

Complex dynamics in the Sherrington-Kirkpatrick game

Desmond Chan^{*}*Department of Informatics, King's College, London WC2B 4BG, United Kingdom*Tobias Galla[†]*Instituto de Física Interdisciplinar y Sistemas Complejos, IFISC, Edifici Instituts Universitaris de Recerca, Campus Universitat de les Illes Balears, E-07122 Palma, Mallorca, Spain*

(Received 2 July 2026; accepted 7 October 2026; published 28 October 2026)

We study the outcome of adaptive learning among many players engaged in sets of two-strategy, two-player games. We are interested in typical games and therefore generate the payoff matrices randomly at the outset. These payoff matrices then remain fixed throughout the learning process. This provides a game-theoretic foundation for the Sherrington-Kirkpatrick (SK) game recently introduced by Garnier-Brun *et al.* [*Phys. Rev. X* **14**, 021039 (2024)]. The original model studied by these authors corresponds to the special case with no bias toward either strategy. Here, we determine the stability of learning in SK games with general random biases and show that the nature of the stable state is affected by random fields. We also introduce a grand-canonical version of the SK game, in which players may choose to abstain. For this extended model, we again determine the stability of learning. Our analysis confirms that complex situations involving many players are often unlearnable, even if each player chooses between only two actions. The rate at which players forget past payoffs and the competitiveness of the game emerge as key parameters determining whether learning converges to a unique fixed point, whether multiple fixed points arise, or whether the dynamics remain persistently volatile.

DOI: [10.1103/b1qh-k8pn](https://doi.org/10.1103/b1qh-k8pn)

I. INTRODUCTION

While there is no unique definition of a complex system, complexity is generally understood to involve the emergence of unexpected and perhaps unpredictable behavior from the interaction of relatively simple agents. What these agents represent depends on the context. They could be members of a population in a model of the spread of an epidemic, genes and proteins in developmental biology, or traders in a financial market. Many of these systems involve competition for limited resources, strategic decision-making, and adaptation. Since its inception in the 1940s [1], game theory has been an important ingredient in modeling these processes. There has been significant interest within the physics community in evolutionary games in particular and in the emergence of cooperation and the general phenomenology of different games, see, for example, Refs. [2,3].

Authors of the bulk of existing work in game theory, both in physics and in other disciplines, have focused on simple games. By this, we mean games with a small number of players who each only have a small number of actions (pure strategies) at their disposal. Classical examples are the prisoner's dilemma or cyclic games such as rock-paper-scissors.

In these games, only two players interact, and they choose between two actions in the prisoner's dilemma or three actions in the rock-paper-scissors game.

Much of the modeling work on strategic decision-making in economics is based on the assumption of perfectly rational players who have full information about the game and who assume that all other players are also fully rational and fully informed. Under these assumptions, the natural outcomes of a game are the so-called Nash equilibria [4]. These are points in the strategy space such that no player can improve his or her payoff by unilaterally changing his or her strategy. Nash proved that such points always exist in sufficiently simple games: broadly when the number of players is finite and when each player only has a finite number of actions to choose from. Importantly, Nash equilibria can be probabilistic combinations of pure strategies. For example, the Nash point for the rock-paper-scissors game is given by the fully mixed strategy in which each player chooses each of the three actions randomly with probability of 1/3 each.

For simple games such as rock-paper-scissors, the concept of Nash equilibrium is perfectly sensible, as the equilibrium points are often unique, and they are easy to find. In more complex situations, it is more questionable if players could find Nash equilibria. In some complicated games, there can be many Nash points [5–7], and algorithms to find these equilibria typically require a computing time that grows exponentially with the size of the game [8].

With this in mind, it is not surprising that Nash equilibria are not the natural outcome in models of human players learning to play a game. Sato *et al.*, for example, have shown that learning can lead to chaotic dynamics even in

^{*}Contact author: desmond.chan@kcl.ac.uk

[†]Contact author: tobias.galla@ifisc.uib-csic.es

low-dimensional games [9–11]. This casts doubt on the usefulness of the idea of Nash equilibria in more complex games.

Complexity in normal-form games can arise in (at least) two distinct ways: (i) by increasing the number of actions available to each player while keeping the number of players small or (ii) by increasing the number of players while keeping the action space small.

The first route was taken, for example, in Refs. [12–14]. Building on earlier work on high-dimensional replicator models [15], Galla and Farmer [12] used simulations and analytical methods from the theory of disordered systems to demonstrate different dynamical outcomes for experience-weighted attraction (EWA) learning in complicated two-player games. They studied the learning dynamics of two players who engage in a repeated game with many actions and where the payoff matrix is drawn at random at the beginning. One main result is that the learning process remains chaotic and unpredictable for a large range of games. This was later extended to multiplayer games [13], where chaos is even more prevalent. Complexity in these studies resulted mainly from a large strategy space.

To investigate scenario (ii)—situations with finite action spaces but a larger number of players—Garnier-Brun *et al.* [16] introduced the Sherrington-Kirkpatrick (SK) game. In this game, many players repeatedly interact, each choosing between two possible actions. Similar to Ref. [12], the interaction between agents is governed by random interaction coefficients, drawn at the beginning and then fixed as the learning process unfolds. As shown in Ref. [16], learning in this game shows rich collective dynamics, and the players are generally unable to learn the best possible point in the strategy space. This leads to an outcome based on the idea of *satisficing*: players must make do with suboptimal rewards which are nonetheless satisfactory and sufficing.

The term *SK game* highlights similarities with the celebrated SK model of spin-glass physics [17]. In the SK game, players choose between two actions (akin to spin up or down in the SK spin-glass model), and as in the SK spin-glass model, there are quenched random interaction coefficients. In this context, we also note work by Menezes and Sherrington on Minority Games (MGs) with SK interactions [18], where there are also many players, each with only two actions to choose from.

The choice to study ensembles of random payoff matrices in Refs. [12–14,16] follows the ideas proposed, for example, by May [19] to characterize large complex systems. Such systems will generally have many interaction coefficients (typically, if there are N degrees of freedom, there would be order $N \times N$ coefficients). It would be impossible to study the outcome in detail as a function of so many input parameters. Additionally, this would not provide much useful insight either. Instead May therefore introduced a statistical view and determined the outcome as a function of key summary statistics of the interaction coefficients, such as their variance. In the context of learning in complex games, this means looking for key parameters of the ensemble of random games and the learning algorithm that determine if and when learning converges or if it remains volatile. Depending on how competitive the game is and on how quickly players adapt, the authors of Refs. [12,13] found regimes for two-player games with many

actions in which learning converges to a unique stable fixed point (albeit not a Nash equilibrium), phases in which there are many stable fixed points, and a regime in which learning remains chaotic and unpredictable. Garnier-Brun *et al.* [16] found broadly similar phases; they also investigated cyclic or quasicyclic behavior in the SK game.

In this paper, we extend the work in Ref. [16] in several different ways: We provide a bottom-up construction of the SK game dynamics starting from concrete payoff matrices. The game studied in Ref. [16] is in fact a special case of a more general setup. More specifically, Garnier-Brun *et al.* focused on a model without external field. This describes a setup in which no player has any intrinsic bias to any of his or her two strategies. As we will describe, some of the phenomena reported in Ref. [16] are restricted to that special case, and allowing for nonzero random fields changes the behavior and phase diagram. We also introduce a version of the SK game in which some players can choose whether to participate in the game at any one time step but where the other players must participate at each step. Thus, the number of active players varies in time. We call this the *grand-canonical SK game*, in analogy with the grand-canonical ensemble in statistical physics and with the so-called grand-canonical MG [20].

The outline of the paper is as follows: In Sec. II, we describe how the SK game can be constructed from a concrete payoff structure and how this leads to setups with and without random fields. Section III contains a formal analysis, using dynamic mean-field theory, a well-established technique in disordered systems. In Sec. IV, we discuss and compare the predictions from this theory with numerical simulations. We establish the different types of behavior in the games without and with field, respectively. Section V contains the definition and analysis of the grand-canonical SK model. We highlight how this model differs from the SK game in which all players must participate at each step. We present our conclusions in Sec. VII.

II. SK MODEL FORMULATION

A. Mixed strategies and payoff structure

We consider a setup with N players. We label the players by letters i or j in $\{1, \dots, N\}$. In the SK game, each player has two pure strategies (actions) to choose from; we label these actions ± 1 . The game proceeds in discrete rounds. We write $s_i(t) = 1$ if player i takes action 1 at time t and $s_i(t) = -1$ if he or she plays action -1 . The variable $x_i(t)$ denotes the mixed strategy of player i at time t . More precisely, $x_i(t)$ is the probability for player i to play action $+1$ at that time. He or she plays action -1 with probability $1 - x_i(t)$. We write $\mathbf{x} = (x_1, \dots, x_N)$.

The interaction between players occurs through two-player games. These games are simple 2×2 games in the sense that only two players participate in each game, and they each only have two actions to choose from. The complexity in the model arises from the assumption that each player engages in many two-player games with different opponents.

More specifically, in each round, each player plays one two-player game with each of the other players. Each player decides what action to take at any given time step, and that

same action $s_i(t) = \pm 1$ is then applied to all two-player games with the other players. In the next step, the player can take a different action. Each player derives a payoff from these interactions via a set of payoff matrices which we will define below. This setup is combined with a learning dynamic that the players use to update their mixed strategies. This will also be described in more detail below.

We write a_{ij}^{++} for the payoff player i receives from the interaction with player j if i plays action +1 and j plays +1. Similarly, a_{ij}^{+-} is the payoff to i if i plays +1 and j plays -1. The quantities a_{ij}^{-+} and a_{ij}^{--} are defined analogously.

Throughout the paper, we assume that the $a_{ij}^{ss'}$ ($s, s' \in \{\pm 1\}$) variables are drawn randomly at the beginning and then remain fixed as the dynamics unfold. Thus, the payoff matrices are the quenched disorder of the problem. We will define the precise ensemble from which these payoff matrix elements are drawn further below.

B. Sato-Crutchfield learning

In the model, each player makes the decision on what action to play based on a pair of propensities. These describe a tendency for the player to play either action. We label these $q_i^\pm(t)$, such that $q_i^+(t)$ is the propensity for player i to play action +1 at time t , and similarly for $q_i^-(t)$. As described below, these propensities capture the outcome of the learning process of the player. The propensities determine the mixed strategy of the player via a softmax (or logit) rule

$$x_i^\pm(t) = \frac{e^{\beta q_i^\pm(t)}}{e^{\beta q_i^+(t)} + e^{\beta q_i^-(t)}}. \quad (1)$$

By construction, $x_i^+(t) + x_i^-(t) = 1$, for all i and t .

The coefficient $\beta \geq 0$ is a parameter of the learning and decision model; it describes an intensity of choice. We assume that all players use the same intensity of choice. For very large β , each player is very likely to play the action with the higher propensity. For $\beta = 0$ instead, the player plays each of the two actions with probability 1/2, irrespective of the propensities.

To simplify the notation, we introduce the following quantities:

$$\begin{aligned} \Pi_i^+[\mathbf{x}] &= \sum_j a_{ij}^{++} x_j^+ + \sum_j a_{ij}^{+-} x_j^-, \\ \Pi_i^-[\mathbf{x}] &= \sum_j a_{ij}^{-+} x_j^+ + \sum_j a_{ij}^{--} x_j^-. \end{aligned} \quad (2)$$

We have written $x_j^+ = x_j$ here for the probability of player j to use action +1 and $x_j^- = 1 - x_j$ for the player to play action -1. The quantity $\Pi_i^s[\mathbf{x}]$ is the (expected) payoff to player i if i plays action $s \in \{\pm 1\}$ and given the mixed strategies $\{x_j^\pm\}$ of the remaining players. We use the notation $\mathbf{x} = (x_1^+, \dots, x_N^+)$. We will occasionally omit the argument $\mathbf{x}(t)$ of the payoffs to keep the notation compact.

To update the propensities, the players use Sato-Crutchfield learning [9–11], a variant of EWA learning [21]. More specifically, we use a batch-learning model [11,12]. This assumes that players take an average over a large (formally infinite) batch of successive interactions with the other players between updates of their propensities. The updates are carried

out assuming that no player changes his or her mixed strategy within a batch of interactions. The update rules then are

$$\begin{aligned} q_i^+(t+1) &= (1-\alpha)q_i^+(t) + \Pi_i^+[\mathbf{x}(t)], \\ q_i^-(t+1) &= (1-\alpha)q_i^-(t) + \Pi_i^-[\mathbf{x}(t)]. \end{aligned} \quad (3)$$

Thus, the update consists of two components. The first is the addition of $\Pi_i^\pm[\mathbf{x}(t)]$ to the respective propensities $q_i^\pm$. In simple terms, agent i adds to his or her propensity for action +1 the average payoff he or she would have received in a large batch of games if player i had always played action +1 and given the current mixed strategies of the other players. Similarly, $\Pi_i^-[\mathbf{x}(t)]$ is added to q_i^- .

The second component of the update is memory loss, characterized by the parameter $\alpha \in [0, 1]$. For $\alpha = 0$, the update simply accumulates payoffs. The propensities $q_i^\pm(t)$ are then the total (expected) payoffs player i would have received had he or she always played action ± 1 , respectively, up to time t . For $\alpha = 1$, on the other hand, past rounds are immediately forgotten, and only the most recent batch is considered to compute propensities for the next time step. Intermediate values of α describe exponential discounting of past payoffs, with a time scale $1/\alpha$. Different choices for the learning dynamics are possible. For example, the authors of Ref. [16] placed a factor of α in front of the term $\Pi_i^\pm$ in Eq. (3). This is also common in the cognitive sciences (see, e.g., Refs. [22–25]). Except for special cases, this can be absorbed into a redefinition of the intensity of choice β and makes no major difference. We follow the conventions of Ref. [12].

Combining Eqs. (3) and (1), we have

$$x_i^\pm(t+1) = \frac{x_i^\pm(t)^{1-\alpha} e^{\beta \Pi_i^\pm[\mathbf{x}(t)]}}{x_i^+(t)^{1-\alpha} e^{\beta \Pi_i^+[\mathbf{x}(t)]} + x_i^-(t)^{1-\alpha} e^{\beta \Pi_i^-[\mathbf{x}(t)]}}. \quad (4)$$

Given $x_i^+(t) + x_i^-(t) = 1$, for all i, t , we can reduce the dynamical system to one variable per player. We choose these variables as

$$m_i(t) = x_i^+(t) - x_i^-(t), \quad (5)$$

taking values in the interval $-1 \leq m_i \leq 1$. If $m_i(t) = -1$, then player i plays action -1 with probability one at time t . If $m_i(t) = 1$, he or she plays action +1. For $m_i(t) = 0$, the player plays each action with probability 1/2.

Following Ref. [12], we take the continuous-time limit $\alpha, \beta \rightarrow 0$ at fixed ratio $\lambda = \alpha/\beta$. As detailed in Appendix A, we then find

$$\frac{d}{dt} m_i = \frac{1}{2} (1 - m_i^2) \left[\Pi_i^+ - \Pi_i^- - \lambda \ln \frac{1 + m_i}{1 - m_i} \right]. \quad (6)$$

The right-hand side is proportional to $1 - m_i^2$; this makes sure m_i never leaves the interval $[-1, 1]$. The first term in the square bracket is akin to what one would see in a replicator equation. If $\Pi_i^+ > \Pi_i^-$, then this term promotes an increase of m_i , and thus, player i becomes more likely to play action +1. For $\Pi_i^+ < \Pi_i^-$, this term leads to a reduction of m_i , making action -1 more likely.

The term proportional to λ represents memory loss. Recalling that $\lambda \geq 0$, it acts as a force toward a random strategy $m_i = 0$, where the player plays each of the two actions with probability 1/2. For example, when m_i approaches one from

below, the expression $-\lambda \ln \frac{1+m_i}{1-m_i}$ will become large and negative and dominate over the payoff term $\Pi_i^+ - \Pi_i^-$. This will lead to a reduction of m_i .

C. The SK game

We now investigate the payoff structure in more detail and show how exactly the dynamics in Ref. [16] can be constructed.

We introduce

$$A_{ij}^+ = a_{ij}^{++} - a_{ij}^{-+}, \quad A_{ij}^- = a_{ij}^{+-} - a_{ij}^{--}. \quad (7)$$

The quantity A_{ij}^+ is the improvement in payoff player i can achieve in the interaction with j by playing strategy $+1$ as compared with i playing -1 and assuming that j plays action $+1$. The coefficient A_{ij}^- is a similar payoff difference but now assuming j plays -1 .

The time evolution of m_i in Eq. (6) only depends on payoff differences, which can be written as follows:

$$\Pi_i^+ - \Pi_i^- = \frac{1}{2} \sum_j (A_{ij}^+ - A_{ij}^-) m_j + \frac{1}{2} \sum_j (A_{ij}^+ + A_{ij}^-). \quad (8)$$

The term containing $A_{ij}^+ + A_{ij}^-$ in this equation acts as a random field. A nonzero value of this quantity indicates that one of the actions of player i does intrinsically better than the other action when playing against j , and when the other player j acts randomly ($x_j^\pm = 1/2$, i.e., $m_j = 0$). We now distinguish between two different scenarios, leading to dynamics without and with such random fields, respectively. In the following, when we speak of a game, we always mean the collection of coefficients A_{ij}^+ and A_{ij}^- for all pairs i, j .

1. Single-matrix game, no external field

We first focus on games where there is no background advantage to any action. This setup leads to the dynamics as described in Ref. [16].

The background bias vanishes when $A_{ij}^+ = -A_{ij}^-$, for all i, j , and we will assume this is the case throughout this subsection. We simply write A_{ij} for this quantity. We then have the following picture: If player j plays action $+1$, then player i scores A_{ij} more points by playing $+1$ than player i using action -1 . If player j plays -1 , then player i scores A_{ij} points more by playing -1 than playing $+1$. A similar interpretation holds for A_{ji} .

This can be illustrated as follows (we draw attention to a similar table in the Appendix of Ref. [16]):

	$j \text{ plays } +1$	$j \text{ plays } -1$	
$i \text{ plays } +1$	$A_{ij}/2$	$-A_{ij}/2$	
$i \text{ plays } -1$	$-A_{ij}/2$	$A_{ij}/2$	

(9)

where the entries illustrate the payoff to player i . We stress that this is not an actual payoff matrix but is only accurate as far as payoff differences go (payoff to player i when i plays action $+1$ compared with i playing action -1). Player j faces a similar table

	$i \text{ plays } +$	$i \text{ plays } -$	
$j \text{ plays } +$	$A_{ji}/2$	$-A_{ji}/2$	
$j \text{ plays } -$	$-A_{ji}/2$	$A_{ji}/2$	

(10)

The game is now entirely defined by the $N \times N$ matrix A_{ij} . It is important to remember that, unlike in Ref. [12], i and j label players, not actions.

The model we end up with has payoff difference $\Pi_i^+ - \Pi_i^- = \sum_j A_{ij} m_j$, see Eq. (8), and hence dynamics given by (after re-scaling time to absorb a factor of $1/2$)

$$\frac{d}{dt} m_i = (1 - m_i^2) \left(\sum_j A_{ij} m_j - \lambda \ln \frac{1 + m_i}{1 - m_i} \right). \quad (11)$$

This is the continuous-time analog of eq. (7) in Ref. [16], where we recall again the different conventions regarding the memory-loss parameter. A brief further discussion can also be found in Appendix B.

We next discuss correlations between A_{ij} and A_{ji} . Imagine a situation where $A_{ij} = A_{ji}$ for a particular pair of players. If $A_{ij} = A_{ji} > 0$, and setting the other plays aside, the best outcome for the pair i, j occurs if they both play $+1$ or -1 . If $A_{ij} = A_{ji} < 0$, then the best outcome for both players is to take opposite actions (one player plays $+1$, the other -1). Thus, as far as the interaction between i and j is concerned, there is no competition when $A_{ij} = A_{ji}$. If one of the mutually favorable states has been reached, neither player can unilaterally increase his or her payoff from the interaction with this particular other player.

Assume now instead a situation in which $A_{ij} = -A_{ji}$, and $A_{ij} > 0$. Player i will then always want to do what player j does. Player j , on the other hand, will want to do the opposite of what i does. The roles of i and j are reversed if $A_{ij} = -A_{ji} < 0$. This is akin to the matching pennies game (see, e.g., Ref. [26]). Whatever combination of actions is played by i and j ($++$, $+-$, $-+$, or $--$), there is always one player who can do better by unilaterally changing his or her action. Also, the two payoffs (that to i and that to j) always sum to a constant (keeping in mind that the matrices above are not payoff matrices). Thus, up to a constant background payoff, this is a zero-sum game.

As motivated in the introduction, we do not focus on particular instances of the payoff matrices. Instead, we focus on the limit of many players (formally, $N \rightarrow \infty$), and our goal is to characterize the outcome of learning in terms of the statistics of the payoff matrix entries. We therefore assume that the A_{ij} values are drawn from a joint distribution and treat the first and second moments as control parameters. Maximum-entropy arguments [12] then lead to a Gaussian distribution. To not introduce any bias, we assume that the mean of each A_{ij} vanishes. Considering the limiting cases $A_{ij} = A_{ji}$ and $A_{ij} = -A_{ji}$ discussed above, we allow for correlations between A_{ij} and A_{ji} . We then arrive at a Gaussian random matrix ensemble with

$$\overline{A_{ij}^2} = \overline{A_{ji}^2} = \frac{\sigma^2}{N-1}, \quad \overline{A_{ij} A_{ji}} = \frac{\Gamma \sigma^2}{N-1}, \quad (12)$$

for all pairs $i < j$. The overbar denotes averages over the random matrix ensemble. The variances of the matrix elements scales as $(N-1)^{-1}$; this guarantees a well-defined analytical limit for $N \rightarrow \infty$ and allows us to carry out a statistical physics analysis similar to Refs. [12, 16]. We choose a factor of $N-1$ here because players are assumed not to play against themselves. In the limit $N \gg 1$, it makes no difference

whether the variance scales as $(N-1)^{-1}$ or N^{-1} , as is more common in the spin-glass literature.

We note that the parameter σ sets the scale of the payoffs, and given that payoffs are multiplied by the strength of selection [Eq. (4)], only the product of β and σ is relevant. Equivalently, any changes of σ can be absorbed into changes of the parameter $\lambda = \alpha/\beta$. We use $\sigma = 1$ throughout in all figures.

The parameter Γ takes values in the interval from -1 to 1 . When $\Gamma = -1$, we have $A_{ij} = -A_{ji}$ with probability one. This is akin to a zero-sum game. For $\Gamma = 1$, on the other hand, we have $A_{ij} = A_{ji}$. This resembles a potential game, where the incentives of the players are aligned with a common underlying potential. Values $-1 < \Gamma < 1$ interpolate between these two extremes.

2. Bimatrix version, random fields

We can also consider scenarios in which we do not impose $A_{ij}^+ = -A_{ij}^-$ in Eq. (8). In this situation, there can be an intrinsic bias to one of the two actions of player i in the interaction with j . The quantity $B_{ij} = (A_{ij}^+ + A_{ij}^-)/2$ quantifies this bias (with positive values representing a bias toward action $+1$). Writing $A_{ij} = (A_{ij}^+ - A_{ij}^-)/2$, this can be illustrated in the payoff matrix for player i ,

$$\begin{array}{c|cc} & j \text{ plays } +1 & j \text{ plays } -1 \\ \hline i \text{ plays } +1 & (A_{ij} + B_{ij})/2 & (-A_{ij} + B_{ij})/2 \\ i \text{ plays } -1 & (-A_{ij} - B_{ij})/2 & (A_{ij} - B_{ij})/2 \end{array} \quad (13)$$

As before, the matrix in Eq. (13), like the one in Eq. (9), should be understood as encoding only the payoff differences for player i that arise from changing his or her own action. More precisely, suppose first that player j plays $+1$. Then the payoff advantage to player i from playing $+1$ rather than -1 is given by $A_{ij} + B_{ij} = A_{ij}^+$. Similarly, if player j plays -1 , the payoff advantage to player i from playing $+1$ rather than -1 is given by $-A_{ij} + B_{ij} = A_{ij}^-$.

The learning dynamics in Eq. (11) can now be written as

$$\frac{d}{dt}m_i = (1 - m_i^2) \left\{ -\lambda \ln \frac{1 + m_i}{1 - m_i} + \sum_j A_{ij}m_j + h_i \right\}, \quad (14)$$

where $h_i = \sum_j B_{ij}$.

Like A_{ij} , we take B_{ij} to be quenched Gaussian random numbers with zero mean and with

$$\overline{(B_{ij})^2} = \frac{\Delta^2}{N-1}, \quad (15)$$

where Δ controls the variances of B_{ij} . For $\Delta = 0$, we return to Eq. (11). In this limit, there is no intrinsic bias for any player to choose any given action. A positive value of Δ means each player has a random intrinsic bias toward a given action. As Δ^2 becomes large (compared with σ^2), players become increasingly biased toward their favorite actions, and in the limit $\Delta^2 \rightarrow \infty$, they always play this action.

We could also allow the quantities B_{ij} and B_{ji} to be correlated, with $\overline{B_{ij}B_{ji}} = \frac{\gamma\Delta^2}{N-1}$. However, the dynamic mean-field calculation shows that correlations between B_{ij} do not appear in the effective single-player description. As a result, it does not affect the behavior of the system when the number of

players becomes very large $N \rightarrow \infty$. We also note that A_{ij} and B_{ij} are uncorrelated. This is not an extra assumption but follows directly from how these two quantities are defined.

III. GENERATING FUNCTIONAL ANALYSIS

A. Effective single-player dynamics

A dynamic mean-field theory for the SK games with and without field can be derived following standard generating-functional methods [27], see also Refs. [28,29]. Alternatively, the cavity method can be used (see, e.g., Refs. [30,31]). We do not include the full derivation in the main paper, but a summary can be found in Appendix C. We find the following process for an effective representative player:

$$\dot{m} = -\lambda(1 - m^2) \ln \frac{1 + m}{1 - m} + (1 - m^2) \left[\Gamma \sigma^2 \int dt' G(t, t') m(t') + \eta(t) \right]. \quad (16)$$

In this expression, the variable $\eta(t)$ is Gaussian colored noise, and the integral term represents retarded self-interaction. These contributions arise from the average over the disorder. We have the following self-consistency relations:

$$\begin{aligned} C(t, t') &\equiv \langle \eta(t) \eta(t') \rangle = \sigma^2 \langle m(t) m(t') \rangle + \Delta^2, \\ G(t, t') &= \frac{\delta}{\delta \eta(t')} \langle m(t) \rangle. \end{aligned} \quad (17)$$

The angle brackets $\langle \dots \rangle$ represent an average over realizations of this effective process. Equations (16) and (17) define a self-consistent problem for the dynamic order parameters $C(t, t')$ and $G(t, t')$. The statistics of the effective process in Eq. (16) in trajectory space reproduce the statistics of the player-specific paths $m_i(t)$ in the original game learning dynamics.

B. Fixed-point analysis

We now broadly follow the ideas in Ref. [15]. Assuming the dynamics converges to a stable fixed point (independent of initial conditions), we have a well-defined solution for the macroscopic order parameters of the form $C(t, t') = q$ and $G(t, t') = G(t - t')$, with causality implying $G(t, t') = 0$ for $t' > t$. We define $\chi = \int_0^\infty d\tau G(\tau)$. At the fixed point, m and η are static random variables with $\langle m^2 \rangle = q$ and $\langle \eta^2 \rangle = \sigma^2 q + \Delta^2$.

Writing $\eta = \sqrt{q\sigma^2 + \Delta^2} z$, where $z \sim \mathcal{N}(0, 1)$, the fixed-point equation becomes

$$(1 - m^2) \left[-\lambda \ln \frac{1 + m}{1 - m} + \Gamma \sigma^2 \chi m + \sqrt{q\sigma^2 + \Delta^2} z \right] = 0. \quad (18)$$

In addition to the boundary solutions $m = \pm 1$, a nontrivial solution satisfies

$$-\lambda \ln \frac{1 + m}{1 - m} + \Gamma \sigma^2 \chi m + \sqrt{q\sigma^2 + \Delta^2} z = 0. \quad (19)$$

Given that $\lambda > 0$, the term $\lambda \ln \frac{1+m}{1-m} - \Gamma \sigma^2 \chi m$ diverges to $-\infty$ as $m \downarrow -1$ and $+\infty$ as $m \uparrow 1$, ensuring at least one solution $m \in (-1, 1)$ for any given z . There are also solutions

$m = \pm 1$. These correspond to diverging propensities ($m \rightarrow \pm\infty$), which are unphysical for any nonzero memory loss parameter ($\lambda > 0$). The physically meaningful solution $m(z)$ of Eq. (19) is therefore strictly in the interval from -1 to 1 and defines a one-to-one relation between z and m [32].

The corresponding self-consistency equations are given by

$$q = \int Dz m(z)^2, \quad \chi = \frac{1}{\sqrt{q\sigma^2 + \Delta^2}} \int Dz \frac{\partial m(z)}{\partial z}, \quad (20)$$

where $Dz = \frac{dz}{\sqrt{2\pi}} e^{-z^2/2}$.

We now briefly comment on the model without external fields. Our model then comes closest to the work in Ref. [16], where a fixed-point regime with all $m_i = 0$ was reported (a paramagnetic state). We also find such a state in our model when there are no external fields. Specifically, the fixed-point solution to Eq. (18) is then given by $m(z) = 0$, for all z , and we have $q = 0$. All players play fully mixed strategies, i.e., they each play actions ± 1 randomly with probability $1/2$ each. The quantity χ as given in Eq. (20) is then to be understood as follows. We write $m(\eta)$ for the solution of

$$-\lambda \ln \frac{1+m}{1-m} + \Gamma \sigma^2 \chi m + \eta = 0, \quad (21)$$

and noting that all probability for η concentrates on $\eta = 0$ at the paramagnetic solution, we then have $\chi = \left. \frac{\partial m(\eta)}{\partial \eta} \right|_{\eta=0}$.

As we will demonstrate below, this paramagnetic solution is specific to the SK game with no intrinsic bias (external field). The fixed-point phase is more involved when $\Delta^2 \neq 0$.

C. Linear-stability analysis

Following the steps described in Refs. [15,29,31], we can identify the onset of a linear instability from the mean-field process. Details are given in the Appendixes. We find that the stability condition can be written as

$$\sigma^2 \left\langle \left(\frac{\partial m}{\partial \eta} \right)^2 \right\rangle < 1. \quad (22)$$

We comment briefly on the case $\Delta = 0$. The self-consistent solution is then $q = 0$, and $m \equiv 0$, $\eta \equiv 0$. Nonetheless, the object $\partial m / \partial \eta$ [with $m(\eta)$ the solution of Eq. (21)] is well defined and will generally be nonzero.

If there is no external field ($\Delta = 0$), the stability condition in Eq. (22) simplifies to $\sigma < 2\lambda - \Gamma \sigma^2 \chi$. One can then show (see again the Appendixes) that the onset of instability occurs when λ takes the value

$$\lambda_{\text{crit}} = (1 + \Gamma)\sigma/2, \quad (23)$$

with instability for $\lambda < \lambda_{\text{crit}}$.

IV. STABILITY DIAGRAM AND DYNAMICAL BEHAVIOR

A. Different dynamical regimes

Similar to the complex two-player games with many actions per player in Ref. [12], the SK game has three dynamical regimes: (i) When λ is large, one finds a phase with a unique stable fixed point for fixed realizations of the payoff matrices;

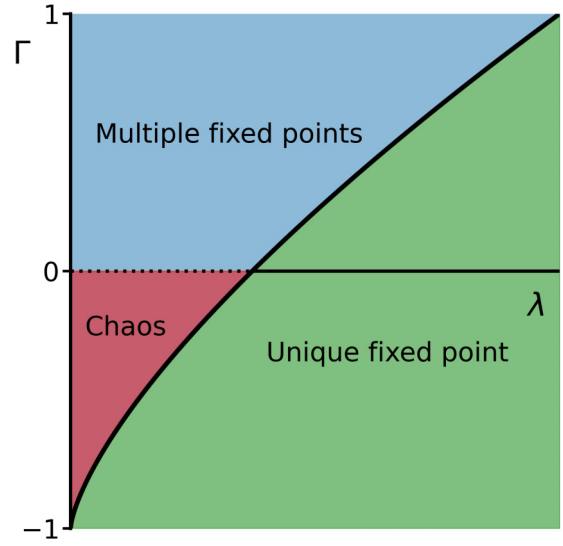


FIG. 1. Illustration of the general structure of the stability diagram of the SK game as a function of the competition parameter Γ and the memory-loss parameter λ . The phase diagram for the grand-canonical SK game (to be defined below) also has this general structure.

(ii) we find a volatile regime for small λ and when $\Gamma < 0$; and finally, (iii) we find a multiple-fixed point regime for small λ when $\Gamma > 0$. Figure 1 shows an illustration of this general structure. We stress that this indicates the dominant behavior that we expect for very large (infinite) systems in each of the phases. In simulations (necessarily conducted at finite N), it is possible that some samples in the formally stable phase remain volatile or that convergence is found for some runs in the chaotic phase. However, these effects generally reduce with an increased number of players.

To illustrate typical trajectories in the stable phase, we focus on one particular instance of a game with no bias toward any action ($\Delta = 0$) in Fig. 2. In the example, there are $N = 100$ players, and the pairwise games are moderately competitive ($\Gamma = -0.5$). In the upper panel, the players are subject to relatively quick memory loss (unique-fixed point phase). The dynamics converge to a state in which all players choose each of their actions at random with probability $1/2$ (i.e., $m_i = 0$, for all i). This is the paramagnetic state described in Ref. [16]. For longer memories (lower panel, unstable phase), the dynamics remain volatile and potentially chaotic.

Figure 3 shows similar trajectories but now for a biased game ($\Delta \neq 0$). For long memories (lower panel), the dynamics is again irregular and does not settle down. When λ is sufficiently large (memory loss sufficiently quick, upper panel), we find again a phase in which each random game has a unique stable fixed point. In contrast with the unbiased game (Fig. 2), the distribution of m_i at the fixed point is spread out in the interval $(-1, 1)$ though. This means that each player plays one of his or her actions more often than the other.

The distribution of $\{m_i\}$ in the fixed-point phase is shown for different values of the bias parameter Δ in Fig. 4. For small biases, the distribution is concentrated on $m = 0$. All players then play each of their strategies with more or less the same probability (mixed strategies are near $x_i = 1/2$). As the

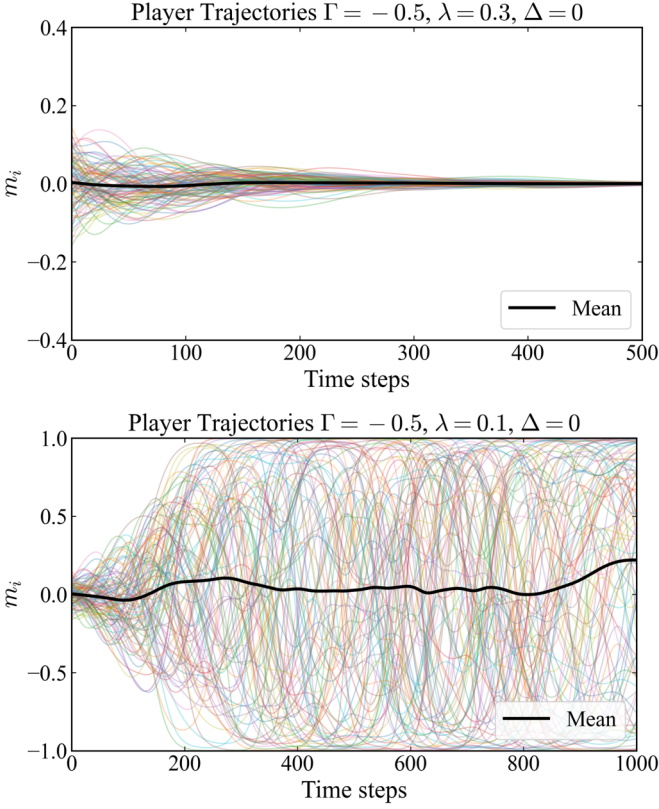


FIG. 2. Simulated trajectories of a game with no intrinsic bias ($\Delta = 0$). Each line represents the time series corresponding to m of each player. The upper panel is for $\lambda = 0.3$ (quick memory loss, stable phase), the lower panel shows the same game but with a learning dynamic with longer memory ($\lambda = 0.1$, chaotic phase). Both games are identical (have the same payoff matrices and parameters), with the exception of λ . In the simulations, there are $N = 100$ players, $\beta = 0.01 \times \sqrt{N} - 1 \approx 0.099$. The game is moderately competitive ($\Gamma = -0.5$). In all figures, we use $\sigma = 1$.

random bias is increased (increasing Δ), the distribution of m_i across players spreads out and eventually becomes bimodal. For large biases, each m_i is either near -1 or near 1 , indicating that each player predominantly plays one of his or her two actions.

B. Stability diagram and role of the random field

We now discuss the stability behavior in more detail. Whether the learning dynamics is stable (in the limit of a large number of players) for a particular parameter combination can be determined by first solving the self-consistency relations in Eq. (20) and then testing the stability condition in Eq. (22). From this, we obtain the stability boundary as a function of Γ , λ , and Δ .

The resulting stability diagram is shown in Fig. 5. The solid black line is the stability boundary determined from the theory. This is where fixed points become linearly unstable. For each fixed value of the bias parameter Δ , the learning dynamics is predicted to converge to a unique fixed point to the right of the respective stability line. To the left, we expect either multiple fixed points or persistently volatile dynamics.

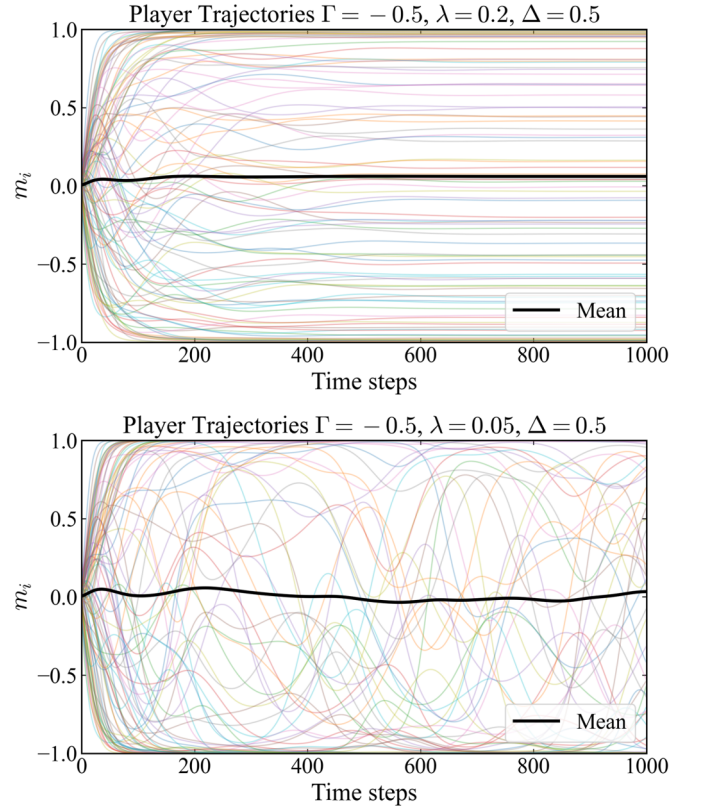


FIG. 3. Simulated trajectories for a game with nonzero bias ($\Delta = 0.5$). Upper panel is for quick memory loss ($\lambda = 0.2$, stable phase), lower panel for longer memories ($\lambda = 0.05$, chaotic phase). Remaining parameters are as in Fig. 3.

These predictions are confirmed by simulation results, also shown in Fig. 5 (background coloring). For each point in parameter space, we compute the fraction of games falling into each category (unique fixed point, multiple fixed points, or persistently volatile behavior) and assign these fractions to the corresponding color channels: red for nonconvergent

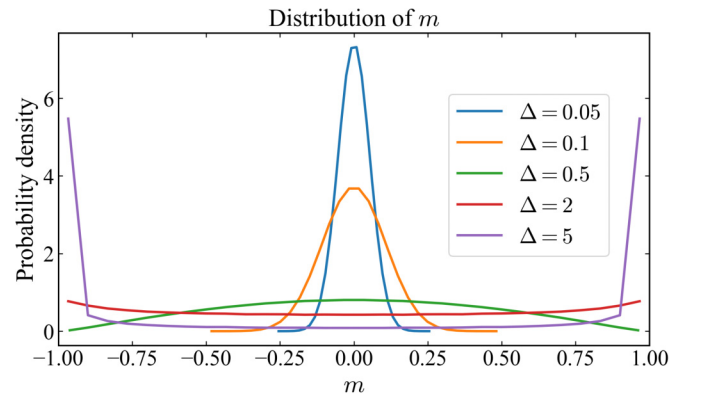


FIG. 4. Distribution of $\{m_i\}$ in the unique fixed-point phase for different magnitudes Δ of the random field. Results are for $\Gamma = -0.5$, $\lambda = 0.5$ and are obtained from the analytical theory by solving the self-consistency relations in the fixed-point phase (see Sec. III B) and then plotting the distribution of the zero $m(z)$'s of Eq. (18), with z a standard Gaussian random variable.

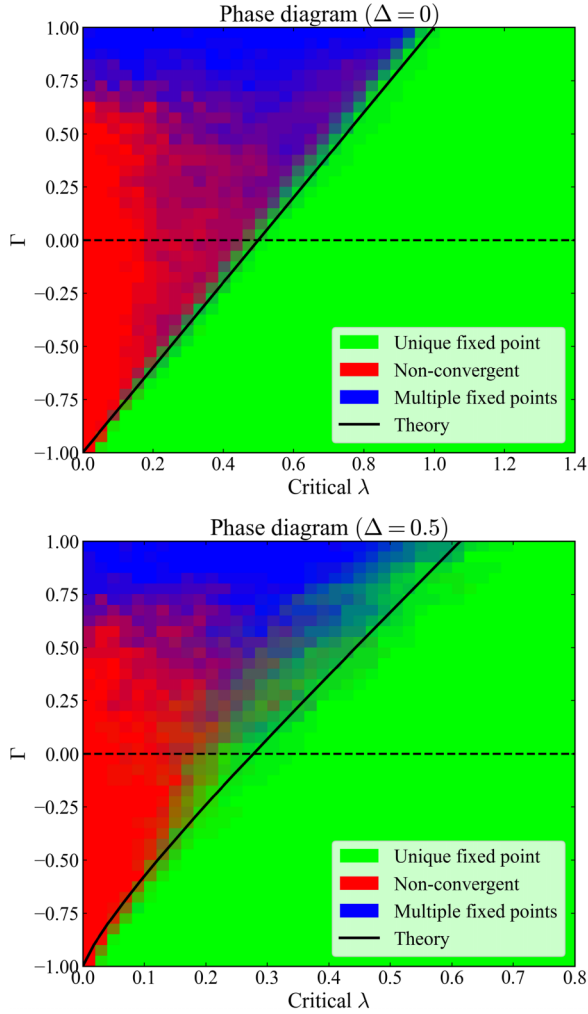


FIG. 5. Empirical phase diagrams for two different choices of Δ . For each parameter setting, we simulate 20 independent games, each with 40 distinct initial conditions. Each game consists of 100 players with $\beta = 0.01 \times \sqrt{99} \approx 0.099$. We note the different scales on the horizontal axis in the two panels. A game is classified as (i) a unique fixed point (green) if all trajectories converge to the same fixed point, (ii) nonconvergent (red) if no trajectories converge, or (iii) multiple fixed points (blue) if trajectories converge to more than one distinct fixed point. Each simulation is run for up to 50 000 steps. A trajectory is deemed converged if, over five consecutive steps, the Euclidean distance between successive updates falls below 10^{-6} . A game is classified as having a unique fixed point if all 20 trajectories converge to the same solution, identified numerically when the pairwise Euclidean distance between fixed points is less than 0.01.

behavior, green for a unique fixed point, and blue for multiple fixed points. The final color at each point is obtained by combining these three components in an RGB color scale, as was previously done for Lotka-Volterra systems with random interactions in Ref. [33]. This leads to a continuous color representation of the empirical phase behavior. Pure colors (red, green, or blue) indicate regions where one type of behavior dominates, while mixed colors correspond to coexistence of different dynamical outcomes across disorder realizations.

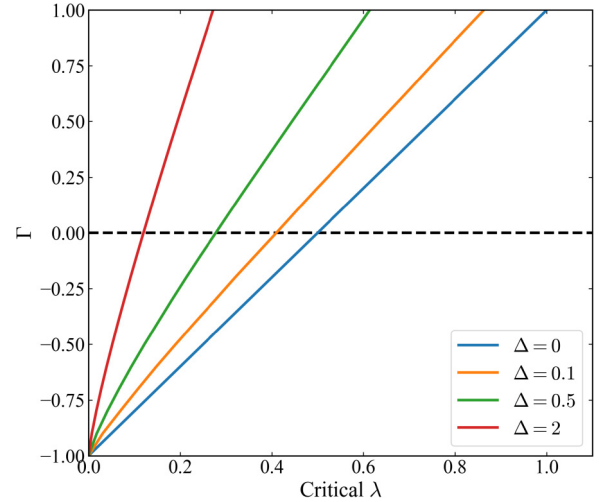


FIG. 6. Stability boundary for the SK game with random field, as a function of the memory-loss and competition parameters λ and Γ . The lines indicate the onset of linear instability as obtained from the theory (see Sec. III C), for different magnitudes Δ of the random field. As can be seen from the figure, the stable region (to the right of the respective lines) becomes larger with increasing field amplitude.

We find that the learning dynamics nearly always converge to unique fixed points to the right of the stability lines (when memory loss is sufficiently quick). When players have longer memories (λ sufficiently small), there are either many fixed points or the dynamics do not settle down. This persistently volatile behavior is mostly observed for sufficiently competitive games ($\Gamma \lesssim 0$ and/or memory loss very weak, red region in Fig. 5), whereas multiple fixed points are mostly seen for cooperation-type games ($\Gamma > 0$, blue region).

Interestingly, increasing the bias parameter Δ makes the learning more stable (the stability line is moved to the left, as shown in Fig. 6, i.e., the stable region becomes larger). A strong bias parameter effectively anchors each player toward a preferred action, which makes volatile motion or the existence of multiple fixed points less likely. This promotes stability.

Mathematically, the stabilizing effect can also be understood as follows. Introducing a random field (i.e., choosing $\Delta > 0$), smears out the distribution of m_i at the fixed point (see Fig. 4), and players become increasingly frozen (i.e., they tend to use only one of their actions; m_i close to $+1$ or -1). The stability condition in Eq. (22) can be written as (see Appendix C 2)

$$\sigma^2 \left\langle \left[\frac{(1 - m^2)}{2\lambda - (1 - m^2)\Gamma\sigma^2\chi} \right]^2 \right\rangle < 1. \quad (24)$$

As the fixed-point values m move toward ± 1 , the numerator $1 - m^2$ tends to zero, whereas the denominator in the average remains finite (for $\lambda > 0$). Thus, the expression on the left-hand side of Eq. (24) reduces, making the system more stable.

V. GRAND-CANONICAL SK GAME

A. Definition

We now take inspiration from the (grand-canonical) MG [20] and define the grand-canonical SK game. In this game,

there are two types of agents. Using the terminology from the MG literature [34] these are “speculators” and “producers”, respectively. Each speculator only has one active strategy and must decide, at each step, whether he or she wants to play or not. If a speculator plays, he or she can only use that single strategy. The only other option is to abstain. A speculator performs an adaptive learning process similar to what we described above. He or she plays if he or she expects his or her payoff to exceed a certain threshold (to be defined later in detail). Producers also only have a single strategy, but they must play no matter what, i.e., they do not have the option of abstaining.

We assume that the total number of players is N . The number of speculators is $N_s = sN$ ($0 \leq s \leq 1$), and the remaining $N_p = (1 - s)N$ agents are producers. Without loss of generality, we assume that the speculators have labels $i = 1, \dots, N_s$, and the producers are the agents $i = N_s + 1, \dots, N$.

We assume that the payoff to agent i is a_{ij} if he or she interacts in a game with player j . For this interaction to occur, both players need to play in that iteration. We write $x_i(t)$ for the probability that player i plays in round t . This quantity can be time dependent (due to the learning dynamics) for $i = 1, \dots, N_s$ (the speculators). The producers $i = N_s + 1, \dots, N$ always have $x_i = 1$.

Each speculator has a propensity to play $q_i(t)$. This propensity determines the probability with which the speculator plays via

$$x_i(t) = \frac{e^{\beta q_i(t)}}{1 + e^{\beta q_i(t)}}. \quad (25)$$

In the limit $\beta \rightarrow \infty$, player i plays with probability one if $q_i > 0$ and with probability zero if $q_i < 0$.

The propensities are updated as follows:

$$q_i(t + 1) = (1 - \alpha)q_i(t) + \Pi_i[\mathbf{x}(t)] - \varepsilon, \quad (26)$$

where α is again memory loss, and where the payoff player i can expect if he or she decides to play (and given the mixed strategies of everybody else) is given by

$$\Pi_i(\mathbf{x}) = \sum_{j=1, \dots, N_s; j \neq i} a_{ij} x_j(t) + \sum_{j=N_s+1}^N a_{ij}. \quad (27)$$

The first sum comes from games with the other speculators, and the second sum is from games with the producers.

The quantity ε can be interpreted as a cost to play, i.e., participating in the game comes with some cost to the player (e.g., a transaction, processing cost, or participation fee), that is to be subtracted from the payoff. The quantity ε therefore describes a minimum (expected) payoff speculators require to decide to play the game (setting memory loss aside, the increment of the propensity is positive if $\Pi_i > \varepsilon$, and negative if $\Pi_i < \varepsilon$).

We assume the following statistics for the payoff matrix elements:

$$\overline{a_{ij}} = 0, \quad \overline{a_{ij}^2} = \frac{\sigma^2}{N-1}, \quad \overline{a_{ij}a_{ji}} = \frac{\Gamma\sigma^2}{N-1}, \quad (28)$$

similar to the two-strategy game.

Taking the continuous-time limit as before and again writing $\lambda = \alpha/\beta$, we find

$$\dot{x}_i = x_i(1 - x_i) \left[-\lambda \ln \frac{x_i}{1 - x_i} + \sum_{j=1}^{N_s} a_{ij} x_j + \sum_{j=N_s+1}^N a_{ij} - \varepsilon \right]. \quad (29)$$

The term from the interaction with the producers acts as an external field $h_i = \sum_{j=N_s+1}^N a_{ij}$, and we have $\overline{h_i^2} = (1 - s)\sigma^2$. We also have $\overline{h_i h_j} = \Gamma\sigma^2/(N - 1)$, but this is subleading and does not contribute to the dynamics in the limit $N \rightarrow \infty$.

B. Generating-functional analysis

Carrying out a standard generating-functional calculation (see Appendix D), we find the dynamic mean-field process

$$\dot{x} = x(1 - x) \left[-\lambda \ln \frac{x}{1 - x} + s\Gamma\sigma^2 \times \int dt' G(t, t') x(t') - \varepsilon + \eta(t) \right], \quad (30)$$

with

$$\langle \eta(t) \eta(t') \rangle = \sigma^2 [s \langle x(t) x(t') \rangle + (1 - s)], \quad (31)$$

and with the response function

$$G(t, t') = \left\langle \frac{\delta x(t)}{\delta \eta(t')} \right\rangle. \quad (32)$$

The notation $\langle \dots \rangle$ again stands for an average over realizations of the effective process (i.e., over realizations of noise trajectories $\{\eta(t)\}$).

The model parameters are $\lambda, s, \Gamma, \sigma^2$, and ε , but fixed points and their stability only depend on λ/σ , ε/σ , Γ , and s . Through all figures for this model, we always use $\sigma = 1$.

We conduct a similar analysis as in the previous sections, first by assuming the dynamics have a stable fixed point; x and η then become static random variables. This time setting $\eta = \sigma\sqrt{sq + (1 - s)}z$, we obtain

$$-\lambda \ln \frac{x}{1 - x} + s\Gamma\sigma^2 \chi x - \varepsilon + \sigma\sqrt{qs + (1 - s)}z = 0. \quad (33)$$

The corresponding self-consistency equations are given by [with $x(z)$ the solution of Eq. (33)]

$$q = \int Dz x(z)^2, \quad \chi = \frac{1}{\sigma\sqrt{qs + (1 - s)}} \int Dz \frac{\partial x(z)}{\partial z}, \quad (34)$$

where we write again $Dz = \frac{dz}{\sqrt{2\pi}} e^{-z^2/2}$.

Conducting a similar linear stability analysis as in the previous section, we find the fixed point is linearly stable when

$$\sigma^2 s \left\langle \left(\frac{\partial x}{\partial \eta} \right)^2 \right\rangle < 1. \quad (35)$$

A full derivation can be found in Appendix D.

C. Stability diagram and dynamical behavior

Similar to the previous setup, we find that the grand-canonical SK game has three dynamical regimes: (i) a phase

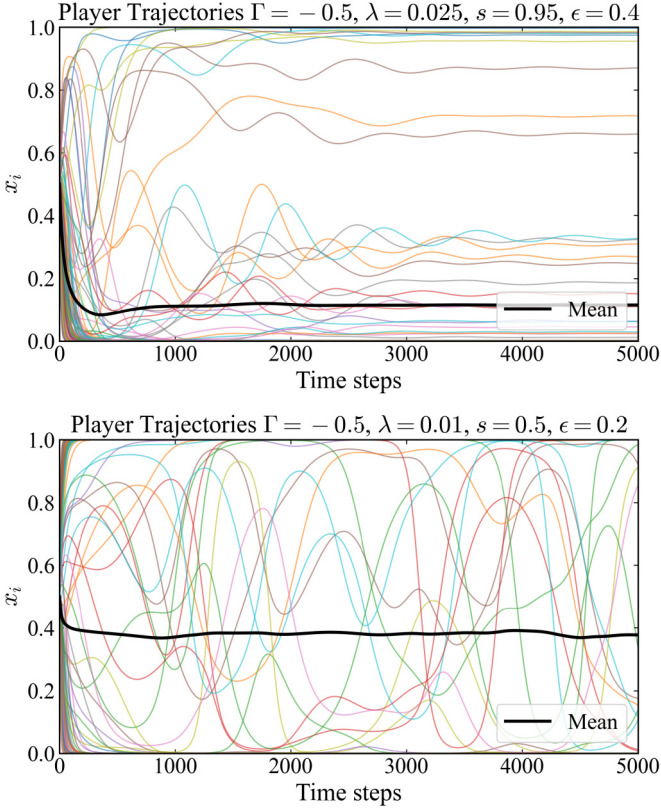


FIG. 7. Simulated trajectories of two different grand-canonical games with different parameters with one converging to the unique fixed point and the other exhibiting chaotic dynamics. Each simulation consists of 100 speculators, and each line represents the time series corresponding to the likelihood each speculator will do an action the given time ($\beta = 0.1$, $\sigma = 1$).

with a unique stable fixed point for fixed instances of the game, (ii) a volatile regime, and (iii) a regime in which each game has many fixed points. Although the critical λ is different and depends on additional parameters, the qualitative picture in terms of Γ and λ for any given choice of parameters remains similar to that shown in Fig. 1. Figure 7 illustrates two instances of learning in the the grand-canonical SK game: one in which the dynamics converges and another in which it exhibits volatile dynamics.

Similar to the previous sections, we can determine the stability diagram by solving the self-consistency equations in Eq. (34) and then testing the self-consistency condition in Eq. (35). Results in the plane spanned by the memory-loss parameter λ and the competition parameter Γ are shown in Fig. 8 for different values of the payoff threshold ε and different fractions s of speculators in the population of players.

We note, when $\varepsilon = 0$ (upper panel of Fig. 8), the phase diagram for the grand-canonical SK game takes a similar shape to the SK game with a random field. We find again stability for sufficiently quick memory loss (sufficiently large λ). Making the game more competitive (reducing Γ) makes learning more stable similar to what was found for complex two-player games in Ref. [12]. We also find that a reduction of the proportion of speculators s leads to a larger stable region. This is consistent with the first relation in Eqs. (17)

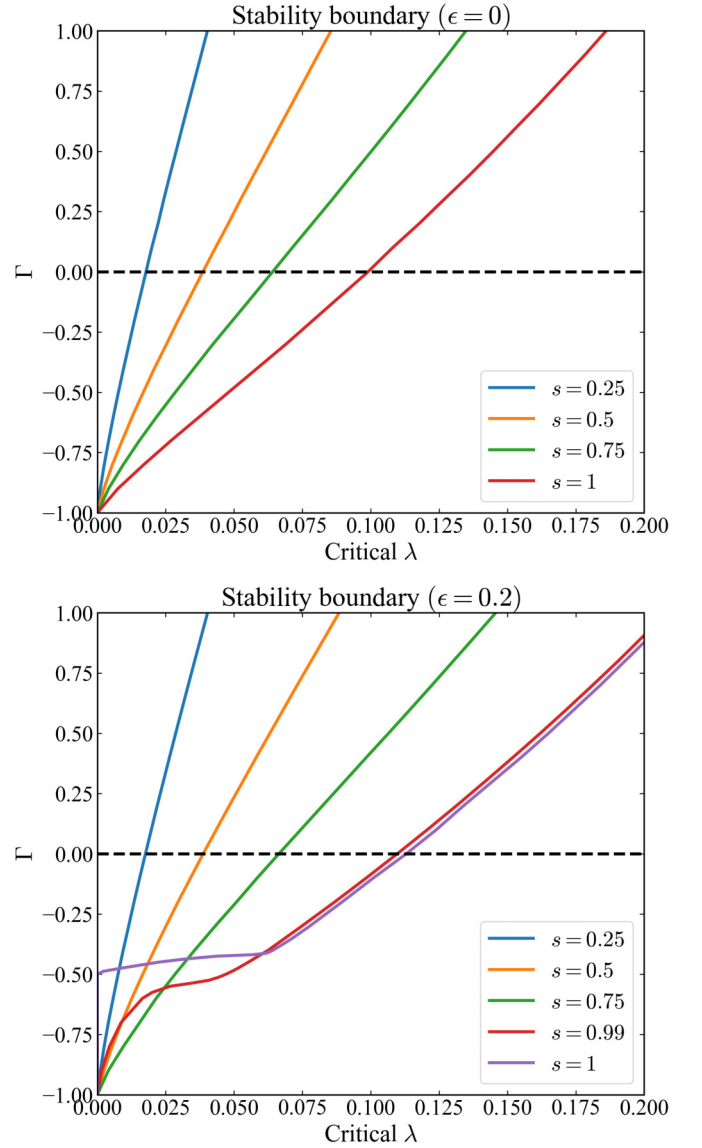


FIG. 8. Stability boundary of the grand-canonical SK game in the (λ, Γ) plane (the memory-loss and competition parameters, respectively), as predicted from the theory for varying s for $\varepsilon = 0$ (upper panel) and $\varepsilon = 0.2$ (lower panel). The learning dynamics converge to unique fixed points to the right of the relevant line. To the right, we find either chaotic dynamics or the learning process has many stable fixed points.

and (31) from which one can see that $1 - s$ is analogous to the field amplitude Δ . An increase of the fraction of producers is similar to an increase of the external field and stabilizes the dynamics.

When $\varepsilon > 0$ (lower panel of Fig. 8), we find that the shape of the stability boundary changes near $s = 1$. A kink emerges in the stability curve for negative Γ . For values of Γ near -1 , the system remains stable for very small λ , but as Γ increases, there is a point where there is a sharp increase in the critical λ . For larger Γ , the relation between Γ and λ returns to being approximately linear. A consequence of this kink is that, for some choices of Γ and ε , increasing the fraction of speculators s can help stabilize the game. We will discuss this in more

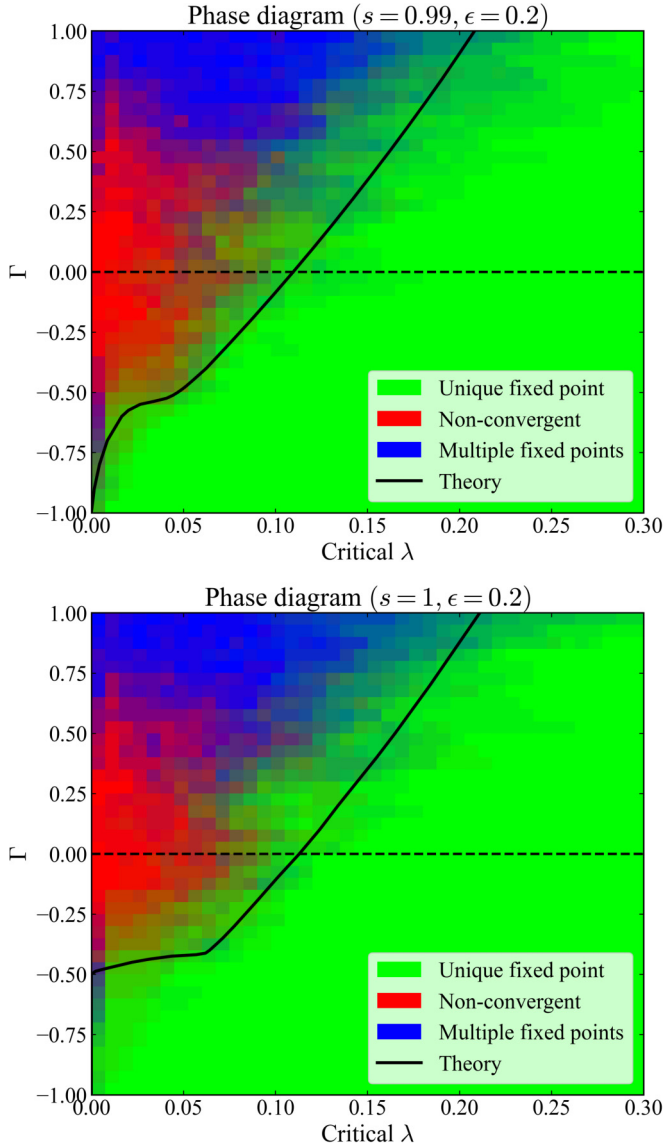


FIG. 9. Empirical phase diagrams for two different choices of s and ε . For each parameter setting, we simulate 20 independent games, each with 40 distinct initial conditions. Each game consists of 100 players with $\beta = 0.01 \times \sqrt{99} \approx 0.099$. A game is classified as (i) a unique fixed point (green) if all trajectories converge to the same fixed point, (ii) nonconvergent (red) if no trajectories converge, or (iii) multiple fixed points (blue) if trajectories converge to more than one distinct fixed point. We classify the dynamical regimes using the same approach as in Fig. 5.

detail in the next section. We note that the kink does not appear to be a numerical artifact, as we verify this behavior with simulations in Fig. 9.

D. Role of the fraction s of speculators and the payoff threshold ε on stability

An interesting picture emerges if we plot the stability diagram in the (s, ε) plane at fixed λ and Γ . We recall that s is the fraction of speculators among the players and ε is the minimum payoff speculators expect to receive to participate in

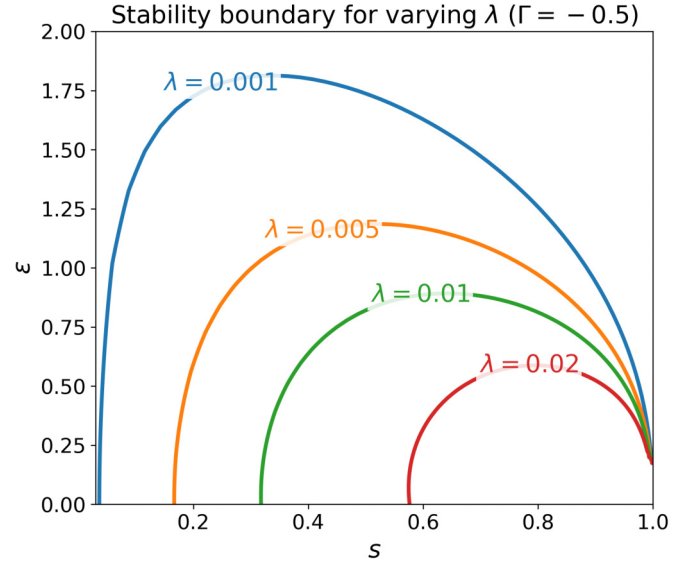


FIG. 10. Stability boundary for the grand-canonical SK game of varying λ for $\Gamma = -0.5$ as a function of s and ε . The stability line is obtained from the analytical theory. The dynamics is predicted to be stable above the stability line.

the game. We show such a stability plot in Fig. 10 for different values of the memory-loss parameter λ . The learning process is predicted to be stable (converge to a unique fixed point for any fixed random game) above the stability line. This means that, at a fixed fraction of speculators s , we find stability for sufficiently large values of ε . Increasing the payoff threshold ε reduces how often each speculator participates in the game. This reduced activity in turn stabilizes the learning dynamics.

As can also be seen from Fig. 10, increasing the fraction of speculators s can have a nonmonotonic effect on stability. Starting from a small fraction of speculators, adding more speculators tends to destabilize the system. However, once s becomes sufficiently large, a further increase can instead stabilize the learning process. This causes a re-entry behavior in Fig. 10. If we fix a value of ε that is not too large and start at low s , we are in the stable phase. Moving horizontally (constant ε) and to the right, we then enter the unstable phase. Continuing to move to the right (i.e., increasing the fraction of speculators further), we eventually return to the stable phase. This is a reflection of the kink in Fig. 8 for s close to one.

We tested these predictions in simulations, results are shown in Fig. 11. The heatmap shows the fraction of samples for which learning converges. The data confirm the re-entry behavior and are broadly in line with the stability line from the theory. Above the line, learning converges for most samples; below the line, convergence is rare.

To understand this further, we have measured the speculator activity. Results are shown in the upper panel of Fig. 12, where we plot the fraction of activity from the speculators relative to the total activity. This is the quantity $s\langle x \rangle / [s\langle x \rangle + (1-s)]$, where $\langle x \rangle$ is the mean probability to play among speculators. We find that this quantity is relatively low when learning is stable (above the stability line in the figure). A larger proportion of speculator activity appears to be associated with instability. In the lower panel, we show

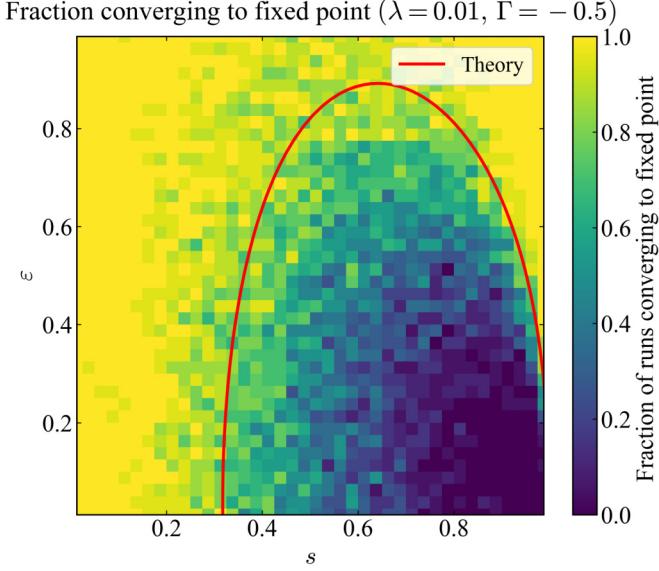


FIG. 11. Numerical confirmation of the stability diagram of the grand-canonical SK game and of the re-entry behavior as a function of the fraction of speculators s . The stability line from the theory is shown in red; learning is predicted to converge above the line. The heatmap shows the number of samples that converged in numerical simulations. For each data point, we ran 20 random games for 50 000 steps ($N = 100$, $\beta = 0.01 \times \sqrt{99} \approx 0.099$, $\Gamma = -0.5$, $\lambda = 0.01$) and recorded how often the learning dynamics converged to a fixed point.

$\langle x \rangle$ for comparison. Naturally, this quantity decreases with an increasing threshold payoff ε . An increased proportion of speculators in the population of players also seems to reduce the mean activity of each speculator. One possible explanation for this latter observation is that increasing the fraction of speculators means that there is a lesser background of producers from which the speculators could profit.

Mathematically, the re-entrance behavior as a function of s is driven by the competition between two contributions in the stability condition in Eq. (35). We write $F(s) = \sigma^2 s \langle (\partial x / \partial \eta)^2 \rangle$ for the expression on the left-hand side. Taking the derivative with respect to s , we find

$$\frac{\partial F}{\partial s} = \sigma^2 \left[\left\langle \left(\frac{\partial x}{\partial \eta} \right)^2 \right\rangle + s \frac{\partial}{\partial s} \left\langle \left(\frac{\partial x}{\partial \eta} \right)^2 \right\rangle \right]. \quad (36)$$

The first term in the square brackets captures a direct effect of increasing the number of speculators, assuming the behavior of all players remains unchanged. It indicates how strongly the activity of any one speculator changes with external perturbations (irrespective of the direction of this change toward higher or lower activity). The term is positive: Speculators are susceptible to perturbations (while producers are not). As s is increased, this has an increasing effect on the left-hand side of the stability condition in Eq. (35) and works against stability.

The second term in the square brackets in Eq. (36) $\frac{\partial}{\partial s} \langle (\frac{\partial x}{\partial \eta})^2 \rangle$ captures how the susceptibility of speculators to external perturbations changes when the composition of the player population is changed. This is the result of interactions between different speculators. This term can be negative, and if that is the case, the effects act to promote stability.

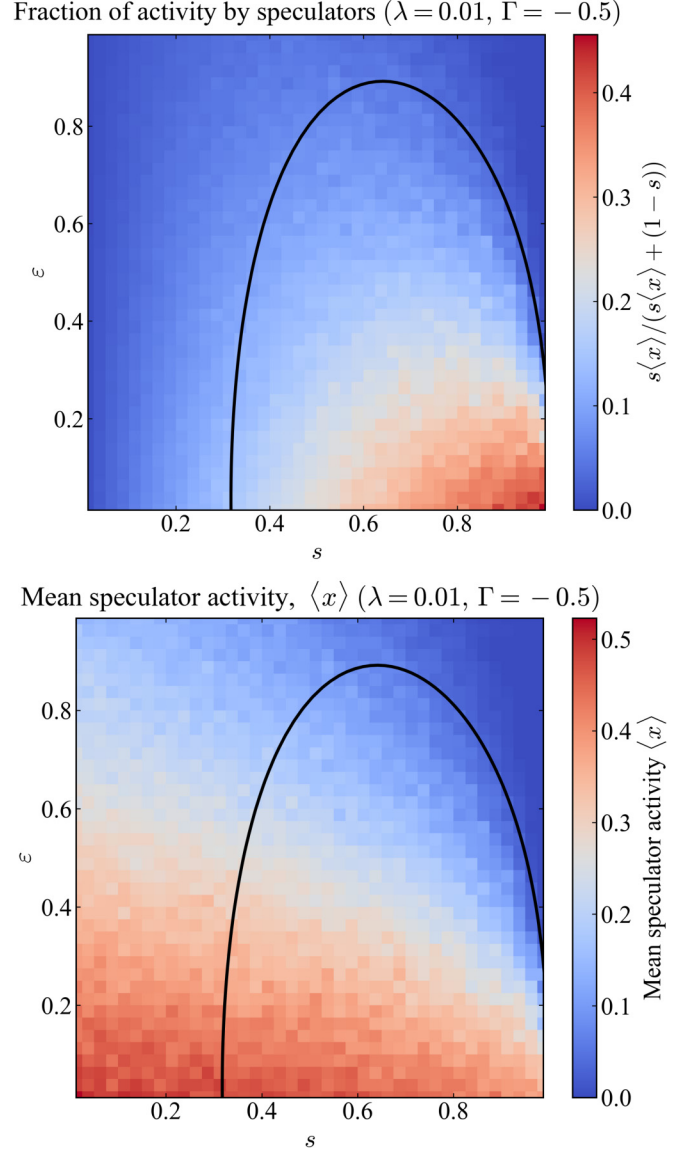


FIG. 12. Plot of the fraction of total activity by speculators (upper panel) and mean speculator activity (lower panel) as a function of s and ε . The stability line from the theory is now shown in black. The quantities for each grid point are obtained via numerical simulation averaged of 20 games, using the the same parameter choices as Fig. 11. The simulations run for 20 000 steps with measurements taken from 10 000 steps onward.

For small s , this second term is small compared with the first. Most speculators are adjusting their strategies to play against the producers, not other speculators; so the first term in Eq. (36) dominates, and the system destabilizes with increasing s . If s is already large to begin with, the second term can dominate, and increasing s further can restabilize the system.

VI. VOLATILITY CLUSTERING

As a final aspect of our analysis, we briefly comment on volatility clustering in the SK game. Volatility is known to cluster in time in financial markets (see, e.g., Refs. [35,36]). This means that there are periods of time in which prices move

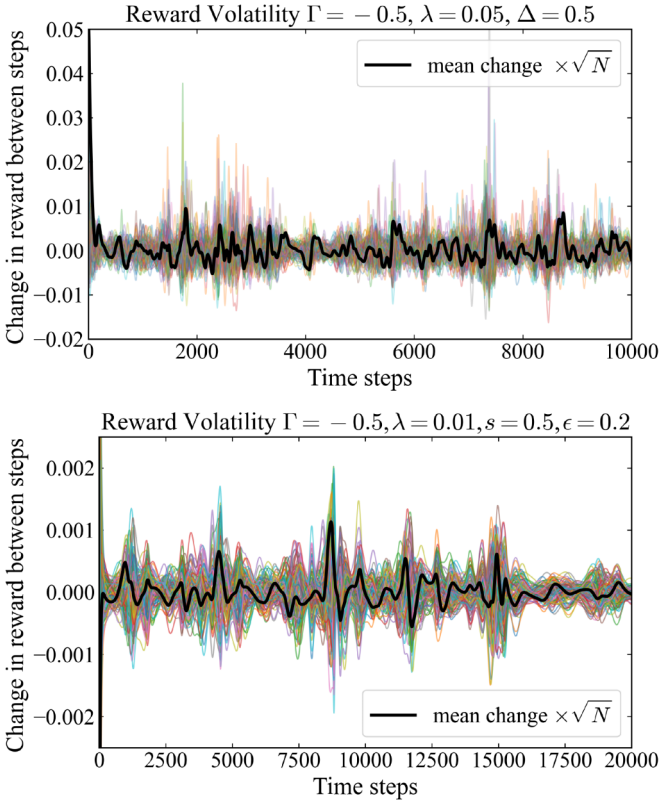


FIG. 13. Reward volatility, defined as the change in the rewards received by each player between consecutive time steps, is shown for two representative games. The upper panel shows the reward volatility for the SK game with a random field. This corresponds to the game shown in the lower panel of Fig. 3 simulated over an extended time period. The lower panel shows the reward volatility of the speculator in the grand-canonical SK game. This corresponds to the game shown in the lower panel of Fig. 7 again simulated over an extended time period. The black lines shows the mean change in reward, rescaled by $\sqrt{N}$ so that its fluctuations are comparable in variance with those of a typical agent.

relatively little and other times where prices move a lot. In the words of Mandelbrot, “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes” [35].

While the SK game is not a direct model of a market and while payoffs are not an equivalent of prices, we find signs of clustering in simulations of the SK game. To this end, we have measured the change of payoff to individual players i from one round to the next, i.e., $\Pi_i[\mathbf{x}(t+1)] - \Pi_i[\mathbf{x}(t)]$. We plot such time series in Fig. 13. The upper panel is for an instance of the SK game with random external fields in the unstable phase. The lower panel is for a grand-canonical SK game, also in the unstable phase. In the figure, one can clearly see periods of increases payoff movements interspersed with more quiet periods. We also plot a rescaled version of the mean reward across players $N^{-1/2} \sum_i \{\Pi_i[\mathbf{x}(t+1)] - \Pi_i[\mathbf{x}(t)]\}$. The prefactor is chosen such that this quantity has fluctuations of the same order as the representative individual $\Pi_i[\mathbf{x}(t+1)] - \Pi_i[\mathbf{x}(t)]$. Clustered activity can also be seen in this quantity.

A possible general mechanism for the emergence of volatility clustering in models with active and inactive states

was proposed in Ref. [37]. This is based on the switching of an agent between the active and inactive states being governed by a random walk for the underlying propensities. This random-walk behavior leads to power-law-distributed residence times. We have carried out some analysis along those lines for the simulations for the grand-canonical SK game in Fig. 13. We found no direct evidence for this proposed mechanism. One possible reason is that the players in our model hold mixed strategies which evolve smoothly in time. We also note that, for small β , there are no abrupt effects on the strategy of a player if q_i crosses zero.

VII. CONCLUSIONS AND DISCUSSION

To summarize, we have studied the outcome of adaptive learning in the SK game. Nash equilibria are the natural outcome for simple games; authors of previous work had shown that adaptive learning does not converge in many complex games with few players but large strategy spaces [12,13]. In the SK game, introduced in Ref. [16], a different route to adding complexity is taken. The number of pure strategies available to each player is small (two pure actions), but the number of interacting players is large. In this scenario, games also remain unlearnable, and players have to make do with suboptimal outcomes.

Our main contributions in this work are to (i) construct the SK game from a concrete payoff structure, (ii) to show that the situation in Ref. [16] is a special case in which no player has any intrinsic bias toward any of his or her two actions and to implement generalization to random external fields, (iii) to establish where in parameter space the three different dynamical regimes (unique fixed points, many fixed points, or persistently volatile motion) occur in the SK games with and without field, and (iv) we introduce a grand-canonical version of the SK game in which there are two types of players (speculators and producers) and where speculators have the option to abstain. We establish the stability behavior for this game and show that there are interesting re-entrance effects as the fraction of speculator in the population of players is varied.

We think our study is of interest in several different ways. It adds to the body of work on the statistical physics of adaptive agent learning in complex situations. This program goes back at least 25 years, when there was significant activity on the MG [20,28], and where tools from disordered systems were used to identify a transition between information efficient and inefficient phases. Researchers, partly predating the MG era, have addressed similar questions [5,6,12,13,15,38]—What is the typical outcome of complex evolutionary dynamics?—but focused situations with a large strategy space. In the present work, along with that in Ref. [16], we instead focus on scenarios in which many players learn to play Gaussian random games with only two actions per player. It is perhaps surprising that the general phenomenology is rather similar to that of the multi-action two-player games in Ref. [12].

This observation makes the work interesting in a more general way. One question that was raised in Refs. [12,13] is if there are general quantifiers which can indicate when particular learning dynamics would fail to converge for a particular ensemble of games. This question is motivated by

observations in fluid dynamics, where the Reynolds number can be used to make approximate predictions as to whether a particular fluid flow can be expected to be turbulent. It is now becoming increasingly clear that the degree of competition and the time scale of memory loss in the learning dynamics are key determinants for the stability of learning in a variety of different complex games. There is also increasing evidence that chaos is not an exception but rather the norm in different competitive games.

ACKNOWLEDGMENTS

T.G. thanks Jean-Philippe Bouchaud and Jérôme Garnier-Brun for discussions about their work [16]. The work is partially supported by the Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (FEDER, UE) under project COSASTI (No. PID2024-157493NB-C22), the María de Maeztu programme for Units of Excellence, No. CEX2021-001164-M, funded by No. MCIN/AEI/10.13039/501100011033. This work is also partially supported by a Ph.D. studentship from the Faculty of Natural, Mathematical & Engineering Sciences, King's College London.

DATA AVAILABILITY

Data and code for the figures are available online [39].

APPENDIX A: DERIVATION OF EQ. (6)

For small α and β , we have $x^{1-\alpha}e^{\beta\Pi} = x + x(-\alpha \ln x + \beta\Pi + \dots)$. We use this expansion in the numerator and denominator of the right-hand side of Eq. (4) and the fact that $x_i^+(t) + x_i^-(t) = 1$. We then arrive at

$$x_i^\pm(t+1) - x_i^\pm(t) \approx x_i^\pm(t) \{-\alpha \ln x_i^\pm(t) + \beta \Pi_i^\pm(\mathbf{x}(t)) - \kappa_i(\mathbf{x}(t))\}, \quad (\text{A1})$$

where $\kappa_i(\mathbf{x}) = -\alpha(x_i^+ \ln x_i^+ + x_i^- \ln x_i^-) + \beta \bar{\Pi}_i(\mathbf{x})$. Thus, after rescaling time and writing $\lambda = \alpha/\beta$, we obtain

$$\frac{d}{dt}x_i^\pm = x_i^\pm[-\lambda \ln x_i^\pm(t) + \Pi_i^\pm(\mathbf{x}) - v_i(\mathbf{x})], \quad (\text{A2})$$

with $v_i(\mathbf{x}) = -\lambda(x_i^+ \ln x_i^+ + x_i^- \ln x_i^-) + \bar{\Pi}_i(\mathbf{x})$.

APPENDIX B: BRIEF COMPARISON WITH THE MODEL IN REF. [16]

The learning dynamics in Ref. [16] can be summarized as follows (see eqs. (1), (3), and (5) in Ref. [16]):

$$Q_i(t+1) = (1-\alpha)Q_i(t) + \alpha \sum_j J_{ij}S_j(t), \quad (\text{B1})$$

where Q_i is the analog of a propensity, and S_j takes the values ± 1 with probabilities $\frac{1}{2}\{1 \pm \tanh[\beta Q_j(t)]\}$, respectively. The J_{ij} variables are quenched random couplings.

Taking the batch limit, we replace $S_j(t)$ with $m_j(t) = \langle S_j(t) \rangle$. Rescaling time (to assume that each learning step takes time α) and going to a continuous-time limit ($\alpha \ll 1$), we find

$$\dot{Q}_i = -Q_i + \sum_j J_{ij}S_j. \quad (\text{B2})$$

Using Eq. (2) in Ref. [16] and the chain rule, we then find

$$\dot{m}_i = \beta(1-m_i^2) \left(-Q_i + \sum_j J_{ij}m_j \right). \quad (\text{B3})$$

We next use $Q_i = \frac{1}{\beta} \text{artanh } m_i = \frac{1}{2\beta} \ln \frac{1+m_i}{1-m_i}$. We then have

$$\dot{m}_i = (1-m_i^2) \left(-\frac{1}{2} \ln \frac{1+m_i}{1-m_i} + \beta \sum_j J_{ij}m_j \right). \quad (\text{B4})$$

APPENDIX C: FURTHER DETAILS OF THE ANALYSIS OF THE SK GAME

In this Appendix, we provide details of the generating-functional analysis of the learning process of the SK game. This follows the lines of Refs. [12,15]. Further details of this type of calculation can also be found in Ref. [29]. Alternative approaches based on the cavity method are used (for related Lotka-Volterra systems) in Refs. [30,31].

1. Generating-functional calculation

The generating function associated with the SK game with a field from Eq. (14) is given by

$$\begin{aligned} Z(\psi) = & \int \mathcal{D}\mathbf{m} \mathcal{D}\hat{\mathbf{m}} \exp \left\{ i \sum_i \int dt \hat{m}_i(t) \left[\frac{\dot{m}_i(t)}{1-m_i(t)^2} + \lambda \ln \left(\frac{1+m_i}{1-m_i} \right) \right] \right\} \\ & \times \exp \left[i \sum_i \int dt \hat{m}_i(t) \left(\sum_j A_{ij}m_j(t) + B_{ij} \right) \right] \exp \left(i \sum_i \int dt m_i(t) \psi_i(t) \right). \end{aligned} \quad (\text{C1})$$

Taking the disorder average in Eq. (D1), we have

$$\begin{aligned} & \overline{\left\{ \exp \left[i \sum_i \int dt \hat{m}_i(t) \left(\sum_j A_{ij}m_j(t) + B_{ij} \right) \right] \right\}} \\ & = \exp \left(\frac{\Delta^2}{2} \sum_i \int dt dt' \hat{m}_i(t) \hat{m}_i(t') \right) \times \exp \left(\frac{\gamma \Delta^2}{2} \sum_{ij} \int dt dt' \hat{m}_i(t) \hat{m}_j(t') \right) \end{aligned}$$

$$\begin{aligned} & \times \exp \left(\frac{\sigma^2}{2} \sum_{ij} \int dt dt' \hat{m}_i(t) \hat{m}_i(t') \hat{m}_j(t) \hat{m}_j(t') \right) \times \exp \left(\frac{\Gamma \sigma^2}{2} \sum_{ij} \int dt dt' \hat{m}_i(t) \hat{m}_j(t') \hat{m}_i(t') \hat{m}_j(t) \right) \\ & = \exp \left(\frac{N}{2} \int dt dt' [\Delta^2 L(t, t') + \gamma \Delta^2 P(t) P(t') + \sigma^2 L(t, t') C(t, t') + \Gamma \sigma^2 K(t, t') K(t', t)] \right), \end{aligned} \quad (C2)$$

where

$$P(t) = \frac{1}{N} \sum_i \hat{m}_i(t), \quad C(t, t') = \frac{1}{N} \sum_i m_i(t) m_i(t'), \quad K(t, t') = \frac{1}{N} \sum_i m_i(t) \hat{m}_i(t'), \quad L(t, t') = \frac{1}{N} \sum_i \hat{m}_i(t) \hat{m}_i(t'). \quad (C3)$$

We can now rewrite the disordered averaged generating functional as follows:

$$\overline{Z(\Psi)} = \int \exp \{ N[\Psi + \Phi + \Omega + O(N^{-1})] \} \mathcal{D}[P, \hat{P}, C, \hat{C}, K, \hat{K}, L, \hat{L}], \quad (C4)$$

where

$$\begin{aligned} \Psi &= i \int dt \hat{P}(t) P(t) + i \int dt dt' (\hat{C}(t, t') C(t, t') + \hat{K}(t, t') K(t, t') + \hat{L}(t, t') L(t, t')), \\ \Phi &= -\frac{\sigma^2}{2} \int dt dt' [L(t, t') C(t, t') + \Gamma K(t, t') K(t', t)] - \frac{\Delta^2}{2} \int dt dt' [L(t, t') + \gamma P(t) P(t')], \\ \Omega &= N^{-1} \sum_i \ln \int D[m_i, \hat{m}_i] \left[\exp \left(i \int dt m_i \psi_i(t) \right) \exp \left\{ i \sum_i \int dt \hat{m}_i(t) \left[\frac{\dot{m}_i(t)}{1 - m_i(t)^2} + \lambda \ln \left(\frac{1 + m_i}{1 - m_i} \right) \right] \right\} \right. \\ &\quad \times \exp \left(-i \int dt dt' [\hat{C}(t, t') m_i(t) m_i(t') + \hat{L}(t, t') \hat{m}_i(t) \hat{m}_i(t') + \hat{K}(t, t') m_i(t) \hat{m}_i(t')] \right) \\ &\quad \left. \times \exp \left(-i \int dt \hat{P}(t) m(t) \right) \right]. \end{aligned} \quad (C5)$$

In the limit of $N \rightarrow \infty$, we can evaluate the integral in Eq. (C4) via the saddle-point method. Setting the variation with respect to the variables P, C, K, L to zero, we obtain

$$\begin{aligned} i\hat{P}(t) &= \gamma P(t), \quad i\hat{C}(t, t') = \frac{\sigma^2}{2} L(t, t'), \\ i\hat{K}(t, t') &= \Gamma \sigma^2 K(t', t), \quad i\hat{L}(t, t') = \frac{1}{2} [C(t, t') + \Delta^2]. \end{aligned} \quad (C6)$$

Extremizing with respect to $\hat{P}, \hat{C}, \hat{K}$, and $\hat{L}$, we find P and $L = 0$. This leads to the following effective process:

$$\begin{aligned} \dot{m} &= -\lambda(1 - m^2) \ln \frac{1 + m}{1 - m} + (1 - m^2) \\ &\quad \times \left[\Gamma \sigma^2 \int dt' G(t, t') m(t') + \eta(t) \right], \end{aligned} \quad (C7)$$

with the self-consistency relations

$$\begin{aligned} C(t, t') &\equiv \langle \eta(t) \eta(t') \rangle = \sigma^2 \langle m(t) m(t') \rangle + \Delta^2, \\ G(t, t') &= \frac{\delta}{\delta \eta(t')} \langle m(t) \rangle. \end{aligned} \quad (C8)$$

2. Linear stability analysis for SK game with random field

For clarity, we denote quantities evaluated at the fixed point with a superscript $*$ in this Appendix. We follow the procedure in Ref. [15]. Linearizing about the fixed point involves

introducing small perturbations about m^* and η^* ,

$$m \rightarrow m^* + \hat{m}, \quad \eta \rightarrow \eta^* + \hat{\eta}. \quad (C9)$$

We obtain the linearized equations about the fixed point

$$\begin{aligned} \frac{d}{dt} \hat{m}(t) &= -2\lambda \hat{m} + (1 - (m^*)^2) \\ &\quad \times \left(\Gamma \sigma^2 \int dt' G(t - t') \hat{m}(t') + \hat{\eta}(t) \right). \end{aligned} \quad (C10)$$

We next carry out a Fourier transform

$$\begin{aligned} \hat{m}(t) &\rightarrow \tilde{m}(\omega), \quad \frac{d}{dt} \hat{m}(t) \rightarrow i\omega \tilde{m}(\omega), \\ \hat{\eta}(t) &\rightarrow \tilde{\eta}(\omega), \quad \hat{G}(t - t') \rightarrow G(\omega). \end{aligned} \quad (C11)$$

We then obtain the following relation:

$$\frac{i\omega + 2\lambda - [1 - (m^*)^2] \Gamma \sigma^2 G(\omega)}{1 - (m^*)^2} \tilde{m}(\omega) = \tilde{\eta}(\omega). \quad (C12)$$

From this, we find

$$\begin{aligned} \langle |\tilde{m}(\omega)|^2 \rangle &= \left\langle |\tilde{\eta}(\omega)|^2 \left| \frac{[1 - (m^*)^2]}{i\omega + 2\lambda - [1 - (m^*)^2] \Gamma \sigma^2 G(\omega)} \right|^2 \right\rangle \\ &= \frac{1}{\sigma^2} \langle |\tilde{\eta}(\omega)|^2 \rangle, \end{aligned} \quad (C13)$$

where the last identity is obtained from self-consistency relation $\langle \hat{\eta}(t) \hat{\eta}(t') \rangle = \sigma^2 \langle \hat{m}(t) \hat{m}(t') \rangle$. We note the random field (parametrized by Δ) does not contribute to the linearized

dynamics about the fixed point (but the random field does affect the fixed point itself).

As with similar problems [12,15], we assume the low frequencies $\omega \rightarrow 0$ are the most unstable. In the zero-frequency limit $\omega \rightarrow 0$, perturbing m^* corresponds to a static displacement of the fixed point. Following the lines of Refs. [12,15], we can then conclude that the fixed point is stable when the following relation is fulfilled:

$$\sigma^2 \left\langle \left[\frac{1 - (m^*)^2}{2\lambda - (1 - (m^*)^2)\Gamma\sigma^2\chi} \right]^2 \right\rangle = \sigma^2 \left\langle \left(\frac{\partial m^*}{\partial \eta^*} \right)^2 \right\rangle < 1. \quad (\text{C14})$$

3. Case without random field

When there is no external field ($\Delta = 0$), the fixed-point distribution becomes concentrated at $m^* = 0$. Substituting this into Eq. (C14), the stability condition becomes

$$\sigma \left[\frac{1}{2\lambda - \Gamma\sigma^2\chi} \right] < 1, \quad (\text{C15})$$

which simplifies to

$$1 < 2\lambda/\sigma - \Gamma\sigma\chi. \quad (\text{C16})$$

We will now show $\chi = 1/\sigma$ at the stability boundary. We note the following relation, valid at the fixed point:

$$-\lambda \ln \frac{1 + m^*}{1 - m^*} + \Gamma\sigma^2\chi m^* + \eta^* = 0. \quad (\text{C17})$$

Differentiating with respect to η^* , we find

$$\left(\frac{-2\lambda}{1 - (m^*)^2} + \Gamma\sigma^2\chi \right) \frac{\partial m^*}{\partial \eta^*} + 1 = 0. \quad (\text{C18})$$

From this, we obtain

$$\chi = \frac{\partial m^*}{\partial \eta^*} = \frac{1 - (m^*)^2}{2\lambda - [1 - (m^*)^2]\Gamma\sigma^2\chi}, \quad (\text{C19})$$

where this is to be evaluated at the fixed point $m^* = 0$. This means that $\chi = (2\lambda - \Gamma\sigma^2\chi)^{-1}$. We use this in Eq. (C16) and find $\chi = 1/\sigma$ at the stability boundary. Substituting this into the relation $\chi = (2\lambda - \Gamma\sigma^2\chi)^{-1}$, which we derived earlier, we find the following condition for the stability boundary:

$$\lambda_{\text{crit}} = (1 + \Gamma)\sigma/2. \quad (\text{C20})$$

Instability occurs for $\lambda < \lambda_{\text{crit}}$.

4. Approach based on eigenvalues of random matrices

One can also obtain the stability boundary for the SK game without random field using an approach based on the spectra of random matrices. This circumvents the dynamical mean-field theory analysis.

The time evolution of $\{m_i\}$ is given by

$$\frac{d}{dt}m_i = (1 - m_i^2) \left(\sum_j A_{ij}m_j - \lambda \ln \frac{1 + m_i}{1 - m_i} \right). \quad (\text{C21})$$

We now introduce the following variables:

$$y_i = \ln \frac{1 + m_i}{1 - m_i}. \quad (\text{C22})$$

While the m_i values are restricted to the interval $(-1, 1)$, y_i can take values on the entire real axis. This simplifies the analysis. The dynamics for the y_i are given by

$$\frac{d}{dt}y_i = 2 \left(\sum_j A_{ij}m_j - \lambda y_i \right). \quad (\text{C23})$$

Expressing m_j on the right in terms of y_j , this can be written in the form

$$\frac{d}{dt}\mathbf{y} = \mathbf{F}(\mathbf{y}). \quad (\text{C24})$$

Suppose we are now at a fixed point of the dynamics, written as $m_i = m_i^*$, i.e., $y_i = y_i^*$. A perturbation of the vector $\mathbf{y}$ will then evolve as

$$\frac{d}{dt}\delta\mathbf{y} = \underline{\underline{J}}\delta\mathbf{y}, \quad (\text{C25})$$

with a Jacobian matrix $\underline{\underline{J}}$ given by

$$J_{ij} = \left. \frac{\partial F_i}{\partial y_j} \right|_{\mathbf{y}=\mathbf{y}^*} = \begin{cases} -2\lambda, & \text{if } i = j, \\ A_{ij}[1 - (m_i^*)^2], & \text{if } i \neq j, \end{cases} \quad (\text{C26})$$

In the absence of a random field (i.e., for $\Delta = 0$), the system reaches the paramagnetic fixed point $m_i^* = 0$, for all i . We then have

$$\underline{\underline{J}} = -2\lambda \underline{\underline{I}} + \underline{\underline{A}}, \quad (\text{C27})$$

where $\underline{\underline{I}}$ is the identity matrix of size $N \times N$.

We can then determine stability of the paramagnetic fixed point from the eigenvalues of $\underline{\underline{J}}$. The matrix $\underline{\underline{A}}$ is a Gaussian matrix; from Girko's elliptic law [40], we know the eigenvalues (written as μ) of $\underline{\underline{A}}$ (in the $N \rightarrow \infty$ limit) are distributed in an ellipse given by

$$\frac{(\text{Re}\mu)^2}{(1 + \Gamma)^2} + \frac{(\text{Im}\mu)^2}{(1 - \Gamma)^2} \leq \sigma^2. \quad (\text{C28})$$

Examples of such spectra are shown in Fig. 14. The eigenvalue of $\underline{\underline{J}}$ with the largest real part is therefore

$$\mu_{\mathbf{J},\text{max}} = \frac{\sigma}{1 + \Gamma} - 2\lambda. \quad (\text{C29})$$

The fixed point is stable when this quantity is negative and unstable if it is positive. This leads to the instability condition

$$\lambda_{\text{crit}} = (1 + \Gamma)\sigma/2. \quad (\text{C30})$$

At the onset of instability, the eigenvalue that crosses the imaginary axis has an imaginary part equal to zero. This corresponds to the $\omega \rightarrow 0$ mode being the first to destabilize.

Introducing a random field Δ means that the m_i^* values are no longer all zero at the fixed point in the stable phase. The m_i^* values are determined by the matrix $\underline{\underline{A}}$. It is then no longer straightforward to use existing results on the spectra of random matrices to determine the stability boundary. In Fig. 15 few show the eigenvalue spectra for examples of games with a random field. However, we note that $m^* \neq 0$ weakens the magnitude of the entries in the Jacobian J_{ij} , $i \neq j$ since they are proportional to $1 - (m^*)^2$. This means a lower λ would be required to stabilize the system. This gives intuition of how introducing Δ stabilizes the system.

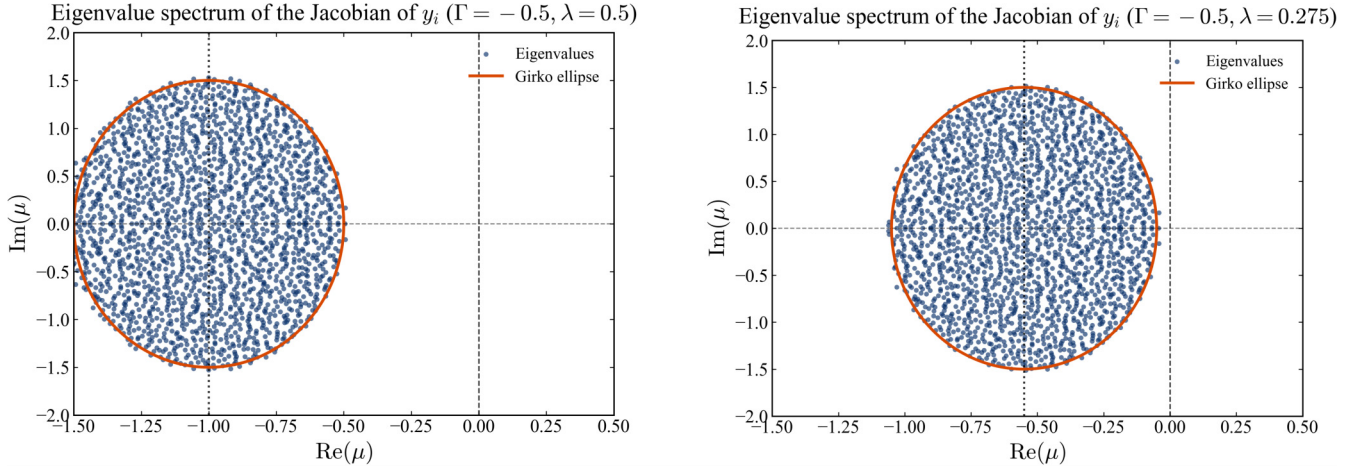


FIG. 14. Plots of the eigenvalues of Jacobian corresponding to y_i at the fixed point for two identical SK games for different values of λ . The parameters of the game are given by $\Gamma = -0.5$, $N = 2000$, $\beta = 0.001$, $\lambda = 0.5, 0.275$, with $\lambda_{\text{crit}} = 0.25$. Games are run up until convergence to the fixed point. At the fixed point, we perturb each y_i to compute the corresponding Jacobian. We plot the eigenvalues of the Jacobian (in blue) against the predicted boundary from random matrix theory (red). These eigenvalues obtained via simulations have been rescaled by a factor of 2, corresponding to the rescaling of time used in Eq. (11). This is necessary because the theoretical calculations are performed in rescaled time. We note the real part of the (rescaled) eigenvalues are centered at -2λ .

APPENDIX D: GENERATING-FUNCTIONAL CALCULATION FOR THE GRAND-CANONICAL SK GAME

1. Dynamic mean-field theory

The generating functional associated with the grand-canonical SK game is given by

$$Z(\psi) = \int \mathcal{D}\mathbf{x} \mathcal{D}\hat{\mathbf{x}} \exp \left\{ i \sum_i \int dt \hat{x}_i(t) \left[\frac{\dot{x}_i(t)}{x_i(t)(1-x_i(t))} + \lambda \ln \left(\frac{x_i}{1-x_i} \right) \right] \right\} \\ \times \exp \left[i \sum_i \int dt \hat{x}_i(t) \left(\sum_j^{N_s} A_{ij} x_j(t) + h_i \right) \right] \times \exp \left(i \sum_i \int dt x_i(t) \psi_i(t) \right). \quad (\text{D1})$$

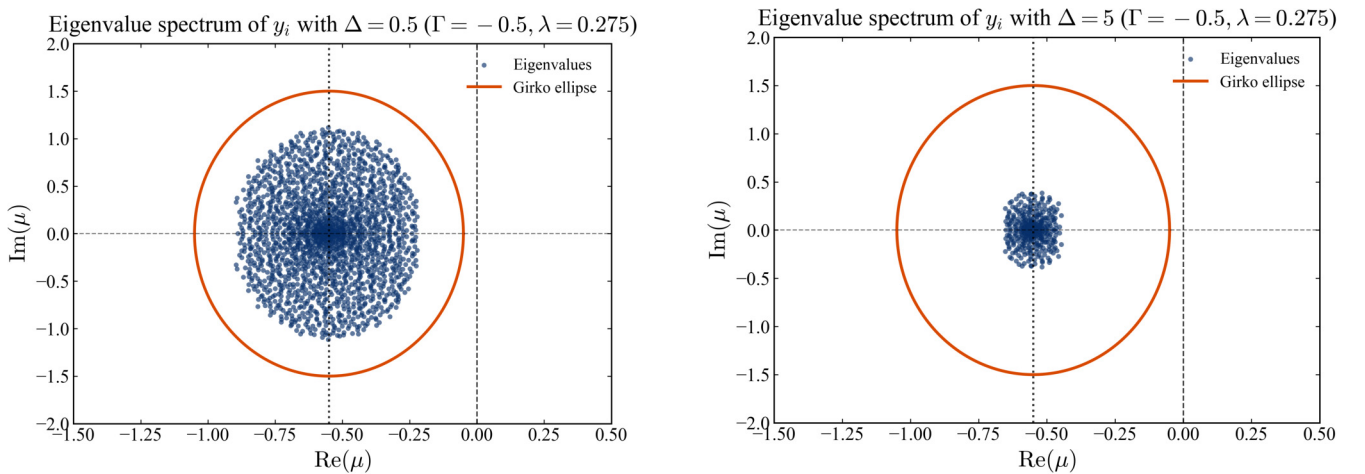


FIG. 15. Plots of the eigenvalues of Jacobian corresponding to y_i for two selected cases with a random field $\Delta = 0.5, 5$. The parameters here are similar to the right plot in Fig. 14. We note that, increasing the strength, the random field decreases the spread of the eigenvalue bulk. Since instability occurs when the largest eigenvalue crosses the real axis, increasing the random field requires a lower λ .

Averaging over the disorder, we find

$$\exp \left[i \sum_i \int dt \hat{x}_i(t) \left(\sum_j^{Ns} A_{ij} x_j(t) + h_i \right) \right] = \exp \left(\frac{Ns\sigma^2}{2} \int dt dt' [(1-s)L(t, t') + sL(t, t')C(t, t') + s\Gamma K(t, t')K(t', t)] \right), \quad (D2)$$

where

$$C(t, t') = \frac{1}{Ns} \sum_i x_i(t)x_i(t'), \quad K(t, t') = \frac{1}{Ns} \sum_i x_i(t)\hat{x}_i(t'), \quad L(t, t') = \frac{1}{Ns} \sum_i \hat{x}_i(t)\hat{x}_i(t'). \quad (D3)$$

At the saddle point, we send $Ns \rightarrow \infty$. Setting the variation with respect to the variables C , K , and L to zero, we find

$$i\hat{C}(t, t') = \frac{s\sigma^2}{2}L(t, t'), \quad i\hat{K}(t, t') = s\Gamma\sigma^2K(t, t'), \quad i\hat{L}(t, t') = \frac{\sigma^2}{2}[sC(t, t') + (1-s)]. \quad (D4)$$

This gives the following effective process:

$$\dot{x} = x(1-x) \left[-\lambda \ln \frac{x}{1-x} + s\Gamma\sigma^2 \int dt' G(t, t')x(t') - \varepsilon + \eta(t) \right], \quad (D5)$$

with

$$\langle \eta(t)\eta(t') \rangle = \sigma^2[s\langle x(t)x(t') \rangle + (1-s)]. \quad (D6)$$

2. Linear stability analysis for the grand-canonical game

We again carry out a linear stability analysis to obtain the stability criterion for the grand-canonical SK game. We find that fixed points are stable when

$$\sigma^2 s \left\langle \left[\frac{x^*(1-x^*)}{\lambda - x^*(1-x^*)s\Gamma\sigma^2\chi} \right]^2 \right\rangle = \sigma^2 s \left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle < 1. \quad (D7)$$

As shown in Fig. 10, when $\Gamma < 0$, increasing the fraction of speculators can destabilize the system when s is small, while for large s , increasing s further makes the dynamics more stable. In the following, we give some further mathematical reasoning as to why this occurs.

a. Increasing the fraction of speculators destabilizes learning when s is small

We introduce the following quantity [see Eq. (D7)]:

$$F(s) = \sigma^2 s \left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle. \quad (D8)$$

As $F(s)$ increases, the system becomes more unstable, with the phase transition occurring when $F(s) = 1$. Differentiating with respect to s , we find

$$\frac{\partial F(s)}{\partial s} = \sigma^2 \left[\left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle + s \frac{\partial}{\partial s} \left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle \right]. \quad (D9)$$

The second term is proportional to s and can be neglected for sufficiently small s . We then have

$$\left. \frac{\partial F(s)}{\partial s} \right|_{s \rightarrow 0} \approx \sigma^2 \left[\left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle \right] > 0. \quad (D10)$$

Therefore, for small s , increasing s destabilizes the system.

b. Adding speculators can stabilize when s is already large

Now we look at the second term in Eq. (D9). We have

$$\frac{\partial}{\partial s} \left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle = \frac{\partial}{\partial s} \left\langle \left[\frac{1}{\lambda/(x^*(1-x^*)) - s\Gamma\sigma^2\chi} \right]^2 \right\rangle. \quad (D11)$$

We focus on $\Gamma < 0$. We have $\lambda/x^*(1-x^*) > 0$, and $\Gamma\chi < 0$ (the susceptibility χ comes out positive). We will show that the denominator in Eq. (D11) increases as $s \rightarrow 1$. Naively, we might assume it is driven by the $-s\Gamma\sigma^2\chi$ term increasing with s . Although that might be true, it is driven by the other term $\lambda/x^*(1-x^*)$ blowing up. We will show this in the next segment.

From Fig. 12, we note $\langle x^* \rangle$ decreases when s and ϵ increase. We can get intuition on how this occurs by looking at the fixed-point equation

$$-\lambda \ln \left(\frac{x^*(z)}{1-x^*(z)} \right) + s\Gamma\sigma^2\chi x^*(z) - \varepsilon + \eta^*(z) = 0. \quad (D12)$$

When s increases, to maintain the equality, $x^*(z)$ has to decrease; likewise, when ε increases. Thus, for some sufficiently large epsilon $\epsilon > \epsilon_0$, as we increase $s \rightarrow 1$, we effectively have $\langle x^* \rangle \rightarrow 0$, which causes the denominator in Eq. (D11) to blow up. Therefore, $\frac{\partial}{\partial s} \left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle$ is negative for large s , and as s increases, we are placing more weight on the strictly negative term in Eq. (D9),

$$\begin{aligned} \frac{\partial F(s)}{\partial s} &= \sigma^2 \left[\underbrace{\left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle}_{>0(\text{dominant at small } s)} + \underbrace{s \frac{\partial}{\partial s} \left\langle \left(\frac{\partial x^*}{\partial \eta^*} \right)^2 \right\rangle}_{<0(\text{can become dominant at larger } s)} \right]. \end{aligned} \quad (D13)$$

Thus, increasing s can decrease $F(s)$ and therefore stabilize the system.

- [1] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, 1944).
- [2] C. Hauert and G. Szabó, Game theory and physics, *Am. J. Phys.* **73**, 405 (2005).
- [3] A. Traulsen and C. Hauert, Stochastic evolutionary game dynamics, *Rev. Nonlinear Dyn. Complexity* **2**, 25 (2009).
- [4] J. F. Nash, Equilibrium points in n -person games, *Proc. Natl. Acad. Sci. USA* **36**, 48 (1950).
- [5] J. Berg and A. Engel, Matrix games, mixed strategies, and statistical mechanics, *Phys. Rev. Lett.* **81**, 4999 (1998).
- [6] J. Berg and M. Weigt, Entropy and typical properties of Nash equilibria in two-player games, *Europhys. Lett.* **48**, 129 (1999).
- [7] A. McLennan and J. Berg, Asymptotic expected number of Nash equilibria of two-player normal form games, *Games Econ. Behav.* **51**, 264 (2005).
- [8] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, The complexity of computing a Nash equilibrium, *Commun. ACM* **52**, 89 (2009).
- [9] Y. Sato, E. Akiyama, and J. D. Farmer, Chaos in learning a simple two-person game, *Proc. Natl. Acad. Sci. USA* **99**, 4748 (2002).
- [10] Y. Sato and J. P. Crutchfield, Coupled replicator equations for the dynamics of learning in multiagent systems, *Phys. Rev. E* **67**, 015206(R) (2003).
- [11] Y. Sato, E. Akiyama, and J. P. Crutchfield, Stability and diversity in collective adaptation, *Physica D* **210**, 21 (2005).
- [12] T. Galla and J. D. Farmer, Complex dynamics in learning complicated games, *Proc. Natl. Acad. Sci. USA* **110**, 1232 (2013).
- [13] J. B. Sanders, J. D. Farmer, and T. Galla, The prevalence of chaotic dynamics in games with many players, *Sci. Rep.* **8**, 4902 (2018).
- [14] M. Pangallo, T. Heinrich, and J. Doyne Farmer, Best reply structure and equilibrium convergence in generic games, *Sci. Adv.* **5**, eaat1328 (2019).
- [15] M. Oppen and S. Diederich, Phase transition and $1/f$ noise in a game dynamical model, *Phys. Rev. Lett.* **69**, 1616 (1992).
- [16] J. Garnier-Brun, M. Benzaquen, and J.-P. Bouchaud, Unlearnable games and “satisficing” decisions: A simple model for a complex world, *Phys. Rev. X* **14**, 021039 (2024).
- [17] D. Sherrington and S. Kirkpatrick, Solvable model of a spin-glass, *Phys. Rev. Lett.* **35**, 1792 (1975).
- [18] P. C. Menezes and D. Sherrington, Minority game with SK interactions, *J. Phys. A: Math. Theor.* **46**, 505004 (2013).
- [19] R. M. May, Will a large complex system be stable? *Nature (London)* **238**, 413 (1972).
- [20] D. Challet, A. De Martino, M. Marsili, and I. Perez Castillo, Minority games with finite score memory, *J. Stat. Mech.* (2006) P03004.
- [21] C. Camerer and T.-H. Hua, Experience-weighted attraction learning in normal form games, *Econometrica* **67**, 827 (1999).
- [22] R. A. Rescorla and A. R. Wagner, A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement, in *Classical Conditioning II: Current Research and Theory*, edited by A. H. Black and W. F. Prokasy (Appleton-Century-Crofts, New York, 1972), pp. 64–99.
- [23] R. S. Sutton, A. G. Barto, and A. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, 1998), Vol. 1.
- [24] M. Pessiglione, B. Seymour, G. Flandin, R. J. Dolan, and C. D. Frith, Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans, *Nature (London)* **442**, 1042 (2006).
- [25] O. Dan, O. Plonsky, and Y. Loewenstein, Behavior engineering using quantitative reinforcement learning models, *Nat. Commun.* **16**, 4109 (2025).
- [26] A. Traulsen, T. Röhl, and H. G. Schuster, Stochastic gain in population dynamics, *Phys. Rev. Lett.* **93**, 028701 (2004).
- [27] C. De Dominicis, Dynamics as a substitute for replicas in systems with quenched random impurities, *Phys. Rev. B* **18**, 4913 (1978).
- [28] A. C. C. Coolen, *The Mathematical Theory of Minority Games: Statistical Mechanics of Interacting Agents* (Oxford University Press, Oxford, 2005).
- [29] T. Galla, Generating-functional analysis of random Lotka-Volterra systems: A step-by-step guide, [arXiv:2405.14289](https://arxiv.org/abs/2405.14289).
- [30] G. Bunin, Ecological communities with Lotka-Volterra dynamics, *Phys. Rev. E* **95**, 042414 (2017).
- [31] F. Roy, G. Biroli, G. Bunin, and C. Cammarota, Numerical implementation of dynamical mean field theory for disordered systems: Application to the Lotka-Volterra model of ecosystems, *J. Phys. A: Math. Theor.* **52**, 484001 (2019).
- [32] There can be multiple solutions of Eq. (19) in the interval $(-1, 1)$ in certain circumstances. This arises when $\Gamma > 0$ and $\lambda < 2\Gamma\sigma^2\chi$. Although we cannot formally demonstrate that this never occurs for the parameter ranges considered in this work, we have not encountered any multivaluedness. We therefore hypothesize that, for any combination of parameters in the stable fixed-point regime, there is a unique solution $m(z)$ of Eq. (19).
- [33] L. Sidhom and T. Galla, Ecological communities from random generalized Lotka-Volterra dynamics with nonlinear feedback, *Phys. Rev. E* **101**, 032101 (2020).
- [34] D. Challet, M. Marsili, and Y.-C. Zhang, Modeling market mechanism with Minority Game, *Physica A* **276**, 284 (2000).
- [35] B. Mandelbrot, The variation of certain speculative prices *J BUS* **36**, 394 (1963).
- [36] R. N. Mantegna and H. E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 1999).
- [37] J.-P. Bouchaud, I. Giardina, and M. Mezard, On a universal mechanism for long ranged volatility correlations, *Quantitative Finance* **1**, 212 (2001).
- [38] H. Rieger, Solvable model of a complex ecosystem with randomly interacting species, *J. Phys. A: Math. Gen.* **22**, 3447 (1989).
- [39] Data and codes: https://github.com/Desmondccw/complex_dynamics_in_sk_game.
- [40] S. O’Rourke, D. Renfrew, A. Soshnikov, and V. Vu, Products of independent elliptic random matrices, *J. Stat. Phys.* **160**, 89 (2015).