SCIENTIFIC REP♺RTS

OPEN

Nash Equilibrium of Social-Learning Agents in a Restless Multiarmed Bandit Game

Kazuaki Nakayama[1], Masato Hisakado[2] & Shintaro Mori[3]

We study a simple model for social-learning agents in a restless multiarmed bandit (rMAB). The bandit has one good arm that changes to a bad one with a certain probability. Each agent stochastically selects one of the two methods, random search (individual learning) or copying information from other agents (social learning), using which he/she seeks the good arm. Fitness of an agent is the probability to know the good arm in the steady state of the agent system. In this model, we explicitly construct the unique Nash equilibrium state and show that the corresponding strategy for each agent is an evolutionarily stable strategy (ESS) in the sense of Thomas. It is shown that the fitness of an agent with ESS is superior to that of an asocial learner when the success probability of social learning is greater than a threshold determined from the probability of success of individual learning, the probability of change of state of the rMAB, and the number of agents. The ESS Nash equilibrium is a solution to Rogers' paradox.

One of the differences between human beings and other animals is that the former transfer their predecessors' experience and wisdom in the form of knowledge[1]. Social learning—learning from the experience of others— is advantageous compared to individual learning[2–4]. Without social learning everybody would have to learn everything for themselves[2]. In other words, individual learning costs more than social learning does[2–4]. Therefore, Rogers' finding that social learning is not necessarily more advantageous than individual learning is counterintuitive[5]. This is now called Rogers' paradox.

Rogers' conclusion seems very strange in light of our experience[4]. Several attempts have been made to solve Rogers' paradox in social learning. Boyd and Richerson[2] pointed out that Rogers' paradox is *not* a paradox when the only benefit of social learning is to avoid learning costs. Further, on analysing two models where social learning reduces individual-learning costs and improves the information obtained through the latter, they concluded that social learning can be adaptive. Enquist *et al.*[3] advocated a learning form called critical social learning, which is social learning supplemented by individual learning. They discussed using rate equations and succeeded in solving the paradox. Rendell *et al.*[4] studied the relative merits of several learning strategies by using a spatially explicit stochastic model.

The concept of adaptive information filtering[3, 6] has been proposed as key to the effective working of social learning. It indicates that each member effectively learns good-quality information provided by other members. For example, in a famous tournament by Rendell *et al.*[6], discountmachine that did the most effective social learning won over the other strategies that combined individual learning and social learning.

In this study, we propose a stochastic model to solve Rogers' paradox in the framework of a restless multi-armed bandit (rMAB) used in that tournament. The objective of this study is to analyse equilibrium social learning in an rMAB. An rMAB is analogous to the "one-armed bandit" slot machine but with multiple "arms", each with a distinct payoff. We call an arm with a high payoff a good arm. The term "restless" means that the payoffs change randomly. Agents maximise their payoffs by exploiting an arm, searching for a good arm at random (individual learning), or copying an arm exploited by other agents (social learning). Because rMAB is simple in structure and its generality, we believe that it is an appropriate framework to consider Rogers' paradox.

As a model for social-learning collectives, Bolton and Harris studied an agent system in a multi-armed bandit[7]. They assumed that the agents could know all information of other agents and obtained a socially optimal experiment (learning) strategy. In the present study, we consider the bounded rationality of agents, who can access the

[1]Department of Mathematics, Faculty of Science, Shinshu University, Asahi 3-1-1, Matsumoto, Nagano, 390-8621, Japan. [2]Fintech Lab. LLC Meguro, Tokyo, 153-0051, Japan. [3]Department of Physics, Faculty of Science, Kitasato University, Kitasato 1-15-1, Sagamihara, Kanagawa, 252-0373, Japan. Correspondence and requests for materials should be addressed to K.N. (email: nakayama@math.shinshu-u.ac.jp)
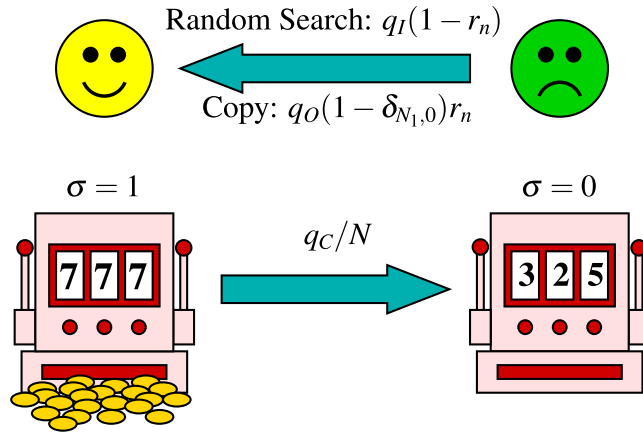
Figure 1. An agent and an rMAB. The number of agents who know a good arm is denoted as N_1.

results of their respective choices only. In addition, we assume that the environment (i.e., the rMAB) changes randomly. We obtain the socially optimal and equilibrium learning strategies.

Model

We make the model as simple as possible and incorporate the property of adaptive filtering of information into it. A mathematical overview of the model is given in the Methods section.

The rMAB has only one good arm and infinitely many bad arms. There are N agents labeled by $n = 1, \ldots, N$. In each turn, an agent (say, agent n) is randomly chosen. He/she exploits his/her arm and obtains payoff 1 if he/she knows a good arm. If he/she does not know a good arm, he/she randomly searches for it (individual learning) with probability $1 - r_n$, or copies the information of other agents' good arms (social learning) with probability r_n. In the random search, the probability that he/she successfully finds a good arm is denoted as q_I. On the other hand, we assume that the copy process succeeds with probability q_O[8] if there is at least one agent who knows a good arm, and fails if no agent knows a good arm. Then, with probability q_C/N, the good arm changes to a bad one and another good arm appears. If a good arm changes to a bad one, the agents who knew the arm are forced to forget it and to know a bad one. See Fig. 1. The difference with our previous model[8] is that there are M good arms in the previous model, whereas in the present model there is only one good arm.

Let σ_n be a random variable defined by

$$\sigma_n = \begin{cases} 1, & \text{if agent } n \text{ knows a good arm,} \\ 0, & \text{if agent } n \text{ does not know a good arm.} \end{cases} \tag{1}$$

This is simply the payoff for agent n. For each turn t, we have a joint probability function $P(\sigma_1, \ldots, \sigma_N | t)$, which evolves in t according to the aforementioned rule. To exclude trivial results, we assume that q_C, q_I and q_O are positive and that r_ns are less than 1[8]. Then, in the long run, we have the unique steady probability function $P(\sigma_1, \cdots, \sigma_N) = \lim_{t \to \infty} P(\sigma_1, \cdots, \sigma_N | t)$. Now, we shall introduce the expected payoff for each agent in the steady state,

$$w_n = E[\sigma_n] = \sum_{\sigma_1=0,1} \cdots \sum_{\sigma_N=0,1} P(\sigma_1, \cdots, \sigma_N) \sigma_n, \quad n = 1, \cdots, N. \tag{2}$$
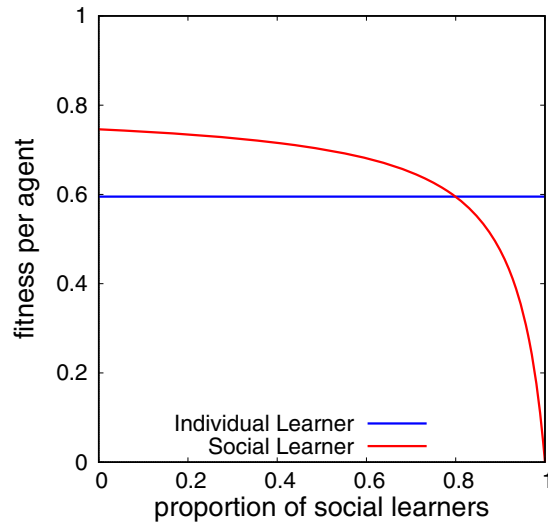
This quantity depends on parameters N, q_C, q_I, q_O, and r_ns. We regard w_n mainly as a function of r_ns. We denote this function by $w(r_n, \bar{r}_n)$, where

$$w(r, \bar{r}) = \frac{1}{a + q_I + (q_O - q_I)r} \left\{ q_I + (q_O - q_I)r - \frac{a q_O r}{a + (N-1)q_I(1 - \bar{r}) + q_I(1 - r)} \right\}, \tag{3}$$

$$a = \frac{q_C}{1 - q_C/N}, \tag{4}$$

$$\bar{r}_n = \frac{1}{N-1} \sum_{k \neq n} r_k = \frac{1}{N-1} \left\{ \sum_{k=1}^{N} r_k - r_n \right\}. \tag{5}$$

Thus, we have $w_n = w(r_n, \bar{r}_n)$ for each $n = 1, \ldots, N$.
In this study, we treat w_n as the fitness for agent n.

Figure 2. Plots of w_I and w_S. Parameters: $N = 10$, $q_C = 0.2$, $q_I = 0.3$, $q_O = 0.8$.

Results and Discussion

Pure Strategies and Rogers' Paradox.   In the present study, the strategy of agent n refers to the social learning probability, r_n. We call $r_n = 0, 1$ as pure strategies and $0 < r_n < 1$ as mixed strategies.

First, we confirm that Rogers' paradox occurs when agents adopt pure strategies. We shall divide N agents into two groups. The first group consists of N_I individual learners ($r_k = 0, k = 1, \ldots, N_I$). The second group consists of $N_S = N - N_I$ social learners ($r_k = 1, k = N_I + 1, \ldots, N_I + N_S$). The corresponding fitness per agent, which we denote respectively as w_I and w_S, are given by

$$w_I = \frac{q_I}{a + q_I}, \qquad w_S = \frac{N_I q_I q_O}{(a + N_I q_I)(a + q_O)},$$ (6)

where a is defined in equation (4).

When $q_O \leq q_I$, we have $w_I > w_S$. Therefore, in this case, individual learning is always favourable over social learning.

Now, we consider the $q_O > q_I$ case. Figure 2 is the plot of w_I and w_S for sufficiently large N.

When the proportion of social learners is small, social learning is effective. However, as the proportion of social learners increases, w_S monotonically decreases and tends to zero. Thus, Rogers' paradox occurs.

It is important to note that $w_I < w_S$ is true when N_I/N is finite, with a sufficiently large N. This is because, as $N \to \infty$, we have $w_I \to q_I/(q_C + q_I)$ and $w_S \to q_O/(q_C + q_O)$.

Nash Equilibrium and Rogers' Paradox.   Let us assume that each agent adopts a mixed strategy, that is, for each $n = 1, \ldots, N$, the social-learning probability, r_n, is an arbitrary number between 0 and 1. This means that agent n performs social learning with probability r_n and individual learning with probability $1 - r_n$. The learning mode that he/she chooses would be decided stochastically and automatically.

We consider the N-tuple, $(r_1, \ldots, r_N)$, of the social-learning probabilities. This is a point in the N-dimensional unit cube $J = [0, 1] \times \ldots \times [0, 1]$. J is regarded as the space of N-tuples of mixed strategies. For each point in J, a joint probability function $P(\sigma_1, \ldots, \sigma_N)$ is determined and an N-tuple, $(w_1, \ldots, w_N)$, of the fitness functions of the agents is calculated.

Now, imagine that agent n maximises w_n by adjusting r_n for fixed r_ks ($k \neq n$). It is not difficult to show that the maximum point is unique (Fig. 3) and expressed as
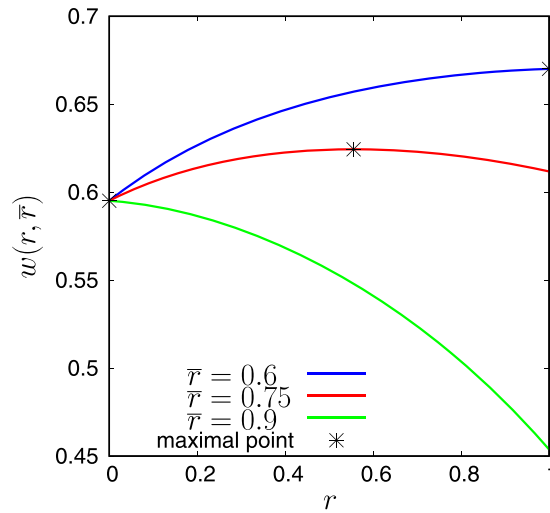
$$r_n = f(\bar{r}_n),$$ (7)

where

$$f(r) = \begin{cases} 0 & q_O \leq q_I, \\ \min(1, \ \max(0, \bar{f}(r))), & q_O > q_I, \end{cases}$$ (8)

$$\bar{f}(r) = \frac{-\zeta + \sqrt{q_O(N-1)(1-r)}\sqrt{\zeta}}{q_O - q_I},$$ (9)

$$\zeta \equiv a + q_I + q_I(N-1)(1-r).$$ (10)

Figure 3. Plots of the function $w(r, \bar{r})$ versus r for various values of $\bar{r}$. Parameters: $N = 10$, $q_C = 0.2$, $q_I = 0.3$, $q_O = 0.8$.

We note that $f(r) \to 0$ as $q_O \to q_I + 0$. Next, we introduce the function,

$$F(r_1, \cdots, r_N) = (f(\bar{r}_1), \cdots, f(\bar{r}_N)). \tag{11}$$

This is a continuous function mapping from the N-dimensional unit cube J into itself. As shown in the Methods section, the fixed point of F is unique and is on the diagonal line of J,

$$r_1 = \cdots = r_N = r_{Nash}, \tag{12}$$

where r_{Nash} is a function of q_C, q_I, q_O and N. The value of r_{Nash} is explicitly given by

$$r_{Nash} = \begin{cases} 1 - \eta, & (q_O - q_I)N > a + q_O, \\ 0, & (q_O - q_I)N \leq a + q_O, \end{cases} \tag{13}$$

where

$$\eta \equiv \frac{2(a + q_O)^2}{(q_O - q_I N)(a + q_O) + (aN + q_O)(q_O - q_I) + \sqrt{D_1}}, \tag{14}$$

$$D_1 = (N - 1)q_O\{ -(4aN + 3q_O N + q_O)q_I^2 \\ + 2(3aq_O N + 2q_O^2 N - 2a^2 - 3aq_O)q_I + aq_O(aN + 3a + 4q_O)\}. \tag{15}$$

The entity r_{Nash} has the following properties (see the Methods section): (i) $0 \leq r_{Nash} < 1$, (ii) $r_{Nash} \to 0$ as $(q_O - q_I)N - (a + q_O) \to 0$, and (iii) the fixed point $(r_{Nash}, \cdots, r_{Nash})$ is the unique Nash equilibrium point in J. Figure 4 is a schematic explanation of the Nash equilibrium point.

Moreover, the corresponding mixed strategy is an evolutionarily stable strategy (ESS)[9] because the fixed point is a Nash equilibrium point in the strong sense,

$$w(r_{Nash}, r_{Nash}) > w(r, r_{Nash}), \quad \text{for all } r \neq r_{Nash}. \tag{16}$$

Further, it is an ESS in the sense of Thomas[10], because the inequality,

$$w(r_{Nash}, r) > w(r, r), \quad \text{for all } r \neq r_{Nash}, \tag{17}$$

is true.

Now, we consider the two fitness functions, w_I and $w_N = w(r_{Nash}, r_{Nash})$. As shown in the Methods section, the inequality $w_N > w_I$ is correct if and only if $(q_O - q_I)N > a + q_O$. See also Fig. 5. The Nash equilibrium point is usually regarded as a stable point in the sense that no agent has an intention to change his/her strategy. Therefore, this inequality claims that the mixed strategy $r_n = r_{Nash}$ ($n = 1, \ldots, N$) can outperform the pure strategy of individual learning. This solves Rogers' paradox. We note that the Nash equilibrium point is realised as a mixed strategy of social learning and individual learning.

Pareto Optimality. Pareto optimality is an important concept alongside Nash equilibrium. Thus, we consider Pareto optimality in our model. We shall adopt a *natural* definition of the Pareto-optimal point in J as the
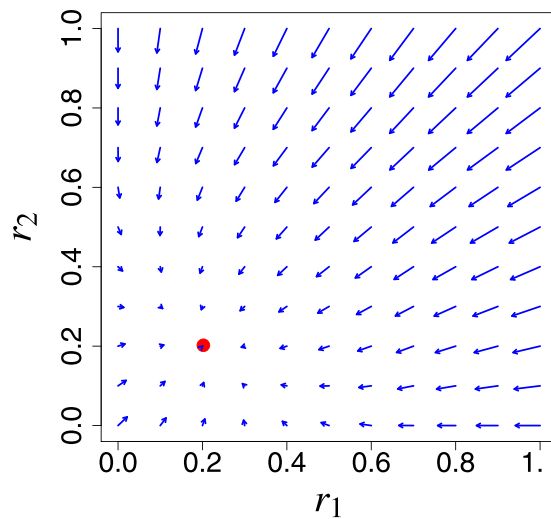
Figure 4. Plot of the vector field $(f(\bar{r}_1) - r_1, f(\bar{r}_2) - r_2)$. The red spot is the Nash equilibrium point. Parameters: $q_C = 0.1$, $q_I = 0.1$, $q_O = 0.8$, and $N = 2$.
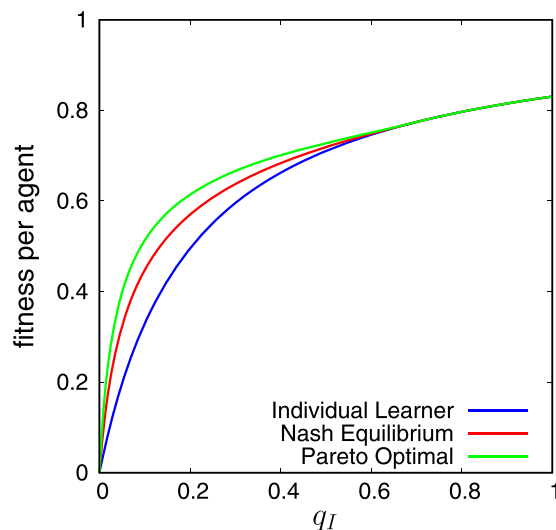
Figure 5. Plots of w_I, w_N, and w_P as functions of q_I. Parameters: $N = 10$, $q_C = 0.2$, $q_O = 0.8$. Thus, $((N - 1)q_O - a)/N \simeq 0.7$.

maximum point of the function, $\sum_{k=1}^{n} w_k$. We can show that the maximum point is unique and is on the diagonal line of J,

$$r_1 = \cdots = r_N = r_{\text{Pareto}}, \tag{18}$$

where r_{Pareto} is a function of q_C, q_I, q_O, and N. The value of r_{Pareto} is explicitly given by

$$r_{\text{Pareto}} = \begin{cases} \dfrac{(a + q_I N)X - (a + q_I)Y}{q_I NX + (q_O - q_I)Y}, & (q_O - q_I)N > a + q_O, \\ 0, & (q_O - q_I)N \leq a + q_O, \end{cases} \tag{19}$$

where

$$X = \sqrt{(N - 1)(a + q_O)(q_O - q_I)}, \tag{20}$$

$$Y = \sqrt{(a + Nq_I)Nq_O}. \tag{21}$$

Further, r_{Pareto} has the following properties: (i) $0 \leq r_{Pareto} < 1$, (ii) $r_{Pareto} \to 0$ as $(q_O - q_I)N - (a + q_O) \to 0$, (iii) $r_{Pareto} < r_{Nash}$ if and only if $(q_O - q_I)N > a + q_O$ (see the Methods section), and (iv) the point $(r_{Pareto}, \ldots, r_{Pareto})$ is the Pareto-optimal point in J. Here, by Pareto optimality, we imply that the statement "if an agent succeeds to increase his/her fitness by changing his/her social-learning probability from r_{Pareto} to $r_{Pareto} + \delta r$ by $\delta r \neq 0$, then another agent's fitness certainly decreases" is true. Such a δr exists when $r_{Pareto} > 0$ and no δr exists when $r_{Pareto} = 0$. The statement is correct in both cases.

We define the Pareto fitness function, $w_P = f(r_{Pareto}, r_{Pareto})$. Then, we have the inequality $w_P > w_N$ if and only if $(q_O - q_I)N > a + q_O$ (see Fig. 5 and the Methods section). This is trivial by the definition of the Pareto-optimal point. Thus, we have established the relation among fitness functions,

$$\begin{cases} w_P > w_N > w_I, & (q_O - q_I)N > a + q_O, \\ w_P = w_N = w_I, & (q_O - q_I)N \leq a + q_O. \end{cases} \tag{22}$$

Concluding Remarks

We have proposed a stochastic model of N agents and an rMAB. The unique Nash equilibrium point in the mixed strategy space J has been presented and shown to be an ESS in the sense of Thomas[10]. The corresponding fitness w_N per agent is greater than the fitness w_I for an individual learner. This solves Rogers' paradox.

In this study, we concentrated on steady states. This is valid if the system relaxes quickly to the steady state (see the Methods section). However, if r_ns change faster than the relaxation to the steady state, it is an introduction of non-trivial dynamics. It may be possible that our system has a nice dynamics possessing the stable Nash equilibrium point.

As a future research subject, we propose an experimental study of human collectives in rMAB. There have been several attempts in this direction[11–13], whose target has been the improvement of performance by social learning, that is, collective intelligence effect. Since we have shown that there is an ESS Nash equilibrium in the social-learning agents system in rMAB, it is interesting to experimentally examine whether the prediction is realised. As a first step, the interactive rMAB game might be a suitable environment where one human competes with many other mixed-strategy agents and $r = r_{Nash}$. We can check whether the social-learning rate of people is the same with r_{Nash}. Second, when many people compete, the Nash equilibrium emerges as the model parameter q_I changes. Meanwhile, we might be able to detect some phase-transitive behaviour[8].

As for theoretical research, the stage of our analysis is far from mature. In the present work, we have studied the game of rMAB in the steady state of the system. However, when the relaxation time of the system discussed in the Methods section is not small enough, the assumption of steadiness is unrealistic in the laboratory experiment. Thus, we need to develop a t-dependent theory. It might be a difficult problem. We believe that the research direction is fruitful.

Methods
Mathematical Overview of the Model. For simplicity we use the following notation,

$$\vec{\sigma} = (\sigma_1, \cdots, \sigma_N), \quad \vec{0} = (0, \cdots, 0), \quad \vec{e}_n = (0, \cdots, 0, \overset{n}{\check{1}}, 0, \cdots, 0), \quad \delta_{\vec{\sigma}', \vec{\sigma}} = \prod_{n=1}^{N} \delta_{\sigma_n', \sigma_n}. \tag{23}$$

Our model develops in t according to an agent action and the subsequent state change of the rMAB. This is a Markov process[14]. The probability of change $\vec{\sigma} \to \vec{\sigma}'$ is described by the transition probability matrix[14],

$$T(\vec{\sigma}'|\vec{\sigma}) = \left(1 - \frac{q_C}{N}\right)\left\{\left(1 - \sum_{n=1}^{N} p_n(\vec{\sigma})\right)\delta_{\vec{\sigma}', \vec{\sigma}} + \sum_{n=1}^{N} p_n(\vec{\sigma})\delta_{\vec{\sigma}', \vec{\sigma} + \vec{e}_n}\right\} + \frac{q_C}{N}\delta_{\vec{\sigma}', \vec{0}}, \tag{24}$$

where

$$p_n(\vec{\sigma}) = \frac{\delta_{\sigma_n, 0}}{N}\left\{r_n(1 - \delta_{N_1, 0})q_O + (1 - r_n)q_I\right\}, \quad N_1 = \sum_{n=1}^{N} \sigma_n. \tag{25}$$

The joint probability function $P(\vec{\sigma}|t) = P(\sigma_1, \ldots, \sigma_N|t)$ satisfies the Chapman-Kolmogorov equation[14],

$$P(\vec{\sigma}|t + 1) = \sum_{\vec{\sigma}'} T(\vec{\sigma}|\vec{\sigma}')P(\vec{\sigma}'|t). \tag{26}$$

Our assumption is that $q_C, q_I, q_O > 0$ and $r_n < 1$ ($n = 1, \ldots, N$). In this case, the matrix T is shown to be irreducible and primitive[15]. Then, the Perron-Frobenius theory[15] ensures that (i) $\lambda_1 = 1$ is an eigenvalue of T of multiplicity 1 and the steady probability function $P(\vec{\sigma})$ is a corresponding eigenvector, (ii) the set $\{|\lambda_i|\}_{i \geq 2}$ of absolute values of eigenvalues of T other than λ_1 has an upper bound $\rho < 1$. When r_ns are fixed, we have the time-homogeneous Markov process[14], that is, the matrix T does not depend on t. Therefore, for any initial probability function $P(\vec{\sigma}|0)$, we have the unique limit $P(\vec{\sigma}) = \lim_{t \to \infty} P(\vec{\sigma}|t)$. Then, it is not difficult to derive equation (3) using $P(\vec{\sigma})$.
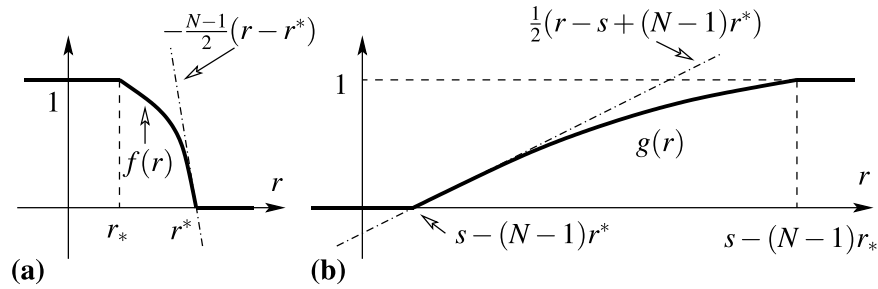
Figure 6. (**a**) Plot of the function $f(r)$. (**b**) Plot of the function $g(r)$.

The convergence $P(\vec{\sigma}|t) \to P(\vec{\sigma})$ is exponential, $|P(\vec{\sigma}|t) - P(\vec{\sigma})| \sim \rho^t$. This means that the relaxation time is $\tau = -1/\log\rho^{-1}$. Thus, when no agent changes his/her social learning probability over a much longer period than τ, the fitness per agent per turn is almost exactly equal to the value of the function w in equation (3).

Existence of a Fixed Point of F. Since the N-dimensional cube $J = [0, 1] \times \ldots \times [0, 1]$ is a compact, convex set and F is a continuous function mapping from J into itself, Brouwer's fixed-point theorem[16] guarantees that there exists a fixed point of F in J.

A Fixed Point of F is a Nash Equilibrium Point, and Vice Versa. Let $(r_1, \ldots, r_N)$ be a fixed point of F, that is, $r_n = f(\bar{r}_n)$ for each $n = 1, \ldots, N$. Since $r = f(\bar{r}_n)$ is the unique maximal point of $w(r, \bar{r}_n)$, we have $w(r_n + \delta r, \bar{r}_n) < w(r_n, \bar{r}_n)$ for each $n = 1, \ldots, N$ when $\delta r \neq 0$. Thus, $(r_1, \ldots, r_N)$ is a Nash equilibrium point. Conversely, let $(r_1, \ldots, r_N)$ be a Nash equilibrium point, that is, $r = r_n$ is a maximal point of $w(r, \bar{r}_n)$ for each $n = 1, \ldots, N$. Since $r = f(\bar{r}_n)$ is the unique maximal point of $w(r, \bar{r}_n)$ (see Fig. 3), we have $r_n = f(\bar{r}_n)$. Thus, $(r_1, \ldots, r_N)$ is a fixed point of F.

Uniqueness of the Fixed Point of F. When $q_O \leq q_I$, we have the unique fixed point $(0, \ldots, 0)$.

Next, we consider the $q_O > q_I$ case.

Let $(r_1, \ldots, r_N)$ be a fixed point of F. Since $\bar{r}_n = (s - r_n)/(N - 1)$, $s = \sum_{k=1}^{N} r_k$, all the r_ns satisfy the common relation,

$$r = g(r) \equiv f\left(\frac{s - r}{N - 1}\right).$$
(27)

Figure 6(b) is a plot of the function $g(r)$.

This is a strictly increasing concave function for $s - (N-1)r^* \leq r \leq s - (N-1)r_*$, where

$$r^* = 1 - \frac{a + q_I}{(N - 1)(q_O - q_I)},$$
(28)

$$r_* = 1 - \frac{-(q_O(a - q_I) - 2aq_I) + \sqrt{D_2}}{2(N - 1)q_I(q_O - q_I)},$$
(29)

$$D_2 = (q_O(a - q_I) - 2aq_I)^2 + 4q_I(q_O - q_I)(a + q_O)^2.$$
(30)

It is not difficult to show that $r_* < r^* < 1$. The maximum value of the derivative $g'(r)$ is $1/2$, which is realised at $r = s - (N-1)r^*$. Thus, $\tilde{g}(r) = r - g(r)$ is a strictly increasing function such that $\tilde{g}(0) \leq 0 \leq \tilde{g}(1)$. Therefore, there is only one zero, r_0, of the function $\tilde{g}(r)$ in the interval $0 \leq r \leq 1$. Then, we conclude that $r_1 = \ldots = r_N = r_0$.

Now we have $s = Nr_0$. Therefore, r_0 is a solution of the equation,

$$r = f(r).$$
(31)

Figure 6(a) is a plot of the function $f(r)$. The function $f(r)$ is a decreasing function. Thus, $h(r) = r - f(r)$ is a strictly increasing function such that $h(0) \leq 0 \leq h(1)$. Therefore, the function $h(r)$ possesses only one zero, r_{Nash}, such that $0 \leq r_{Nash} < 1$. Thus, we have $r_0 = r_{Nash}$. This proves the uniqueness of the Nash equilibrium point.

Inequality $w_P > w_N > w_I$. It is sufficient to consider the $(q_O - q_I)N > a + q_O$ case. Then, r_{Nash} satisfies $r = \bar{f}(r)$. We introduce the following function,

$$k(u) = (a + q_O)u^2 - \{(q_O - q_I N)(a + q_O) + (aN + q_O)(q_O - q_I)\}u + (q_O - q_I)(q_O - q_I N^2).$$
(32)

It is not difficult to check that $1/(1 - r_{Nash})$ is the larger root of $k(u)$. We note that $k(1) < 0$.

Next, we define r_I as

$$r_I = 1 - \frac{a + q_O}{(q_O - q_I)N}.$$

(33)

This entity has the following properties: (i) $0 < r_I < 1$, (ii) $w(r_I, r_I) = w_I$, and (iii) $k\,(1/(1 - r_I)) > 0$. On the other hand, it is elementary to show that $k\,(1/(1 - r_{Pareto})) < 0$. Thus, we conclude that $r_{Pareto} < r_{Nash} < r_I$.

Now, r_{Pareto} is the maximal point of $w(r, r)$. Therefore, we have the inequality $w(r_{Pareto}, r_{Pareto}) > w(r_{Nash}, r_{Nash}) > w(r_I, r_I)$, that is, $w_P > w_N > w_I$.

References

1. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (University of Chicago Press, Chicago, 1985).
2. Boyd, R. & Richerson, P. J. Why does culture increase human adaptability? *Ethol. Sociobiol.* **16**, 125–143, doi:10.1016/0162-3095(94)00073-G (1995).
3. Enquist, M., Eriksson, K. & Ghirlanda, S. Critical social learning: A solution to Rogers's paradox of nonadaptive culture. *Am. Anthropol.* **109**, 727–734, doi:10.1525/aa.2007.109.4.727 (2007).
4. Rendell, L., Fogarty, L. & Laland, K. N. Rogers' paradox recast and resolved: Population structure and the evolution of social learning strategies. *Evolution* **64**, 534–548, doi:10.1111/j.1558-5646.2009.00817.x (2010).
5. Rogers, A. R. Does biology constrain culture? *Am. Anthropol.* **90**, 819–831, doi:10.1525/aa.1988.90.4.02a00030 (1988).
6. Rendell, L. *et al*. Why copy others? Insights from the social learning strategies tournament. *Science* **328**, 208–213, doi:10.1126/science.1184719 (2010).
7. Bolton, P. & Harris, C. Strategic experimentation. *Econometrica* **67**, 349–374, doi:10.1111/1468-0262.00022 (1999).
8. Mori, S., Nakayama, K. & Hisakado, M. Phase transition of social learning collectives and the echo chamber. *Phys. Rev. E* **94**, 052301, doi:10.1103/PhysRevE.94.052301 (2016).
9. Maynard-Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, 1982).
10. Thomas, B. Evolutionary stability: states and strategies. *Theor. Popul. Biol.* **26**, 49–67 (1984).
11. Toyokawa, W., Kim, H. & Kameda, T. Human collective intelligence under dual exploration-exploitation dilemmas. *PloS One* **9**, e95789, doi:10.1371/journal.pone.0095789 (2014).
12. Kameda, T. & Nakanishi, D. Cost-benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects. *Evol. Hum. Behav.* **23**, 373–393, doi:10.1016/S1090-5138(02)00101-0 (2002).
13. Yoshida, S., Hisakado, M. & Mori, S. Interactive restless multi-armed bandit game and swarm intelligence effect. *New Generat. Comput.* **34**, 291–306, doi:10.1007/s00354-016-0306-y (2016).
14. Stroock, D. W. *An Introduction to Markov Processes* (Springer-Verlag, Heidelberg, 2014).
15. Meyer, C. D. *Matrix Analysis and Linear Algebra* (SIAM, 2000).
16. Granas, A. & Dugundji, J. *Fixed Point Theory* (Springer-Verlag, New York, 2003).

Acknowledgements

Author Contributions

S.M. and M.H. conceived the model. K.N. performed a theoretical analysis. All authors contributed to analysing and interpreting the results and to writing the manuscript.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.