \quad (4)$$

for all  $t \in \mathbb{R}_+$ ,  $i \in N$ , and  $s_i \in S_i$ . Let  $\hat{\sigma}^0$  denote the initial value of  $\hat{\sigma}^t$ .

Proposition 1 states that the limit of the state of the game  $\sigma^{\eta,k}$  for any sequence of  $\eta$ 's and  $k$ 's such that  $\eta \rightarrow 0$  and  $k\eta \rightarrow t$  is almost surely given by the state  $\hat{\sigma}^t$  of the dynamical system (4) at any finite time  $t \in \mathbb{R}_+$ .

**Proposition 1** *Suppose that for all  $0 < \eta \leq 1$  it holds that  $\sigma^{\eta,0} = \hat{\sigma}^0$  with probability 1. Consider some  $t$  with  $0 \leq t < \infty$  and let  $\eta \rightarrow 0$  and  $k\eta \rightarrow t$ . Then  $\sigma^{\eta,k}$  converges in probability to  $\hat{\sigma}^t$ .*

**Proof.** To prove Proposition 1, we use Theorem 1.1 in Chapter 8 of Norman (1972). Norman's theorem is a general central limit theorem for discrete-time Markov processes with infinite state space, which describe adjustment by small steps. In fact, Proposition 1 follows directly from parts (A) and (B) of Norman's theorem and therefore it is sufficient to verify the conditions imposed by the theorem.

We apply Norman's theorem to the adjustment process  $\{\sigma^{\eta,k}\}_{k \in \mathbb{N}_0}$  and we regard  $\sigma^{\eta,k}$  as an element of  $[0, 1]^{\Sigma_{i \in N} |S_i|} \subset \mathbb{R}^{\Sigma_{i \in N} |S_i|}$ . Norman's condition (a.1) requires that the state spaces of the processes  $\{\sigma^{\eta,k}\}_{k \in \mathbb{N}_0}$  approximate the convex hull of these spaces as  $\eta$  goes to zero. This condition is trivially satisfied because the state space of the adjustment process  $\{\sigma^{\eta,k}\}_{k \in \mathbb{N}_0}$  is equal to  $\Delta$ , which is independent of  $\eta$ . Conditions (a.2) and (a.3) consider the functions  $v : \mathbb{R}^{\Sigma_{i \in N} |S_i|} \times \mathbb{R} \rightarrow \mathbb{R}^{\Sigma_{i \in N} |S_i|}$  and  $w : \mathbb{R}^{\Sigma_{i \in N} |S_i|} \times \mathbb{R} \rightarrow \mathbb{R}^{(\Sigma_{i \in N} |S_i|) \times (\Sigma_{i \in N} |S_i|)}$ , defined by

$$v(\sigma, \eta) := E \left[ \frac{\sigma^{\eta,k+1} - \sigma^{\eta,k}}{\eta} \middle| \sigma^{\eta,k} = \sigma \right]$$

and

$$\begin{aligned} w(\sigma, \eta) &:= E \left[ \left( \frac{\sigma^{\eta, k+1} - \sigma^{\eta, k}}{\eta} \right)^2 \middle| \sigma^{\eta, k} = \sigma \right] - E \left[ \frac{\sigma^{\eta, k+1} - \sigma^{\eta, k}}{\eta} \middle| \sigma^{\eta, k} = \sigma \right]^2 \\ &= Var \left[ \frac{\sigma^{\eta, k+1} - \sigma^{\eta, k}}{\eta} \middle| \sigma^{\eta, k} = \sigma \right]. \end{aligned}$$

Here,  $Var \left[ \frac{\sigma^{\eta, k+1} - \sigma^{\eta, k}}{\eta} \middle| \sigma^{\eta, k} = \sigma \right]$  denotes the  $n$ -dimensional variance-covariance matrix of the random variable  $\frac{1}{\eta} (\sigma^{\eta, k+1} - \sigma^{\eta, k})$  conditional on the event that the state of the game  $\sigma^{\eta, k} = \sigma$ . Similarly to condition (a.1), the conditions require that there exist functions  $v(\cdot)$  and  $w(\cdot)$  that approximate  $v(\cdot, \eta)$  and  $w(\cdot, \eta)$  when  $\eta$  is small. Since for our adjustment process functions  $v(\cdot, \eta)$  and  $w(\cdot, \eta)$  are, again, independent of  $\eta$ , conditions (a.2) and (a.3) are satisfied, respectively.

Norman's conditions (b.1), (b.2), and (b.3) require  $v$  to be differentiable, the derivative of  $v$  to be bounded, and the derivative of  $v$  to be Lipschitz continuous, respectively. Condition (b.4) states that the function  $w$  should be Lipschitz continuous as well. These conditions hold since the functions  $v$  and  $w$  are polynomial functions with a compact domain.

Finally, Norman's condition (c) requires the function  $r : \mathbb{R}^{\Sigma_{i \in N} |S_i|} \rightarrow \mathbb{R}$ , defined by

$$r(\sigma) := E \left[ \left| \frac{\sigma^{\eta, k+1} - \sigma^{\eta, k}}{\eta} \right|^3 \middle| \sigma^{\eta, k} = \sigma \right],$$

where

$$|x|^3 := \sum_{l=1}^{\Sigma_{i \in N} |S_i|} |x_l|^3$$

for  $x \in \mathbb{R}^{\Sigma_{i \in N} |S_i|}$ , to be bounded from above. Condition (c) is satisfied because the function  $r$  is a polynomial function with a compact domain as well.

Norman's theorem ensures that the expectations of the random variables  $\sigma^{\eta, k}$  converge to the expectation of  $\hat{\sigma}^t$  as  $\eta \rightarrow 0$  and  $k\eta \rightarrow t$ . Since convergence of expectations implies convergence in probability this concludes the proof.  $\square$

According to Proposition 1, if  $\eta$  and  $k\eta$  are close to 0 and  $t$ , respectively, then, with large probability,  $\sigma^{\eta, k}$  will be close to  $\hat{\sigma}^t$ . Thus, frequent play and slow adjustment make it possible to apply a law of large numbers argument which states that actual adjustment and expected movement of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  coincide. Consequently, the behavior predicted by the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  in finite time can to some extent be analyzed by means of the dynamical system (4).

The first step in an analysis of the dynamical system (4) consists of the identification of the steady states. For this purpose, we introduce the notion of best-reply matching equilibrium in finite normal-form games.

**Definition 2** Let  $G = \langle N, (S_i)_{i \in N}, (\Sigma_i)_{i \in N} \rangle$  be a game. A mixed-strategy profile  $\sigma \in \Delta$  is a best-reply matching equilibrium if for every player  $i \in N$  and  $s_i \in S_i$  it holds that

$$\sigma_i(s_i) = \sum_{\{s_{-i} \in S_{-i} \mid s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \sigma_{-i}(s_{-i}). \quad (5)$$

The set of best-reply matching equilibria of a game  $G$  is denoted by  $BRM(G)$ .

Definition 2 says that in a best-reply matching equilibrium, every player  $i \in N$  plays a pure strategy  $s_i \in S_i$  with a probability which is strongly related to the probability that  $s_i$  is a best reply to a pure-strategy profile of his opponents. In case the pure strategy  $s_i$  is a unique best reply, the probability that player  $i$  assigns to it equals the probability that  $s_i$  is a best reply to a pure-strategy profile of player  $i$ 's opponents. As becomes clear from expression (5), this equality no longer holds true in case of multiple best replies. Namely, player  $i$  will then equally distribute the probability that a pure-strategy profile of his opponents is realized over all best replies to that profile.

Droste *et al.* (2000) contains an extensive treatment of best-reply matching equilibria, showing, among other things, that for every finite game  $G$  the set of best-reply matching equilibria  $BRM(G)$  is nonempty, and that the set of pure-strategy best-reply matching equilibria coincides with the set of strict Nash equilibria of a game. Moreover, for two-player games the dimension of the (compact, convex) set  $BRM(G)$  is equal to the number of minimal curb sets of the game minus one. The new equilibrium notion is furthermore illustrated by means of several well-known games.

For example, in a  $2 \times 2$  coordination game like the Battle of the Sexes (see Figure 1), the set of best-reply matching equilibria can be calculated as  $BRM(G) = \{\sigma \in \Delta \mid \sigma_1(A) = \sigma_2(A)\}$ , meaning that a strategy profile is a best-reply matching equilibrium if and only if both players play the same mixed strategy. Notice that the two pure strategy best-reply matching equilibria  $(A, A)$  and  $(B, B)$  are exactly the strict Nash equilibria of the game. There exists another Nash equilibrium in mixed strategies, which depending on the payoffs of the game is symmetric or not. If, e.g.,  $a < b$ , then this Nash equilibrium is not symmetric and hence not an element of  $BRM(G)$ . The reason is intuitive. Equalizing expected payoffs over both pure strategies leads the row-player to put more probability on strategy B than she believes the other player does. At the same time, the column-player puts more probability on A than he believes the other player does. This is not in accordance with matching behavior. In a best-reply matching equilibrium a player wants to do exactly the same as the opponent. If one chooses A with some probability, the other will choose A with the same probability.

FIGURE 1

While other games, like Matching Pennies, possess a unique best-reply matching equilibrium, in case of a coordination game the set of best-reply matching equilibria is rather large. This has, of course, an effect on the predictive power of the equilibrium concept. However, as we will see in Section 6, it is exactly the adjustment process of the present model that leads to a refinement of the equilibrium based on its asymptotic prediction. For more details concerning best-reply matching in games we refer to the above mentioned paper.<sup>2</sup>

Proposition 3 gives a motivation for the concept of best-reply matching equilibrium by stating that the steady states of the dynamical system (4) are exactly the best-reply matching equilibria of the underlying normal-form game.

**Proposition 3** *Consider a game  $G = \langle N, (S_i)_{i \in N}, (\succeq_i)_{i \in N} \rangle$ . Then the set of steady states of the dynamical system (4), which is characterized by  $\dot{\sigma}_i(s_i) := \frac{d\hat{\sigma}_i^t(s_i)}{dt} = 0$  for all  $i \in N$  and  $s_i \in S_i$ , coincides with  $BRM(G)$ , i.e., the set of best-reply matching equilibria of  $G$ .*

**Proof.** Proposition 3 follows immediately from comparing the conditions for a steady state of the dynamical system (4) and Definition 2.  $\square$

We next come to the second step in an analysis of the dynamical system (4), which consists of a stability analysis of the steady states.

## 5 Stability Analysis

To establish stability properties of single steady states or sets of steady states of the continuous time limit introduced in Section 4, we use both the eigenvalue analysis of the Jacobian and the direct Lyapunov method.

The eigenvalue analysis is based on the well-known result that a steady state of a linear system of differential equations is stable if the real parts of the eigenvalues of the Jacobian, evaluated at the steady state, are nonpositive and every eigenvalue, which has a zero real part, has single multiplicity. Furthermore, a steady state is asymptotically stable if and only if all eigenvalues have negative real parts. Finally, a steady state is unstable if there exists an eigenvalue of the Jacobian with a positive real part. With respect to the direct Lyapunov method we use Theorem 6.4 in Weibull (1995).

For general finite normal-form games  $G$  it is not possible to derive explicit results with respect to the stability properties of steady states of the dynamical system (4). For that reason, we consider  $2 \times 2$  bimatrix games and coordination games.<sup>3</sup>

<sup>2</sup>Copies can be obtained from the corresponding author (kosfeld@iew.unizh.ch).

<sup>3</sup>In addition to our results a general convergence result is available for two-player games, showing that the set of best-reply matching equilibria is globally asymptotically stable for the dynamical system (4). See appendix.

### 5.1. $2 \times 2$ Bimatrix Games

A  $2 \times 2$  bimatrix game is a two-player normal-form game where the pure-strategy set of both players consists of two elements. Denoting  $S_1 = \{T, B\}$  and  $S_2 = \{L, R\}$ , the dynamical system (4) corresponding to a  $2 \times 2$  bimatrix game becomes

$$\begin{aligned}\dot{\hat{\sigma}}_1(T) &= \theta \left( \left( \sum_{\{s_2 \in S_2 | T \in B_1(s_2)\}} \frac{1}{|B_1(s_2)|} \hat{\sigma}_2^t(s_2) \right) - \hat{\sigma}_1^t(T) \right) \\ \dot{\hat{\sigma}}_2(L) &= \theta \left( \left( \sum_{\{s_1 \in S_1 | L \in B_2(s_1)\}} \frac{1}{|B_2(s_1)|} \hat{\sigma}_1^t(s_1) \right) - \hat{\sigma}_2^t(L) \right),\end{aligned}\tag{6}$$

for all  $t \in \mathbb{R}_+$ . The equations for  $\dot{\hat{\sigma}}_1(B)$  and  $\dot{\hat{\sigma}}_2(R)$  are redundant since  $\hat{\sigma}_1^t(B) = 1 - \hat{\sigma}_1^t(T)$  and  $\hat{\sigma}_2^t(R) = 1 - \hat{\sigma}_2^t(L)$ . Let  $\hat{\sigma} \in \Delta$  denote a best-reply matching equilibrium of a  $2 \times 2$  bimatrix game, or equivalently a steady state of the dynamical system (6). Proposition 4 gives the stability properties of a best-reply matching equilibrium  $\hat{\sigma} \in \Delta$ .

**Proposition 4** *Let  $G = \langle N, (S_i)_{i \in N}, (\succeq_i)_{i \in N} \rangle$  be a  $2 \times 2$  bimatrix game. Then every best-reply matching equilibrium  $\hat{\sigma} \in \Delta$  is Lyapunov stable under the dynamical system (6). If the game  $G$  is not a coordination or a Chicken game, then every best-reply matching equilibrium is even asymptotically stable.*

**Proof.** To prove Proposition 4, we use the eigenvalue analysis. The Jacobian  $\mathbf{J}_{\hat{\sigma}^t}$  of the dynamical system (6) is given by

$$\mathbf{J}_{\hat{\sigma}^t} = \begin{bmatrix} \frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_1^t(T)} & \frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)} \\ \frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)} & \frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_2^t(L)} \end{bmatrix} = \begin{bmatrix} -\theta & \frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)} \\ \frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)} & -\theta \end{bmatrix}.$$

Consider the entry  $\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)}$ . Depending on the best-reply structure of the game  $G$  this entry equals  $-\theta$ ,  $-\frac{\theta}{2}$ ,  $0$ ,  $\frac{\theta}{2}$ , or  $\theta$ . To illustrate why this holds, we consider the case where T is the unique best reply to L and both T and B are best replies to R. For this best-reply structure it holds that the first equation of the dynamical system (6) equals

$$\begin{aligned}\dot{\hat{\sigma}}_1(T) &= \theta \left( \hat{\sigma}_2^t(L) + \frac{1}{2} \hat{\sigma}_2^t(R) - \hat{\sigma}_1^t(T) \right) \\ &= \theta \left( \hat{\sigma}_2^t(L) + \frac{1}{2} (1 - \hat{\sigma}_2^t(L)) - \hat{\sigma}_1^t(T) \right).\end{aligned}$$

It follows immediately that in this case the entry  $\frac{d\dot{\hat{\sigma}}_1(T)}{d\hat{\sigma}_2^t(L)} = \frac{\theta}{2}$ . Considering the other eight possible best-reply structures of a  $2 \times 2$  bimatrix game then gives the required result. Similarly, it can easily be shown that the entry  $\frac{d\dot{\hat{\sigma}}_2(L)}{d\hat{\sigma}_1^t(T)}$  also equals  $-\theta$ ,  $-\frac{\theta}{2}$ ,  $0$ ,  $\frac{\theta}{2}$ , or  $\theta$ .



We are left with the actual eigenvalue analysis of the Jacobian  $\mathbf{J}_{\hat{\sigma}^t}$ . In determining the eigenvalues we distinguish five cases. First, let either  $\frac{d\dot{\hat{\sigma}}_1(\text{T})}{d\dot{\hat{\sigma}}_2(\text{L})}$  or  $\frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})}$  be equal to 0. This implies that the Jacobian  $\mathbf{J}_{\hat{\sigma}^t}$  is triangular and that the eigenvalues are therefore given by the entries on the diagonal, i.e.,  $\lambda_{1,2} = -\theta < 0$ . Consequently, a steady state  $\hat{\sigma} \in \Delta$  is asymptotically stable.

Second, let  $\frac{d\dot{\hat{\sigma}}_1(\text{T})}{d\dot{\hat{\sigma}}_2(\text{L})} = -\theta$ . In this case the eigenvalues of the Jacobian  $\mathbf{J}_{\hat{\sigma}^t}$  are given by

$$\lambda_{1,2} = -\theta \pm \sqrt{-\theta \frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})}}.$$

When  $\frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})} = -\theta$ , we find that  $\lambda_1 = -2\theta$  and  $\lambda_2 = 0$  which implies a steady state  $\sigma \in \Delta$  to be Lyapunov stable, but not asymptotically stable. In all other cases, i.e.,  $\frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})}$  equal to  $-\frac{\theta}{2}$ ,  $\frac{\theta}{2}$ , or  $\theta$ , it holds that  $\text{Re}(\lambda_{1,2}) < 0$ , which implies that a steady state  $\hat{\sigma} \in \Delta$  is asymptotically stable.

Third, let  $\frac{d\dot{\hat{\sigma}}_1(\text{T})}{d\dot{\hat{\sigma}}_2(\text{L})} = -\frac{\theta}{2}$ . A straightforward calculation shows that the resulting eigenvalues equal

$$\lambda_{1,2} = -\theta \pm \sqrt{-\frac{\theta}{2} \frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})}}.$$

Furthermore, it follows immediately that for all possible values of  $\frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})}$  it holds that  $\text{Re}(\lambda_{1,2}) < 0$ , i.e., a steady state  $\hat{\sigma} \in \Delta$  will be asymptotically stable.

Fourth, let  $\frac{d\dot{\hat{\sigma}}_1(\text{T})}{d\dot{\hat{\sigma}}_2(\text{L})} = \frac{\theta}{2}$ . In this case the eigenvalues are given by

$$\lambda_{1,2} = -\theta \pm \sqrt{\frac{\theta}{2} \frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})}}.$$

Again, this implies that  $\text{Re}(\lambda_{1,2}) < 0$  for all possible values of  $\frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})}$ . Consequently, a steady state  $\hat{\sigma} \in \Delta$  is asymptotically stable.

Finally, let  $\frac{d\dot{\hat{\sigma}}_1(\text{T})}{d\dot{\hat{\sigma}}_2(\text{L})} = \theta$ . This gives rise to the following eigenvalues

$$\lambda_{1,2} = -\theta \pm \sqrt{\theta \frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})}}.$$

From the above expression it can easily be deduced that  $\text{Re}(\lambda_{1,2}) < 0$  and that a steady state  $\hat{\sigma} \in \Delta$  is asymptotically stable, except when  $\frac{d\dot{\hat{\sigma}}_2(\text{L})}{d\dot{\hat{\sigma}}_1(\text{T})} = \theta$ . In the latter case, we find that  $\lambda_1 = -2\theta$  and  $\lambda_2 = 0$  which implies a steady state  $\hat{\sigma} \in \Delta$  to be Lyapunov stable, but not asymptotically stable.

Thus, we find that if either

$$\frac{d\hat{\sigma}_1(\text{T})}{d\hat{\sigma}_2^t(\text{L})} = \frac{d\hat{\sigma}_2(\text{L})}{d\hat{\sigma}_1^t(\text{T})} = \theta \quad \text{or} \quad \frac{d\hat{\sigma}_1(\text{T})}{d\hat{\sigma}_2^t(\text{L})} = \frac{d\hat{\sigma}_2(\text{L})}{d\hat{\sigma}_1^t(\text{T})} = -\theta, \quad (7)$$

the equilibrium is Lyapunov stable. In all other cases, the equilibrium is even asymptotically stable. Clearly, condition (7) is equivalent to the game being a coordination game or a Chicken game, respectively.  $\square$

Proposition 4 determines the stability of a best-reply matching equilibrium in  $2 \times 2$  bimatrix games. As shown in the appendix, for general two-player games a further result is available. In this case, the set of best-reply matching equilibria is globally asymptotically stable under the dynamical system (4) and each equilibrium is Lyapunov stable.<sup>4</sup> This result reveals a great difference between the dynamics (4) and the standard best-response dynamics, where players play a myopic best response to the empirical distribution of the opponent's play. It is known, for example, that the best-response dynamics need not converge to Nash equilibrium in games that are of a structure like Shapley's  $3 \times 3$  game, i.e., where players use more than two strategies in the mixed Nash equilibrium (Shapley, 1964; Krishna and Sjöström, 1998). In contrast, the present result implies that also in these games the dynamics (4) converge to a best-reply matching equilibrium of the game.

For the special case of two-player coordination games we can prove asymptotic stability of the set of best-reply matching equilibria using the direct Lyapunov method.

## 5.2. Coordination Games

A two-player game is a coordination game if both players  $i \in N = \{1, 2\}$  have the same set of pure strategies, i.e.,  $S_1 = S_2 =: X$ , and the unique best reply to a pure strategy of the opponent is to play the same pure strategy. In the case of a two-player coordination game the dynamical system (4) therefore reduces to

$$\dot{\hat{\sigma}}_i(s_i) = \theta \left( \hat{\sigma}_j^t(s_i) - \hat{\sigma}_i^t(s_i) \right) \quad (8)$$

for all  $t \in \mathbb{R}_+$ ,  $i, j \in \{1, 2\}$ ,  $j \neq i$ , and  $s_i \in X$ . It follows directly from the dynamical system (8) and Proposition 3 that the set of best-reply matching equilibria  $BRM(G)$  of a two-player coordination game is given by  $\{\hat{\sigma} \in \Delta \mid \hat{\sigma}_1(s_i) = \hat{\sigma}_2(s_i) \text{ for all } s_i \in X\}$ . Using the direct Lyapunov method, Proposition 5 gives the stability properties of the set of best-reply matching equilibria and each single best-reply matching equilibrium of a two-player coordination game.

---

<sup>4</sup>Note that in most  $2 \times 2$  bimatrix games Proposition 4 guarantees even asymptotic stability of a single equilibrium.

**Proposition 5** *Let  $G = \langle N, (S_i)_{i \in N}, (\succeq_i)_{i \in N} \rangle$  be a two-player coordination game. Then the set of best-reply matching equilibria  $BRM(G)$  is asymptotically stable under the dynamical system (8). Furthermore, each single best-reply matching equilibrium  $\hat{\sigma} \in \Delta$  is Lyapunov stable under the dynamical system (8).*

**Proof.** First, consider the set of all steady states of the dynamical system (8), i.e., let  $A = \{\hat{\sigma} \in \Delta \mid \hat{\sigma}_1(s_i) = \hat{\sigma}_2(s_i) \text{ for all } s_i \in X\}$ . It is easily seen that the set  $A$  is closed. Let the Lyapunov function  $v : \mathbb{R}^{|X|} \times \mathbb{R}^{|X|} \rightarrow \mathbb{R}_+$  be given by

$$v(\hat{\sigma}) = \frac{1}{2} \sum_{s_i \in X} (\hat{\sigma}_2(s_i) - \hat{\sigma}_1(s_i))^2.$$

Obviously,  $v(\hat{\sigma})$  is continuously differentiable and  $v(\hat{\sigma}) = 0$  if and only if  $\hat{\sigma} \in A$ . Furthermore, it holds that

$$\sum_{i \in N} \sum_{s_i \in X} \frac{dv(\hat{\sigma})}{d\hat{\sigma}_i(s_i)} \dot{\hat{\sigma}}_i(s_i) = -2\theta \sum_{s_i \in X} (\hat{\sigma}_2(s_i) - \hat{\sigma}_1(s_i))^2 < 0$$

for all  $\hat{\sigma} \notin A$ . According to the direct Lyapunov method, the set  $A$  is asymptotically stable.

Second, let  $A$  be the closed and connected set consisting of a single steady state, say  $(\sigma_1, \sigma_1) \in \Delta$ , of the dynamical system (8), i.e.,  $A = \{(\sigma_1, \sigma_1)\}$ . Let the Lyapunov function  $v : \mathbb{R}^{|X|} \times \mathbb{R}^{|X|} \rightarrow \mathbb{R}_+$  be given by

$$v(\hat{\sigma}) = \frac{1}{2} \sum_{s_i \in X} (\hat{\sigma}_1(s_i) - \sigma_1(s_i))^2 + (\hat{\sigma}_2(s_i) - \sigma_1(s_i))^2.$$

Again, it is obvious that  $v(\hat{\sigma})$  is continuously differentiable,  $v(\hat{\sigma}) = 0$  if and only if  $\hat{\sigma} \in A$ , and

$$\sum_{i \in N} \sum_{s_i \in X} \frac{dv(\hat{\sigma})}{d\hat{\sigma}_i(s_i)} \dot{\hat{\sigma}}_i(s_i) = -\theta \sum_{s_i \in X} (\hat{\sigma}_2(s_i) - \hat{\sigma}_1(s_i))^2 \leq 0$$

for all  $\hat{\sigma} \notin A$ . Consequently, according to the direct Lyapunov method, each single steady state of the dynamical system (8) is Lyapunov stable.  $\square$

Notice that the Lyapunov function  $v$  as used in the proof of Proposition 5 is not only properly defined on a small neighborhood of the set of best-reply matching equilibria, but even on the entire state space  $\Delta$ . Hence, the set of best-reply matching equilibria is globally asymptotically stable, meaning that convergence to that set will be observed independent of the initial state of the dynamical system (8).

Before turning to the stochastic adjustment process again, it is worth mentioning that an analogous result to Proposition 5 can be proven for the  $(2 \times 2)$  game Chicken. In this case, the Lyapunov function can be taken as

$$v(\hat{\sigma}) = \frac{1}{2} \sum_{s_i \in \{s_1, s_2\}} (1 - \hat{\sigma}_2(s_i) - \hat{\sigma}_1(s_i))^2.$$

## 6 Asymptotic Behavior

This section deals with the asymptotic long-run behavior of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  as  $k \rightarrow \infty$ . We specify asymptotic behavior and thereby show that it may be different from the behavior in finite time as discussed in Sections 4 and 5.

We closely follow Norman's (1968, 1972) work on distance-diminishing Markov processes. To study the asymptotic behavior, we need the following notation. Recall that the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  is a discrete-time Markov process with infinite compact state space  $\Delta$ . For  $\sigma \in \Delta$  and Borel set  $A \subset \Delta$  we denote the transition probabilities of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  in the first period by the stochastic kernel  $K(\sigma, A) = \Pr(\sigma^1 \in A \mid \sigma^0 = \sigma)$ . Analogously, the transition probabilities of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  over the first  $k$  periods are given by the stochastic kernel  $K^k(\sigma, A) = \Pr(\sigma^k \in A \mid \sigma^0 = \sigma)$ . A stochastic kernel is a function that is a probability measure in its second variable for each value of its first, and measurable in its first variable for each value of its second. We let  $K^0(\sigma, A)$  be equal to 1 if  $\sigma \in A$  and 0 otherwise. A sequence  $\{K^k\}_{k \in \mathbb{N}_0}$  of stochastic kernels is said to converge uniformly to a stochastic kernel  $K^\infty$  if for each Borel subset  $A$  of  $\Delta$  and  $\varepsilon > 0$ , there exists an integer  $M$  such that

$$K^\infty(\sigma, \text{int}(A)) - \varepsilon \leq K^k(\sigma, A) \leq K^\infty(\sigma, \text{cl}(A)) + \varepsilon$$

for all  $k \geq M$  and  $\sigma \in \Delta$ . In the above expression,  $\text{int}(A)$  and  $\text{cl}(A)$  denote the interior and the closure of set  $A$ .

The analysis of asymptotic behavior of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  is conducted through several results and examples. The first proposition shows uniform convergence of a stochastic kernel that is strongly related to the adjustment process.

**Proposition 6** *The stochastic kernel  $\frac{1}{k+1} \sum_{j=0}^k K^j$  of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  converges uniformly as  $k \rightarrow \infty$  to a stochastic kernel  $K^\infty$ .*

**Proof.** Theorem 2.1 in Norman (1968) is a convergence result for distance-diminishing discrete-time Markov processes with infinite state space that implies Proposition 6. It is therefore sufficient to verify that the conditions (H1)-(H8), which are imposed by Theorem 2.1 in Norman (1968), are satisfied by the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$ .

Condition (H1) states that the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  should be a Markov process, which is obviously satisfied. Condition (H2) requires that  $S_i$  is finite for all  $i \in N$ , which by assumption is true as well. Norman's condition (H3) requires  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  to be memory-less and temporally homogeneous, meaning that the probabilities of all possible pure-strategy profiles at time  $k$  depend only on the state of the game at time  $k$ , i.e.,  $\sigma^k$ , and not on earlier states

or pure-strategy profiles or even on the number of repetitions of the game. This condition is clearly satisfied.

Conditions (H4) and (H5) require the state space to be a compact metric space, which are satisfied because  $\Delta_i$  is a compact subset of  $\mathbb{R}^{|S_i|-1}$  for all  $i \in N$ . As the function  $\varphi_s : \Delta \rightarrow \mathbb{R}$  defined by

$$\varphi_s(\sigma^k) := \prod_{i=1}^N \sigma_i^k(s_i)$$

is differentiable on the state space  $\Delta$ , condition (H6) only requires the absolute value of each component of  $\frac{d\varphi_s(\sigma^k)}{d\sigma^k}$  to be finite. In fact, it can easily be verified that the absolute value of each component of  $\frac{d\varphi_s(\sigma^k)}{d\sigma^k}$  is an element of  $[0, 1]$ .

Norman's condition (H7) requires the following. Consider any time  $k$  and let  $\sigma^k$  and  $\tilde{\sigma}^k$  be two possible states of the game at that time. Consider also a fixed pure-strategy profile  $s \in S$ . Denote by  $\sigma^{k+1}$  and  $\tilde{\sigma}^{k+1}$  the states of the game that are reached if the preceding states were  $\sigma^k$  and  $\tilde{\sigma}^k$ , respectively, and pure-strategy profile  $s$  was realized at time  $k$ . (H7) requires that

$$\|\sigma^{k+1} - \tilde{\sigma}^{k+1}\| \leq \|\sigma^k - \tilde{\sigma}^k\|,$$

where  $\|\cdot\|$  denotes the Euclidean norm. A straightforward calculation shows that for the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  it holds that

$$\|\sigma^{k+1} - \tilde{\sigma}^{k+1}\| = (1 - \theta) \|\sigma^k - \tilde{\sigma}^k\|.$$

Since  $0 < \theta < 1$ , condition (H7) is clearly satisfied.

Finally, condition (H8) requires the weak inequality  $\|\sigma^{k+1} - \tilde{\sigma}^{k+1}\| \leq \|\sigma^k - \tilde{\sigma}^k\|$  to be strict in certain special cases. Since  $0 < \theta < 1$ , the inequality is always strict with respect to the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$ . Norman's condition (H8) is therefore automatically satisfied.  $\square$

Given an initial state of the game, the stochastic kernel  $\frac{1}{k+1} \sum_{j=0}^k K^j$  specifies for each Borel subset  $A$  of  $\Delta$  the average probability of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  being in  $A$  during the first  $k+1$  periods. Notice that Proposition 6 only asserts convergence of these average probabilities and is therefore not meant to suggest that there exists a mixed-strategy profile  $\sigma^\infty$  to which  $\sigma^k$  converges with probability 1 as  $k \rightarrow \infty$ . Though such convergence may very well occur, it does not occur in general, as we will see below.

To study properties of the stochastic kernel  $K^\infty$ , we need some further notation. Recall that for every player  $i \in N$  and each pure-strategy profile  $s_{-i} \in S_{-i}$  of his opponents, we denote the set of best replies by

$$B_i(s_{-i}) = \{s_i \in S_i \mid \exists \tilde{s}_i \in S_i : (\tilde{s}_i, s_{-i}) \succ_i (s_i, s_{-i})\}.$$

Let  $B(s) = \prod_{i \in N} B_i(s_{-i})$  and for every set  $C = \prod_{i \in N} C_i \subset S$  let  $B_i(C) = \cup_{s \in C} B_i(s_{-i})$  and  $B(C) = \prod_{i \in N} B_i(C)$ . A nonempty set  $C \subset S$  is said to be closed under best replies, or equivalently a curb set, if  $B(\prod_{i \in N} C_i) \subset C$ . Curb is short for closed under rational behavior. A set is curb if all best replies against all these pure strategies are contained in the set. Such a set  $C$  is called a minimal curb set if it does not properly contain a curb set. It is straightforward to show that  $B(\prod_{i \in N} C_i) = C$  for every minimal curb set. For a game  $G$  let  $\mathcal{C}(G)$  denote the set of minimal curb sets of that game. Since the joint pure-strategy set  $S$  is always a curb set, every game  $G$  has at least one minimal curb set, i.e.,  $\mathcal{C}(G) \neq \emptyset$ . The notion of curb set was introduced by Basu and Weibull (1991).

In addition, for every nonempty set  $C_i \subset S_i$  let  $\Delta_i(C_i) \subset \Delta_i$  be the face of the simplex  $\Delta_i$  spanned by  $C_i$ , i.e.,

$$\Delta_i(C_i) = \{\sigma_i \in \Delta_i \mid \text{supp}(\sigma_i) \subset C_i\},$$

where  $\text{supp}(\sigma_i) = \{s_i \in S_i \mid \sigma_i(s_i) > 0\}$  is the support of  $\sigma_i \in \Delta_i$ . Likewise, for every collection  $C = \prod_{i \in N} C_i \subset S$  of nonempty sets  $C_i \subset S_i$ , one for every player  $i \in N$ , we let the closed and convex set  $\Delta(C)$  be the face of the mixed-strategy set  $\Delta$  spanned by  $C$ , i.e.,  $\Delta(C) = \prod_{i \in N} \Delta_i(C_i)$ . Finally, for a set of mixed strategies  $A_i \subset \Delta_i$  of player  $i$  let  $\text{supp}(A_i) = \{s_i \in S_i \mid \sigma_i(s_i) > 0 \text{ for some } \sigma_i \in A_i\}$  be the support of  $A_i$ . For  $A = \prod_{i \in N} A_i$  with nonempty  $A_i \subset \Delta_i$ , let  $\text{supp}(A) = \prod_{i \in N} \text{supp}(A_i)$ .

Our next results characterize the relation between curb sets and absorbing sets of the adjustment process. We then prove absorption of the process in minimal curb sets of the game (Proposition 9). Finally, we look at absorbing states of the process (Proposition 10 and 11). A Borel set  $A \subset \Delta$  is an absorbing set of the adjustment process, if for all  $\sigma \in A$  and all  $k \in \mathbb{N}$ ,  $K^k(\sigma, A) = 1$ . An absorbing state is a singleton absorbing set.

**Lemma 7** *If  $C$  is a curb set,  $\Delta(C)$  is an absorbing set. On the other hand, if  $A$  is an absorbing set,  $\text{supp}(A)$  is a curb set.*

**Proof.** The first claim of Lemma 7 is obvious as the dynamics of individual strategy adjustment involve best replies only. The second claim is proven by contradiction. Denote  $C := \text{supp}(A)$ . Suppose there exists a player  $i \in N$  with strategy  $s_i \in S_i$  such that  $s_i \in B_i(s_{-i})$  for some  $s_{-i} \in C_{-i}$  and  $s_i \notin C_i$ . Since  $s_{-i}$  receives positive probability in  $A$ , because of the adjustment rule (1)  $s_i$  must receive positive probability, as well. Hence,  $s_i \in C_i$ , yielding the contradiction.  $\square$

Because the full strategy space  $\Delta$  is also an absorbing set, we are rather interested in a subcollection of these sets, i.e., the collection of minimal closed absorbing sets. A Borel set  $A \subset \Delta$  is a minimal closed absorbing set, if it is (topologically) closed, absorbing, and does

not properly contain a closed absorbing subset. The existence of a minimal closed absorbing set is guaranteed by compactness of the space  $\Delta$ . This follows from an application of Zorn's lemma and is proven, e.g., in Norman (1972, Lemma 4.1, p.52).<sup>5</sup> An absorbing state is obviously a minimal closed absorbing set.

**Lemma 8** *Let  $A$  be a minimal closed absorbing set. Then  $A \subset \Delta(C)$ , where  $C$  is a minimal curb set.*

**Proof.** By Lemma 7,  $\text{supp}(A)$  is a curb set. We have to show that  $\text{supp}(A)$  is indeed a minimal curb set, which is proven by contradiction. Suppose there exists a subset  $D \subset \text{supp}(A)$ ,  $D \neq \text{supp}(A)$ , that is curb. By Lemma 7,  $\Delta(D)$  is absorbing. Consider  $\Delta(D) \cap A$ . Because the adjustment process is distance-diminishing (conditions (H6) to (H8) in Norman, 1968; see also the proof of Proposition 6),  $\Delta(D) \cap A$  cannot be empty. More precisely, let  $\sigma^0 \in \Delta(D)$  and  $\tilde{\sigma}^0 \in A$  be two initial starting points of play, such that  $\sigma^0 \neq \tilde{\sigma}^0$  (otherwise  $\sigma^0 = \tilde{\sigma}^0 \in \Delta(D) \cap A$ ), and both  $\sigma^0(s^0) > 0$  and  $\tilde{\sigma}^0(s^0) > 0$  for some outcome  $s^0 \in D$ . Such an outcome and hence such profiles exist because  $D \subset \text{supp}(A)$ . Define  $\sigma^1$  and  $\tilde{\sigma}^1$  as the corresponding adjusted mixed strategy profiles based on the learning rule (1) and the realization of  $s^0$ . I.e. for each  $i \in N$ ,  $\sigma_i^1$  is defined as

$$\sigma_i^1(s_i) = \begin{cases} (1 - \theta) \sigma_i^0(s_i) + \frac{\theta}{|B_i(s_{-i}^0)|} & \text{if } s_i \in B_i(s_{-i}^0), \\ (1 - \theta) \sigma_i^0(s_i) & \text{otherwise,} \end{cases}$$

and  $\tilde{\sigma}^1$  is defined respectively. Now, iterate this procedure. For every  $k \in \mathbb{N}$  choose an outcome  $s^k \in D$  such that  $\sigma^k(s^k) > 0$  and  $\tilde{\sigma}^k(s^k) > 0$ , and define  $\sigma^{k+1}$  and  $\tilde{\sigma}^{k+1}$  as the corresponding adjusted mixed strategy profiles after realization of  $s^k$ . Hence, for every  $i \in N$

$$\sigma_i^{k+1}(s_i) = \begin{cases} (1 - \theta) \sigma_i^k(s_i) + \frac{\theta}{|B_i(s_{-i}^k)|} & \text{if } s_i \in B_i(s_{-i}^k), \\ (1 - \theta) \sigma_i^k(s_i) & \text{otherwise,} \end{cases}$$

and  $\tilde{\sigma}^{k+1}$  is defined respectively. Again, this is possible for every  $k$  because  $D \subset \text{supp}(A)$  and both  $\Delta(D)$  and  $A$  are absorbing. Sequences  $(\sigma^k)_{k \in \mathbb{N}_0}$  and  $(\tilde{\sigma}^k)_{k \in \mathbb{N}_0}$  are two different paths of play generating the same sequence of outcomes  $(s^k)_{k \in \mathbb{N}}$ . Because the adjustment process is distance diminishing it follows that  $\|\sigma^k - \tilde{\sigma}^k\| \rightarrow 0$  as  $k \rightarrow \infty$ . Since both  $\Delta(D)$  and  $A$  are closed, the limit points lie in  $\Delta(D) \cap A$ . Therefore,  $\Delta(D) \cap A \neq \emptyset$ . As  $\Delta(D) \cap A$  is clearly a closed absorbing set and properly contained in  $A$ ,  $A$  cannot be minimal, which yields the contradiction.  $\square$

<sup>5</sup>Norman (1972) uses the terminology of an ergodic kernel to refer to a minimal closed absorbing set.

To see that faces of minimal curb sets are not necessarily minimal closed absorbing sets themselves, consider the game represented in Figure 2. The state  $\sigma = \left(\left(\frac{1}{2}, \frac{1}{2}\right), (1, 0)\right)$  is the unique absorbing state of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  that is associated with that game. (Cf. Proposition 10, the absorbing state is also the unique best-reply matching equilibrium of the game.) At the same time  $\{T, B\} \times \{L\}$  is the unique minimal curb set of the game.

FIGURE 2

By compactness the number of minimal closed absorbing sets is at least one. For the class of so-called compact Markov processes Norman (1972, Section 3.4) shows that the number is also finite. Moreover, for the same class of processes he proves convergence to the union of minimal closed absorbing sets. As the adjustment process is shown to be compact we can apply these results to our case. Denote  $E = \cup_{l=1}^d A_l$  the union of minimal closed absorbing sets of the adjustment process. Clearly,  $E$  is a Borel set. Lemma 8 guarantees that  $E$  is contained in the union of the faces of minimal curb sets. Hence, using Norman's convergence result we obtain convergence of play to the minimal curb sets of a game.

**Proposition 9** *Consider a game  $G = \langle N, (S_i)_{i \in N}, (\succeq_i)_{i \in N} \rangle$ . Then for each initial state  $\sigma \in \Delta$  it holds that*

$$\sum_{C \in \mathcal{C}(G)} K^\infty(\sigma, \Delta(C)) = 1.$$

**Proof.** Following Chapter 3 of Norman (1972) a Markov process is compact if it has a compact metric state space and the corresponding transition operator  $U$ , defined by

$$Uf(\sigma) = \int_{\Delta} K(\sigma, d\tilde{\sigma})f(\tilde{\sigma}),$$

is a so-called Doeblin-Fortet operator on the space of real-valued bounded Lipschitz functions on  $\Delta$ . The latter condition puts some bound on  $U$  as well as on the powers  $U^k$  that correspond to subsequent periods of the process controlled by kernels  $K^k$ ,  $k \geq 1$  (cf. Definition 1.2, Norman, 1972, p.35). In particular, Theorem 1.2 in Chapter 2 of Norman (1972) shows that the transition operator of any distance diminishing Markov process is a Doeblin-Fortet operator. Hence, the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  is compact.

Theorem 4.1 and 4.2 in Section 3.4 of Norman (1972) show that the following holds for a compact Markov process. To every minimal closed absorbing set  $A_l$  ( $l = 1, \dots, d$ ) there exists a unique distribution  $\mu_l$  that has  $A_l$  as its support, i.e.,  $A_l$  is the smallest closed set with  $\mu_l(A_l) = 1$ . Each of these distributions is stationary, i.e.,

$$\int_{\Delta} \mu_l(d\sigma)K(\sigma, B) = \mu_l(B)$$



for any Borel set  $B$ . Moreover, for all  $\sigma \in \Delta$  and Borel sets  $B$

$$K^\infty(\sigma, B) = \sum_{l=1}^d g_l(\sigma) \mu_l(B), \quad (9)$$

where the  $g_l$  are functions on  $\Delta$  which satisfy

$$\sum_{l=1}^d g_l(\sigma) = 1 \quad (10)$$

for all  $\sigma \in \Delta$ . For Borel set  $E = \cup_{l=1}^d A_l$ , (9) and (10) imply that  $K^\infty(\sigma, E) = 1$ . Using Lemma 8 from above the proposition follows,

$$\sum_{C \in \mathcal{C}(G)} K^\infty(\sigma, \Delta(C)) \geq K^\infty\left(\sigma, \bigcup_{C \in \mathcal{C}(G)} \Delta(C)\right) \geq K^\infty(\sigma, E) = 1.$$

□

Together with Proposition 6, Proposition 9 implies that play according to the adjustment process is asymptotically absorbed in a minimal curb set. In case the game has several minimal curb sets, Proposition 9 typically does not specify which curb set will characterize asymptotic behavior. In fact, if  $\sigma \in \text{int}(\Delta)$ , then every face  $\Delta(C)$ , with  $C \in \mathcal{C}(G)$ , has a positive probability of characterizing asymptotic behavior. This follows, e.g., from applying the analysis in Section 7.2 of Bush and Mosteller (1955) (or similarly the results in Section 3.4 in Norman, 1972) to the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$ .

The next proposition describes the absorbing states of the adjustment process.

**Proposition 10** *Consider a game  $G = \langle N, (S_i)_{i \in N}, (\succeq_i)_{i \in N} \rangle$ . Then every pure-strategy best-reply matching equilibrium of  $G$  is an absorbing state of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$ . Furthermore, every absorbing state of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  is a best-reply matching equilibrium, but possibly not in pure strategies.*

**Proof.** As a pure-strategy best-reply matching equilibrium is a singleton minimal curb set the first statement follows from Lemma 7. To prove the second claim, suppose that  $\sigma$  is an absorbing state. Then for  $i \in N$ ,  $s_i \in S_i$  either

$$\sigma_i(s_i) = (1 - \theta)\sigma_i(s_i) + \frac{\theta}{|B_i(s_{-i})|} \quad (11)$$

for some  $s_{-i} \in S_{-i}$ , or

$$\sigma_i(s_i) = (1 - \theta)\sigma_i(s_i). \quad (12)$$

Equation (12) implies that  $\sigma_i(s_i) = 0$ . Equation (11), which can arise only if  $\sigma_i(s_i) > 0$ , implies that

$$\sigma_i(s_i) = \frac{1}{|B_i(s_{-i})|} \quad (13)$$

for each  $s_{-i}$  that is played with positive probability. Thus, an absorbing state is where every player always plays a best-reply, and in case of multiple best-replies each action receives the same weight. If best replies are unique,  $\sigma$  is a best-reply matching equilibrium in pure strategies. If  $\sigma$  is an absorbing state that puts positive probability on more than one action for at least one player  $i \in N$ , it still holds for every player  $i \in N$  and  $s_i \in S_i$  that

$$\sigma_i(s_i) = \frac{1}{|B_i(s_{-i})|} \quad \text{for all } s_{-i} \text{ such that } \sigma_{-i}(s_{-i}) > 0.$$

Hence,

$$\begin{aligned} \sigma_i(s_i) &= \frac{1}{|B_i(s_{-i})|} \sum_{\{s_{-i} | \sigma_{-i}(s_{-i}) > 0\}} \sigma_{-i}(s_{-i}) \\ &= \sum_{\{s_{-i} | \sigma_{-i}(s_{-i}) > 0\}} \frac{1}{|B_i(s_{-i})|} \sigma_{-i}(s_{-i}), \end{aligned}$$

where the last equality follows from the fact that the number of best replies must be the same for each profile  $s_{-i}$ , since  $\sigma$  is an absorbing state and therefore (13) is true. As only best replies are played in an absorbing state it follows that

$$\{s_{-i} | \sigma_{-i}(s_{-i}) > 0\} = \{s_{-i} | s_i \in B_i(s_{-i})\}$$

and therefore the right-hand side of the last equation is equal to

$$\sum_{\{s_{-i} | s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \sigma_{-i}(s_{-i}),$$

which means that  $\sigma$  is a best-reply matching equilibrium. This concludes the proof.  $\square$

**Proposition 11** *Consider a game  $G = \langle N, (S_i)_{i \in N}, (\sum_i)_{i \in N} \rangle$ . Then for every  $C \in \mathcal{C}(G)$  it holds that  $\Delta(C) \subset \Delta$  contains a best-reply matching equilibrium.*

**Proof.** For every  $i \in N$ ,  $s_i \in S_i$ , and  $\sigma_{-i} \in \Delta_{-i}$ , we define

$$r_i(s_i, \sigma_{-i}) := \sum_{\{s_{-i} \in S_{-i} | s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \sigma_{-i}(s_{-i}),$$

where the empty sum is zero by definition. Now let  $i \in N$  and  $\sigma \in \Delta(C)$ . Notice that

$$\begin{aligned}
\sum_{s_i \in C_i} r_i(s_i, \sigma_{-i}) &= \sum_{s_i \in C_i} \sum_{\{s_{-i} \in C_{-i} \mid s_i \in B_i(s_{-i})\}} \frac{1}{|B_i(s_{-i})|} \sigma_{-i}(s_{-i}) \\
&= \sum_{s_{-i} \in C_{-i}} \sum_{s_i \in B_i(s_{-i})} \frac{1}{|B_i(s_{-i})|} \sigma_{-i}(s_{-i}) \\
&= \sum_{s_{-i} \in C_{-i}} \sigma_{-i}(s_{-i}) \\
&= 1.
\end{aligned}$$

Consequently, the mapping

$$\begin{aligned}
r : \Delta(C) &\rightarrow \Delta(C) \\
\sigma &\mapsto r(\sigma)
\end{aligned}$$

with  $r(\sigma)_i(s_i) = r_i(s_i, \sigma_{-i})$  is well-defined. In the definition of the function  $r_i$  neither the index set in the summation sign nor the number  $|B_i(s_{-i})|$  of best replies depends on the mixed-strategy profile  $\sigma$ . Hence, this mapping is obviously continuous. Application of Brouwer's fixed point theorem yields the existence of a strategy profile  $\sigma \in \Delta(C)$  such that  $\sigma = r(\sigma)$ , which is a best-reply matching equilibrium.  $\square$

**Remark.** For two-player games it is possible to show that the face of a minimal curb set contains a unique best-reply matching equilibrium. Consider the function  $r$  defined in the proof of Proposition 11. In the case of a two-player game this function is linear and can be written in the following form.<sup>6</sup> For  $s_i \in S_1$  and  $s_j \in S_2$ ,

$$\begin{aligned}
r(\sigma)_1(s_i) &= \sum_j a_{ij} \sigma_2(s_j) \\
r(\sigma)_2(s_j) &= \sum_i b_{ji} \sigma_1(s_i),
\end{aligned}$$

where  $a_{ij} = 1$  if  $s_i$  is the unique best reply of player one to strategy  $s_j$  of player two,  $a_{ij} = \frac{1}{k}$  if  $s_i$  is one of  $k$  different best replies to  $s_j$ , and  $a_{ij} = 0$  if  $s_i$  is not a best reply to  $s_j$ .  $b_{ji}$  is defined respectively. Regarding  $\sigma$  as an element of  $\mathbb{R}^{|S_1|+|S_2|}$ ,  $r$  can be written in the following matrix notation:

$$\begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix} =: \Phi \tag{14}$$

with  $A = (a_{ij})$  and  $B = (b_{ji})$ . A best-reply matching equilibrium is a fixed point of  $\Phi$ , i.e. a vector  $\sigma \in \mathbb{R}^{|S_1|+|S_2|}$  such that  $\Phi\sigma = \sigma$ . Now, uniqueness follows directly from the theory of finite Markov chains. Because the transposed matrix  $\Phi^T$  is a stochastic matrix, i.e. its rows

---

<sup>6</sup>Note the close relation to Josef Hofbauer's proof in the appendix.

sum up to 1, a best-reply matching equilibrium is in fact an invariant distribution for the stochastic matrix  $\Phi^T$ . Such a distribution exists and is unique, whenever the matrix  $\Phi^T$  is irreducible, i.e. all states communicate with each other, or, in other words, the full state space is the unique communicating class. In our setting states correspond to the pure strategies of the players and two states communicate with each other if and only if there exists a sequence of best replies that connects one strategy with the other. Hence, a communicating class is a curb set and the matrix is irreducible if and only if the full state space is a minimal curb set.

The result indicates a close relation between the number of minimal curb sets of a game and the set of best-reply matching equilibria. Indeed, for two-player games it follows immediately that the dimension of the set of best-reply matching equilibria is equal to the number of minimal curb sets minus one (cf. Droste *et al.*, 2000).

We have seen that absorbing states correspond to best-reply matching equilibria which are contained in the faces of the minimal curb sets. However, it may very well be that the face of a minimal curb set does not contain an absorbing state at all. Consider, e.g., the game Matching Pennies as represented in Figure 3. The unique curb set and therefore also the unique minimal curb set of the game is given by the joint pure-strategy set  $S = \{H, T\} \times \{H, T\}$ . Due to the presence of the best-reply cycle an absorbing state of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  does not exist. As a result, the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  never settles down, i.e.,  $\sigma^k$  does not converge to a mixed-strategy profile  $\sigma^\infty$  as  $k \rightarrow \infty$ . Since the unique and asymptotically stable best-reply matching equilibrium of the Matching Pennies game is given by  $\sigma = \left( \left( \frac{1}{2}, \frac{1}{2} \right), \left( \frac{1}{2}, \frac{1}{2} \right) \right)$ , the game also indicates that asymptotic behavior of the deterministic approximation (equation (4)) and asymptotic behavior of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  may be different.

FIGURE 3

If, on the other hand, a minimal curb set  $C$  contains an absorbing state  $\sigma$  the adjustment process will converge to that state once it converges to the set  $C$ . An example for such a situation is the game given in Figure 2 with  $\sigma$  being equal to  $\left( \left( \frac{1}{2}, \frac{1}{2} \right), (1, 0) \right)$  and  $C$  equal to  $\{T, B\} \times \{L\}$ .

We conclude this section with the coordination game in Figure 4, which belongs to the class of coordination games discussed in Section 5.2. The game has two strict Nash equilibria, in which both players play the unique best reply to the opponent's pure strategy,  $(A, A)$  and  $(B, B)$ . It is easy to see that both strict Nash equilibria are minimal curb sets and that there are no other ones. Hence, Proposition 9 implies that play converges to one of the strict

Nash equilibria with probability one. Note that both Nash equilibria are absorbing states of the adjustment process and pure-strategy best-reply matching equilibria of the game (cf. Proposition 10).

FIGURE 4

The coordination game shows again that asymptotic behavior of the deterministic approximation and asymptotic behavior of the adjustment process  $\{\sigma^k\}_{k \in \mathbb{N}_0}$  may differ. To illustrate this point, consider a mixed-strategy profile that is close to the mixed-strategy best-reply matching equilibrium  $\sigma = \left( \left( \frac{1}{2}, \frac{1}{2} \right) \left( \frac{1}{2}, \frac{1}{2} \right) \right)$ . Due to the stability properties of the deterministic approximation, as discussed in Section 5.2, that mixed-strategy profile will stay close to  $\left( \left( \frac{1}{2}, \frac{1}{2} \right) \left( \frac{1}{2}, \frac{1}{2} \right) \right)$  under the dynamical system (8). In particular, it will stay in the interior of the state space in finite time. As shown above, however, asymptotic behavior of the stochastic adjustment process is characterized by one of the two strict Nash equilibria that lie on the boundary of the state space.

## 7 Conclusion

We have analyzed a myopic adjustment process that is based on a particular form of reinforcement learning where the value of a pure strategy to be reinforced depends on the fact of being a best-reply to the experienced outcome of the game. The adjustment process differs from other learning processes in the sense that no explicit belief formation takes place and that mixed strategies are not a result of any optimization procedure. Optimization takes place only on the level of pure strategies, as best replies are the alternatives players focus on. Mixed strategies are then formed by reinforcement of valuable pure strategies. Our results show that the adjustment of strategies leads to best-reply matching behavior, both in finite time and, if play settles down, also in the asymptotic limit of the process. The notion of best-reply matching has been introduced by Droste *et al.* (2000) to model boundedly rational decision-making in non-cooperative games. Best-reply matching does not conflict with rational behavior per se. In particular, if players use pure strategies then best-reply matching equilibrium coincides with strict Nash equilibrium. If strategies are mixed then best-reply matching implies in general non-rational behavior. We show that for  $2 \times 2$  bimatrix games all best-reply matching equilibria are Lyapunov stable steady states of the deterministic version of the adjustment process, and for a large subclass of these games they are also asymptotically stable. In general two-player coordination games the set of equilibria is globally asymptotically stable. These results are obtained by approximating the original stochastic process

by a dynamical system of deterministic differential equations. In the asymptotic limit the stochastic process concentrates on the minimal curb sets of the game. Moreover, absorbing states are best-reply matching equilibria, each being contained in the face of a minimal curb set. Overall we find strong support for the new equilibrium concept.

## Appendix

The following result is due to Josef Hofbauer, University of Vienna. We give his original proof, adjusted to our notation.

**Proposition 12** *For 2 person games, the set of best-reply matching equilibria is globally asymptotically stable for the adjustment process (4).*

**Proof.** For two person games the differential equation (4) is linear and can be written in the form

$$\begin{aligned}\dot{\hat{\sigma}}_1(s_i) &= \sum_j a_{ij} \hat{\sigma}_2(s_j) - \hat{\sigma}_1(s_i) \\ \dot{\hat{\sigma}}_2(s_j) &= \sum_i b_{ji} \hat{\sigma}_1(s_i) - \hat{\sigma}_2(s_j).\end{aligned}\tag{15}$$

[We take the liberty to omit the common factor  $\theta > 0$ .]

Here  $a_{ij} = 1$  if  $s_i$  is the first player's unique best response to strategy  $s_j$  of the second player,  $a_{ij} = \frac{1}{k}$  if  $s_i$  is one of  $k$  different pure best responses to  $s_j$ , and  $a_{ij} = 0$  if  $s_i$  is not a best response to  $s_j$ . Similarly for  $B = (b_{ji})$ . The matrix

$$\begin{pmatrix} 0 & A \\ B & 0 \end{pmatrix}$$

is a stochastic matrix: Its entries are nonnegative and its columns sum up to 1. Hence, by the Perron-Frobenius theorem all its eigenvalues are contained in the unit disk. The matrix corresponding to the above linear differential equation (15) reads

$$\begin{pmatrix} -I & A \\ B & -I \end{pmatrix}.$$

Its eigenvalues are obtained by shifting those of the above stochastic matrix by one unit to the left. Therefore they are contained in the disk with center  $-1$  and radius 1. So they are either zero or have negative real part. Since the solutions of (15) starting in  $\Delta$  are bounded, the theory of linear differential equations tells us that each solution in  $\Delta$  converges to an equilibrium, the set of equilibria is globally asymptotically stable for (15), and each equilibrium is Lyapunov stable.  $\square$

## References

- Basu, K., and Weibull, J.W. (1991). "Strategy Subsets Closed under Rational Behavior," *Economics Letters* **36**, 141-146.
- Benaïm, M., and Weibull, J.W. (1999). "Stochastic Micro Foundations of Deterministic Population Dynamics in Evolutionary Game Theory," mimeo, Stockholm School of Economics.
- Börgers, T., and Sarin, R. (1995). "Naive Reinforcement Learning with Endogenous Aspirations," *International Economic Review*, forthcoming.
- Börgers, T., and Sarin, R. (1997). "Learning Through Reinforcement and Replicator Dynamics," *Journal of Economic Theory* **77**, 1-14.
- Bush, R.R., and Mosteller, F. (1955). *Stochastic Models for Learning*. New York: Wiley.
- Camerer, C. and Ho, T.-H. (1999). "Experience-weighted Attraction Learning in Normal Form Games," *Econometrica* **67**, 827-874.
- Cross, J.G. (1983). *A Theory of Adaptive Economic Behavior*. Cambridge, UK: Cambridge University Press.
- Droste, E., Kosfeld, M., and Voorneveld, M. (2000). "Best-Reply Matching in Games," mimeo, Tilburg University.
- Edwards, W. (1956). "Reward Probability, Amount, and Information as Determiners of Sequential Two-Alternative Decision," *Journal of Experimental Psychology* **52**, 177-188.
- Erev, I., and Roth, A.E. (1998). "Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria," *American Economic Review* **88**, 848-881.
- Estes, W.K. (1950). "Toward a Statistical Theory of Learning," *Psychological Review* **57**, 94-107.
- Hart, S., and Mas-Colell, A. (2000). "A Simple Adaptive Procedure Leading to Correlated Equilibrium," *Econometrica*, **68**, 1127-1150.
- Hurkens, S. (1995). "Learning by Forgetful Players," *Games and Economic Behavior* **11**, 304-329.



- Krishna, V. and Sjöström, T. (1998). "On the Convergence of Fictitious Play," *Mathematics of Operations Research* **23**, 479-511.
- Norman, M.F. (1968). "Some Convergence Theorems for Stochastic Learning Models with Distance Diminishing Operators," *Journal of Mathematical Psychology* **5**, 61-101.
- Norman, M.F. (1972). *Markov Processes and Learning Models*. New York: Academic Press.
- Roth, A.E., and Erev, I. (1995). "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term," *Games and Economic Behavior* **8**, 164-212.
- Sandholm, W.H. (1999). "Evolution and Equilibrium under Inexact Information," mimeo, University of Wisconsin-Madison.
- Shapley, L.S. (1964). "Some Topics in Two-Person Games," in *Advances in Game Theory* (M. Dresher, L.S. Shapley, and A.W. Tucker, Eds.), pp. 1-28, Princeton: Princeton University Press.
- Siegel, S., and Goldstein, D.A. (1959). "Decision Making Behaviour in a Two-Choice Uncertain Outcome Situation," *Journal of Experimental Psychology* **57**, 37-42.
- Suppes, P., and Atkinson, R. (1960). *Markov Learning Models for Multiperson Interaction*. Stanford: Stanford University Press.
- Weibull, J.W. (1995). *Evolutionary Game Theory*. Cambridge, MA: MIT Press.

	A	B
A	$a, b$	$0, 0$
B	$0, 0$	$b, a$

Figure 1: Battle of the Sexes

	L	R
T	1, 1	1, 0
B	1, 1	0, 0

Figure 2: A game

	H	T
H	1, -1	-1, 1
T	-1, 1	1, -1

Figure 3: The Matching Pennies game

	A	B
A	2, 2	0, 0
B	0, 0	1, 1

Figure 4: A coordination game