Simon M. Huttegger
Brian Skyrms

Emergence of Information Transfer by Inductive Learning

Abstract. We study a simple game theoretic model of information transfer which we consider to be a baseline model for capturing strategic aspects of epistemological questions. In particular, we focus on the question whether simple learning rules lead to an efficient transfer of information. We find that reinforcement learning, which is based exclusively on payoff experiences, is inadequate to generate efficient networks of information transfer. Fictitious play, the game theoretic counterpart to Carnapian inductive logic and a more sophisticated kind of learning, suffices to produce efficiency in information transfer.

Keywords: Information, game theory, inductive logic, reinforcement learning, fictitious play.

1. Introduction and background

The traditional epistemic agent is gathering knowledge in isolation. This model of knowledge acquisition is characteristic of a multitude of schools of thought, ranging from Cartesian epistemology to Bayesian epistemology. For the latter, a number of basic results have been established which show that a Bayesian agent can extract probabilistic information from her environment by learning from experience. The works of Carnap on inductive logic [6, 7, 8], as well as the works of de Finetti on subjective probability [10] may be taken as representative work on this branch of inductive logic in the past century; see also Zabell [34]. These results typically assume that the sequence of events that the agent experiences corresponds to multinomial trials.

Our state of knowledge is less advanced for more general models of epistemic agents. Such generalizations would also take into account information transfer between agents, as well as non-stationary environments. The significance of information sharing between agents has been recognized for some time by philosophers working on social epistemology [14] and on the division of cognitive labor [18]. Moreover, game theoretic models of epistemic communities have recently gained some attention [35].

Game theory is of epistemological interest for another reason as well. Repeated games are typically non-stationary. Agents influence each other

Special Issue: Applied Logic in the Philosophy of Science
Edited by **Leon Horsten and Igor Douven**

Studia Logica (2008) 89: 237–256
DOI: 10.1007/s11225-008-9123-8

nonlinearly. Thus, learning from experience in a game theoretic environment raises problems which do not arise in classical inductive logic (on learning in games see, e.g., [9, 13, 33]). Theories of learning in games are of obvious importance to epistemology. In general, we should expect that learning in social interactions requires higher cognitive abilities than learning in a stationary environment. Moreover, social interactions with a quite simple structure might require less cognitive abilities than very complex social interactions (another interesting study in this context is [27]). To get beyond such mere conjectures, we have to study the performance of different kinds of learning in a variety of environments.

In this paper we try to merge these two aspects. We will study a simple game theoretic model of information transfer. This model is taken from Bala and Goyal [2]. It will be introduced in Section 3, where we shall also review some possible applications of the Bala-Goyal game.

To capture the second aspect, we will be analyzing the performance of learning agents in the Bala-Goyal game. There are two kinds of learning we shall consider: reinforcement learning and fictitious play (Section 2). Reinforcement learning algorithms are self-centered learning rules. This means that an agent bases her decisions on nothing but the history of her own payoffs. Fictitious play is more akin to cognitive learning. Players are assumed to have beliefs about what the other players are going to do. In its simplest form, fictitious play is Carnapian inductive logic plus maximization of expected utility (see Skyrms [30]). Both learning rules are well studied, and both have been proposed as models of human learning in games. From a theoretical standpoint, an interesting question to ask is whether a very simple learning rule leads to satisfactory behavior in a given situation. For its simplicity, reinforcement learning thus appears to be a natural first candidate. Our main results suggest that the Bala-Goyal game presents problems to reinforcement learners which they can, in general, not overcome satisfactorily (Section 4). We will also discuss some results on fictitious play in the Bala-Goyal game (Section 5). These results indicate that learning rules which are not purely self-centered perform better in the Bala-Goyal game. A short discussion tries to place our results in philosophical perspective (Section 6).

2. Two models of learning from experience

Preliminaries

$A^j = \{1, 2, \ldots, m_j\}$ denotes the set of player j's pure strategies. The Cartesian product of the other players' strategy choices is denoted by $A^{-j} = A^1 \times$

$\cdots \times A^{j-1} \times A^{j+1} \times \cdots \times A^k$. The set of strategy profiles is $A = A^1 \times \cdots \times A^k$ and is identified with $A = A^j \times A^{-j}$. The set of player j's mixed strategies is denoted by $S^{m(j)}$. $S^{m(j)}$ is the set of all probability distributions over A^j. Thus $S^{m(j)}$ can be identified with the $(m(j)-1)$-dimensional probability simplex. The joint space of mixed strategies is given by $S = S^{m(1)} \times \cdots \times S^{m(k)}$. The joint mixed strategy space of the other players is denoted by S^{-j}. The payoff to player j is given by her utility function $u_j : A \to \mathbb{R}$, where $\mathbb{R}$ denotes the set of real numbers. u_j can be uniquely extended to a multilinear function from $S^{m(j)}$ to $\mathbb{R}$. This extension will also be denoted by u_j. The set of players, their strategies and utility functions defines a game Γ in strategic form.

Reinforcement learning

Reinforcement learning basically means that actions which led to higher payoffs in the past are more likely to be chosen again. A quite straightforward formulation of this principle is associated with Herrnstein in the psychological literature, and with Roth and Erev [12, 28] in the economics literature. Suppose Γ is a k-person game in strategic form which is repeatedly played at times $n = 1, 2, 3, \ldots$. Each player j has $m(j)$ pure strategies. At time n, the vector of propensities $\mathbf{q}^j_n = (q^j_{1n}, \ldots, q^j_{m(j)n})$, where $\mathbf{q}^j_n \in \mathbb{R}^{m(j)}_+$, represents player j's attraction toward each of her actions. That is, q^j_{in} is player j's tendency to choose strategy i at time n, $\mathbf{q}^j_0$ being the player's initial propensities. Setting

$$p^j_{in} = \frac{q^j_{in}}{Q^j_n},$$

where $Q^j_n = \sum_m q^j_{mn}$, induces a probability distribution over j's strategies given by the vector $\mathbf{p}^j_n = (p^j_{1n}, \ldots, p^j_{m(j)n}) \in S^{m(j)}$. The probability of choosing a strategy is thus the relative frequency of the associated propensity.

After having chosen a strategy, player j updates her vector of propensities according to the following rule:

$$q^j_{i(n+1)} = q^j_{in} + \sigma^j_{in} \tag{1}$$

where σ^j_{in} is a random variable defined by

$$\sigma^j_{in} = \begin{cases} u_j(i, a^{-j}_n), & \text{if } i \text{ was played at time } n \\ 0 & \text{otherwise.} \end{cases}$$

Here, a_n^{-j} denotes the strategy profile of the other players at time n. Thus, a reinforcement learner updates her propensity for the strategy she has played by the strategy's payoff while keeping the propensities for all the other strategies fixed.

In the basic model, experiences from the remote past have the same effect on present decisions as experiences which were made just recently. This is unrealistic from a psychological and a methodological standpoint (see Skyrms and Pemantle [31]). It is psychologically implausible since memories fade. It is methodologically implausible since the players are assuming an essentially stationary environment. Discounted reinforcement learning is a straightforward way to meet these problems. The discounted version of the basic model introduces a discount parameter $\phi \in (0, 1)$. The rule by which a player updates her propensities is modified to

$$q^j_{i(n+1)} = \phi q^j_n + \sigma^j_{in}.$$

At each step in time past weights are discounted by ϕ. As $\phi \to 0$, past experiences are forgotten faster. As $\phi \to 1$, the discounted model approaches the basic model.

Fictitious Play

While reinforcement learning falls under the category of behavioral learning rules, fictitious play is a simple formalization of cognitive learning [5]. One way to introduce fictitious play is by modeling players as having beliefs about the choices of the other players. Let $\mathbf{e}^i_n$ be the random unit vector in $\mathbb{R}^{m(i)}$ whose entries are defined by

$$e^i_{kn} = \begin{cases} 1 & \text{if } i \text{ chose } k \text{ at time } n \\ 0 & \text{otherwise.} \end{cases}$$

Player j's probabilities about i's actions are then given by

$$\mathbf{z}^j_{in} = \frac{1}{n + \Lambda_{ji}} \sum_{k=1}^{n} (\mathbf{e}^i_k + \lambda^j_i),$$

$\lambda^j_i = (\lambda^j_{i1}, \ldots, \lambda^j_{im(i)})$ being j's vector of prior weights; i.e. $\lambda^i_{ik} \geq 0$ is j's prior weight that player i chooses her kth action. $\Lambda_{ji} = \sum_k \lambda^j_{ik}$.

Moreover, at each period n player j chooses a best response given her beliefs $\mathbf{z}^j_n \in S^{-j}$. If $\mathbf{y} \in S^{-j}$, then $k \in A^j$ is a *(pure) best response* to $\mathbf{y}$ if $u_j(k, \mathbf{y}) \geq u_j(l, \mathbf{y})$ for all $l \in A^j$. There may be more than one best response. Then a tie-breaking rule has to be specified. We assume that, when confronted with a tie, players choose each best response with positive probability, where the probability distribution stays fixed over time.

Fictitious play can be modified in a number of ways. One possible variant is obtained by restricting the memory of each player to a certain number of rounds m. Taking $m = 1$ yields the naive best response, or Cournot, dynamics. Cournot agents know the last period's strategy profile and play a best response to it.

Following Bala and Goyal [2], we will also consider a modified version of the Cournot dynamics, the myopic best response dynamics with inertia. At each period n, each agent j exhibits inertia with some fixed probability $0 < \mu_j < 1$. This means that with probability μ_j agent j chooses a_n^j again in period $n+1$. With probability $\nu_j = 1 - \mu_j$ player j chooses a best response to a_n^{-j} just like a Cournot agent. If there are multiple best responses to a_n^{-j}, we assume that the agent chooses a best response according to a probability distribution which assigns positive weight to each best response.

3. A simple model of information transfer

The model introduced by Bala and Goyal in [2] involves n agents. Each agent has a piece of information which is valuable to the other agents. For simplicity, it is assumed that the value of each piece of information is normalized to 1. Agents may visit each other to get to the information. A visit bears a cost of c. Throughout this paper we will assume that $0 < c < 1$. (The equilibrium properties of Theorem 3.1 below change if we allow for $c > 1$.) This implies that it always pays off to visit another agent. Each round, a player has to decide what players to visit. Thus a player has

$$\sum_{k=0}^{n-1} \binom{n-1}{k} = 2^{n-1}$$

strategies. Each visit results in an asymmetric transfer of information. This means that if A visits B, then A gets the information of B while B does not get the information of A. A player also gets the information from indirect links: if A_i visits A_{i+1} for $i = 1, \ldots, k$, then A_1 gets the information from $A_2, \ldots, A_k$. These properties make the Bala-Goyal game more challenging from a strategic point of view than the networking model of Skyrms and Pemantle [31].

We may picture the choices of the agents as a directed graph G. The n vertices of G are the players. A directed edge (i, j) means that agent i visits agent j. (Thus, information is flowing from j to i.) A *path* in G from i to j is a sequence of directed edges $(i, k_1), (k_1, k_2), \ldots, (k_{m-1}, k_m), (k_m, j)$. G is *connected* if there is a path from any agent i to any other agent j. G is *minimally connected* if it is connected and if the removal of any directed edge results in a disconnected graph G'. These concepts allow for a convenient formulation of some equlibrium properties of the Bala-Goyal game.

PROPOSITION 3.1 (Bala and Goyal [2]). *Let $0 < c < 1$. A joint strategy in the Bala-Goyal game is a Nash equilibrium in pure strategies if and only if its corresponding graph G is minimally connected. A joint strategy is a strict Nash equilibrium if and only if its corresponding graph G is a connected circle.*

We would like to emphasize that there are also Nash equilibria in mixed strategies. Consider the graph $G = \{(1, 2), (1, 3), (2, 1), (3, 1)\}$ for a *three*-player Bala-Goyal game. G is minimally connected. Suppose that player 2 mixes between $(2, 1)$ and $(2, 3)$ with probabilities p and $(1-p)$. The resulting profile is a Nash equilibrium as long as $p > c$. Nash equilibria of this kind will play an important role in the learning dynamics of the Bala-Goyal game.

The Kula Ring of the Trobriand islands provides a real-world example of a social network were transfer happens on a circle structure. It was first described by B. Malinowski in 1922 [22]. The Kula Ring is a ritualized exchange of two kinds of valuables, necklaces and bracelets. Along with this, goods and informations are exchanged between neighboring islands. Each group of islands is connected to two other groups of islands. Necklaces are exchanged in one direction, and bracelets in the other. They travel along two circles and visit all islands in the Kula Ring until they get back again. (These goods are not supposed to be owned by one and the same person for a long time.) People from one element of the Kula Ring visit a neighboring element to get necklaces, and the other neighboring element to get bracelets. Thus the asymmetry of transfer in the Bala-Goyal game seems to hold. There are a number of other examples of circle structures in anthropology (see, e.g., [23]).

4. Information transfer by reinforcement learning

The basic model

The cognitive requirements for reinforcement learning are quite low. An agent just has to keep track of the sums of her past reinforcements for each choice. In particular, a reinforcement learner needs no information about her environment. Nonetheless, it has been shown quite recently that the basic model converges to optimal actions in stationary environments [3, 17]. Suppose the agent chooses from a finite set of actions $A = \{a_1, \dots, a_m\}$ at times $n = 1, 2, 3, \dots$. Suppose in addition that the agent updates the weights $q_n = (q_{1n}, \dots, q_{mn})$ according to the basic model (1). The sequence of the agent's choices up to time n generates the σ-field $\mathcal{F}_n$. If X denotes a random variable, $\mathbb{E}[X|\mathcal{F}_n]$ is the conditional expectation of X given $\mathcal{F}_n$.

Proposition 4.1 (Beggs [3]). *Let $\gamma > 1$. If for some i and for all n*

$$\mathbb{E}[u_{n+1}|\mathcal{F}_n, a_i \textit{ chosen at } n+1] > \gamma \mathbb{E}[u_{n+1}|\mathcal{F}_n, a_j \textit{ chosen at } n+1], \tag{2}$$

where u_{n+1} denotes the decision maker's payoff at time $n+1$, then the probability that the decision maker chooses a_j converges to zero as $n \to \infty$ almost surely.

Proposition 4.1 tells us that a reinforcement learner will learn to avoid suboptimal actions. E.g. in a multi-armed bandit problem where one slot machine pays off with higher probability than any other, a reinforcement learner will choose to play this slot machine in the long run.

Proposition 4.1 does, of course, not settle the issue of reinforcement learning in games. Some results in this direction are to be found in [3, 17, 19, 29]. A frequently employed technique to analyze stochastic learning models such as the basic model is stochastic approximation theory (see Benaïm [4]). Stochastic approximation theory asserts that the basic model (1) is associated with a system of ordinary differential equations (ODE) whose limiting behavior is closely related to the long-run behavior of the basic model. This result is quite intuitive. Think of the basic model as an urn process. Each strategy of a player is represented by balls of a particular color. When a player chooses a particular strategy, she adds balls of the corresponding color according to the payoff she receives. Assuming that payoffs are positive, the number of balls is growing. Hence, adding balls makes less difference at time n than at previous times. The process becomes increasingly "deterministic" as time proceeds.

The ODE associated with the basic model can be obtained by looking at the expected change in p^j_{in}:

$$\mathbb{E}[p^j_{i(n+1)} - p^j_{in}|\mathcal{F}_n] = \frac{p^j_{in}(u_j(a_i, \mathbf{p}^{-j}_n) - u_j(\mathbf{p}))}{Q^j_n} + O\left(\frac{1}{(Q^j_n)^2}\right)$$

The first term of the right side of this equation is reminiscent of the evolutionary replicator dynamics (see Hofbauer and Sigmund [15]): the expression

$p_i^j(u_j(a_i, \mathbf{p}_n^{-j}) - u_j(\mathbf{p}))$ can be interpreted as the rate of change in p_i^j which is given by its current value and the payoff difference between the payoff player j gets from playing strategy a_i and the average payoff $u_j(\mathbf{p})$ (which j gets when playing the mixed strategy $\mathbf{p}^j$ against $\mathbf{p}^{-j}$). Indeed, a variant of the replicator dynamics turns out to be the ODE associated with the basic model. To account for the different step sizes $1/Q_n^j$, a new variable μ_n^j is introduced for each player by setting $\mu_n^j = n/Q_n^j$ (see [17]). The expected change in p_{in}^j can then be rewritten as

$$\begin{aligned} \mathbb{E}[p_{i(n+1)}^j - p_{in}^j|\mathcal{F}_n] &= \frac{1}{n}\mu_n^j(p_{in}^j\left(u_j(a_i, \mathbf{p}_n^{-j}) - u_j(\mathbf{p})\right)) + O\left(\frac{1}{n^2}\right) \\ \mathbb{E}[\mu_{n+1}^j - \mu_n^j|\mathcal{F}_n] &= \frac{1}{n}\mu_n^j\left(1 - \mu_n^j u(\mathbf{p})\right) + O\left(\frac{1}{n^2}\right) \end{aligned}$$

The introduction of the new variables turns the basic model into a stochastic algorithm with deterministic step size $1/n$. Hopkins and Posch [17] show that the ODE associated with the basic model is given by the following system of differential equations:

$$\dot{p}_i^j = \mu^j p_i^j(u_j(a_i, \mathbf{p}_n^{-j}) - u_j(\mathbf{p})), \quad \dot{\mu}^j = \mu^j(1 - \mu^j u_j(\mathbf{p})). \tag{3}$$

The first set of equations for the rate of change of p_i^j is nothing but the replicator dynamics rescaled by the factor μ^j.

The long-run behavior of solutions to the system (3) for the Bala-Goyal game can be studied by using standard tools from dynamical systems theory. We state a first result in this direction after the following lemma.

Lemma 4.2. *If $(\bar{\mathbf{p}}, \bar{\mu})$ with $\bar{\mu}^j = 1/u_j(\bar{\mathbf{p}})$ is a rest point of (3), then the only eigenvalue of the matrix $d\dot{\mu}/d\mu$ is -1. Consequently, this eigenvalue has multiplicity n.*

Proof. The entries of the matrix $d\dot{\mu}/d\mu$ are given by the partial derivatives

$$\frac{\partial}{\partial \mu^i}\mu^j(1 - \mu^j u_j(p)).$$

This expression is 0 if $i \neq j$. If $i = j$ it is equal to $1 - 2\mu^j u_j(\mathbf{p})$. At $(\bar{\mathbf{p}}, \bar{\mu})$ this equals -1. Thus, at the rest point $d\dot{\mu}/d\mu$ is a diagonal matrix with all diagonal entries equal to -1. ■

Theorem 4.3. *Let Γ be a n-person Bala-Goyal game. Then each strategy profile that corresponds to a circle is asymptotically stable for (3).*

PROOF. Notice first that all strategy profiles $\bar{\mathbf{p}}$ corresponding to a circle are rest points for the replicator dynamics. By Lemma 2 of [17], $(\bar{\mathbf{p}}, \bar{\mu})$ with $\mu^j = 1/u_j(\bar{\mathbf{p}})$ is a rest point for (3). To determine the stability of $(\bar{\mathbf{p}}, \bar{\mu})$, we look at the Jacobian matrix of (3) evaluated at $(\bar{\mathbf{p}}, \bar{\mu})$. The Jacobian matrix is of the form

$$J = \begin{pmatrix} L & 0 \\ d\dot{\mu}/d\mathbf{p} & d\dot{\mu}/d\mu \end{pmatrix}. \tag{4}$$

Here, L is the Jacobian matrix of the system $\dot{p_i^j} = \mu^j p_i^j(u_j(a_i, \mathbf{p}_n^{-j}) - u_j(\mathbf{p}))$. Due to the zero-entries in the upper-right corner, the eigenvalues of J are given by the eigenvalues of L and $d\dot{\mu}/d\mu$. At the rest point $(\bar{\mathbf{p}}, \bar{\mu})$, L is identical to the Jacobian of the *adjusted replicator dynamics*, which is given by

$$\dot{p_i^j} = \frac{p_i^j(u_j(a_i, \mathbf{p}_n^{-j}) - u_j(\mathbf{p}))}{u_j(\mathbf{p})}.$$

A well-known result in evolutionary game theory states that all eigenvalues of the adjusted replicator dynamics at a strict Nash equilibrium are negative. Lemma 4.2 implies that, at $(\bar{\mathbf{p}}, \bar{\mu})$, $d\dot{\mu}/d\mu$ has only negative eigenvalues. Hence, all eigenvalues of J at $(\bar{\mathbf{p}}, \bar{\mu})$ are negative. Therefore, $(\bar{\mathbf{p}}, \bar{\mu})$ is asymptotically stable. ■

Theorem 4.3 has also consequences for the learning process as given by the basic model (1). Since each strategy profile corresponding to a circle is asymptotically stable for (3), the basic model converges to each of these profiles with positive probability. This can be shown by applying a result of the type of Theorem 7.3 of Benaïm [4].

One might wonder if the circle is the only stable strategy configuration of the Bala-Goyal game for the basic model. The next theorem provides a partial answer to this question. To give a heuristic explanation, consider the graph G from Section 3. G is a Nash equilibrium. As it turns out, G is also stable under the dynamics (3) as will be shown in the proof of Theorem 4.4. More precisely, this configuration is Lyapunov stable for (3), which means that nearby solutions stay nearby for all future times. This implies that there exists an open set O of initial states close to G such that solutions starting in O will not converge to the circle.

THEOREM 4.4. *Let Γ be a n-player Bala-Goyal game. Then there exist open sets of initial conditions which do not converge to a strategy profile which corresponds to a circle for (3).*

Proof. Suppose that Γ is a three-player Bala-Goyal game. Denote by p_i^j the probability with which player j chooses her ith strategy, where $j = 1, 2, 3$ and $i = 1, 2, 3, 4$. Let s_1^j be the strategy of visiting noone, s_2^j the strategy of visiting $j+1$ modulo 3, s_3^j the strategy of visiting $j+2$ modulo 3, and s_4^j the strategy of visiting $j+1$ and $j+2$ modulo 3. Suppose that $\bar{p}_4^1 = 1$, $\bar{p}_2^2 = 1$, $\bar{p}_2^3 = \lambda$ and $\bar{p}_3^3 = 1 - \lambda$ where $0 \leq \lambda \leq 1$ (c being the cost parameter). Denote the corresponding joint state by $\bar{\mathbf{p}}$. Then it is easy to verify that $(\bar{\mathbf{p}}, \bar{\mu})$ with $\bar{\mu}^j = 1/u_j(\bar{\mathbf{p}})$ is a rest point for (3). Indeed,

$$\dot{p}_4^1 = \bar{p}_4^1(u_j(s_4^1, \bar{\mathbf{p}}^{-j}) - u_j(\bar{\mathbf{p}})) = 0$$

and similarly for the other components of $\bar{\mathbf{p}}$. Thus, by Lemma 2 of [17], $(\bar{\mathbf{p}}, \bar{\mu})$ is a rest point of (3). It remains to show the stability of this rest point. By Lemma 4.2, the eigenvalues of $d\dot{\mu}/d\mu$ evaluated at $(\bar{\mathbf{p}}, \bar{\mu})$ are negative. The matrix L of the Jacobian (4) has only non-positive eigenvalues. To see this, denote by supp$(\bar{\mathbf{p}})$ the support of $\bar{\mathbf{p}}$, i.e. the strategies s_i^j with $\bar{p}_i^j > 0$. Consider $s_i^j \notin$ supp$(\bar{\mathbf{p}})$. The eigenvalues with respect to these pure strategies are called transversal eigenvalues and are given by $u_j(s_i^j, \bar{\mathbf{p}}^{-j}) - u_j(\bar{\mathbf{p}})$ (see Hofbauer and Sigmund [15] for a derivation). All transversal eigenvalues are negative as long as $\lambda > p$. Notice first that $u_j(s_1^j, \mathbf{x}^{-j}) = 0$ regardless of $\mathbf{x}^{-j}$. Hence all transversal eigenvalues relative to s_1^j are negative. Moreover $u_1(s_2^1, \bar{\mathbf{p}}^{-1}) = 1 - c < 2(1 - c) = u_1(\bar{\mathbf{p}})$. If $\lambda < p$ then $u_1(s_3^1, \bar{\mathbf{p}}^{-1}) < 2(1 - c)$ and $u_2(s_2^2, \bar{\mathbf{p}}^{-1}) < 2 - c$. Finally, for $j = 2, 3$ $u_j(s_4^j, \bar{\mathbf{p}}^{-j}) = 2(1 - c) < 2 - c = u_j(\bar{\mathbf{p}})$. Consider the eigenvalues with respect to $s_i^j \in$ supp$(\bar{\mathbf{p}})$ next. Since all points $\bar{\mathbf{p}}$ with $0 \leq \lambda \leq 1$ are rest points, the part of state space spanned by supp$(\bar{\mathbf{p}})$ is a linear manifold of rest points. Thus all eigenvalues with respect to $s_i^j \in$ supp$(\bar{\mathbf{p}})$ are 0. Hence if $\lambda < p$ all eigenvalues of the Jacobian for (3) are non-positive with all transversal eigenvalues being negative. Since the zero eigenvalues only apply to the linear manifold of rest points, the points $(\bar{\mathbf{p}}, \bar{\mu})$ with $\lambda < p$ are Liapunov stable, which means that trajectories close to $(\bar{\mathbf{p}}, \bar{\mu})$ stay close to it. This implies that there exist open sets of initial conditions which include each $(\bar{\mathbf{p}}, \bar{\mu})$ such that solutions starting in one of these open sets do not converge to the circle. The above argument can obviously be extended to a Bala-Goyal game with more than three players. Just consider the state where player 1 visits all the other players while they visit 1. ■

In the case of Theorem 4.4 conclusions about the behavior of the corresponding stochastic process are not as easy to draw as for Theorem 4.3. The reason for this is that the components of Nash equilibria which are Liapunov stable are not attractors like the circle (if S denotes state space then

$A \subset S$ is an attractor of a dynamics ϕ if A is Liapunov stable and if there exists a neighborhood U of A such that $\phi_t(x) \to A$ as $t \to \infty$ for all $x \in U$). If N denotes such a component of Nash equilibria, then points in the relative interior of N are Liapunov stable, but points on the relative boundary of N are (second order) unstable. Thus, even if we assume that N attracts an open set of initial conditions, there does not exist a neighborhood U of N that would meet the requirements in the definition of an attractor. But the existence of such a neighborhood is essential for proving a result like Theorem 7.3 of Benaïm [4].

Thus, proving whether the basic model converges to points in a component like N is an open problem. Numerical simulations suggest that convergence occurs. To be more specific, we carried out simulations with three and four players. Even after 10^6 runs a considerable amount of the simulations ended up in a state very much like a point in N (the exact amount depends on c). Thus, given the heuristic result of Theorem 4.4, we may conjecture that the basic model does indeed converge to components such as N. But one has to be cautious with conclusions of this kind, since the slow speed of the process near N casts doubt on whether convergence occurs in the infinite limit (see [1, 25, 26] for similar points). However, our simulations, together with Theorem 4.4, show: even if convergence occurs, the time for convergence exceeds the lifetime of any simulation and the lifetime of any relevant physical system. Therefore, in the time horizon we are interested in, we may conclude that the basic model fails to robustly converge to the circle in the Bala-Goyal game.

Discounted basic model

One might suspect that in the Bala-Goyal game players who come close to a suboptimal state will eventually forget this state when choosing strategies with higher payoffs. The next theorem shows that this does not hold. Instead, the weights for playing their other strategies might be forgotten.

Theorem 4.5. *The discounted basic model does not converge to the circle of the Bala-Goyal game with positive probability.*

Proof. Suppose there are n players. Let $\mathbf{P}$ be the matrix which has the visiting probabilities of the players as entries. To be more specific, p_{ij} is the probability that agent i visits agent j for $j = 1, \ldots, n$, p_{ik} is the probability that agent i visits one of the $\binom{n-1}{2}$ pairs of agents, suitably numbered, and so on. $p_{i,2^{n-1}}$ is the probability that agent i visits all the other players. We regard p_{ii} as the probability that i visits nobody. Given $\varepsilon > 0$ we associate a directed graph $G = G(\mathbf{P})$ with $\mathbf{P}$ as follows. The directed edge $(i, j) \in G$ if $p_{ik} > \varepsilon$ for at least one k that corresponds to a subset of $\{1, \ldots, n\}$ that has j as a member. We assume that $\varepsilon < (2(2^{n-1} - 1))^{-1}$. We will show that G will with positive probability just consist of the edges $(1, 2), (1, 3), \ldots, (1, n), (2, 1), (3, 1), \ldots, (n, 1)$ in the limit. Let us denote this graph by G^*.

The first step consists in showing that for any $\mathbf{P}$ we can get from $G(\mathbf{P})$ to G^* with probability at least $\delta_1 > 0$. There exists a number N such that if 1 chooses her (2^{n-1})st strategy and j chooses strategy 1, $j = 2, \ldots, n$ for N consecutive rounds, then $G(\mathbf{P}(N)) = G^*$. If we choose ε small enough, then each round this probability is at least ε^n. Taking $0 < \delta_1 < \varepsilon^{nN}$ completes the first step.

The second step consists in showing that with probability at least δ_2, $G(\mathbf{P}(t)) = G^*$ for all $t \geq N$. By the definition of G^*

$$\sum_{j \neq 2^{n-1}} p_{1j} \leq (2^{n-1} - 1)\varepsilon \quad \text{and} \quad \sum_{j \neq 1} p_{ij} \leq (2^{n-1} - 1)\varepsilon, i = 2, \ldots, n.$$

Since $\varepsilon < 2(2^{n-1} - 1)^{-1}$ all these sums are at most $1/2$. By letting the players choose according to G^* for k rounds each sum is bounded by $1/2\phi^k$, where ϕ is the discount parameter. Therefore, the probability that all players choose according to G^* for N consecutive rounds is at least

$$\prod_{k=0}^{N-1} \left(1 - \frac{1}{2}\phi^k\right)^n.$$

Since $0 < \phi < 1$, $\sum_k \phi^k < \infty$. Hence letting $N \to \infty$ yields an infinite product greater than 0. Taking δ_2 less than this value, the probability that the players' choices converge to G^* is at least $\delta_1 \delta_2$. ■

5. Information transfer by fictitious play

Unlike reinforcement learning, fictitious play assumes that players have a quite accurate model of the other agents. While reinforcement learners gather information indirectly by their payoff history, fictitious play learners gather information about the world directly by observation. The last property also holds for naive best response learning.

Bala and Goyal [2] study the myopic best response dynamics with inertia in the Bala-Goyal game. Unlike reinforcement learners, agents choosing according to this version of the Cournot dynamics converge to the circle almost surely.

Theorem 5.1 (Bala and Goyal [2]). *If all players of the n-player Bala-Goyal game choose according to the myopic best response dynamics with inertia, then the strategy profile converges to a circle of the Bala-Goyal game with probability* 1.

Fictitious play learners are somewhat more sophisticated since they take account of the whole history of the repeated game. A simple, but useful result follows from the fact that the circle is a strict Nash equilibrium (see, e.g., Fudenberg and Levine [13, Proposition 2.1]). For this and all the statements up to, and including Theorem 5.7 we assume that players learn according to fictitious play. (Other simplifying assumptions for the statements of the rest of this section will be introduced after the next theorem.)

Theorem 5.2. *If the joint strategies of the players correspond to a circle of the n-player Bala-Goyal game at time t, then this joint strategy is played at all subsequent times.*

To make the model more tractable, we will look at the case of three players whose initial beliefs are uniform over the other agents' actions. Moreover, the statements of our results will be simplified by assuming that $1/2 < c < 1$ (see Remark 5.4 below). The strictly dominated strategy of visiting nobody can, and will, be ignored.

Let v_{ij}^n be the number of times i has visited j by time n. Then $v_i = n - v_{ij}^n - v_{ik}^n$ $(k \neq j)$ is the number of times i has visited both of the other agents by time n. λ_{ij} and λ_i are the respective prior weights. We assume that $\lambda_{ij} = \lambda_i = 1$ for all i, j. p_{ij}^{n+1} denotes i's probability that j visits her at time $n+1$, and it is given by

$$p_{ij}^{n+1} = \frac{v_{ji}^n + 1}{n+3}.$$

We denote the probability that i visits both of the other players by q_i. Then $q_i = 1 - p_{ij} - p_{ik}$. V_{ij}^n denotes the event that i has chosen to visit j at time n, and V_i^n denotes the event that i has chosen to visit j and k at time n. We shall suppress the time-index whenever it is convenient to do so. The expected values for i's actions are given by

$$\mathbb{E}_n[V_{ij}^{n+1}] = 2 - c - p_{ij}^n, \quad \mathbb{E}_n[V_{ik}^{n+1}] = 2 - c - p_{ik}^n \quad \text{and} \quad \mathbb{E}_n[V_i^{n+1}] = 2(1-c),$$

where $\mathbb{E}_n$ is the conditional expectation relative to the σ-algebra $\mathcal{F}_n$ generated by the variables V_{ij}^m, V_i^m, $m = 1, \ldots, n$.

To formulate the next series of results we introduce the directed graphs $G = \{(i,j), (i,k), (j,i), (k,i)\}$ and $G' = \{(i,j), (i,k), (j,k), (k,i)\}$. Moreover, let H be a directed graph that corresponds to a profile where at least two players visit both other players, and let $K = \{(i,k),(j,i),(k,i)\}$ and $L = \{(i,j),(i,k),(j,k),(k,j)\}$. We note that for $1/2 < c < 1$ and initial weights of 1, the players start by choosing a profile K or a circle profile.

Lemma 5.3. *If $1/2 < c < 1$, then the joint strategies of three players do not switch to G without passing through a profile different from K. For G' either the same conclusion as for G holds, or G' leads to a circle profile.*

Proof. Suppose that before period $n+1$ the players have always chosen K. In period $n+1$ they choose according to G only if

$$p_{ji}, p_{ki} \leq c \leq p_{ij}, p_{ik} \text{ and } p_{ji} \leq p_{jk}, p_{ki} \leq p_{kj}.$$

This follows by a simple inspection of the expected values. We assume in addition, and without loss of generality, that $p_{ij} \leq p_{ik}$. This set of inequalities implies

(i) $v_{ji}, v_{ki} \geq v_{ij}, v_{ik}$

(ii) $v_{ki} \leq v_{ik}$, $v_{ji} \leq v_{ij}$ and $v_{ji} \leq v_{ki}$

since $v_{ik} = n - v_{ij}$, $v_{jk} = n - v_{ji}$ and $v_{kj} = n - v_{ki}$. These inequalities imply that $v_{ij} = n/2$ for all i, j. Hence, for all i, j,

$$p_{ij} = \frac{n+2}{2(n+3)}.$$

Since $p_{ij} = c$ we must have

$$\frac{n+2}{n+3} = 2c$$

which is impossible for $c > 1/2$. To prove the part of the lemma for G', observe that the same inequalities as for G hold except that $p_{ik} \leq c$ and $p_{ik} \leq p_{ji}$. If $p_{ik} = p_{ji}$, then the same arguments as for G apply. If $p_{ik} < p_{ji}$, then calculating the v_{ij} for the next round shows that the players will choose according to the circle after a finite number of rounds. ■

Remark 5.4. If $0 < c < 1/2$, profiles corresponding to G are possible but unlikely. If $p_{ij} = c$ for all i, j, then

$$n = \frac{6c-2}{1-2c} = \varphi(c).$$

φ is a strictly decreasing negative function of c for $1/2 < c < 1$, but it is a strictly increasing positive function for $0 < c < 1/2$. As $c \uparrow 1/2$, $\varphi(c) \uparrow \infty$.

By the intermediate value theorem, $\varphi(c)$ takes on any positive integer n as value. But notice that the set $\{c : \varphi(c) = n$ for some integer $n\}$ has measure zero since φ is strictly increasing. For instance, $\varphi(c) < 2$ if, and only if, $c < 2/5$. So for all $c < 2/5$ the conclusion of Lemma 5.3 also holds. The event that the inequalities of the proof of Lemma 5.3 are satisfied may thus be considered as exceptional and not robust to small perturbations of c.

Lemma 5.5. *If* $1/2 < c < 1$*, then three players will never choose profile* H*.*

Proof. Suppose i and j visit both other players in period n. This is the case only if $p_{ik}, p_{jk} \geq c$. Hence $p_{ik} + p_{jk} \geq 2c > 1$ which is impossible since $p_{ik} + p_{jk} \leq 1$. ■

Lemma 5.6. *If* $1/2 < c < 1$*, then the joint strategies of three players do not switch to* L *without passing through a profile different from* K*.*

Proof. Suppose the players choose with positive probability according to $(i,j),(i,k),(j,k),(k,j)$ for the first time in period $n+1$. Then at time n profile $(i,k),(j,i),(k,i)$ (j and k may be interchanged) must have occurred with positive probability. Thus, $v_{jk}^{n-1} \geq (n-1)/2$ and $v_{kj}^{n-1} \geq (n-1)/2$. Hence, $v_{ji}^{n} = n - v_{jk}^{n} \leq (n+1)/2$ and $v_{ki}^{n} \leq (n+1)/2$. Therefore, at time $n+1$,

$$p_{ij}^{n+1} + p_{ik}^{n+1} \leq 2\frac{(n+1)/2+1}{n+3} = 1 < 2c.$$

This implies that i does not choose to visit j and k at time $n+1$ (just consider the corresponding expected values). ■

By the foregoing lemmata, the players will always choose a profile equivalent to K or the circle since they start with one of these profiles. This does not imply convergence to the circle: the players may switch strategies forever. Our next result shows that this will not happen.

Theorem 5.7. *If* $1/2 < c < 1$*, then the joint strategies of three players converge to a profile corresponding to a circle with probability* 1*.*

Proof. K implies that $v_{ji} \geq v_{ki}$, $v_{ij} \leq v_{kj}$ and $v_{ik} \leq v_{jk}$. If $v_{kj} > v_{ij}$, then after a finite number of periods $v_{ji} = v_{ki}$ or $v_{ik} = v_{jk}$. If the latter, then there is a probability of at least $\varepsilon > 0$ that the players get to the circle. If not, they get to a profile equivalent to K. If $v_{ji} = v_{ki}$ first, then i chooses randomly until j or k also choose randomly. The result of these choices is again a circle or a profile equivalent to K. If $v_{kj} = v_{ij}$, then j will choose randomly until i or k also choose randomly. The result is, again, that the players get to the circle or to a profile equivalent to K. Hence, starting from a profile equivalent to K leads, within a finite number of periods, to a period n where the players get to the circle with probability at least $\varepsilon > 0$, or they get back to a profile equivalent to K. To finish the argument, consider a sequence of choices where each time (i) at least one player chooses randomly, (ii) these random choices are independent of choices at all other times, and (iii) the random choices lead to the circle with probability at least $\varepsilon > 0$. If E_n denotes the event that the players get to the circle at time n, then clearly

$$\sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty.$$

Since the events E_n are independent, the second Borel-Cantelli Lemma implies that E_n will happen infinitely often with probability 1 (see, e.g., Durrett [11, Section 1.6]). Therefore, the players will, with probability 1, randomly choose according to the circle at some finite period n if they have always returned to K before n. ■

Numerical simulations show that convergence to the circle takes place quite rapidly and for all $c \in (0, 1)$. Numerical simulations for four players led to the same result. Since nothing in the basic structure of the pure Nash equilibria changes when we consider more than three players, we conjecture that n players will also converge to the circle if each player starts with weights initialized to 1.

When players start with random initial weights, numerical simulations with three and four players indicate that they do not always converge to the circle. The reason for this is quite simple. If initially λ_{ij}, λ_{ik}, λ_{jk} and λ_{kj} are very high, the players will choose G. By playing G, the number of visits v_{ji}, v_{ki} increases, while v_{ij}, v_{ik}, v_{kj} and v_{jk} remain constant. As a result, the players will continue to choose according to G.

Notice that if j or k were to choose the payoff-equivalent strategies (j, k) and (k, j) sufficiently often, the players might be able to get away from G. To make this idea precise, let us consider a modification of fictitious play (a similar version of fictitious play was studied in [21, 32]). For this modification of fictitious play, the belief formation process is the same as for standard fictitious play. The choice rule is changed into a rule that tells the players to choose an ε-best response with positive probability; i.e., if x_n^j denotes the maximum conditional expected value of player j at time n, then, for some $\varepsilon > 0$, i chooses any strategy with expected value $\mathbb{E}_{n-1}[V_{ij}^n] \geq x_n - \varepsilon$ with some fixed positive probability δ. As our next result shows, a state such as G will not persist under this modified version of fictitious play.

Theorem 5.8. *If players use an ε-best response rule, then they will almost surely not converge to a profile such as G.*

Proof. Suppose the players converge to a profile G. Then there exists some period N such that we have (i, j), (i, k) and (k, i) for all periods $n \geq N$. Let us denote the initial weights of j that i or k will visit her by λ_{ij} and λ_{kj}, respectively. Let Λ denote the sum of j's initial weights. Then

$$p_{ji}^{n+1} = \frac{v_{ij}^n + \lambda_{ij}}{n + \Lambda} \quad \text{and} \quad p_{jk}^{n+1} = \frac{v_{kj}^n + \lambda_{kj}}{n + \Lambda}.$$

For $n \geq N$, v_{ij}^n and v_{kj}^n remain constant. Hence for any $\varepsilon > 0$ there exists a $m \geq N$ such that $p_{jk}^m - p_{ji}^m < \varepsilon$. Thus, if i and k continue in choosing according to G, the event that j will always play (j, i) from some time onward has probability zero. This follows because $|p_{jk}^k - p_{ji}^k| < \varepsilon$ for all $k > m$ and from the second Borel-Cantelli Lemma. ■

Theorem 5.8 is consistent with numerical simulations. In fact, simulations with three and four players always converge to the circle when players use the modified best response rule which treats approximate ties as ties.

It should be emphasized that tie-breaking according to a probability distribution which assigns positive weight to all (approximate) best responses is essential for our proofs. The Bala-Goyal game is degenerate, since it allows for continua of Nash equilibria. It has been shown by Monderer and Sela [24] that there exist tie-breaking rules such that fictitious play does not converge in degenerate 2×2 games. However, many tie-breaking rules should work, as we either get back to a tie, or get away from the suboptimal equilibrium.

6. Discussion

As Theorem 4.1 shows, reinforcement learning is remarkably effective in achieving efficient solutions in some problems. This does not only apply to bandit problems, or more general sequences of events which are exemplified by condition (2). Reinforcement learning is also performing well in some classes of games, such as, e.g., non-degenerate partnership games, i.e. games where the interests of the players largely coincide (see [17]). However, as our results suggest, reinforcement learning seems to be unable to reliably generate efficient communication networks in the Bala-Goyal game.

In contrast, very simple forms of cognitive learning with best, or approximate best responses suffice to achieve efficiency. With respect to the Kula Ring, this leaves us with a puzzle: according to the account of Malinowski, the participants of the Kula Ring do clearly not have the kind of global information fictitious play learners use. But the Kula Ring itself is quite efficient nonetheless.

This puzzle poses one set of problems for future research. One possible line of research is to find algorithms "in between" reinforcement learning and fictitious play. Camerer and Ho [5] show that there exists a continuum of learning algorithms between reinforcement learning and fictitious play. Fictitious play as presented in Section 2 is equivalent to hypothetical reinforcement learning where players reinforce the weights for all actions by the payoff they would have gotten had they played the corresponding strategy. Reinforcement learning assigns zero weight to these hypothetical reinforcements, while fictitious play assigns a weight of one to hypothetical and actual reinforcements. Using these weights as parameter sets yields a parametrized family of algorithms. Notice, however, that Camerer-Ho learning also uses some kind of global information, in that hypothetical reinforcement for actions not taken must be calculated. One could get around this by letting the players learn these hypothetical payoffs; i.e. the player keeps track of the payoff she got from playing each action and uses this as an estimate for the hypothetical payoff. This form of learning is studied in Leslie and Collin's model-free version of fictitious play (see [20]).

Another way to overcome the problem of global information is to take into account growth of networks. Circles may start small. Global information may still be feasible in these small networks. Circles then grow by expansion or coalescing when two rings come into contact. But this sort of model remains to be explored.

Hopkins [16] also compares reinforcement learning and fictitious play and finds striking similarities between perturbed reinforcement learning and stochastic fictitious play in two-person games. The crucial difference seems to be that stochastic fictitious play gives rise to faster learning. What this means for our findings is no straightforward matter since Hopkins uses modified algorithms. Convergence might not work for Camerer-Ho learning or for stochastic fictitious play. Another road for future research consists in extending the analysis to other signaling interactions and more complicated signaling networks.

Acknowledgements. This paper was presented at conferences in Bristol and Liverpool, and at seminars in Irvine, Vienna, Innsbruck and Salzburg. We wish to thank the audiences at the conferences and seminars. We are particularly grateful to U. Berger and J. Hofbauer for helpful suggestions. Special thanks go to L. Narens for suggesting the Kula Ring as a topic.

S. Huttegger's research was supported by the Konrad Lorenz Institute for Evolution and Cognition Research.

References

[1] Argiento, R., R. Pemantle, B. Skyrms, and S. Volkov, 'Learning to Signal: Analysis of a Micro Level Reinforcement Model', *Stochastic Processes and their Applications* (forthcoming).

[2] Bala, V., and S. Goyal, 'A Noncooperative Model of Network Formation', *Econometrica* 68 (2000), 1181–1129.

[3] Beggs, A. W., 'On the Convergence of Reinforcement Learning', *Journal of Economic Theory* 122 (2005), 1–36.

[4] Benaïm, M., 'Dynamics of Stochastic Approximation Algorithms', in *Le Seminaire de Probabilites, Lecture Notes in Mathemtics*, vol. 1709, Springer-Verlag, New York, 1999, pp. 1–68.

[5] Camerer, C., and T. Ho, 'Experience-weighted attraction learning in normal form games', *Econometrica* 67 (1999), 827–874.

[6] Carnap, R., *Logical Foundations of Probability*, Chicago University Press, Chicago, 1950.

[7] Carnap, R., *The Continuum of Inductive Methods*, Chicago University Press, Chicago, 1952.

[8] Carnap, R., 'A Basic System of Inductive Logic', in Rudolf Carnap, and Richard C. Jeffrey (eds.), *Studies in Inductive Logic and Probability I*, University of California Press, Los Angeles, 1971, pp. 33–31.

[9] Cesa-Bianchi, N., and G. Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, Cambridge, 2006.

[10] De Finetti, B., 'Foresight: Its Logical Laws, its Subjective Sources', in Henry E. Kyburg, and Howard E. Smokler (eds.), *Studies in Subjective Probability*, John Wiley and Sons, New York, 1964, pp. 93–158.

[11] Durrett, R., *Probability: Theory and Examples*, Duxbury Press, Belmont, CA, 1996.

[12] Erev, I., and A. E. Roth, 'Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria', *American Economic Review* 88 (1998), 848–880.

[13] Fudenberg, D., and D. K. Levine, *The Theory of Learning in Games*, MIT Press, Cambridge, Mass., 1998.

[14] Goldman, A., *Knowledge in a Social World*, Oxford University Press, Oxford, 1999.

[15] Hofbauer, J., and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, 1998.

[16] Hopkins, E., 'Two Competing Models of How People Learn in Games', *Econometrica* 70 (2002), 2141–2166.

[17] Hopkins, E., and M. Posch, 'Attainability of Boundary Points under Reinforcement Learning', *Games and Economic Behavior* 53 (2005), 110–125.

[18] Kitcher, P., 'The Division of Cognitive Labor', *Journal of Philosophy*, June (1990), 5–22.

[19] Laslier, J.-F., R. Topol, and B. Walliser, 'A Behavioral Learning Process in Games', *Games and Economic Behavior* 37 (2001), 340–366.

[20] Leslie, D. S., and E. J. Collins, 'Convergent Multiple-Times-Scales Reinforcement Learning Algorithms in Normal Form Games', *The Annals of Applied Probability* 13 (2003), 1231–1251.

[21] Leslie, D. S., and E. J. Collins, 'Generalized Weakened Fictitious Play', *Games and Economic Behavior* 56 (2006), 285–298.

[22] Malinowski, B., *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventures in the Archipelagoes of Melanesian New Guinea*, Routledge and Kegan Paul, London, 1922.

[23] McKinnon, S., *From a Shattered Sun. Hierarchy, Gender, and Alliance in the Tanimbar Islands*, The University of Wisconsin Press, Madison, 1991.

[24] Monderer, D., and A. Sela, 'A 2 × 2 game without the fictitious play property', *Games and Economic Behavior* 14 (1994), 144–148.

[25] Pemantle, R., and B. Skyrms, 'Reinforcment Schemes may take a long Time to Exhibit Limiting Behavior', Preprint (2001).

[26] Pemantle, R., and S. Volkov, 'Vertex Reinforced Random Walk on $\mathbb{Z}$ has Finite Range', *Annals of Probability* 48 (2004), 1368–1388.

[27] Posch, M., A. Pichler, and K. Sigmund, 'The Efficiency of Adapting Aspiration Levels', *Proceedings of the Royal Society London B* 266 (1999), 1427–1436.

[28] Roth, A., and I. Erev, 'Learning in Extensive Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term', *Games and Economic Behavior* 8 (1995), 164–212.

[29] Rustichini, A., 'Optimal Properties of Stimulus Response Learning Models', *Games and Economic Behavior* 29 (1999), 244–273.

[30] Skyrms, B., 'Carnapian Inductive Logic for Markov Chains', *Erkenntnis* 35 (1991), 439–460.

[31] Skyrms, B., and R. Pemantle, 'A dynamic model of social network formation', *Proceedings of the National Academy of Sciences* 97 (2000), 16, 9340–9346.

[32] Van der Genugten, B., 'A Weakened Form of Fictitious Play in Two-Person Zero-Sum Games', *International Game Theory Review* 2 (2000), 307–328.

[33] Young, H. P., *Strategic Learning and its Limits*, Oxford University Press, Oxford, 2004.

[34] Zabell, S., *Symmetry and Its Discontents*, Cambridge University Press, Cambridge, 2006.

[35] Zollman, K. J. S., 'The Epistemic Benefit of Transient Diversity', *Philosophy of Science* (forthcoming).

Simon M. Huttegger
Konrad Lorenz Institute for
Evolution and Cognition Research
Adolf Lorenz Gasse 2
A-3422 Altenberg, Austria
simon.huttegger@kli.ac.at

Brian Skyrms
Department of Logic and Philosophy of Science
University of California at Irvine
3151 Social Science Plaza A
Irvine, CA 92697-5100, USA
bskyrms@uci.edu