

# SCIENTIFIC REPORTS



OPEN

## Stochastic evolution in populations of ideas

Robin Nicole<sup>1</sup>, Peter Sollich<sup>1</sup> & Tobias Galla<sup>2</sup>

Received: 10 November 2016

Accepted: 07 December 2016

Published: 18 January 2017

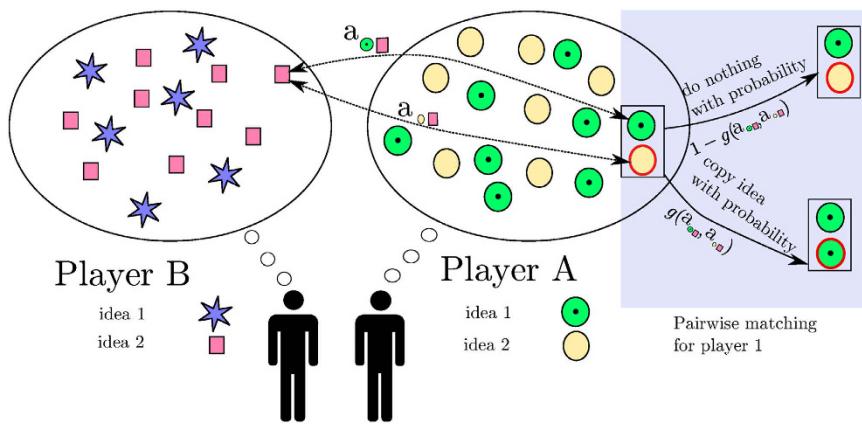
**It is known that learning of players who interact in a repeated game can be interpreted as an evolutionary process in a population of ideas. These analogies have so far mostly been established in deterministic models, and memory loss in learning has been seen to act similarly to mutation in evolution. We here propose a representation of reinforcement learning as a stochastic process in finite ‘populations of ideas’. The resulting birth-death dynamics has absorbing states and allows for the extinction or fixation of ideas, marking a key difference to mutation-selection processes in finite populations. We characterize the outcome of evolution in populations of ideas for several classes of symmetric and asymmetric games.**

The study of games in non-cooperative game theory has traditionally focused on the analysis of their equilibrium points, in particular the celebrated Nash equilibria<sup>1,2</sup>. These are the points in strategy space that fully rational players choose, based on full information of the game and assuming that their opponents act fully rationally as well. At a Nash point no player can increase their payoff by *unilaterally* changing their strategy. These ideas provide a natural first approach to the analysis of games, and they are mathematically convenient as they do not involve any actual dynamics. On the other hand the scope of such equilibrium concepts is naturally limited. The question of how players would find optimal points in strategy space is not asked, let alone answered. Experiments in behavioural economics show that real-world players do not behave fully rationally in repeated games, and suggest that inductive learning from past experience may be a better model than the assumption of full rationality<sup>3,4</sup>.

In many models of dynamic learning, players do not find the mutually optimal strategy immediately; in fact they potentially never do. Instead they initially try out the different actions available to them, and attempt to learn from past experience. Players assess the success or otherwise of individual strategies and then choose those that worked well in the past. Their opponents adapt as well, and strategies that may have performed well previously can become less successful when the opponents’ propensities have changed. This generates a coupled dynamics between the players, and it is not clear a-priori if and when such dynamics converge to Nash points. Indeed, work on games of low and high complexity has suggested that learning may result in chaotic motion<sup>5–8</sup>, in some cases with very high dimensional attractors. Situations in which systems of this type settle down to unique well-defined fixed points then seem to be the exception rather than the rule.

Learning and adaptation based on past experience can be interpreted as an evolutionary process of ‘ideas’ in the minds of the players. Börgers and Sarin, for example, write<sup>9</sup> ‘*Decision makers are usually not completely committed to just one set of ideas [...]. Rather [...] several possible ways of behaving are present in their minds simultaneously. Which of these predominate, and which are given less attention, depends on the experiences of the individual. The change which the “population of ideas” in the decision maker’s mind undergoes may be analogous to biological evolution*’. Similar approaches have also been used in models of language evolution; see e.g. Blythe *et al.*<sup>10</sup>. In the context of a game the evolutionary process in a population of ideas broadly works as follows: each player carries in his or her mind a mixed populations of ideas. These represent the different actions (pure strategies) he or she can take in the game. Different ideas will be present in the player’s mind in different proportions. At each instance of the game each player pulls out one idea (action) out of their mind at random, and uses it in the game. The ideas that are more frequent in the player’s mind will be used more often than those which are present less in the population. The composition of the player’s mind thus represents their mixed strategy. Over time the player learns from past experience, and the population of ideas in their mind undergoes an evolutionary process: less successful ideas are displaced by more successful strategies. This is illustrated in Fig. 1, and akin to well-known birth-death processes in evolutionary dynamics<sup>11</sup>. It is hence no surprise that the equations governing multi-player learning can be very similar to those used to model evolutionary dynamics<sup>9,12</sup>.

<sup>1</sup>Department of Mathematics, King’s College London, Strand, London, WC2R 2LS, United Kingdom. <sup>2</sup>Theoretical Physics, School of Physics and Astronomy, The University of Manchester, Manchester M13 9PL, United Kingdom. Correspondence and requests for materials should be addressed to R.N. (email: robin.nicole@kcl.ac.uk) or P.S. (email: peter.sollich@kcl.ac.uk) or T.G. (email: tobias.galla@manchester.ac.uk)



**Figure 1. Illustration of the evolutionary process that occurs in a population of ideas: two ideas in the mind of player A are selected ( $\odot$  and  $\circlearrowleft$ ) as indicated by the rectangle on the right.** Both ideas play against the same randomly chosen adversary idea (here  $\square$ ) in the population of ideas of player B and the relevant payoffs are recorded, here denoted  $a\odot\square$  and  $a\circlearrowleft\square$ . Idea  $\odot$  is switched to  $\circlearrowleft$  with probability  $g(a\odot\square, a\circlearrowleft\square)$  depending on these payoffs. An analogous process occurs in the population of ideas of player B. The non-negative function  $g(\cdot, \cdot)$  is increasing in the first argument, and decreasing in the second. It defines the mechanics of the evolutionary process. See also the text in Secs. 1 and 2 for further details.

Most existing analogies between learning and evolutionary dynamics are at the level of deterministic differential equations though, formally describing the dynamics of infinite populations. At the same time, evolutionary dynamics in finite populations shows several phenomena that arise solely from intrinsic stochasticity. These effects include noise-driven fixation and extinction, which are not captured by deterministic approaches. A substantial amount of work is available on the dynamics of stochastic birth-death processes, including an analytical formalism to compute fixation probabilities and the times to fixation<sup>11,13–15</sup>.

The main purpose of the present work is to develop a microscopic representation of reinforcement learning as a stochastic evolutionary process in a finite population of ideas. Ideas in this description are members of a finite populations, and undergo a birth-death process. This approach allows us to establish the analogy between learning and evolution at the level of *stochastic* population dynamics. More specifically we will define the transition rates of a birth-death process in a population of ideas, such that the deterministic description in the limit of infinite populations reproduces the so-called Sato-Crutchfield differential equations<sup>12,16</sup>. We show that the notion of reproductive fitness needs to be augmented by an entropic restoring force to capture weak decision preferences and/or memory loss in game learning. These restoring forces play a role similar to that of mutation in evolutionary dynamics. Crucially, however, the birth-death dynamics in finite populations of ideas has absorbing states so that ideas can go extinct or reach fixation. This marks a key difference compared to mutation-selection dynamics, where there are no absorbing states.

The remainder of the paper is organized as follows. In Sec. 1 we briefly summarize the mathematics of the standard replicator dynamics and of the reinforcement learning dynamics we use as a basis for the evolution of ideas. In Sec. 1 we then introduce the birth-death process for finite populations of ideas, and we study its properties for simple symmetric games. In Sec. 2 we extend the analysis to two-player learning in asymmetric games. Finally in Sec. 13 we collect our conclusions and present an outlook towards future work. Further technical details of our analysis can be found in the Supplementary Material.

## Deterministic evolutionary dynamics and adaptive learning

**Evolutionary dynamics and replicator equations.** *Single-population replicator equations.* The evolutionary dynamics of interacting individuals in infinite populations is frequently described by replicator or replicator-mutator equations. These are deterministic ordinary differential equations. We focus on a population of individuals of  $S$  different types,  $i = 1, \dots, S$ , and write  $x_i(t)$  for the fraction of individuals of type  $i$  in the population at time  $t$ , and  $\mathbf{x} = (x_1, \dots, x_S)$ . At all times  $\sum_i x_i(t) = 1$ . We assume that individuals interact in a symmetric two-player normal form game<sup>17</sup>. This is specified by a payoff matrix  $\mathbf{A} = (a_{ij})$ . The entry  $a_{ij}$  is the payoff to an individual of type  $i$  in an interaction with an individual of type  $j$ . The setup of a symmetric game is not to be confused with a game for which the payoff matrix is symmetric, i.e. its own transpose.

The average payoff per game to an individual of type  $i$  in a population of composition  $\mathbf{x}$  is given by  $\pi_i(\mathbf{x}) = \sum_j a_{ij} x_j$ . In order to keep the notation compact, we will omit the argument  $\mathbf{x}$  in the following. The standard replicator equations are then given by<sup>17</sup>

$$\dot{x}_i = x_i(\pi_i - \bar{\pi}), \quad (1)$$

with  $\bar{\pi} = \sum_j x_j \pi_j$ . These dynamics can be derived from a birth-death process in the limit of an infinite population. This will be discussed in more detail below.

*Two-population replicator dynamics.* The case of asymmetric games refers to situations in which different individuals take on different roles, e.g. male and female in Dawkin's battle of the sexes<sup>18</sup>, or buyers and sellers in a stock market. In this case individuals belonging to different populations. In two-population replicator systems the fitness of individuals in population  $A$  is determined by their interaction with individuals in population  $B$ , and vice versa. Selection and evolution then occur within each population<sup>17</sup>. This leads to the following two-population replicator dynamics:

$$\dot{x}_i^A = x_i^A (\pi_i^A - \pi^A), \quad (2)$$

$$\dot{x}_i^B = x_i^B (\pi_i^B - \pi^B), \quad (3)$$

where  $x_i^A$  is the frequency with which individuals of type  $i$  occur in population  $A$ , and  $x_i^B$  the frequency with which the  $i$ -th type occurs in population  $B$ . It is important to note that the label  $i$  in either population is a simple numbering of pure strategies, e.g. in Dawkin's battle of the sexes  $i=1, 2$  in the populations of males may refer to 'faithful' and 'philanderer', and in the population of females the same labels may refer to 'coy' and 'fast'<sup>18,19</sup>.

In the above equations we have used the shorthands,

$$\pi_i^A = \sum_j a_{ij} x_j^B, \quad (4)$$

$$\pi_i^B = \sum_j b_{ij} x_j^A, \quad (5)$$

as well as  $\pi^A = \sum_i \pi_i^A x_i^A$  and similarly  $\pi^B = \sum_i \pi_i^B x_i^B$ .

**Discrete-time Sato-Crutchfield learning.** Following Sato and Crutchfield's approach<sup>12,16</sup> we consider two players, labelled  $A$  and  $B$  repeatedly playing an asymmetric game with payoff matrices  $\mathbf{A}$  ( $\mathbf{B}$ ) for player  $A$  ( $B$ ). For simplicity, we will assume that both players have the same number  $S$  of actions available, but the extension to the more general case is straightforward<sup>16</sup>. Hence  $\mathbf{A}$  and  $\mathbf{B}$  will be  $S \times S$  matrices, with entries denoted  $a_{ij}$  and  $b_{ji}$ ,  $i, j = 1, \dots, S$ . As implied in (4) and (5) above,  $a_{ij}$  is the payoff to player  $A$  if she chooses action  $i$  while player  $B$  plays action  $j$ ;  $b_{ji}$  is the payoff to player  $B$  in this situation.

At each instance of the game, each player  $\mu \in \{A, B\}$  will choose one action. In order to monitor the relative success of the different actions, each player holds an 'attraction' for each action. We will write  $Q_i^\mu(t)$  for the attraction player  $\mu$  has for action  $i$  at time  $t$ . Sato-Crutchfield learning assumes a soft-max (or logit) rule to convert a set of attractions  $Q_1^\mu, \dots, Q_S^\mu$  into a mixed strategy,

$$x_i^\mu = \frac{\exp(\Gamma Q_i^\mu)}{\sum_j \exp(\Gamma Q_j^\mu)}. \quad (6)$$

The parameter  $\Gamma \geq 0$  represents the intensity of choice<sup>3,5,20</sup>. When  $\Gamma = 0$  attractions play no role and players choose their actions with equal probability. In the limit  $\Gamma \rightarrow \infty$  players play a pure strategy that always chooses the action with the highest attraction.

In the original articles by Sato and Crutchfield<sup>12,16</sup>, the preferences for the different actions are updated in discrete time. It is also assumed that a large (formally infinite) number of rounds of the game is played in between such updates, and that player  $A$  observes player  $B$ 's actions and vice versa. Each agent then has full knowledge of the other agent's mixed strategy. This is a simplification of the model, which was made for convenience by Sato *et al.*<sup>16</sup> and results in a full deterministic dynamics. The learning dynamics remains stochastic if the number of observations made between updates is finite<sup>21,22</sup>.

Proceeding on the basis of a deterministic dynamics, Sato-Crutchfield learning takes the form

$$Q_i^A(t+1) = (1 - \alpha) Q_i^A(t) + \sum_j a_{ij} x_j^B(t), \quad (7)$$

$$Q_i^B(t+1) = (1 - \alpha) Q_i^B(t) + \sum_j b_{ij} x_j^A(t). \quad (8)$$

The parameter  $\alpha$  describes geometric discounting over time. For  $\alpha = 0$  the players have full memory of the past, and the attraction  $Q_i^\mu(t)$  represents the total payoff player  $\mu \in \{A, B\}$  would have achieved up to time  $t$  given the other player's actions, and if  $\mu$  had always used action  $i$ . For positive values of  $\alpha$  more recent rounds contribute more to the attraction than iterations of the game in the distant past. The parameter  $\alpha$  is restricted to the range  $0 \leq \alpha \leq 1$ .

**Continuous-time limit and modified replicator equations.** Combining Eqs (6, 7 and 8) one finds

$$x_i^\mu(t+1) = \frac{[x_i^\mu(t)]^{1-\alpha} \exp(\Gamma \pi_i^\mu)}{\sum_j [x_j^\mu(t)]^{1-\alpha} \exp(\Gamma \pi_j^\mu)}. \quad (9)$$

In order to derive a continuous-time limit we formally rescale the time step of learning to be  $\Delta t$  (so that  $t+1$  on the LHS of Eq. (9) becomes  $t+\Delta t$ ). We also rescale the model parameters and write  $\alpha\Delta t$  instead of  $\alpha$ , and  $\Gamma\Delta t$  instead of  $\Gamma$ . Then taking the limit  $\Delta t \rightarrow 0$  we find

$$\dot{x}_i^\mu = \Gamma x_i^\mu \left[ \left( \pi_i^\mu - \sum_j \pi_j^\mu x_j^\mu \right) - \lambda \left( \ln x_i^\mu - \sum_j x_j^\mu \ln x_j^\mu \right) \right], \quad (10)$$

where  $\lambda = \alpha/\Gamma$ . The first term on the right-hand side is the expression known from the standard multi-population replicator dynamics in Eqs (2) and (3). The term proportional to  $\lambda$  exerts a force towards a uniformly mixed strategy,  $x_i^\mu = 1/S$ . This ‘entropic’ force will be strong when either the intensity of choice is low (players tend to choose their actions at random), or when memory loss is quick (propensities do not become sufficiently different to discriminate effectively between actions).

We conclude this section by two brief, but consequential observations. First, the flow of the replicator Eqs (2) and (3) can be towards stable fixed points at which one or several of the actions are not played (i.e.  $x_i^\mu = 0$ ). This cannot occur in the Sato-Crutchfield equations when  $\lambda > 0$ . Any attracting fixed points must be in the interior of strategy space. Secondly we note that the Sato-Crutchfield eq. (10) can be written in the form of conventional replicator equations

$$\dot{x}_i^\mu = \Gamma x_i^\mu \left[ f_i^\mu - \sum_j x_j^\mu f_j^\mu \right] \quad (11)$$

by introducing a modified fitness as

$$f_i^\mu = \pi_i^\mu - \lambda \ln x_i^\mu. \quad (12)$$

This will be the starting point for our construction of an individual-based model for the evolution of a population of ideas.

### Stochastic dynamics in finite populations: the case of symmetric games

**Birth-death dynamics.** To briefly recall the main features of simple birth death processes<sup>11,13</sup> we consider a population of  $N$  individuals, each of which can be of one of two types,  $i=1, 2$ . We write  $n$  for the number of individuals of type 1; the remaining  $N-n$  individuals are of type 2. Evolution proceeds in this population via a continuous-time Markov process with transition rates  $T_n^+$  from state  $n$  to state  $n+1$ , and  $T_n^-$  from state  $n$  to state  $n-1$ . In the context of evolutionary games these rates are of the general form<sup>23</sup>:

$$T_n^+ = \frac{n(N-n)}{N} g(\pi_1, \pi_2), \quad (13)$$

$$T_n^- = \frac{n(N-n)}{N} g(\pi_2, \pi_1), \quad (14)$$

where  $\pi_1 = [a_{11}n + a_{12}(N-n)]/N$  is the fitness of an individual of type 1 in the population, with an analogous expression for  $\pi_2$ . The rates scale linearly with the population size  $N$  – this is a standard choice<sup>11,23</sup>, which implies that time is effectively measured in units of generations. From these rates a deterministic dynamics is obtained in the limit  $N \rightarrow \infty$ <sup>11</sup>. For large (formally infinite) populations and writing  $x = n/N$ , one finds

$$\dot{x} = x(1-x)[g(\pi_1, \pi_2) - g(\pi_2, \pi_1)]. \quad (15)$$

A commonly used choice for the function  $g(\cdot, \cdot)$  is the so-called linear pairwise comparison process<sup>11,23</sup>,

$$g(\pi_1, \pi_2) = \frac{1}{2}[1 + \Gamma(\pi_1 - \pi_2)] \quad (16)$$

where the parameter  $\Gamma \geq 0$  is chosen small enough to ensure that  $g \geq 0$  for all  $x$ . The duplicate use of  $\Gamma$  is intentional, as will become clear shortly. With the above choice of  $g$  one obtains

$$\dot{x} = \Gamma x(1-x)(\pi_1 - \pi_2), \quad (17)$$

Modulo the constant pre-factor  $\Gamma$  this is easily shown to be the replicator eq. (1) with  $S=2$ .

**Interpretation of fitness in the linear pairwise comparison process.** We digress briefly in this subsection to discuss how individuals in the above birth-death dynamics have access to their fitness, i.e. their average payoff.

A common interpretation of fitness functions of the type  $\pi_i = \sum_j a_{ij}x_j$  requires a fast interaction time scale on which individuals face each other in the game<sup>21,22,24</sup>. The evolutionary dynamics is assumed to be a (much) slower process; it can therefore draw on knowledge of  $\pi_i$  as defined above.

One particular advantage of the linear pairwise comparison process (16) is that it does not require such a separation of time scales between interaction and evolution. Instead one can construct the evolutionary process as follows: for any (potential) birth-death event an ordered triplet of individuals from the population is picked (with replacement). We refer to the individuals in this triplet as “primary”, “secondary” and “adversary”, and

denote their types by  $i_1, i_2, i_a$ . Once a triplet has been picked, the primary and secondary individual both play against the adversary and receive payoffs  $a_{i_1 i_a}$  and  $a_{i_2 i_a}$ , respectively. The secondary individual ( $i_2$ ) is then replaced by an individual of the primary type ( $i_1$ ) – a combined death-birth event – with probability  $g(a_{i_1 i_a}, a_{i_2 i_a})$ ; otherwise the system is left unchanged. For the choice of  $g$  as in Eq. (16) the Markov chain governing this process is then that described by the rates in Eq. (13, 14). This is easily demonstrated for  $S=2$ . With appropriate scaling of the rates with  $N$  we find

$$T_n^+ = \frac{n_1 n_2}{N^2} [n_1 g(a_{11}, a_{21}) + n_2 g(a_{12}, a_{22})]. \quad (18)$$

The term in square brackets effectively averages over the choice of adversary. Using the specific form of the linear pairwise comparison process in Eq. (16), this can be written as

$$T_n^+ = \frac{n_1 n_2}{N} g\left(\frac{n_1}{N} a_{11} + \frac{n_2}{N} a_{12}, \frac{n_1}{N} a_{21} + \frac{n_2}{N} a_{22}\right), \quad (19)$$

which demonstrates the equivalence.

**Birth-death dynamics in a finite population of ideas.** We now construct an individual-based representation of Sato-Crutchfield dynamics. Motivated by Eq. (12) we introduce the modified fitness

$$f_i = \sum_j a_{ij} \frac{n_j}{N} - \lambda \ln \frac{n_i}{N}, \quad (20)$$

which can be seen as ‘entropically’ penalizing ideas that occur very frequently, and favouring rarer types. Focusing on the simplest case  $S=2$  we use birth-death rates

$$T_n^+ = \frac{n(N-n)}{N} g(f_1, f_2), \quad T_n^- = \frac{n(N-n)}{N} g(f_2, f_1). \quad (21)$$

with  $g$  as defined in Eq. (16). This is a representation of Sato-Crutchfield learning in the sense that it leads to the dynamics

$$\dot{x} = \Gamma x(f_1 - f) = \Gamma x(\pi_1 - \pi) - \Gamma \lambda x(\ln x + s), \quad (22)$$

in the limit of infinite populations. We have written  $s = -[x \ln x + (1-x) \ln(1-x)]$ , and  $f = xf_1 + (1-x)f_2$ . Our main focus from now on will be the behaviour of this birth-death process in *finite* populations.

The parameters  $\Gamma$  and  $\lambda$  need to be chosen such that all transition rates  $T_n^\pm$  are non-negative. Written out explicitly the transition rates in Eq. (21), with the definition (16) are

$$T_n^\pm = \frac{n(N-n)}{N} \frac{1}{2} \left[ 1 \pm \Gamma \left( \Delta\pi - \lambda \ln \frac{N-n}{n} \right) \right]. \quad (23)$$

Thus, we require

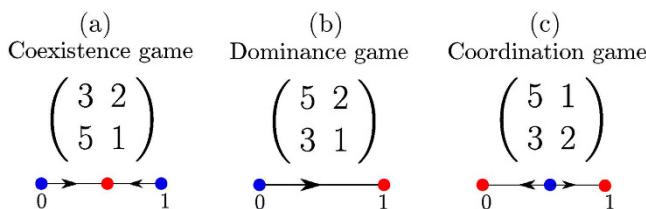
$$\Gamma \left| \Delta\pi - \lambda \ln \frac{N-n}{n} \right| \leq 1 \quad (24)$$

for all  $n=1, \dots, N-1$ . At fixed  $\Gamma$ , this imposes a constraint  $\lambda < \lambda_c$  where  $\lambda_c = \mathcal{O}(1/\ln N)$  is weakly dependent on population size; see the Supplementary Material for details. Alternatively, one could choose a manifestly positive function  $g(\cdot, \cdot)$ , such as  $g(f_1, f_2) = [1 + \exp(-2\Gamma(f_1 - f_2))]^{-1}$ . The resulting dynamics is known as the Fermi process<sup>15,23</sup>. While the fixed points of the resulting deterministic dynamics are the same as for the linear comparison process, the dynamics themselves are quantitatively different from Sato-Crutchfield dynamics. We therefore do not pursue this route.

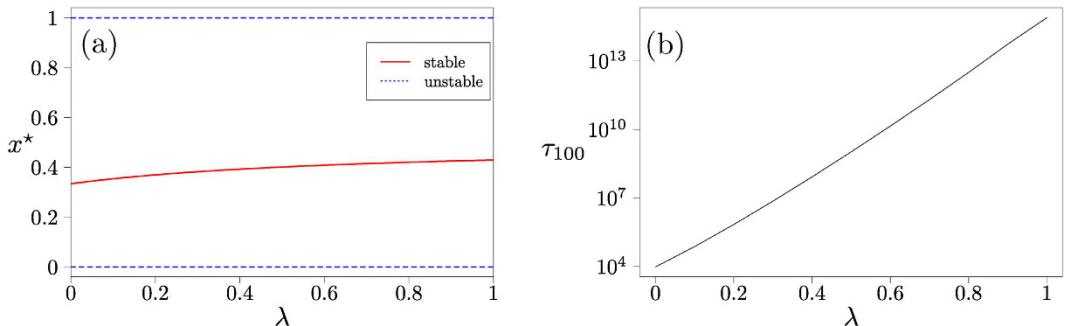
The expressions in Eq. (23) imply  $T_n^+ = T_n^- = 0$ , keeping in mind that  $\lim_{n \rightarrow 0} n \ln n = 0$ . The states  $n=0$  and  $n=N$  are therefore absorbing. Accordingly, the birth-death dynamics in the population of ideas shows fluctuation-induced extinction of ideas (or equivalently fixation). In the remainder of this section we study these fixation phenomena in the context of simple  $2 \times 2$  games.

**Application to symmetric two-player two-strategy games.** We focus on three common types of games<sup>15</sup> that cover the qualitatively distinct deterministic flow patterns available under replicator dynamics. The corresponding payoff matrices are given in Fig. 2, along with illustrations of the respective replicator flow ( $\lambda=0$ ). The points  $x=0$  and  $x=1$  are fixed points for all games for all values of  $\lambda$ .

Note that it is only asymmetric games as defined in Sec. 1 for which the deterministic dynamics has a natural interpretation in terms of Sato-Crutchfield learning for a two-player game. Our study of symmetric games, where the only notion of game play is in the pairwise interaction of the *individuals* in a population—rather than between two distinct populations representing players in the sense of Sato-Crutchfield—is primarily a warm-up. It will help us identify some important mechanisms of the fixation dynamics, such as deterministic relaxation and activation, that will be helpful in our analysis of asymmetric games in Sec. 2.



**Figure 2.** Payoff matrices  $A$  of the three main types of two-strategy two-player symmetric games, and their flow diagrams in  $x \in [0, 1]$  under replicator dynamics.



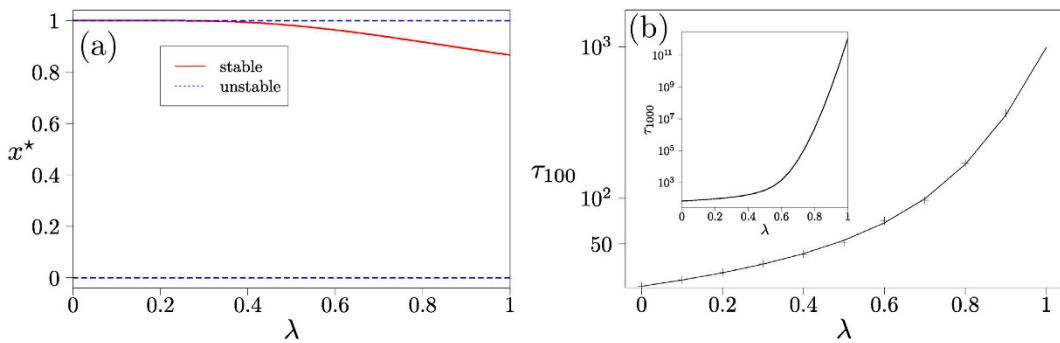
**Figure 3. Co-existence game.** (a) Location of fixed points of single-population Sato-Crutchfield learning, Eq. (22). (b) Mean fixation time as a function of  $\lambda$  in a finite population of size  $N = 200$ , starting at initial condition  $n = 100$ . The line is obtained using the known closed-form solution for simple birth-death processes<sup>11</sup>. Intensity of choice is  $\Gamma = 0.1$ .

In numerical evaluations we keep the intensity of choice at  $\Gamma = 0.1$  throughout. This value is commonly used in replicator dynamics<sup>5,25</sup> as it is a good compromise between best response dynamics ( $\Gamma \sim 1$  and above) and the weak selection limit<sup>15</sup> ( $\Gamma \ll 1$ ). We mostly use population size  $N = 200$ , which is large enough to understand qualitative aspects of the fixation behaviour using the deterministic flow diagram, but small enough for finite size effects to be easily observable. Sometimes larger  $N = 1000$  or  $2000$  are required to see activation effects (see below) clearly.

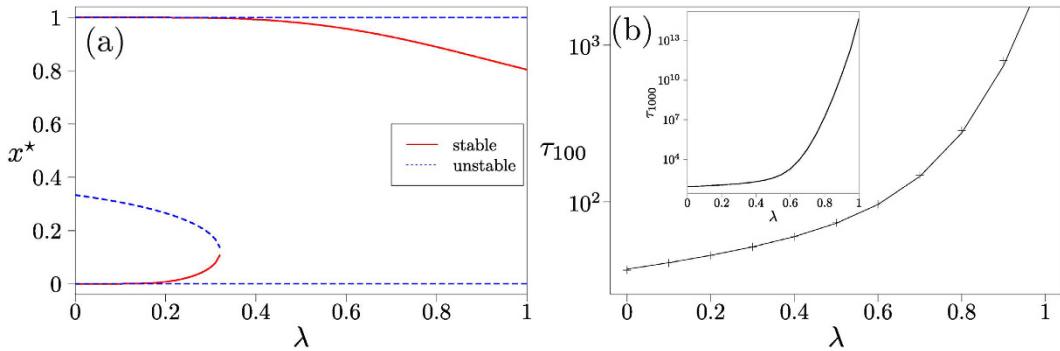
**Co-existence games.** The boundary fixed points ( $x = 0, x = 1$ ) are unstable for co-existence games under replicator flow, and there is a stable interior fixed point  $x^*$  where both types of ideas coexist. The memory-loss term in the Sato-Crutchfield equation ( $\lambda > 0$ ) does not change the qualitative features of the flow; its main effect is to move the stable fixed point closer the centre of the state space, as shown in Fig. 3(a). For very quick memory loss ( $\lambda \gg 1$ ) the fitness  $f_i$  of either type of individual is entirely dominated by the entropic term, and both types of individuals are present with equal frequency.

The path to fixation in finite population coexistence games consists of two parts: (i) an initial relaxation to the vicinity of the interior fixed point; (ii) activation to one of the two absorbing states, driven by fluctuations; see also the Supplement for further discussion. Eyring-Kramers theory<sup>26–28</sup> indicates that the typical time required for such an activation event grows exponentially with the height of the relevant activation barrier, and with the inverse variance of the noise,  $N$ . The height of the activation barrier is affected by the restoring force of the entropic term. Accordingly, the fixation time shown in Fig. 3(b) shows a strong dependence of fixation times on the model parameter  $\lambda$  at fixed  $N$ . The functional form is approximately exponential, suggesting a linear increase in the activation barrier with  $\lambda$ . This is intuitively plausible in the limit of large  $\lambda$ : the entropic term will dominate the dynamics, and it is linear in  $\lambda$ .

**Dominance games.** In this type of game one idea is dominant, and always has a higher payoff than the other type of idea. The replicator flow has constant sign; for the choice of payoff matrix in Fig. 2(b) it has an unstable fixed point at  $x = 0$ , and a stable fixed point at  $x = 1$ . The Sato-Crutchfield dynamics at  $\lambda > 0$  has an additional stable interior fixed point  $x^*$ , which approaches unity as  $\lambda \rightarrow 0$ , see Fig. 4(a). In finite populations the dynamics is similar to that of the coexistence game when  $\lambda > 0$ . After an initial relaxation towards the interior fixed point, noise drives the system to fixation. Given that the fixed point is located close to  $x = 1$  for small and moderate  $\lambda$ , fixation will mostly occur at the upper absorbing boundary. As before fixation times increase with  $\lambda$  but are rather shorter than in the coexistence game, see Fig. 4(b). Exponential dependence of the fixation time on  $\lambda$  is only seen when  $\lambda$  is sufficiently large so that the internal fixed point is well separated from the absorbing states, or when the population size is large enough for the activation barrier to show. For small and moderate values of  $\lambda$  the activation barrier is too shallow relative to the noise strength for Eyring-Kramers theory to apply.



**Figure 4. Dominance game.** (a) Location of fixed points of single-population Sato-Crutchfield learning, Eq. (22). (b) Mean fixation time as a function of  $\lambda$  in a finite population of size  $N = 200$ , starting at initial condition  $n = 100$ , comparing theory (continuous line) to direct numerical simulations of the dynamics (markers) using the Gillespie algorithm. Intensity of choice is  $\Gamma = 0.1$ . In the inset of panel (b) we show the mean fixation time starting from  $n = 1000$  for a population of size 2000, where the crossover to an exponential dependence on  $\lambda$  is visible at large  $\lambda$ .



**Figure 5. Coordination game.** (a) Location of fixed points of single-population Sato-Crutchfield learning. (b) Mean fixation time as a function of  $\lambda$  in a finite population of size  $N = 200$ , starting at initial condition  $n = 100$ , comparing theory (continuous line) to numerical simulations of the dynamics (markers). Intensity of choice is  $\Gamma = 0.1$ . In the inset of panel (b) we show the mean fixation time starting from  $n = 1000$  for a population of size 2000.

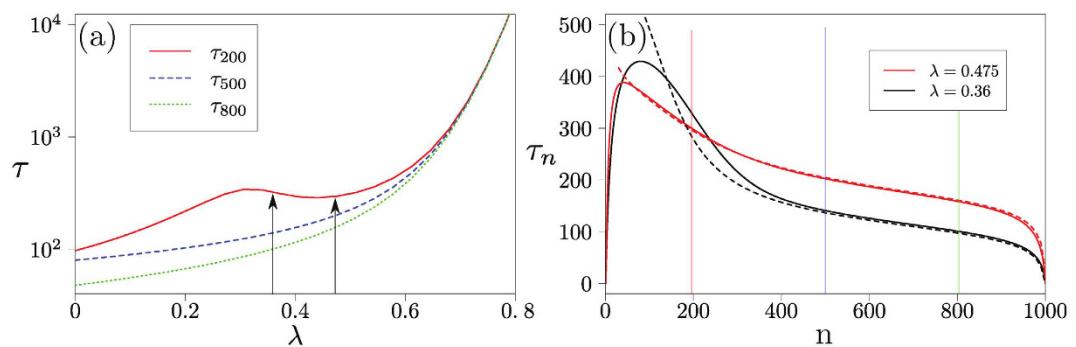
**Coordination games.** In addition to the trivial fixed points at the boundaries, the replicator dynamics of the coordination game has an unstable interior fixed point  $x_0^*$ . With memory-loss ( $\lambda > 0$ ) the dynamics develops a more intricate structure, see Fig. 5. At small but non-zero  $\lambda$  there are five fixed points. As  $\lambda$  is increased, two of these fixed points merge in a saddle-node bifurcation; we denote the corresponding value of  $\lambda$  by  $\lambda_c$ . For stronger memory loss there are three fixed points, but with reversed stability compared to the situation at  $\lambda = 0$ : unstable fixed points at  $x = 0$  and  $x = 1$ , and a stable interior fixed point whose location depends on  $\lambda$ .

For  $\lambda < \lambda_c$  and initial conditions  $n/N = x > x_0^*$ , i.e. above the unstable fixed point in the lower left of Fig. 5(a), fixation takes place as in the dominance game by deterministic relaxation to the stable fixed point near  $x = 1$ , followed by noise-driven absorption. The increase of the fixation time with  $\lambda$  is shown in Fig. 5(b) and is qualitatively similar to the behaviour for the dominance game as plotted in Fig. 4(b).

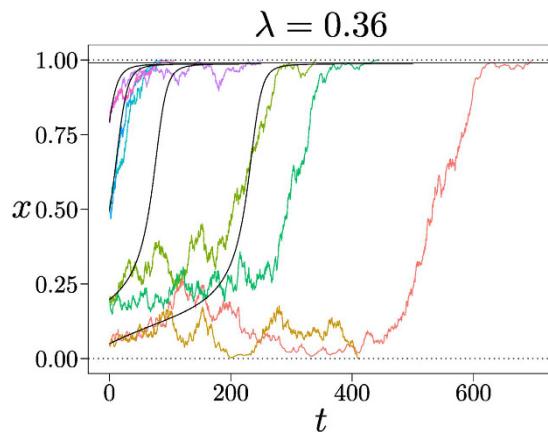
For initial conditions with  $x < x_0^*$ , the behaviour of the system and the resulting fixation time is more intricate, as shown in Fig. 6. Panel (a) demonstrates that the fixation time can now exhibit a *non-monotonic* dependence on the strength of memory loss  $\lambda$ , provided the starting point is sufficiently close to the location of the saddle-node bifurcation. The data in panel (b) show that the starting point has a non-trivial influence on fixation time.

This dependence on the initial condition  $x$  for  $\lambda < \lambda_c$  can be understood as follows. If  $x$  is smaller than the unstable (interior) fixed point at the given  $\lambda$ , deterministic relaxation will be to the stable fixed point at *lower*  $x$ , and activation from there will accordingly be to  $x = 0$  rather than  $x = 1$ . A more detailed analysis for large  $N$  can be found in the Supplement. This shows that close to the bifurcation, activation towards  $x = 0$  is slower—exponentially in  $N$ —than across the barrier to the stable fixed point at large  $x$ , so the system follows the latter route and eventually reaches  $x = 1$ . We emphasize that this is a non-trivial prediction for the dynamics in finite populations; it cannot be deduced from the deterministic Sato-Crutchfield dynamics.

Moving beyond the bifurcation ( $\lambda > \lambda_c$ ), the situation is simpler again. For sufficiently large  $N$  one predicts fixation by relaxation directly to the stable fixed point close to  $x = 1$ , and activation to  $x = 1$  from there. In Fig. 7 one can see that the system relaxes to the stable fixed point close to  $x = 1$  following the deterministic dynamics,



**Figure 6. Coordination game.** (a) Mean fixation time in a finite population (with  $\Gamma = 0.1$ ) as a function of  $\lambda$ , the memory-loss parameter, and for different initial conditions  $n$ . (b) Mean fixation time as a function of the initial condition for two fixed values  $\lambda$  indicated by arrows in (a). We show data for a larger population size  $N = 1000$  to reveal the non-monotonies in  $\lambda$ . In (b), vertical lines indicate the initial conditions used in (a). Also shown are the times taken under the deterministic dynamics (dashed lines) to get from the initial condition to within  $c/N$  of the stable fixed point; the order unity constant  $c$  is chosen to give a good description of the actual fixation times for initial conditions near the fixed point.



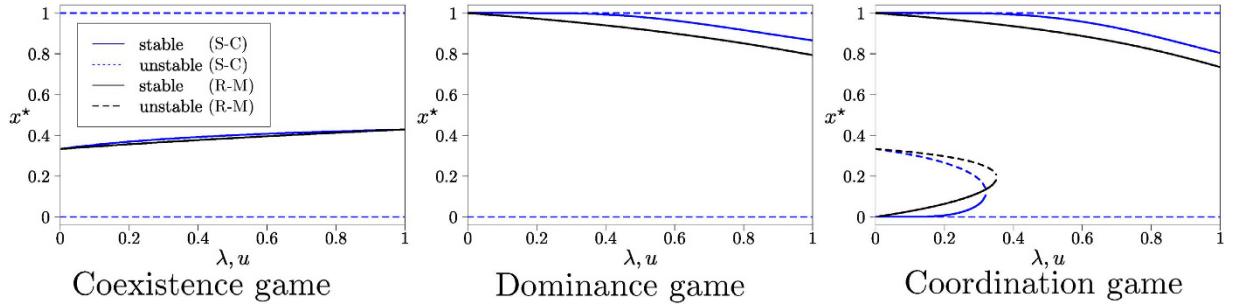
**Figure 7. Sample trajectories of a coordination game, for different initial conditions;  $\Gamma = 0.1$ ,  $\lambda = 0.36$  and  $N = 1000$  (coloured curves).** The black curves show the trajectory for the deterministic dynamics (11) starting from the same set of initial conditions. The trajectories follow the deterministic dynamics fairly closely for initial conditions  $x = 0.5$  and  $0.8$ . For initial condition  $x = 0.2$ , fluctuations determine how fast the system escapes from the initial region of relatively weak deterministic flow. For initial  $x = 0.05$ , this effect is even stronger. One of the two trajectories shown also illustrates direct activation, to fixation at  $x = 0$ , against the deterministic flow.

then fixation occurs by activation. For small  $N$  and close to the bifurcation threshold, the system might initially stay in a region of relatively weak deterministic flow. A detailed analysis of this phenomenon is deferred to the Supplement.

**Comparison with replicator-mutator dynamics.** The effect of the entropic term in the Sato-Crutchfield equations is akin to that of mutation<sup>29,30</sup> in evolutionary processes. Both mutation and entropic terms describe forces that act towards the centre of strategy space and drive the population away from states in which one species (or one idea) dominates, and we here include a brief comparison. We choose the replicator-mutator equation of the form discussed by Bladon *et al.*<sup>23</sup>

$$\dot{x} = \left(1 - \frac{u}{2}\right)x(1-x)(\pi_1 - \pi_2) - \frac{u}{2}\left(x - \frac{1}{2}\right), \quad (25)$$

where  $u > 0$  is the mutation rate. In order to compare the effects of mutation with those of memory loss in the learning process, we show the bifurcation diagrams of the replicator-mutator dynamics along with those of Sato-Crutchfield learning in Fig. 8, for the three classes of symmetric games we have considered. The main difference between the two flows is that Sato-Crutchfield dynamics has additional fixed points at  $x = 0$  and  $x = 1$ . As these are unstable for  $\lambda > 0$ , they do not lead to qualitative differences in the long-time deterministic dynamics. However, for finite  $N$  the difference is significant: replicator-mutator dynamics does not have absorbing states, so the question of fixation does not arise.



**Figure 8.** Fixed point diagrams of Sato-Crutchfield (S-C) learning (22), (blue and red lines), and replicator-mutator (R-M) dynamics (25) (black lines) for our three types of symmetric  $2 \times 2$  games. The full black lines show the stable fixed points of the replicator-mutator dynamics and the dashed line its unstable fixed points.

### Asymmetric games and multiple populations of ideas

**Birth-death dynamics for multiple populations of ideas.** In this section we extend the stochastic dynamics for populations of ideas to games with multiple populations. We focus on the simplest case of two-player two-strategy games, though the approach easily extends to more general games. Our starting point are the Sato-Crutchfield eq. (10), which simplify to

$$\dot{x}^A = \Gamma x^A (\pi_1^A - \pi^A) - \Gamma \lambda x^A (\ln x^A + s^A), \quad (26)$$

$$\dot{x}^B = \Gamma x^B (\pi_1^B - \pi^B) - \Gamma \lambda x^B (\ln x^B + s^B), \quad (27)$$

where  $\pi_1^A = a_{11}x^B + a_{12}(1 - x^B)$ ,  $\pi_2^A = a_{21}x^B + a_{22}(1 - x^B)$ , with analogous expressions for  $\pi_1^B$  and  $\pi_2^B$ . We have also written  $\pi^A = x^A\pi_1^A + (1 - x^A)\pi_2^A$ , and  $s^A = -[x^A \ln x^A + (1 - x^A) \ln(1 - x^A)]$ . Similar definitions apply to  $\pi^B$  and  $s^B$ . The variable  $x^A$  denotes the probability with which player A chooses their action 1 and similarly for  $x^B$ .

The stochastic evolutionary dynamics now occurs in two finite populations of ideas, one for either player, each consisting of  $N$  individuals. We write  $n$  for the number of ideas of type 1 in population A, and similarly  $m$  for the number of ideas of type 1 in population B. The dynamics is defined by the rates for birth-death transitions in population A,  $(n, m) \rightarrow (n \pm 1, m)$ ,

$$T_{(n,m)}^{A+} = \frac{1}{2} \frac{n(N-n)}{N} \left[ 1 + \Gamma \left( \pi_1^A - \pi_2^A - \lambda \ln \frac{n}{N-n} \right) \right], \quad (28)$$

$$T_{(n,m)}^{A-} = \frac{1}{2} \frac{n(N-n)}{N} \left[ 1 + \Gamma \left( \pi_2^A - \pi_1^A - \lambda \ln \frac{N-n}{n} \right) \right], \quad (29)$$

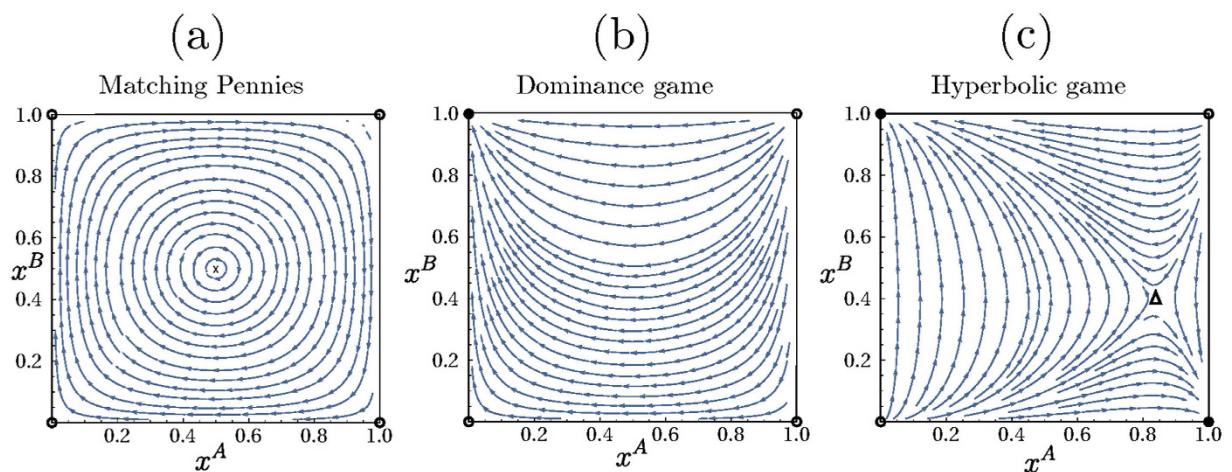
and analogous rates for transitions  $(n, m) \rightarrow (n, m \pm 1)$  in population B

$$T_{(n,m)}^{B+} = \frac{1}{2} \frac{m(N-m)}{N} \left[ 1 + \Gamma \left( \pi_1^B - \pi_2^B - \lambda \ln \frac{m}{N-m} \right) \right], \quad (30)$$

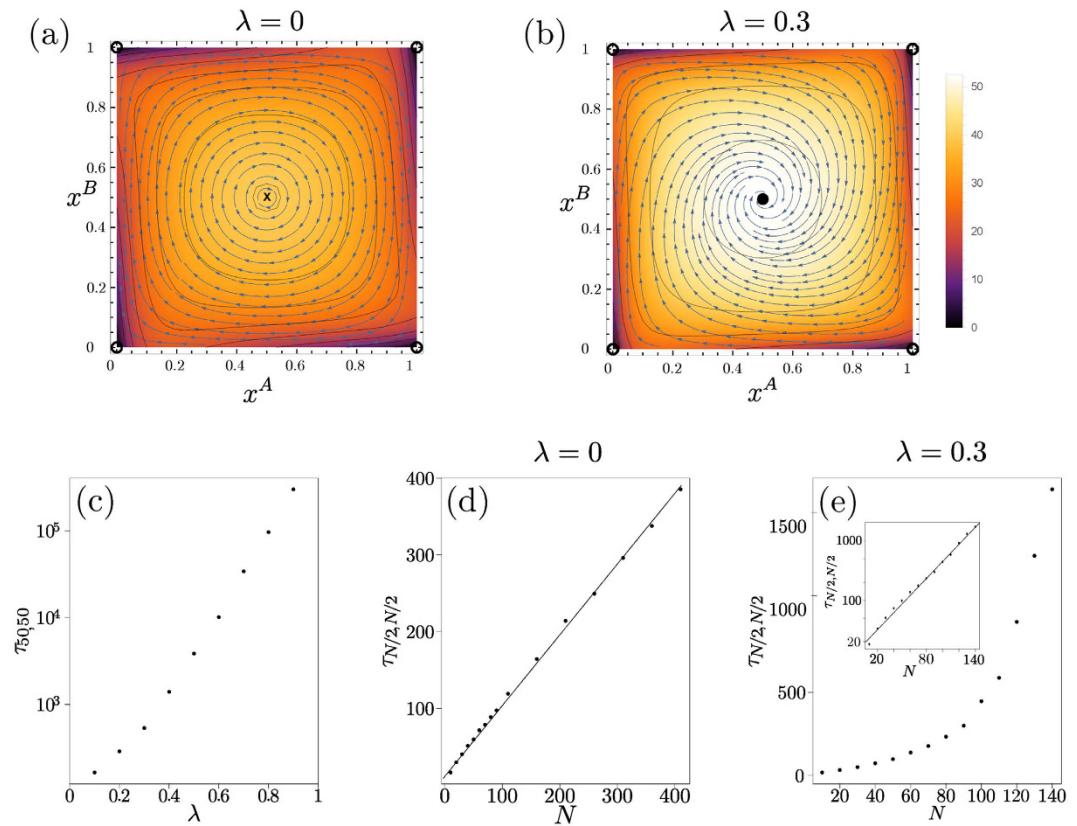
$$T_{(n,m)}^{B-} = \frac{1}{2} \frac{m(N-m)}{N} \left[ 1 + \Gamma \left( \pi_2^B - \pi_1^B - \lambda \ln \frac{N-m}{m} \right) \right]. \quad (31)$$

The two-population birth-death dynamics has four absorbing states,  $(n, m) = (0, 0), (0, N), (N, 0), (N, N)$  in finite populations. In the limit  $N \rightarrow \infty$  and writing  $x^A = n/N$  as well as  $x^B = m/N$ , this process leads to the deterministic two-population Sato-Crutchfield eqs (26) and (27).

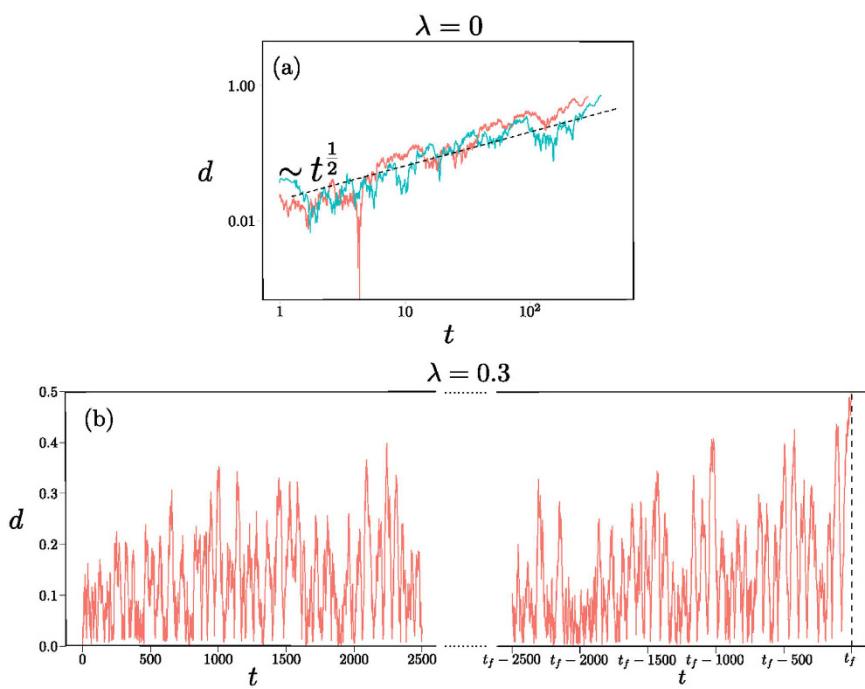
**Examples of two-player two-strategy asymmetric games.** We now study the corresponding fixation properties, focusing on a few key examples of asymmetric two-player games, chosen from the different categories of possible two-population replicator flows<sup>17</sup>: (i) the so-called Matching Pennies game, also known as Dawkin's Battle of the Sexes<sup>18</sup>; (ii) games in which one player has an action that strictly dominates the alternative action; and (iii) games in which the replicator flow has a hyperbolic interior fixed point. The three cases are illustrated in Fig. 9. Note that in our analysis, we calculate fixation times for asymmetric games using the backward master equation for a system of size  $N=30$ . This is smaller than the value we use for symmetric games ( $N=100$ ) for computational reasons: we now need to find the fixation time for  $N^2$  rather than  $N$  different initial condition, from a system of linear equations (the backward master equation) of correspondingly larger size.



**Figure 9.** Two-population replicator flows for different types of  $2 \times 2$  asymmetric games: (a) the Matching Pennies game is a zero-sum game; the replicator dynamics has a conserved quantity and exhibits cyclic trajectories. The game (b) has one pure-strategy fixed point while (c) has a hyperbolic fixed point. Stable fixed points are labeled by full dots, saddles (fixed points with one unstable and one stable direction) by triangles, unstable fixed points (two unstable directions) by empty dots and finally cyclic fixed points (whose Jacobian eigenvalues are purely imaginary) by a cross.



**Figure 10. Matching Pennies game.** (a,b) Flow under deterministic Sato-Crutchfield learning for  $\lambda = 0$  and  $\lambda = 0.3$ , respectively. Overlaid is a heat map indicating the fixation time as a function of the starting point (obtained from the backward master equation<sup>31</sup> for a system of size  $N = 30$ ). (c) Fixation time from simulations as a function of  $\lambda$ , for population size  $N = 100$  and  $(n, m) = (N/2, N/2)$  as initial condition. Panels (d,e) show fixation time  $\tau_{N/2,N/2}$  against  $N$  for  $\lambda = 0$  and  $\lambda = 0.3$ , respectively. Panel (d) shows linear scaling of fixation time with  $N$  (solid line) consistent with fixation by radial diffusion, whereas panel (e) displays approximately exponential scaling (see log-linear plot in inset) as fixation now requires activation against the flow.



**Figure 11. Matching Pennies game.** Sample trajectories in a population of size  $N=500$  and with  $\Gamma=0.1$ , for (a)  $\lambda=0$  and (b)  $\lambda=0.3$ . We show the distance  $d$  of  $(x^A, x^B)$  from the fixed point at  $(0.5, 0.5)$  versus time  $t$  in log-linear scale, to focus on the radial motion. Note the difference between diffusive dynamics in (a) – the dashed line shows the expected power law  $1/2$  for a diffusive process – and activation in (b). For the latter we plot the beginning of the trajectory, showing how the system reaches a metastable steady state where it fluctuates around the centre of the state space ( $d=0$ ), and on the right the end of the fixation trajectory where a fluctuation takes the system to one of the four absorbing states at time  $t_f$ .

**Matching Pennies game.** This game is represented by the following payoff bi-matrix

$$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (32)$$

In addition to the trivial fixed points at the corners of phase space the replicator dynamics ( $\lambda=0$ ) has the fixed point  $\mathbf{x}^* = x^A, x^B = (0.5, 0.5)$ . Trajectories that start elsewhere will form closed periodic orbits around the fixed point as shown in Fig. 10(a). Fixation in one of the four corners in *finite* populations will therefore be due to radial diffusion. Diffusion distances generally grow as  $\sqrt{Dt}$ . As the diffusion constant is  $D \sim 1/N$  in our case, covering a radial distance of order unity to reach one of the two corners requires time  $t \sim N$ . This linear growth of fixation time with population size is shown in Fig. 10(d).

The effect can be seen as an analogue of the trapping in regions of low flow discussed in the Supplement, but here the (radial) flow is zero over an extended region rather than at a single point, causing a stronger fixation time growth ( $N$  versus  $\ln N$ ) with population size.

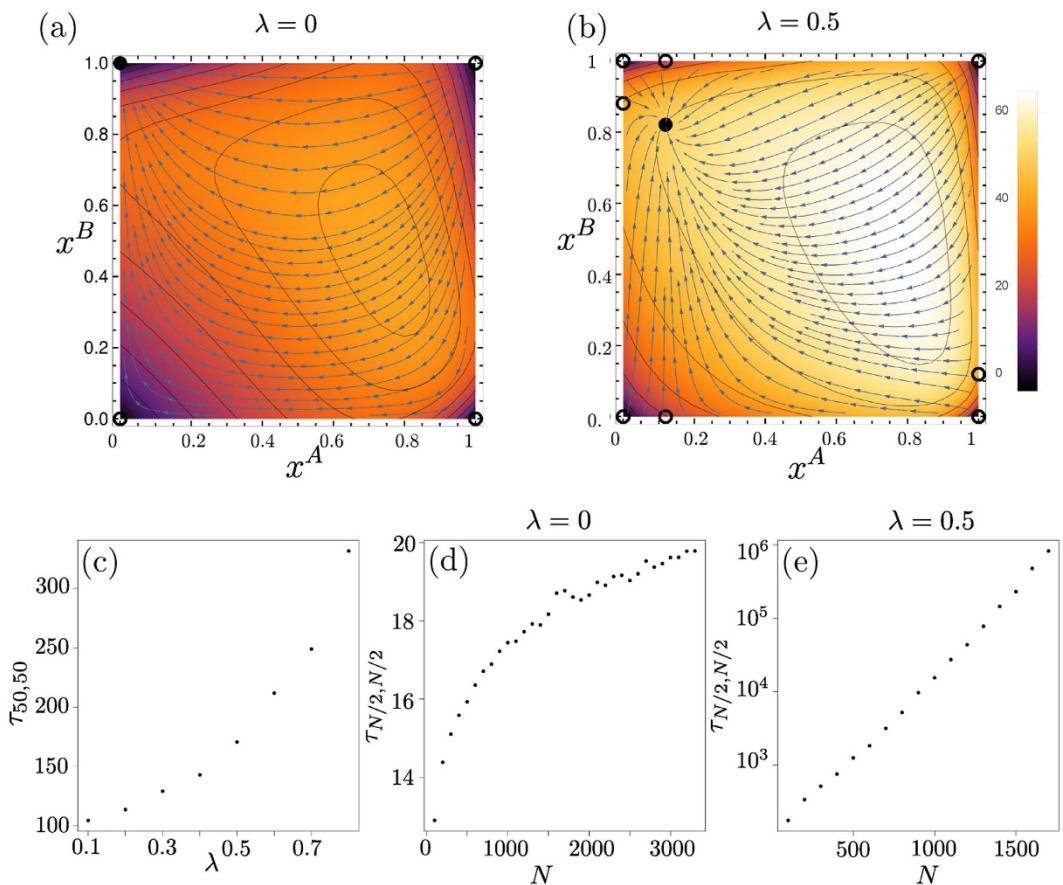
As soon as one has nonzero memory loss  $\lambda$ , the point  $\mathbf{x}^*$  becomes an attractor of the dynamics, with the whole state space as basin of attraction as shown in Fig. 10(b). As before, fixation will therefore proceed along the sequence of relaxation to this fixed point followed by activation to one of the absorbing states. The activation phase again requires a time scaling exponentially with the population size  $N$ . This change in scaling is clear by comparing Fig. 10(d,e) and emphasizes that the addition of the entropic term in the fitness has qualitative consequences for the fixation dynamics. The sample trajectories in Fig. 11 further illustrate this.

When  $\lambda$  becomes large, the flow and hence the activation barrier becomes proportional to  $\lambda$  to leading order, producing fixation times that scale exponentially with  $\lambda$  as can be seen in Fig. 10(c).

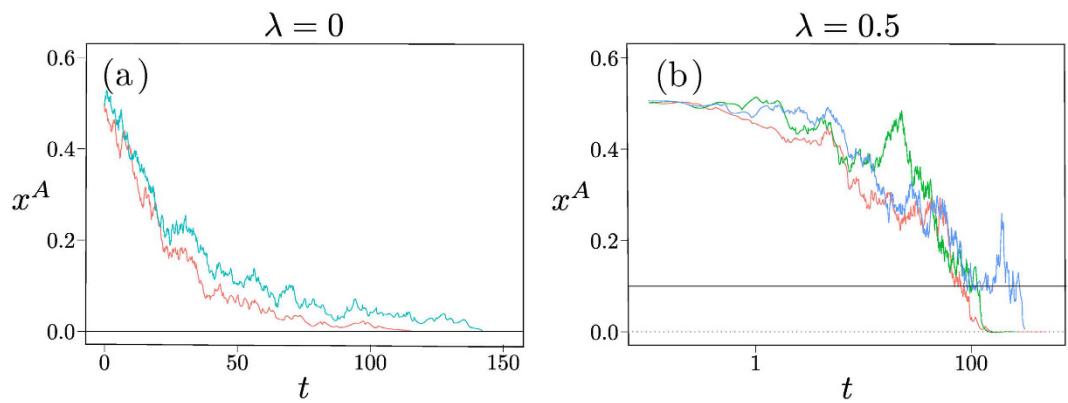
**Dominance game.** An example of this case is defined by the payoff structure

$$\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (33)$$

Its Sato-Crutchfield dynamics for  $\lambda=0$  has four fixed points in the corners of the state space, one of which is stable. Fixation will then typically proceed by deterministic relaxation to this fixed point. For infinite  $N$  this would take infinite time as the approach to the fixed point is exponential. At finite  $N$ , one expects that fixation takes place once this exponential approach gets within distance  $1/N$  – the grid spacing in the  $(x^A, x^B)$ -plane – of the fixed point. The fixation time should then scale logarithmically with  $N$ ; the data in Fig. 12(d) are consistent with this.

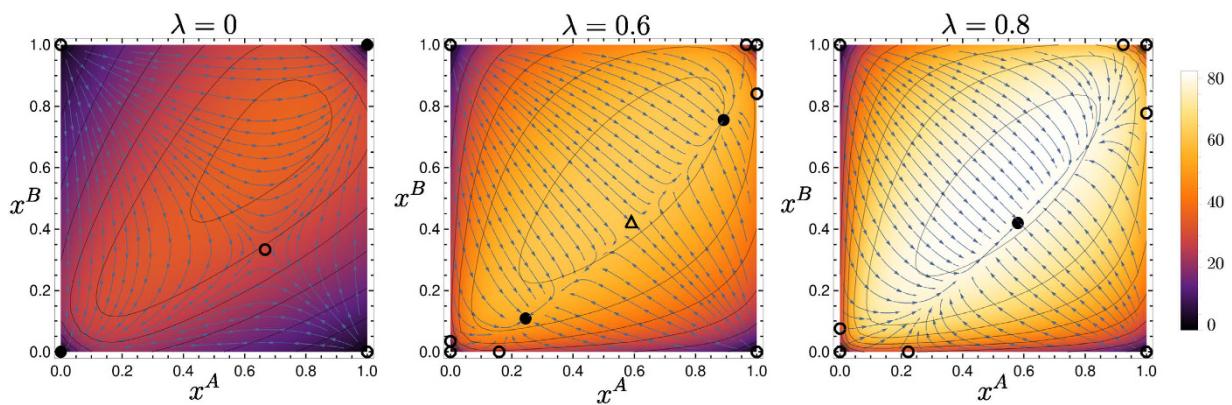


**Figure 12. Dominance game.** (a,b) Flow under deterministic Sato-Crutchfield learning for  $\lambda=0$  and  $\lambda=0.5$ , respectively. Overlaid is a heat map indicating the mean fixation time as a function of the starting point (obtained from the backward master equation for a system of size  $N=30$ ); (c) Fixation time from Gillespie simulations as a function of  $\lambda$ , for population size  $N=100$  and  $(n, m)=(N/2, N/2)$  as initial condition. (d,e) Fixation time  $\tau_{N/2, N/2}$  against  $N$  for  $\lambda=0$  and  $0.5$ , respectively. The fixation time in (d) exhibits logarithmic scaling with  $N$  resulting from the exponential approach to the stable fixed point. The scaling of the fixation time in (e) is approximately exponential with  $N$  because fixation involves activation.

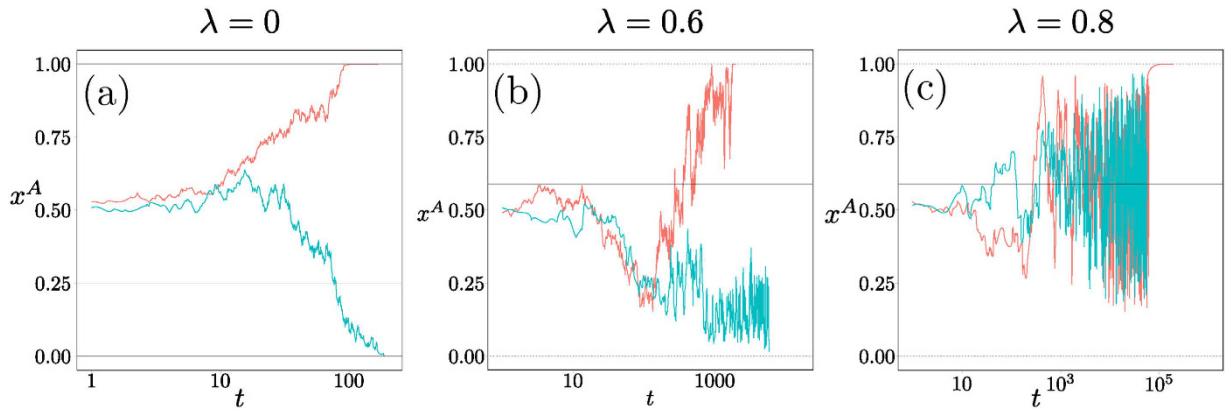


**Figure 13. Dominance game.** Sample trajectories in a population of size  $N=500$  and with  $\Gamma=0.1$ , for (a)  $\lambda=0$  and (b)  $\lambda=0.5$ . We show  $x^A$  against time (linear axis in (a), logarithmic axis in (b)). The full and dashed horizontal lines show the  $x^A$ -coordinate of the stable and unstable fixed points of the deterministic dynamics, see also Fig. 12.

As the memory-loss parameter  $\lambda$  is increased from zero, the stable fixed point moves continuously towards the centre of the state space, with all four corners then unstable fixed points. (There are also two additional saddle points on the boundary near the original stable fixed point.) Fixation will take place by relaxation followed by activation, resulting in exponential growth of fixation times with  $N$  (Fig. 12(e)) and, at large  $\lambda$ , also with  $\lambda$



**Figure 14. Hyperbolic game.** Flow under deterministic Sato-Crutchfield learning for  $\lambda = 0, 0.6$  and  $0.8$ , respectively. Overlaid is in each panel a heat map showing the mean fixation time as a function of starting point in a system of size  $N = 30$ . The three chosen values of  $\lambda$  show different fixed point structures as indicated by the symbols.



**Figure 15. Hyperbolic game.** Sample trajectories in a population of size  $N = 500$  and with  $\Gamma = 0.1$ , for (a)  $\lambda = 0$ , (b)  $0.6$  and (c)  $0.8$ , in the same representation as in Fig. 13. Panel (a) shows relaxation to the region around the saddle point, with fluctuations then determining at which boundary fixed point fixation occurs. The trajectories in (b) start similarly but then are driven to one of two *interior* stable fixed points, from which fixation proceeds by activation to the nearest boundary. In (c), all trajectories go to the single interior fixed point, from which fixation by activation occurs to one of two boundary fixed points (top right and bottom left in right-hand panel of Fig. 14).

(Fig. 12(c)). The sample trajectories in Fig. 13 illustrate the qualitative differences between the fixation dynamics for  $\lambda = 0$  and  $\lambda > 0$ .

**Hyperbolic game.** An example of this class of games is given by the payoff matrices

$$\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad (34)$$

For  $\lambda = 0$ , the Sato-Crutchfield dynamics has one saddle point in the interior of the state space, two stable fixed points in two opposite corners of the state space, and two unstable fixed points in the remaining corners; cf. Fig. 14. As for the dominance game, fixation will proceed by deterministic relaxation, leading to exponential approach to one of the two stable fixed points. Logarithmic growth with  $N$  of fixation times should again result, though we have not verified this explicitly.

Each of the two stable fixed points has its own basin of attraction. This is a new feature compared to the dominance game. For  $N \rightarrow \infty$ , the location in strategy space where fixation occurs will be entirely determined by which basin the system starts off in. For finite  $N$ , fluctuation effects will then make the choice of fixation location stochastic.

With increasing  $\lambda$ , the two stable fixed points in the corners move to the interior of the state space. At a critical value  $\lambda_c$ , these two fixed points merge with the saddle point into a single stable fixed point. (This is the consequence of a symmetry in our payoff matrices; without this, the saddle would annihilate with one stable fixed point

and the other would survive.) The presence of this bifurcation would suggest, by analogy with the results for the coordination game, a non-monotonic dependence of the fixation time on  $\lambda$  near  $\lambda_c$ . Presumably the values of  $N$  required to see this will be large again, however, and we were unable to reach them in the two-population case with reasonable computational effort. Nonetheless, Fig. 15 illustrates clearly that as  $\lambda$  varies, the different fixed point structures of the deterministic dynamics cause qualitative changes in the fixation trajectories.

## Summary and outlook

We have interpreted learning in games as a pairwise comparison process within a population of ideas. In the limit of large population size, the dynamics is described by the deterministic Sato-Crutchfield equations. While these equations for learning have been widely studied, there has (to our knowledge) not been any systematic derivation from a birth-death process in finite populations. Such individual-based foundations are only available for simpler replicator (or replicator-mutator) dynamics<sup>11,19,23</sup>. We fill this gap by defining such an individual-based process in a finite population of ideas. The construction in Sec. 1 and 2 involves augmenting the standard fitness function by a term proportional to the information content ( $-lnx_i$ ) of species  $i$ . While the behaviour of deterministic Sato-Crutchfield learning in continuous time is fairly similar to the outcome of replicator-mutator dynamics in infinite populations, there are marked differences between their stochastic representations in finite systems. Mutation processes prevent fixation or extinction, but these phenomena can and will occur in finite populations of ideas, even at non-zero memory loss.

In order to develop some intuition for the general phenomena that can occur in finite populations of ideas we first studied three types of symmetric games (Sec. 1). We focused on the dependence of the fixation dynamics on the size of the population and on the memory-loss parameter  $\lambda$ . In our interpretation this latter parameter becomes the strength of the preference for rare ideas. The variety of different behaviours observed could be understood by decomposing the fixation dynamics into a sequence of elementary events, such as relaxation to stable fixed points and activation against the deterministic flow driven by demographic noise. We then broadened our analysis to include asymmetric two-player games (Sec. 2). Further features of the dynamics are then observed, such as fixation by diffusion when the relevant part of the dynamics is not opposed by the deterministic flow.

Most of our results are obtained from direct Gillespie simulations of the stochastic evolution of ideas, or from numerical solutions of the corresponding backward master equation. In the case of symmetric games we have complemented this with an analysis for large population size  $N$  (see the Supplement). This allows one to identify the dominant scaling of fixation times and reveals subtle effects that cannot be deduced from the fixed point structure of the dynamics (see Supplement). For asymmetric games there is in general no mapping to noisy descent on an effective potential energy, because of the lack of detailed balance. However, as discussed e.g. by Bouchet *et al.*<sup>32</sup>, one should – in principle – be able to obtain fixation times for large  $N$  by using Freidlin-Wentzell large deviation theory. This is left to future work.

We think our work will enrich the mathematical theory of learning and evolutionary dynamics, providing a novel interpretation of learning in games with imperfect memory as a pairwise matching process between ideas. Our construction places the dynamics of learning in the context of stochastic population dynamics, and, we hope, it will encourage further studies of learning based on the established toolbox for evolutionary dynamics in finite populations.

## References

1. Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, NJ, 1953).
2. Nash, J. F. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA* **36**, 48–49 (1950).
3. Camerer, C. F. & Ho, T.-H. Experience-weighted attraction learning in normal form games. *Econometrica* **67**, 827–874 (1999).
4. Camerer, C. F. *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press, Princeton, NJ, 2003).
5. Galla, T. & Farmer, J. D. Complex dynamics in learning complicated games. *Proc. Natl. Acad. Sci. USA* **110**, 1232–1236 (2011).
6. Skyrms, B. Chaos in game dynamics. *J. Log. Lang. Inf.* **1**, 111–13 (1992).
7. Brock, W. A. & Hommes, C. H. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *J. Econ. Dyn. Control.* **22**, 1235–1274 (1998).
8. Sato, Y., Akiyama, E. & Farmer, J. D. Chaos in learning a simple two-person game. *Proc. Natl. Acad. Sci. USA* **99**, 4748–4751 (2002).
9. Börgers, T. & Sarin, R. Learning through reinforcement and replicator dynamics. *J. Econ. Theory* **77**, 1–14 (1997).
10. Blythe, R. A. & McKane, A. J. Stochastic models of evolution in genetics, ecology and linguistics. *J. Stat. Mech.: Theor. and Exp.* **2007**, P07018 (2007).
11. Traulsen, A. & Hauert, C. Stochastic evolutionary game dynamics. *Reviews of nonlinear dynamics and complexity* **2**, 25–61 (2009).
12. Sato, Y. & Crutchfield, J. P. Coupled replicator equations for the dynamics of learning in multiagent systems. *Phys. Rev. E* **67**, 015206 (2003).
13. Nowak, M. A. *Evolutionary Dynamics* (Belknap Press, Cambridge, MA, 2006).
14. Antal, T. & Scheuring, I. Fixation of strategies for an evolutionary game in finite populations. *Bull. Math. Biol.* **68**, 1923–1944 (2006).
15. Altrock, P. M. & Traulsen, A. Fixation times in evolutionary games under weak selection. *New J. Phys.* **11**, 013012 (2009).
16. Sato, Y., Akiyama, E. & Crutchfield, J. P. Stability and diversity in collective adaptation. *Physica D* **210**, 21–57 (2005).
17. Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).
18. Dawkins, R. *The Selfish Gene* (Oxford University Press, Oxford, UK, 1976).
19. Traulsen, A., Claussen, J. C. & Hauert, C. Coevolutionary dynamics: from finite to infinite populations. *Phys. Rev. Lett.* **95**, 238701 (2005).
20. Ho, T.-H., Camerer, C. F. & Chong, J.-K. Self-tuning experience weighted attraction learning in games. *J. Econ. Theory* **133**, 177–198 (2007).
21. Galla, T. Intrinsic noise in game dynamical learning. *Phys. Rev. Lett.* **103**, 198702 (2009).
22. Roca, C. P., Cuesta, J. A. & Sánchez, A. Time scales in evolutionary dynamics. *Phys. Rev. Lett.* **97**, 158701 (2006).
23. Bladon, A. J., Galla, T. & McKane, A. J. Evolutionary dynamics, intrinsic noise, and cycles of cooperation. *Phys. Rev. E* **81**, 066122 (2010).
24. Roca, C. P., Cuesta, J. A. & Sánchez, A. Evolutionary game theory: Temporal and spatial effects beyond replicator dynamics. *Phys. Life Rev.* **6**, 208–249 (2009).

25. Realpe-Gómez, J., Szczesny, B., Dall'Asta, L. & Galla, T. Fixation and escape times in stochastic game learning. *J. Stat. Mech.: Theor. and Exp.* **2012**, P10022 (2012).
26. Hänggi, P., Talkner, P. & Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251 (1990).
27. Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **3**, 107–115 (1935).
28. Kramers, H. A. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304 (1940).
29. Komarova, N. L. Replicator-mutator equation, universality property and population dynamics of learning. *J. Theor. Biol.* **230**, 227–239 (2004).
30. Mobilia, M. Oscillatory dynamics in rock-paper-scissors games with mutations. *J. Theor. Biol.* **264**, 1–10 (2010).
31. Norris, J. R. *Markov chains*. Cambridge series in statistical and probabilistic mathematics (Cambridge University Press, Cambridge, UK, 1998).
32. Bouchet, F. & Reygner, J. Generalisation of the Eyring–Kramers transition rate formula to irreversible diffusion processes. *Ann. I. H. Poincaré* **1–34** (2016).

## Acknowledgements

T.G. and P.S. acknowledge support from the Engineering and Physical Sciences Research Council EPSRC (UK), under grant EP/K000632/1.

## Author Contributions

T.G., R.N. and P.S. conceived and designed the experiments. R.N. performed the experiments and analyzed the data. T.G., R.N. and P.S. wrote the paper.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Nicole, R. *et al.* Stochastic evolution in populations of ideas. *Sci. Rep.* **7**, 40580; doi: 10.1038/srep40580 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017