Unlearnable Games and "Satisficing" Decisions: A Simple Model for a Complex World

Jérôme Garnier-Brun [1,2,*] Michael Benzaquen [1,2,3] and Jean-Philippe Bouchaud [1,3,4]

[1]*Chair of Econophysics and Complex Systems, École polytechnique, 91128 Palaiseau Cedex, France*
[2]*LadHyX, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France*
[3]*Capital Fund Management, 23 Rue de l'Université, 75007 Paris, France*
[4]*Académie des Sciences, 23 Quai de Conti, 75006 Paris, France*

As a schematic model of the complexity economic agents are confronted with, we introduce the "Sherrington-Kirkpatrick game," a discrete time binary choice model inspired from mean-field spin glasses. We show that, even in a completely static environment, agents are unable to learn collectively optimal strategies. This is either because the learning process gets trapped in a suboptimal fixed point or because learning never converges and leads to a never-ending evolution of agent intentions. Contrarily to the hope that learning might save the standard "rational expectation" framework in economics, we argue that complex situations are generically *unlearnable* and agents must do with *satisficing* solutions, as argued long ago by Simon [Q. J. Econ. **69**, 99 (1955)]. Only a centralized, omniscient agent endowed with enormous computing power could qualify to determine the optimal strategy of all agents. Using a mix of analytical arguments and numerical simulations, we find that (i) long memory of past rewards is beneficial to learning, whereas overreaction to recent past is detrimental and leads to cycles or chaos; (ii) increased competition (nonreciprocity) destabilizes fixed points and leads first to chaos and, in the high competition limit, to quasicycles; (iii) some amount of randomness in the learning process, perhaps paradoxically, allows the system to reach better collective decisions; (iv) nonstationary, "aging" behavior spontaneously emerges in a large swath of parameter space of our complex but static world. On the positive side, we find that the learning process allows cooperative systems to coordinate around satisficing solutions with rather high (but markedly suboptimal) average reward. However, hypersensitivity to the game parameters makes it impossible to predict *ex ante* who will be better or worse off in our stylized economy. The statistical description of the space of satisficing solutions is an open problem.

I. INTRODUCTION

A. Rationality, complexity, and learning

Classical economics is based on the idea that rational agents make optimal decisions, i.e., optimize their expected utility over future states of the world, weighted by their objective probabilities. Such an idealization of human behavior has been criticized by many (see, e.g., Refs. [1–6]). In particular, assuming that all agents are rational, allowing one to use game-theoretic arguments to build such optimal strategies—often the result of complicated mathematical calculations—is implausible, to say the least.

A way to possibly save the rational expectation paradigm is to posit that agents are able to learn best responses from past experience. Yes, agents are only "boundedly" rational, but they learn and, in the long run, they act "as if" they were rational [7]. This is clearly expressed by Evans and Honkapohja in their review paper on the subject [8]. They note that *[i]n standard macroeconomic models rational expectations can emerge in the long run, provided the agents' environment remains stationary for a sufficiently long period.*

While seemingly reasonable, this proposition is by no means guaranteed to be legitimate. Indeed, the hypothesis that the environment should be stationary over "sufficiently long periods" can be restated in terms of the speed of convergence of the learning process, that should be short enough compared to the correlation timescale of the environment. However, in many circumstances and, in particular, in complex games, the convergence of the learning process to a collectively optimal state can be exceedingly long or may, in fact, *never* take place.

*jerome.garnier-brun@polytechnique.edu

For example, reasonable learning rules can trap the system in some suboptimal regions of the (high-dimensional) solution space; see, e.g., Refs. [9–12]. In other words, the learning process itself can be nonergodic, even if the environment is described by an ergodic, stationary process. Another possibility is that agents' strategies, even probabilistic, evolve chaotically forever, as was found by Galla and J. Farmer [13] in the context of competitive multichoice two-player games or by one of us (J.-P. B.) with R. Farmer in a simple binary choice, multiplayer game [14]. In such cases, the probabilities governing the different possible choices are not fixed but must themselves be described by probabilities.

This is, in fact, a generic feature of "complex systems." As proposed by Parisi [15,16], the description of such systems requires the introduction of *probabilities of probabilities*, as their statistical behaviors themselves (and not only individual trajectories) are highly sensitive to the small changes in parameters, initial conditions, or time. The inability to describe such systems with knowable probabilities was coined "radical complexity" in Ref. [17].

The sensitivity of optimal solutions to the parameters of the problem, or to the algorithm used to find them, has a very real consequence: One can no longer assume that all agents, even fully rational, will make the same decision, since any small perturbation may lead to a completely different solution, although similar in performance. In other words, the "common knowledge" assumption is not warranted [18]. This has already been underlined, for example, in the context of portfolio optimization in Refs. [19,20], or in the context of networked economies [12,21], but is expected to be of much more general scope, as anticipated by Keynes long ago and emphasized by many heterodox economists in the more recent past [1,3,5,6].

Here, we dwell on this issue in the context of a multiplayer binary game—the "Sherrington-Kirkpatrick (SK) game"—understood as an idealization of the economic world where agents strongly interact in such a way that their payoffs depend nontrivially on the action of others. In our setting, some relationships are mutually beneficial, while others are competitive. Agents have to learn how to coordinate to optimize their expected gains, which they do in a standard reinforcement way by observing the payoff of their past actions and adapting their strategies accordingly.

We find that our stylized model gives rise to a very wide range of dynamical behavior: suboptimal fixed points (also known as Nash equilibria), but also limit cycles when learning is too fast, and "chaos," meaning that individual mixed strategies never converge. Chaos can be deterministic, when learning is noiseless, or stochastic when random noise is introduced in the learning process. In such chaotic situations, the environment of each agent changes over time, not because of exogenous shocks but because of the endogenous learning dynamics of other agents, as, in fact, anticipated in Ref. [22]: Complex economies are always

out of equilibrium [23]. This was also argued recently in Ref. [12]: A purely static network of interactions can generate an apparent never-ending evolution of the world in which agents live, although it may appear stationary for very long periods of time. In such a case, one speaks of "aging," which we address in a dedicated section below.

Even when learning is efficient and fixed points are reached, the average payoff of agents—although better than random—is noticeably lower than the maximum possible value that could be obtained if a perfectly informed social planner with colossal computing power was dictating their strategies to each agent (and provided they accepted to follow its advice).

Hence, our model provides an explicit example of Simon's *satisficing* principle [1]: With reasonable, boundedly rational methods, agents facing complex problems can reach only some satisfying and sufficing but suboptimal solutions. In fact, in the absence of the omniscient and omnipotent social planner alluded to above, agents can hardly do better—the world in which they live in is *de facto unlearnable*. This is the main message of our paper.

A further interesting twist of the model is that the fixed point reached by the learning process depends sensitively on the initial condition and/or the detailed structure of the interaction network, an effect called "parameter chaos" in the context of disordered systems. As a consequence, which agents will do better than average and which will do worse is totally unpredictable, even when the economic interactions between agents are fixed and known. Correspondingly, small changes in the interaction network (as a result of—say—small exogenous shocks or as some agents die and other are born [14]) can lead to very large changes in individual strategies that agents would have to relearn from scratch.

This feature actually corresponds to another possible definition of a "complex" system [16]: Small perturbations can lead to disproportionally large reconfigurations of the system—think of sand (or snow) avalanches as an archetype of this type of behavior: A single grain of sand might leave the whole slope intact or, occasionally, trigger a landslide [24] (see also, e.g., Refs. [3,5,21,22,25–32] for early and more recent discussions of complexity ideas in the context of economics). We believe that all the above features, which are present in our model, are, in fact, common to many other specifications, such as those proposed in Refs. [12,13,33], and are expected to be universally present as soon as the situation faced by learning agents is "sufficiently complex."

B. Related ideas

Our contribution follows the footsteps of several important contributions on the subject of learning in economics [8], in particular, the complex two-player game of Galla and J. Farmer [13] and its multiplayer generalization in [33], in which the seminal idea of *unlearnable*

games is introduced. Our work goes beyond these papers in several respects. First, our framework allows one to give a precise meaning to the idea of *satisficing solutions*. To the best of our knowledge, the SK game is the first model in which this longstanding concept is concretely illustrated, as it *spontaneously* emerges from the dynamics of boundedly rational agents. Second, we emphasize the conceptual importance of parameter chaos and the fragility of attainable fixed points. Third, we insist on the corresponding generic nonstationary, aging nature of the dynamics even when the world in which agents live is completely static. None of these important aspects have been covered in Refs. [13,33]. Finally, our framework allows us to leverage many exact results that have been obtained for the SK model, such as the value of the optimal average rewards or the full distribution of individual rewards, to which the result of learning can be compared. Such comparisons are difficult to make within the context of Galla and J. Farmer and are, in fact, essential in our opinion, as they allow us to expand the notion of unlearnability beyond purely dynamical considerations. Our game is unlearnable not only when its dynamics is chaotic, but also when it reaches some fixed points, as those are *demonstrably* suboptimal.

Hence, we believe that our SK game is conceptually and technically simpler than the random payoff games introduced by Galla and J. Farmer. This should ease the dissemination of the concepts of unlearnable games and satisficing solutions, which we believe to be of great relevance to describe realistic situations.

Another relevant work is the study of Grandmont [2] on the stability of economic equilibrium under different learning rules. Although his general conclusions are somewhat similar to ours, there are important differences, due to the fact that complexity, in our model, arises from the strong interaction between agents that are absent in Grandmont's framework. The main differences are the following.

(i) While our agents can in some cases (i.e., when interactions are reciprocal) converge to a locally stable equilibrium, this equilibrium is rather inefficient compared to the best theoretical one, which is unattainable without the help of a central planner with superpowers. Furthermore, instead of a single equilibrium in the Grandmont case, the equilibrium reached by our learning agents is one among an exponential number of possible equilibria that are each sensitively dependent on the parameters of the model. In other words, which equilibrium is reached by the agents is totally unpredictable.

(ii) As soon as some small amount of noise is present, Grandmont-style instabilities set in. In our case, however, there is no change in stability *per se*; the system is simply exploring the set of possible equilibria following random fluctuations, albeit in a slow, nonergodic way: The effective stability of the visited equilibria increases with the age of the system.

(iii) In the case where interactions are sufficiently non-reciprocal, there are no fixed points to the learning dynamics of the SK game, except the trivial one where agents play randomly at each turn (like "rock paper scissors"). However, this is *not* what agents agree to do. They keep holding strong beliefs at each time step, such beliefs evolving chaotically in time—but now in an ergodic way [34]. Whereas there is now a unique "equilibrium" as in Grandmont's model, it is not reached due to purely dynamical effects, unrelated to the (in)stability of the fixed point.

Somewhat related ideas can be found in the recent paper of Hirano and Stiglitz [37], where there can be a plethora of equilibrium trajectories, converging neither to a steady state or even to a limit cycle, what they call "wobbly" macrodynamics. We should finally mention the work of Dosi *et al.* [38], which showcases that even sophisticated learning rules can fail at improving the individual and collective outcomes when interacting agents are heterogeneous. Although the context is quite different, such a finding appears to be consistent with the saturation of the average reward at suboptimal satisficing levels that we observe when the memory loss rate of our learning agents vanishes.

C. Outline of the paper

The layout of the paper is as follows. In Sec. II, we introduce our spin-glass-inspired game. We then review the model's main features and present what we believe are the most important conceptual results in the socio-economic context in Sec. III. In Sec. IV, we enter the technical analysis of the SK game by discussing the existence and abundance of fixed-point solutions. A similar analysis is conducted for short limit cycles when there are no fluctuations in the system in Sec. V. Having determined when these solutions may or may not exist, we take interest in the $N \to \infty$ dynamics in Sec. VI by writing the dynamical mean-field theory of the problem. Combining the static and dynamic pictures, we focus on the fully rational (or "zero-temperature") limit and, in particular, on the role of learning on the collective dynamics in Sec. VII. Section VIII finally considers the effect of fluctuations stemming from the bounded rationality of agents in their decision-making process. In Sec. IX, we summarize our findings and discuss future perspectives as well as the relevance of our model for other complex systems such as biological neural networks.

II. A SIMPLE MODEL FOR A COMPLEX WORLD

A. Setup of the model

As a minimal, stylized model for decision making in a complex environment of interacting agents, we restrict ourselves to binary decisions, as in many papers on models

with social interactions; see, e.g., Refs. [14,39–42]. At every time step t, each agent i plays $S_i(t) = \pm 1$, with $i = 1, \ldots, N$, which can be thought of, for example, as the decision of an investor to buy or to sell the stock market or the decision of a firm to increase or to decrease production, etc. The difference of incentive to play $S_i(t) = +1$ rather than $S_i(t) = -1$ is denoted as $Q_i(t)$ and is defined as the agent's estimate of the payoff associated to $S_i(t) = +1$ compared to that of $S_i(t) = -1$. The actual decision of agent i is probabilistic and drawn using the classic "logit" rule in discrete decision theory [43]; i.e., it is sampled from a Boltzmann distribution over choices:

$$\mathbb{P}[S_i(t) = \pm 1] = \frac{e^{\pm \beta Q_i(t)}}{e^{\beta Q_i(t)} + e^{-\beta Q_i(t)}} = \frac{1}{2}[1 \pm \tanh(\beta Q_i(t))],$$
(1)

or, equivalently, the expected choice m_i (or "intention") of agent i at time t is given by

$$m_i(t) := \langle S_i(t) \rangle = \tanh(\beta Q_i(t)).$$
(2)

Parameter β, assumed to be independent of i henceforth, is analogous to the inverse temperature in statistical physics and represents the agent's *rationality* or *intensity of choice* [44]. The limit $\beta \to \infty$ corresponds to perfectly rational agents, in the sense that they systematically pick the choice corresponding to their preference [given by the sign of $Q_i(t)$], while setting $\beta = 0$ gives erratic agents that randomly pick either decision with probability $1/2$ regardless of the value of $Q_i(t)$. This, in fact, turns out to be the case, in our model, in a whole region of parameter space: When β is smaller than some critical value β_c, all intentions m_i do converge to zero, leading to a random string of decisions.

The evolution of the preference $Q_i(t)$ is where the learning takes place. We resort to so-called "Q learning" [45], i.e., reinforcement learning with a memory loss parameter α. Given the (yet unspecified) reward $R_i^{\pm}(t)$ associated to making the choice $S_i(t) = \pm 1$, the evolution of incentives (and, in turn, beliefs) is given by

$$Q_i(t+1) = (1 - \alpha)Q_i(t) + \alpha[R_i^+(t) - R_i^-(t)].$$
(3)

This map amounts to calculating an exponentially weighted moving average (EWMA) on the history of rewards $R_i^{\pm}(t)$. Taking $\alpha = 0$, the agent's preferences are fixed at their initial values, and we thus restrict ourselves to $\alpha > 0$. When $\alpha \to 0$, $Q_i(t)$ is approximately given by the average reward over the last α^{-1} time steps. Note here that this *averaging* of past rewards is not exactly the same as the *accumulation* rule [where the reward would not be multiplied by α in Eq. (3)] appearing in some forms of "experience weighted attraction" that are popular in the socio-economic context [46]. In fact, one can always rescale the prefactor α into β, so our choice just means that Q does not

artificially explode as $\alpha \to 0$, which would impose that β effectively diverges in that limit.

Now, the missing ingredient is the specification of the rewards that encodes heterogeneity and nonreciprocity of interactions. Inspired by the theory of spin glasses, in particular, by the SK model [47,48], we define the rewards associated to choices $\pm$ as

$$R_i^{\pm}(t) = \pm \frac{1}{2} \left(\sum_{j=1}^{N} J_{ij} S_j(t) + h_i(t) \right).$$
(4)

Here, the matrix elements J_{ij} specify the mutually beneficial or competitive nature of the interactions between i and j, while $h_i(t)$ can be interpreted as the personal preference of the agent toward the choice corresponding to the sign of this "external field." (Note that J_{ij} measures the impact of the decision of j on the reward of i.) We henceforth set $h_i(t) = 0$, but note that the presence of such idiosyncratic preference may be relevant, as it breaks the global $\mathbf{S}(t) \to -\mathbf{S}(t)$ symmetry of the game. In a game-theoretic language, this extra term indeed allows one to have non-symmetric payoff matrices for all pairwise interactions in between the agents, as clearly visible when describing the game in the standard "normal-form" representation; see Appendix A. More complicated, multiagent-dependent payoffs could also be considered, which corresponds to p-spin models in the spin-glass literature and that we cursorily discuss in the conclusion [49].

In the context of firm networks, a client-supplier relation would correspond to $J_{ij} > 0$, whereas two firms i and j competing for the same clients would correspond to $J_{ij} < 0$. In the so-called "Dean problem," $J_{ij} > 0$ means that agents i and j get along well, whereas $J_{ij} < 0$ means that they are in conflict [50]. The sign of S_i determines in which of the two available rooms agent i should sit in order to minimize the number of possible conflicts. A predator-prey situation is when $J_{ij} \times J_{ji} < 0$, meaning that if i makes a gain, j makes a loss and vice versa.

Note that reward $R_i(t)$ depends on the actual (realized) decision of other players and not their expected decisions or intentions. In other words, agents resort to *online* learning, which differs from *offline* learning where other players' decisions $S_i(t)$ are averaged over large batch sizes during which their inclinations would be assumed constant and replaced by their expectation $m_i(t)$.

Based on the learning dynamics, agents thus make a decision based on an imperfectly learned approximation of what other players are likely to do. Indeed, using Eq. (4) to express $Q_i(t)$ as the time-weighted sum (EWMA) of past rewards, plugging Eq. (3), and finally replacing in Eq. (2), one finds that the evolution of agent i's intention is written

$$m_i(t+1) = \tanh\left(\beta \sum_{j=1}^{N} J_{ij} \tilde{m}_j^{\alpha}(t) \right),$$
(5)

where $\tilde{m}_i^\alpha(t)$ can be interpreted as the estimate of agent j's expected decision at time $t + 1$ based on its past actions up to time t:

$$\tilde{m}_i^\alpha(t) = \alpha \sum_{t' \le t} (1 - \alpha)^{t - t'} S_i(t'). \tag{6}$$

Expressed in this form, it is clear that there is characteristic timescale $\tau_\alpha \sim 1/\alpha$ over which past choices contribute to the moving average $\tilde{m}_i^\alpha(t)$. Note that offline learning would correspond to a different evolution equation, namely,

$$m_i(t + 1) = \tanh\left(\beta \sum_{j=1}^N J_{ij} m_j(t)\right), \tag{7}$$

although the two coincide in the $\alpha \to 0$ limit.

At this stage, it may be useful to compare and contrast our prescription with previous work. On the one hand, the learning procedure closely resembles the original proposition by Sato, Akiyama, and Crutchfield [51] and its treatment by Galla [13,52,53] and others [54–57]; however, these authors considered games comprising only two players with many strategies. The subsequent case explored by Galla and J. Farmer considering a larger number of players [33] therefore lies closer to our setting, but still considers many strategies and a correspondingly more complex payoff structure, while the similar binary decision models proposed by Semeshenko et al. [58,59] are restricted to perfectly rational agents and homogeneous interactions $J_{ij} = J_0 > 0 \ \forall \ i, j$. Note that all these works also consider accumulated rewards and offline learning, in contrast with our averaged rewards and online learning (for a comparison between the two and other learning dynamics, see Ref. [57]). On the other hand, replicator models with random nonsymmetric interactions between a large number of species [60,61] share many features with the system at hand, but the prescribed dynamics are inherently linked to evolutionary principles such as extinction that are not present in our model. We can finally mention other Ising-inspired games, such as that introduced in Refs. [62–64], which lead to similar fixed-point equations. So far, however, these models have been studied without heterogeneities and, therefore, do not present the "radical complexity" related to the presence of a very large number of possible solutions discussed hereafter.

B. The interaction matrix

In order to rely on known results about the SK model, we assume in the following that all the agents randomly interact with one another, meaning that all elements of the matrix $\mathbf{J}$ are nonzero. Sparse matrices, corresponding to low-connectivity interaction matrices, would probably be more realistic in an economic context. However, we expect that many of the conclusions reached below will

qualitatively hold in such cases as well. Such a universality is strongly supported by the preliminary numerical simulations on random regular graphs with a much smaller connectivity presented in Appendix B.

We choose interactions J_{ij} between i and j to be random Gaussian variables of the order of $N^{-1/2}$ [65], with J_{ij}, in general, different from J_{ji}, accounting for possible nonreciprocity of interactions. More precisely, we introduce the parameter ε and write the interaction matrix as

$$J_{ij} = \left(1 - \frac{\varepsilon}{2}\right) J_{ij}^S + \frac{\varepsilon}{2} J_{ij}^A, \tag{8}$$

with $\mathbf{J}^S$ a symmetric matrix and $\mathbf{J}^A$ an antisymmetric matrix. The entries of both these matrices are independent and sampled from a Gaussian distribution of mean 0 and variance σ^2/N (the case of a nonzero average value of $\mathbf{J}^S$ is briefly discussed below). This defines what we call the SK game henceforth.

In the following, we set $\sigma = 1$ without loss of generality. The resulting variance of J_{ij} is, thus, given by

$$v(\varepsilon) := N\mathbb{V}[J_{ij}] = 1 - \varepsilon + \frac{1}{2}\varepsilon^2. \tag{9}$$

The specific cases $\varepsilon = \{0, 1, 2\}$, hence, correspond to fully symmetric ($J_{ij} = J_{ji}$), asymmetric (i.e., J_{ij} and J_{ji} independent), and antisymmetric ($J_{ij} = -J_{ji}$) interactions, respectively. We can, thus, also characterize the correlation between J_{ij} and J_{ji} through parameter η:

$$\eta = \frac{\overline{J_{ij}J_{ji}}}{\overline{J_{ij}^2}} = \frac{1 - \varepsilon}{v(\varepsilon)}, \tag{10}$$

where overlines indicate an average over the disorder.

It may actually be insightful to allow for a nonzero average value to the interaction parameters and define the rewards $R_i^\pm(t)$ as

$$R_i^\pm(t) = \pm\frac{1}{2}\left(\sum_{j=1}^N J_{ij}S_j(t) + J_0 M(t)\right);$$

$$M(t) := \frac{1}{N}\sum_{j=1}^N S_j(t). \tag{11}$$

Note that if only the J_0 term is present, the game becomes simple, in the sense that either $J_0 > 0$ and $\beta J_0 > 1$, and all agents converge to the same strategy $m_i \equiv m$ with m a nonzero solution of $m = \tanh(\beta J_0 m)$, or they converge to the random rock-paper-scissor strategy $m_i \equiv 0$.

As mentioned above, one may think that agents have some idiosyncratic preferences or different costs associated to the two possible decisions $S_i = \pm 1$, in the form of the possibly time-dependent field $h_i(t)$. Naturally, h_i favors

$S_i = +1$ if positive and $S_i = -1$ if negative. In the present paper, we restrict to $h_i(t) \equiv 0$, $\forall\ i, t$, but one expects from the literature on spin glasses that the main results discussed below would still hold for small enough h_i's. Beyond some threshold value, on the other hand, agents end up aligning to their *a priori* preference, i.e., $m_i h_i > 0$.

III. BASIC INTUITION AND MAIN RESULTS

In this section, we review the most salient properties of the SK game' introduced above and why these results may be of interest in the context of socio-economic modeling. This section is intended to be give a broad, nontechnical overview, whereas the following sections delve in more details, using the language and methods of statistical physics. Readers more interested in conceptual results and less in technical details are invited to read this section, skip the subsequent ones, and jump directly to the final discussion in Sec. IX.

In the SK game, the payoff of each agent is a random function of the decisions of all other agents. Hence, learning the optimal strategy (in terms of the probability for agent i to play $+1$ or -1) is bound to be extremely difficult. Defining the average excess reward [66] per agent as

$$\mathcal{R}_N := \frac{1}{N} \sum_{i,j=1}^{N} S_i J_{ij} S_j \qquad (12)$$

and noting that $\sum_{i,j=1}^{N} S_i J_{ij} S_j = [1 - (\varepsilon/2)] \sum_{i,j=1}^{N} S_i J_{ij}^S S_j$, the largest possible average reward $\mathcal{R}_\infty$ for $N \to \infty$ can be exactly computed using the celebrated Parisi solution of the classical SK model and reads [48,67]

$$\lim_{N \to \infty} \mathcal{R}_N := \mathcal{R}_\infty = 0.7631\ldots \times (2 - \varepsilon). \qquad (13)$$

However, the strategies corresponding to such an optimal value can be determined only by algorithms that need a time exponential in N [68]. Hence, it is expected that simple learning algorithms will inevitably fail to find the true optimal solution. Nevertheless, we also know from the spin-glass folklore [48] (see below for more precise statements) that many configurations of $\{S_i\}$'s correspond to quasioptima or, in the language of Simon, satisficing solutions [1] (see also Ref. [20]). It is in a sense the proliferation of such suboptimal solutions that prevent simple algorithms to find the *optimum optimorum*. Furthermore, if learning indeed converges (which is not the case when ε is too large, i.e., when interactions are not reciprocal enough), the obtained fixed point heavily depends on the initial condition and/or the specific interaction matrix $\mathbf{J}$.

A. Phase diagram in the noiseless limit

Let us first consider the case where agents always choose the action that would have had the best average reward in the past $Q_i(t)$ (assuming that other agents still played what they played). This corresponds to the noiseless learning limit $\beta \to \infty$. In this case, the iteration map Eq. (5) becomes

$$S_i(t+1) = \mathrm{sgn}\left(\sum_{t' \leq t}(1-\alpha)^{t-t'} \sum_{j=1}^{N} J_{ij} S_j(t')\right), \qquad (14)$$

and the model is fully specified by two parameters: α (controlling the memory timescale of the agents) and ε (controlling the reciprocity of interactions). For N not too large, the evolution of Eq. (14) leads to either fixed points, or oscillations, or else chaos. The schematic phase diagram in the plane (α, ε) is shown in Fig. 1.

One clearly sees a region for (α, ε) small where learning reaches a fixed point, in which the average reward is close to but significantly below the theoretical optimum $\mathcal{R}_\infty$ given by Eq. (13); see Fig. 2 [70]. Note that learning definitely helps: For $\varepsilon = 0$, most fixed points are characterized by a typical reward $\mathcal{R} \approx 1.01$ [72], significantly worse than the value of approximately 1.40 reached by our learning agents, extrapolated to $N \to \infty$.

As ε and α are varied, one observes the following features.

(i) When ε is not too large (interactions sufficiently reciprocal) and α increases (shorter and shorter memory), learning progressively ceases to converge and oscillations start appearing: Impatient learning generates cycles. This leads to a sharp decrease of the average reward [see Fig. 2(a)], as agents overreact to new information and are no longer able to coordinate on a mutually beneficial equilibrium. A similar effect is observed in dynamical models of supply chains, where overreaction leads to oscillating prices and production—the so-called bullwhip effect [74] (see also Refs. [21,23]).

(ii) Conversely, when α is small (long memory) and ε increases, the probability to reach a fixed point progressively decreases, and when a fixed point is
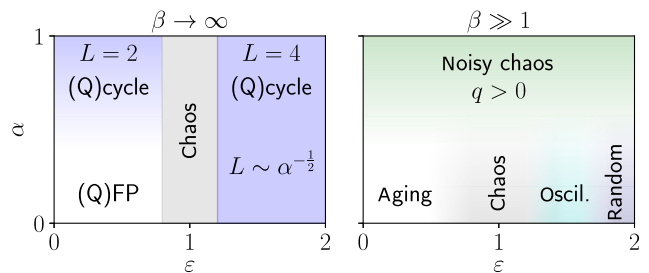

FIG. 1. Qualitative phase diagram in the (ε, α) plane for the SK game in the noiseless (left) and weak noise (right) regimes. FP and Q refer to "fixed point" and "quasi," respectively. "Random" refers to the case where agents pick ± 1 with equal probability $[m_i(t) = 0\ \forall\ t]$, whereas "chaos" means that at a given instant of time players have well-defined intentions m_i, but these evolve chaotically with time.
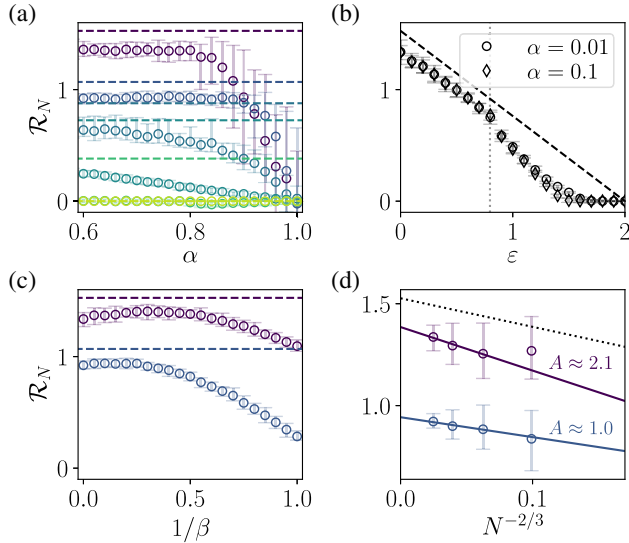
FIG. 2. Evolution of the average reward with (a) the memory loss rate α, for $\varepsilon = \{0, 0.6, 0.85, 1.05, 1.5, 2\}$ from dark purple to light green, $\beta \to \infty$; (b) the asymmetry ε for $\alpha = 0.01$ and $\alpha = 0.1$, $\beta \to \infty$—the vertical dotted line indicates the value ε_c above which the system becomes chaotic [73]; (c) the noise level $1/\beta$ for $\alpha = 0.01$, $\varepsilon = \{0, 0.6\}$ (dark purple and blue, respectively)—note the nonmonotonic behavior; (d) system size N for $\alpha = 0.01$, $\beta \to \infty$, $\varepsilon = \{0, 0.6\}$ (dark purple and blue, respectively)—continuous lines showing fits $\mathcal{R}_N = \mathcal{R}_\infty - AN^{-2/3}$, excluding $N = 32$. The dotted line indicates the best fit for the SK ground state, for which $A \approx 1.5$ [71]. In (a)–(c), $N = 256$ and the dashed line represents the Parisi solution for $\mathcal{R}_\infty$.

reached, the average reward is reduced—see Fig. 2(b). Beyond some threshold value, the dynamics becomes completely chaotic, leading to further loss of reward. Note that "chaos" here means that, although agents have well-defined intentions $m_i(t) \neq 0$ at any given instant of time, these intentions evolve chaotically forever [75].

(iii) Surprisingly, oscillations *reappear* when ε becomes larger than unity, i.e., when interactions are mostly antisymmetric, "predator-prey"-like. Perhaps reminiscent of the famous Lotka-Volterra model, agents' decisions and payoffs become periodic, with a period that scales anomalously as $\alpha^{-1/2}$, i.e., much shorter than the natural memory timescale α^{-1}. Although not the Nash equilibrium $m_i \equiv 0$, this oscillating state allows the average reward to be positive, even when at each instant of time some agents have negative rewards.

(iv) Only in the extreme competition limit $\varepsilon = 2$ [corresponding to a zero sum game, Eq. (13)] and for small α are agents able to learn that the unique Nash equilibrium is to play random strategies $m_i \equiv 0$ (see Fig. 1, bottom-right region) [76].

(v) Finally, in the extreme (and unrealistic) case $\alpha = 1$, where agents choose their strategy based on only the

last reward, the system evolves, as ε increases from zero, from high-frequency oscillations with period $L = 2$ to "weak chaos," to "strong chaos" when $\varepsilon \approx 1$, and finally back to oscillations of period $L = 4$ when $\varepsilon \to 2$.

In order to characterize more precisely such temporal behaviors, it is useful to introduce the two-point autocorrelation function of the *expected* decisions or *intentions*:

$$C(t, t + \tau) = \frac{1}{N} \sum_i \overline{\langle m_i(t) m_i(t + \tau) \rangle}, \qquad (15)$$

where the angular brackets now refer to an average over initial conditions [77]. In cases where the dynamics are assumed to be time-translation invariant, we write $C(\tau)$, which corresponds to the above quantity averaged over time after the system has reached a steady state.

The autocorrelation function corresponding to the different cases described above are plotted in Fig. 3. Note that the signature of oscillations of period L is that $C(nL) \equiv 1$ for all integer n. However, note that, when $\varepsilon < \varepsilon_c$, not all spins flip at each time step. The fact that $C(2n + 1) = 0$ means that half of the spins, in fact, remain fixed in time, while the other half oscillate in sync between $+S_i$ and $-S_i$ [78]. In the chaotic phases, $C(\tau)$ tends to zero for large τ, with either underdamped or overdamped oscillations. Hence, in these cases, the configuration $\{S_i\}$ evolves indefinitely with time and hardly ever revisits the same states.
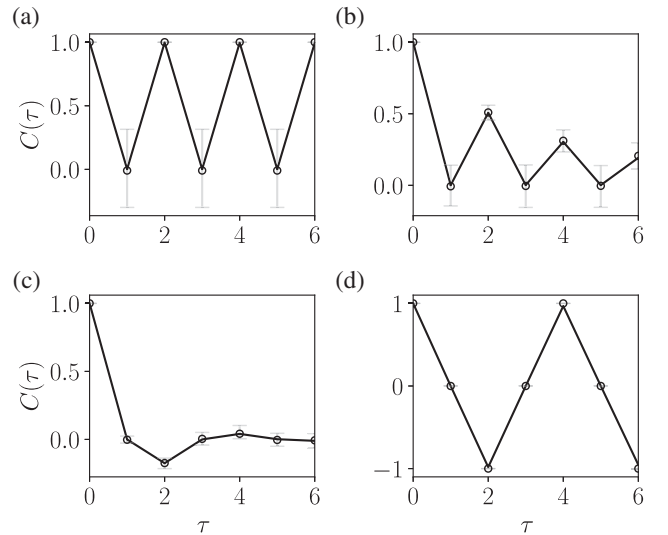

FIG. 3. Evolution of the steady-state two-point correlation function for $\alpha = 1$, $\beta \to \infty$, markers indicating simulations of the game at $N = 256$ averaged over 128 samples of disorder and initial conditions, error bars showing 95% confidence intervals, and continuous lines representing the solution of the DMFT equation [see Eq. (30) below]. (a) $\varepsilon = 0.1$ ($\varepsilon < \varepsilon_c \approx 0.8$), cycles of length $L = 2$. (b) $\varepsilon = 0.85$ ($\varepsilon_c < \varepsilon < 1$), "weakly" chaotic behavior. (c) $\varepsilon = 1.05$ ($1 < \varepsilon < 2 - \varepsilon_c$), "strongly" chaotic behavior. (d) $\varepsilon = 1.5$ ($\varepsilon > 2 - \varepsilon_c$), cycles of length $L = 4$.

B. Finite N, large t versus finite t, large N

The previous results hold in the long-time limit for finite N, i.e., when we take formally the limit $t \to \infty$ before $N \to \infty$. There is, however, another regime for which the $N \to \infty$ limit is taken first, which might be more appropriate in the context of firm networks (say) when the number of firms is very large.

Interestingly, specific analytical tools are available to treat this regime—the so-called dynamic mean-field theory (DMFT) that we use below.

In this infinite N regime, new types of behavior appear that one can call quasifixed points or quasicycles. In the case of quasifixed points, learning does not strictly speaking converge, but actions $S_i(t)$ fluctuate around time-independent averages. In other words, the two-point correlation $C(\tau)$ is not equal to one for all τ (which would be the case for a fixed point) but reaches a positive plateau value for large τ: $C(\tau \to \infty) = C_\infty > 0$; see Ref. [79] for the $\alpha = 1$ case. Only when $\varepsilon = 0$ does one find $C_\infty = 1$. The same holds for quasicycles of length L if one considers the correlation function computed for $\tau = nL$, with n an integer: $C(nL \to \infty) = C_\infty > 0$, with $C_\infty = 1$ when $\varepsilon = 2$.

The schematic phase diagram drawn in Fig. 1 then continues to hold in the large N finite t limit, provided one interprets "fixed points (cycles)" as "quasifixed points (cycles)" in the sense defined above. More details on the subtle role of α, N, and t are provided in the next, technical sections.

C. Noisy learning and "aging"

In the presence of noise, the "convictions" $|m_i|$ of agents naturally decrease and, in fact, become zero (i.e., decisions are totally random) beyond a critical noise level that depends on the asymmetry parameter ε: More asymmetry leads to more fragile convictions (see Fig. 16 below for a more precise description).

When the noise is weak but nonzero, strict fixed points do not exist anymore but are replaced (for ε small enough) by quasifixed points—in the sense that the intentions m_i fluctuate around some plateau value for very long times before evolving to another configuration completely uncorrelated with the previous one. This process goes on forever, albeit at a rate that slows down with time: Plateaus become more and more permanent. This is called "aging" in the context of glassy systems.

In a socio-economic context, it means that a form of quasiequilibrium is temporarily reached by the learning process, but such a quasiequilibrium will be completely disrupted after some time, even in the absence of any exogenous shocks. This is very similar to the quasinonergodic scenario recently proposed in Ref. [14], although in our case the evolution time is not constant but increases with the "age" of the system, i.e., the amount of time the game has been running.

Perhaps counterintuitively, however, the role of noise is on average beneficial when $1/\beta$ is not too large. Indeed, as shown in Fig. 2, the average reward first increases as a small amount of noise is introduced before reaching a maximum beyond which "irrationality" becomes detrimental. The intuition is that, without noise, the system gets trapped by fixed points with large basins of attraction but lower average rewards. A small amount of noise allows agents to reassess their intentions and collectively reach more favorable quasifixed points, much as with simulated annealing or stochastic gradient descent.

When learning leads to a chaotic evolution, i.e., when J_{ij} and J_{ji} are close to uncorrelated ($\varepsilon \sim 1$), noise in the learning process does not radically change the evolution of the system: Deterministic chaos just becomes noisy chaos. However, there is still a distinction between a low-noise phase where, at each instant of time, agents have nonzero expected decisions m_i (that evolve over time) from a high-noise phase where agents always make random choices between ± 1 with probability $1/2$ (see Fig. 1).

Finally, in the case where learning leads to cycles, any amount of noise irremediably disrupts the synchronization process, and cycles are replaced by pseudocycles, with either underdamped or overdamped characteristics. In the limit $\varepsilon \to 2$, fluctuations drive the system to a paramagnetic state where $q = C(0) = 0$ (see Fig. 16 below), meaning the agents remain undecided.

D. Individual rewards

As we have noted above, the average (excess) reward is close to but significantly below the theoretical optimum $\mathcal{R}_\infty$ given by Eq. (13). However, some agents are better off than others, in the sense that the individual excess reward e_i at the fixed point (when fixed points exist) is different from agent to agent. Noting that

$$e_i \coloneqq R_i^{S_i^\star} - R_i^{-S_i^\star} = \left| \sum_j J_{ij} S_j^\star \right|, \qquad (16)$$

where the second equality holds because at the fixed point one must have $S_i^\star = \operatorname{sgn}(R_i^\star)$, it is clear that, in the fully reciprocal case $\varepsilon = 0$, all rewards e_i are positive. The distribution $\rho(e)$ of these rewards over agents is expected to be self-averaging for large N, i.e., independent of the specific realization of the J_{ij} and of the initial condition. Such distribution is shown in Fig. 4. One notices that $\rho(e)$ vanishes linearly when $e \to 0$:

$$\rho(e) \underset{e \to 0}{\approx} \kappa e, \qquad \kappa \approx 1.6,$$

as for the standard SK model, although the value of κ is distinctly different from the one obtained for the true optimal states of the SK model, for which $\kappa_{\mathrm{SK}} \approx 0.6$ [80,81]. Such a discrepancy is expected, since the fixed
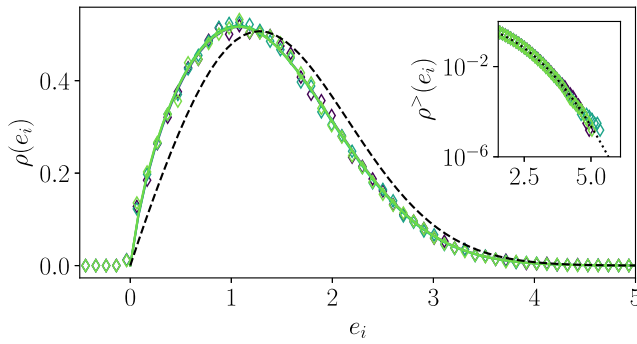
FIG. 4.   Distribution of individual rewards at $\beta \to \infty$ for $N = 512$ and 128 initial conditions and realizations of the disorder in the fully symmetric case $\varepsilon = 0$ for $\alpha = \{0.5, 0.1, 0.01\}$ (dark to light coloring). The dashed line is the Sommers-Dupont analytical solution to the SK model [80]. Inset: associated survival function in a lin-log scale and focusing on the right tail, the dotted line corresponding to a Gaussian fit.
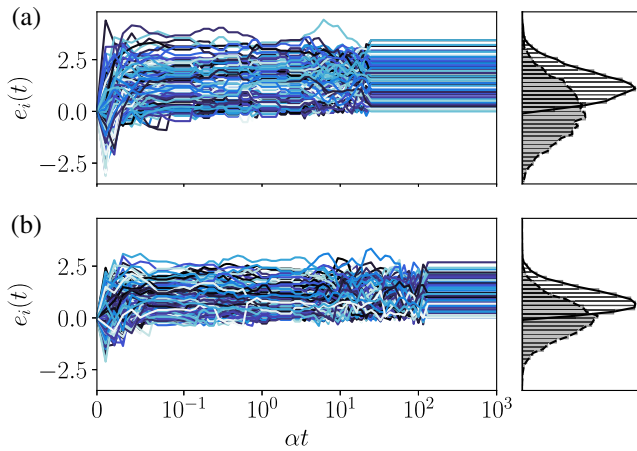

FIG. 5.   Evolution of individual rewards in time for $N = 256$, $\alpha = 0.01$, $\beta \to \infty$, and (a) $\varepsilon = 0$ and (b) $\varepsilon = 0.6$. Right: histogram of the individual rewards after a single time step (shaded) and at the final time (unshaded).

points are obtained as the long-time limit of the learning process—in particular, since $\kappa > \kappa_{\mathrm{SK}}$, the number of poorly rewarded agents is too high compared to what it would be in the truly optimal state. Note that, once α is sufficiently

small for the system to reach a fixed point, its precise value does not seem to have an impact on the distribution of rewards and κ.

Another important remark is that the distribution of rewards $\rho(e)$ does not develop a "gap" for small e, i.e., a region where $\rho(e)$ is exactly zero. In other words, although all agents have positive rewards, some of them are very small. This is associated with the so-called "marginal stability" of the equilibrium state [82], to wit, its fragility with respect to small perturbations, as discussed in more detail in the next subsection.

For very large e, the distribution $\rho(e)$ decreases like a Gaussian (Fig. 4 inset), corresponding to a central limit theorem behavior in that regime, as for the SK model [83]. Figure 5 shows how the rewards of individual agents evolve from an initially random configuration before settling to constant (but heterogeneous) values at the fixed point.

As competitive effects get stronger (i.e., as ε increases) and the system ceases to reach a fixed point, the distribution $\rho(e)$ develops a tail for negative values of e, meaning that some agents make negative gains; see Fig. 6. In the extreme "predator-prey" limit $\varepsilon = 2$, the distribution $\rho(e)$ becomes perfectly symmetric around $e = 0$, as expected—see Fig. 6, rightmost plot. However, note that there is no persistence in time of the winners: The individual reward autocorrelation function eventually decays to zero, possibly with oscillations in the competitive region.

E. Unpredictable equilibria

Now, the interesting point about our model is that the final rewards are highly dependent on the initial conditions and/or the realization of the J_{ij}'s. In other words, successful agents in one realization of the game become the losers for another realization obtained with different initial conditions. A way to quantify this is to measure the cross-sectional correlation of final rewards for two different realizations, i.e.,

$$C_N^\times := \frac{1}{N} \sum_i (e_i^a - \langle e^a \rangle)(e_i^b - \langle e^b \rangle), \quad a \neq b, \quad (17)$$
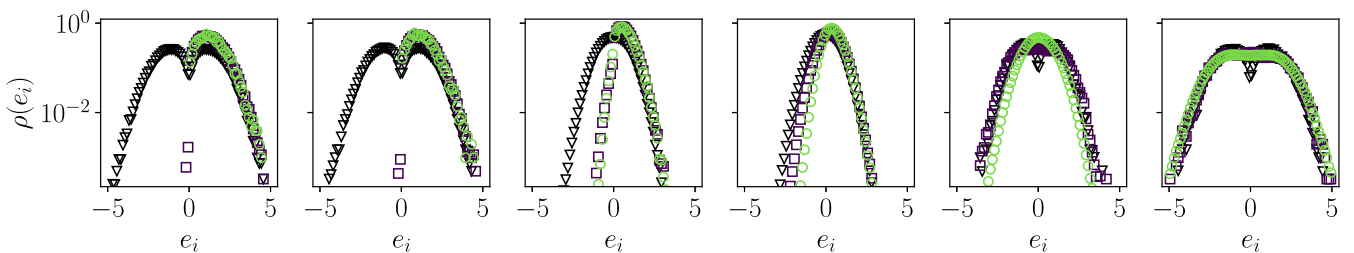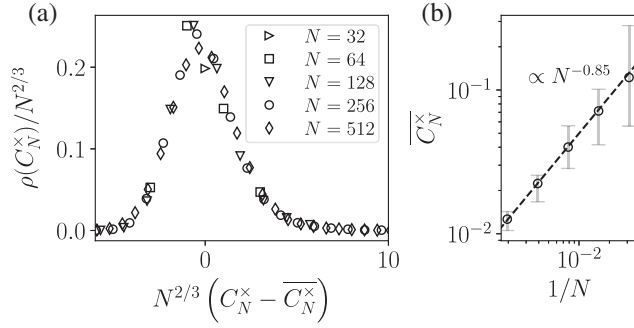

FIG. 6.   Distribution of individual rewards for $N = 256$ and $\alpha = \{1, 0.5, 0.01\}$ represented by black triangles, purple squares, and green circles, respectively, $\beta \to \infty$, measured over 32 initial conditions and realizations of the disorder. From left to right: $\varepsilon = \{0, 0.1, 0.85, 1.05, 1.5, 2\}$.

FIG. 7. Overlap between solutions for different initial conditions and identical draws of interactions, for $\alpha = 0.01$, $\beta \to \infty$, and $\varepsilon = 0$. (a) Distribution of overlaps shifted by the mean and rescaled with the system size N to the power 2/3. (b) Average overlap as a function of system size in log-log coordinates, with the best regression line $N^{-0.85}$. Error bars show the 95% confidence interval over 16 different draws of the disorder.

where a and b correspond to two different initial conditions and $\langle e \rangle$ corresponds to the cross-sectional average reward. As shown in Fig. 7, $C_N^\times$ goes to zero at large N, indicating that the final outcome of the game, in terms of the winners and the losers, cannot be predicted. The dependence on N appears to be nontrivial, with different exponents governing the decay of the mean overlap $\overline{C_N^\times}$ (decaying as $N^{-0.85}$) and its standard deviation (decaying as $N^{-2/3}$).

A similar effect would be observed if instead of changing the initial condition one would randomly change the interaction matrix $\mathbf{J}$ by a tiny amount ϵ. The statement here is that, for any small ϵ, $C_N^\times$ goes to zero for sufficiently large N. This is called "disorder chaos" in the context of spin glasses [87]; by analogy with known results for the SK model, we conjecture that $C_N^\times$ is a decreasing function of $N\epsilon^\zeta$, where ζ is believed to be equal to 3 in the SK case [88]. This means that when $N \gg \epsilon^{-\zeta}$ the rewards between two systems with nearly the same interaction structure, starting with the same initial conditions, are close to independent.

Such a sensitive dependence of the whole equilibrium state of the system (in our case, the full knowledge of the intentions m_i of all agents) prevents any kind of "common knowledge" assumption about what other agents will decide to do in a specific environment. No reasonable learning process can lead to a predictable outcome; even the presence of a benevolent social planner assigning their optimal strategy to all agents would not be able to do so without a perfect knowledge of all interactions between agents and without exponentially powerful (in N) computing abilities. Such a "radically complex" situation leads to "radical uncertainty" in the sense that the behavior of agents, even rational, cannot be predicted. Learning agents can achieve only satisficing solutions that are, furthermore, hypersensitive to details. As we see in Sec. III C, any amount of noise in the learning process makes the whole

system "jump" from one satisficing solution to another in the course of time.

F. Increasing cooperativity

A way to help agents coordinate is to use rewards given by Eq. (11) with $J_0 > 0$, representing a nonzero average cooperative contribution to rewards. This term obviously helps agents finding mutually beneficial strategies. (Note that, with our normalization, the J_0 term is, in fact, $N^{-1/2}$ times smaller than the random interaction terms J_{ij}.)

The impact of such a term is well understood in the context of the SK model for $\varepsilon = 0$ [89], which nicely translates into the current dynamical framework. For $\beta = \infty$, one finds that, whenever $J_0 \leq 1$, the average intention $M(t)$ remains zero for large N, and one expects that the learning process is not affected by such a "nudge." When $J_0 > 1$, on the other hand, the situation changes as all agents start to coordinate on one of the two possible choices. As shown in Fig. 8, the average intention becomes nonzero, although a finite fraction of agents still play opposite to the majority because of their own idiosyncratic rewards.

For $J_0 \gg 1$, radical complexity disappears and learning quickly converges to the obvious optimal strategy where all agents make the same move $S_i = +1$ or $S_i = -1$, $\forall\ i$. In this case, $\mathcal{R}_N = J_0$ as $M(t)$ eventually reaches unity; see Fig. 8. For $\varepsilon > 0$, the same occurs albeit for different values of J_0.

In the case $J_0 < 0$ with $|J_0| \gg 1$, the only solution of Eq. (7) (valid for $\alpha \to 0$) is $m_i = 0$ for all i; i.e., agents cannot coordinate and play random strategies.

G. Habit formation

Up to this point, all results have assumed that there is no self-interaction, $J_{ii} = 0$. Nonetheless, it is interesting to consider the possibility of having an $O(1)$ positive diagonal term in the interaction matrix. In the socio-economic context, such a contribution is relevant, as it represents self-reinforcement of past choices, which is also called
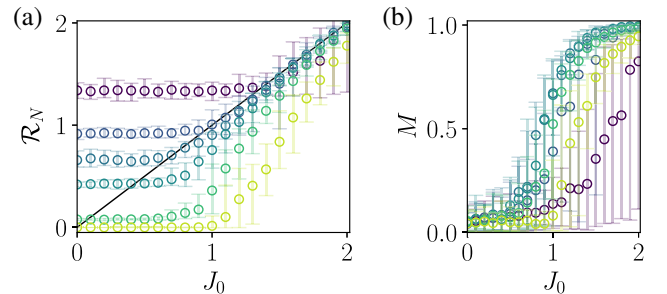

FIG. 8. (a) Evolution of the average reward with the incentive to cooperate J_0 for $N = 256$, $\alpha = 0.01$, $\beta \to \infty$, and $\varepsilon = \{0, 0.6, 0.85, 1.05, 1.5, 2\}$ with colors ranging from purple to light green with increasing ε. (b) Average intention in the long-time limit $M = \lim_{t \to \infty} M(t)$ for the same parameters.

"habit formation" where agents stick to past choices, a popular idea in behavioral science; see, e.g., Refs. [9,11], and references therein.

The introduction of a diagonal contribution has important consequences for the problem. Assuming the self-interaction is identical for all agents, $J_{ii} = J_d > 0$, it will rather intuitively favor the emergence of fixed points, since agents are tempted to stick to past choices. It is, for instance, known that, in the case of fully random interactions $\varepsilon = 1$, fixed points will start to appear when J_d is sufficiently large [90]. Interestingly, these fixed points can be very difficult to reach dynamically with standard Hopfield dynamics ($\alpha = 1$).

Adding such diagonal contribution to our learning dynamics, we observe that the fraction of trajectories converging to seemingly dynamically inaccessible configurations significantly increases, especially when $\alpha \ll 1$. While further work is required to precisely assess the effectiveness of learning with self-reinforcement (particularly as finite size effects appear to play a significant role), such a result is consistent with the overall influence of learning reported here.

H. Core messages

In line with the conclusions of Galla and J. Farmer [13], our multiagent binary decision model provides an explicit counterexample to the idea that learning could save the rational expectation framework (cf. Sec. I and Ref. [8]).

Learning, in general, does not converge to any fixed point, even when the environment (in our case, the interaction matrix **J**) is completely static: Nonstationarity is self-induced by the complexity of the game that agents are trying to learn, as also recently argued in Ref. [12].

When learning does indeed converge (which requires *a minima* a high level of reciprocity between agents) the collective state reached by the system is far from the optimal state, which only a benevolent, omniscient social planner with formidable powers can achieve. In other words, even more sophisticated learning rules would not really improve the outcome: The SK game is *unlearnable* and—as argued by Simon [1]—agents must resort to suboptimal, satisficing solutions.

Furthermore, any small random perturbation (noise in the learning process or slow evolution in the environment) eventually destabilizes any fixed point reached by the learning process and completely reshuffles the collective state of the system: In the long run, agents initially favoring the +1 decision end up favoring −1, and better-off agents end up being the underdogs, and vice versa (much as in the simpler model in Ref. [14]).

Finally, even in the most favorable case of a fully reciprocal game with slow learning, the average reward is, in fact, *improved* when some level of noise (or irrationality) is introduced in the learning rule, before degrading again for large noise.

IV. FIXED-POINT ANALYSIS AND COMPLEXITY

We see that our model displays a wide variety of complex collective dynamics. Only in some cases does learning converge to nontrivial fixed points where strategies are probabilistic but with time-independent probability $p_\pm$ to play ± 1, such that $p_\pm = (1 \pm m_i^\star)/2$ for agent i. Such a steady state would be analogous to an *economic equilibrium* (although it is essential to dissociate this notion from that of a thermodynamic equilibrium, which may exist only in the case of fully reciprocal interactions, $\varepsilon = 0$).

We mostly focus, in the following, on the long memory case $\alpha \ll 1$ which is most relevant for thinking about learning in a (semi)realistic context. In this case, one can show that the exponential moving average on the realized values $S_i(t)$ converges to one on the expected values $m_i(t)$. Indeed, as detailed in Appendix C,

$$\left\langle \left(\alpha \sum_{t' \leq t} (1 - \alpha)^{t-t'} [m_i(t') - S_i(t')] \right)^2 \right\rangle \leq \frac{\alpha}{2 - \alpha} \xrightarrow{\alpha \to 0} 0.$$

(18)

This means that, up to fluctuations of the order of $\sqrt{\alpha}$, we can describe the dynamics of the system through a deterministic iteration on $m_i(t)$, in fact corresponding to offline learning. (We see below that the neglected fluctuations are of the order of $\sqrt{\alpha}/\beta$.)

Further making the *Ansatz* that the mean-field dynamics will eventually reach a fixed point $m_i(t) = m_i^\star \ \forall \ i$ given sufficient time, Eq. (5) then yields

$$m_i^\star = \tanh \left(\beta \sum_j J_{ij} m_j^\star \right).$$

(19)

This equation is known in the spin-glass literature as the naive mean-field equation (NMFE) when $\varepsilon = 0$ and defines a so-called static quantal response equilibrium, similar to its fully mean-field equivalent ($J_{ij} = J/N$) studied in Ref. [62].

To the reader familiar with the physics of disordered systems, this equation is immediately reminiscent of the celebrated Thouless-Anderson-Palmer (TAP) equation [91] describing the mean magnetization in the SK spin glass [47]. Physically, the NMFE is satisfied when minimizing the free energy of a system of N sites comprising $M \to \infty$ binary spins, with sites interacting through an SK-like Hamiltonian [92]. Despite being seemingly simpler than its previously mentioned TAP counterpart, which includes an additional Onsager "reaction term," the NMFE shares many of its properties. Relevant to our problem, both the NMFE and the TAP equations have a paramagnetic phase ($m_i^\star = 0 \ \forall \ i$) for $\beta < \beta_c$, while above this critical value there is a spin-glass phase where $q^\star = N^{-1} \sum_i (m_i^\star)^2 > 0$ and solutions are exponentially abundant in N [92–94]. The NMFE has a critical temperature $1/\beta_c = 2$ as opposed to

$1/\beta_c = 1$ in the TAP case, while the two equations become strictly equivalent in the $\beta \to \infty$ limit.

Using known properties from the spin-glass literature, we can, therefore, already establish that *if* the system reaches a fixed point when interactions are fully reciprocal ($\varepsilon = 0$) and memory is long ranged, it will be either a trivial fixed point where agents continue making random decisions forever ($m_i^\star = 0 \ \forall \ i$), or, when learning is not too noisy ($\beta > \beta_c$), the number of fixed points is $\sim \exp[\Sigma(\beta)N]$, where $\Sigma(\beta)$ is called the "complexity." In this second case, the fixed point actually reached by learning depends sensitively on the initial conditions and the interaction matrix **J**.

How is this standard picture altered when interactions are no longer reciprocal? In such cases, the system cannot be described using the equilibrium statistical mechanics machinery.

A. Critical noise level

In order to extend the notion of critical noise β_c to $\varepsilon > 0$, one can naively look at the linear stability of the paramagnetic solution $m_i^\star = 0 \ \forall \ i$ to Eq. (19). Just as in the TAP case [48], expanding the hyperbolic tangent to the second order and projecting the vector of m_i on an eigenvector of **J**, the stability condition can be expressed with the largest eigenvalue of the interaction matrix. Adapting known results from random matrix theory to our specific problem formulation, the spectrum of **J** can be expressed as an interpolation between a Wigner semicircle on the real axis ($\varepsilon = 0$), the Ginibre ensemble ($\varepsilon = 1$), and a Wigner semicircle on the imaginary axis ($\varepsilon = 2$) [95]. The resulting critical "temperature" is then given by

$$T_c(\varepsilon) = \frac{1}{\beta_c(\varepsilon)} = \frac{1}{2}\frac{(2-\varepsilon)^2}{\sqrt{\upsilon(\varepsilon)}}, \qquad (20)$$

recovering the known result $1/\beta_c = 2$ for the case $\varepsilon = 0$. (We recall that we have set the interaction variance σ^2 to unity throughout the paper. If needed, σ can be reinstalled by the rescaling $\beta \to \beta\sigma$.)

B. The elusive complexity

To determine if there are still an exponential number of fixed points to reach below the candidate critical noise level, i.e., if there is a spin-glass phase, for $\beta > \beta_c(\varepsilon)$ when $\varepsilon > 0$, we should study the complexity, defined for a single realization of the disorder as

$$\Sigma(\beta, \varepsilon) = \lim_{N\to\infty} \frac{1}{N}\log \mathcal{N}_J(N, \beta, \varepsilon), \qquad (21)$$

where $\mathcal{N}_J$ is the number of fixed points in the system for a given interaction matrix. There are then two ways to compute an average of this quantity over the disorder:

the "quenched" complexity, where the mean of the logarithm of the number of solutions is considered, and its "annealed" counterpart, where the logarithm is taken on the mean number of solutions. The former is usually considered to be more representative, as unlikely samples leading to an abnormally large number of solutions can be observed to dominate the latter (see, e.g., Refs. [20,96] for recent examples), but requires a more involved calculation with the use of the so-called "replica trick" [48]. In the TAP case, quenched and annealed complexities coincide for solutions above a certain free-energy threshold [72] (where most solutions lie but importantly not the ground state).

As a matter of fact, even in the annealed case, the computation of the TAP complexity has proved to be a formidable task and has sparked a large amount of controversy, as the original solution computed by Bray and Moore (BM) [72] has been put into question before being (partially) salvaged by the metastability of TAP states in the thermodynamic limit [97]. For a relatively up to date summary of the situation, we refer the reader to Parisi's contribution in Ref. [98].

While the BM approach can be adapted to the NMFE [93,99], several aspects of the calculation remain unclear, particularly as the absence of a subdominant reaction term means that the argument of the metastability of states in the $N \to \infty$ limit is no longer valid *a priori*, although numerical results support the marginally stable nature of NMFE fixed points in the thermodynamic limit [93]. We leave its extension to $\varepsilon > 0$ to a later dedicated work.

Nevertheless, the previously introduced critical β_c and the existing computation of the number of fixed points as a function of ε in the $\beta \to \infty$ limit [73,100] can be used to conjecture the boundaries of the region in (β, ε) space where the complexity Σ is nonvanishing. Indeed, in the zero-temperature case, it has been shown [73] that the annealed complexity can be expressed as a function of the asymmetry parameter η defined in Eq. (10) as

$$\Sigma(\eta) = -\frac{1}{2}\eta x^2 + \log 2 + \log \Phi(\eta x), \qquad (22)$$

with Φ the Gaussian cumulative density and x the solution to

$$x\Phi(\eta x) = \Phi'(\eta x), \qquad \Phi(x) := \frac{1}{2}\mathrm{erfc}\left(-\frac{x}{\sqrt{2}}\right). \qquad (23)$$

The main insight provided by this result is that the complexity vanishes at $\eta = 0$, corresponding to $\varepsilon = 1$, where the paramagnetic fixed point is supposed to be unstable as $\beta_c(\varepsilon = 1) = \sqrt{2}$. As the complexity is a decreasing function of temperature, this, therefore, means that $\varepsilon = 1$ is an upper limit for the existence of fixed points when β is finite. This conjecture is also consistent with the breakdown of fixed-point solutions to the dynamical
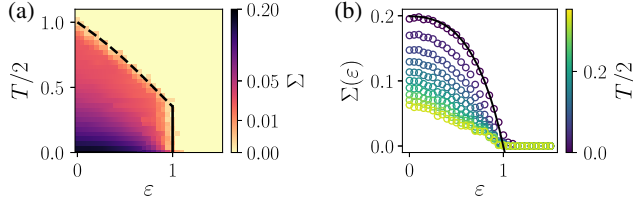
FIG. 9. (a) Annealed complexity of the naive mean-field equation in (T, ε) space, where we recall $T = \beta^{-1}$, measured numerically for $N = 40$. The dashed line represents the critical temperature $T_c(\varepsilon)$ for which the paramagnetic fixed point ceases to be stable, while the continuous line indicates $\varepsilon = 1$ inferred from the $\beta \to \infty$ result. (b) Annealed complexity as a function of ε for varying temperatures $T < 1/\sqrt{2}$, i.e., in the bottom region in (a), where the complexity vanishes at $\varepsilon = 1$. The continuous line represents the $\beta \to \infty$ analytical solution, recovering the result of Tanaka and Edwards [101] $\Sigma \approx 0.1992$ for $\varepsilon = 0$. For $\varepsilon > 1$ and $T = 0$, the only possible fixed point is $m_i^\star = 0$, $\forall\ i$.

mean-field theory below the critical noise level that is discussed in Sec. VIII A, as well as the saddle point equations obtained when adapting the BM calculation to $\varepsilon > 0$.

Combining these two somewhat heuristic delimitations for the existence of a large number of nontrivial fixed points, we obtain the critical lines shown in Fig. 9(a). Overlaying these borders with the annealed complexity measured numerically, we find a very good agreement. In particular, the vanishing of the complexity at $\varepsilon = 1$ in appears to be consistent for $T < T_c(\varepsilon = 1) = 1/\sqrt{2}$, as shown in Fig. 9(b). The agreement with the $\beta \to \infty$ analytical result, represented by the continuous line, also appears to validate our counting method at low temperatures. Note that one can, in fact, show that, for $\varepsilon > 1$ and $N = \infty$, the only fixed point (or Nash equilibrium) is the rock-paper-scissors equilibrium $m_i^\star = 0$, $\forall\ i$.

V. COUNTING LIMIT CYCLES

In the previous section, we establish the region of parameter space where exponentially numerous fixed points exist, which might possibly be reached by learning in the slow limit $\alpha \ll 1$. However, limit cycles of various lengths turn out to also be exponentially numerous when $\varepsilon < 1$, so we need to discuss them as well before understanding the long-term fate of the learning process within our stylized complex world.

A. Cycles without memory ($\alpha = 1$)

In the memoryless limit, the dynamics becomes that of the extensively studied Hopfield model [100,102,103], where the binary variable represents the activation of a neuron evolving as

$$S_i(t + 1) = \text{sgn}\left(\sum_j J_{ij} S_j(t)\right), \tag{24}$$

with *parallel* updates. Counting limit cycles of length L is even more difficult than counting fixed points (which formally correspond to $L = 1$). Some progress has been reported by Hwang *et al.* [73] in the memoryless case $\alpha = 1$. The notion of fixed-point complexity Σ [defined in Eq. (21)] can be extended to limit cycle complexity Σ_L for limit cycles of length L, with $\Sigma_{L=1} \equiv \Sigma$. The results of Hwang *et al.* [73] can be summarized as follows.

(i) When $\varepsilon < 1$, limit cycles with $L = 2$ have the largest complexity, which is exactly twice the fixed-point complexity: $\Sigma_2 = 2\Sigma_1$ (as was, in fact, previously shown by Gutfreund, Reger, and Young [100]).

(ii) The complexities $\Sigma_L(\varepsilon)$ all go to zero when $\varepsilon = 1$.

(iii) When $1 < \varepsilon \leq 2$, limit cycles with $L = 4$ dominate, with $\Sigma_4(\varepsilon) \geq \Sigma_2(2 - \varepsilon)$.

(iv) Close to $\varepsilon = 1$, the cutoff length L_c, beyond which limit cycles become exponentially rare, grows exponentially with N: $L_c \sim e^{aN}$, where a weakly depends on ε.

From this analysis, one may surmise the following.

(a) When a limit cycle is reached by the dynamics, it is overwhelmingly likely to be of length $L = 2$ for $\varepsilon < 1$ and of length $L = 4$ for $1 < \varepsilon \leq 2$.

(b) Even if exponentially less numerous, exponentially long cycles will dominate when $e^{aN} > e^{N\Sigma_2}$, which occurs when $\varepsilon_c < \varepsilon < 2 - \varepsilon_c$, with $\varepsilon_c \approx 0.8$.

These predictions are well obeyed by our numerical data; see Figs. 3 and 11. Note, however, the strong finite N effects that show up in the latter figure, which we discuss in the next sections.

B. Cycles with memory ($\alpha < 1$)

When $\alpha < 1$ and $\beta = \infty$, the update of $S_i(t)$ is given by Eq. (14) and has the same fixed points independently of α, but, of course, different limit cycles, which may, in fact, cease to exist when α is small. In this section, we attempt to enumerate the number of cycles of length L in the spirit of the calculation of Hwang *et al.* [73] for $\alpha < 1$. As detailed in Appendix D, we write the number of these cycles as a sum over all possible trajectories of a product of δ functions ensuring the $\alpha < 1$ dynamics of Q_i are satisfied between two consecutive time steps, while a product of Heaviside step functions enforces $S_i(t) = \text{sgn}[Q_i(t)]$. Introducing the integral representation of the δ function, averaging over the disorder, and taking appropriate changes of variable to decouple the N dimensions, the (annealed) complexity of cycles of length L is written

$$\Sigma_L(\alpha, \eta) = \underset{\hat{R}, \hat{K}, \hat{V}}{\text{saddle}} \left\{ \sum_{s < t} i\hat{R}(t, s)i\hat{K}(t, s) \right.$$

$$\left. - \frac{\eta}{2} \sum_{t, s} \hat{V}(t, s)\hat{V}(s, t) + \log \mathcal{I}_L \right\}, \tag{25}$$
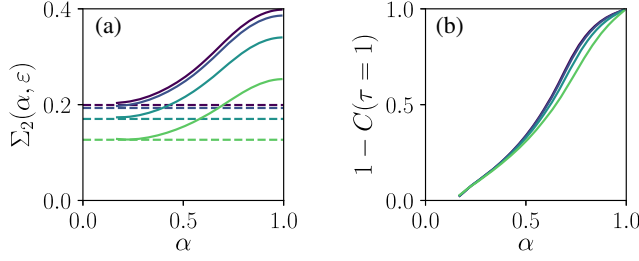
FIG. 10. (a) $L = 2$ cycle complexity from the numerical resolution of the saddle point equations for $\varepsilon = \{0, 0.2, 0.4, 0.6\}$ from dark purple to light green, dashed lines indicating the fixed-point complexity associated to each parameter. (b) Order parameter at the $L = 2$ saddle, showing the nontrivial coalescence of the cycle and fixed-point saddle point as α is decreased. The numerical resolution appears to break down when we get close to $\alpha = 0$.

where $\hat{R}(t, s)$ and $\hat{K}(t, s)$ are symmetric matrices, while $\hat{V}(t, s)$ is not *a priori*, $\mathcal{I}_L$ is explicitly given in Appendix D, and we can notably identify $-i\hat{R}(t, s) = C(|t - s|)$. As a sanity check, one can verify that the $L = 1$ case, corresponding to the fixed-point complexity, is indeed independent of α and is given by the same expression as Eq. (22); see Appendix D 1. In a similar vein, one can recover $\Sigma_2(\alpha = 1) = 2\Sigma_1(\eta)$.

Numerically solving the saddle point equations for decreasing values of α, it appears that $\Sigma_2(\alpha, \eta) \to \Sigma(\eta)$ when $\alpha \to 0^+$; see Fig. 10. While numerical difficulties prevent us from exploring very small values of α, it seems clear that the saddle point corresponding to the $L = 2$ cycles eventually coalesces with the fixed-point saddle (which is known to be a subdominant saddle point when $\alpha = 1$; see Ref. [73] and the discussion above). In any case, and perhaps surprisingly, there does not appear to be a critical value of α below which fixed points become more abundant than cycles. We, therefore, expect a progressive crossover and not a sharp transition.

VI. DYNAMICAL MEAN-FIELD THEORY

We, thus, see that both fixed points and limit cycles are exponentially numerous. However, the question remains as to what happens *dynamically*, as the existence of a large number of fixed points or limit cycles by no means guarantees that these will be reached at long times.

In fact, the number of agents N is expected to play a major role in determining the long-term fate of the system. In particular, there are strong indications that the time τ_r needed to reach a fixed point or a limit cycle grows itself exponentially with N, at least when $\alpha = 1$ [103]. More precisely,

$$\tau_r \sim N^s e^{NB(\varepsilon)}, \tag{26}$$

where s is an exponent (possibly dependent on ε) and $B(\varepsilon)$ an effective barrier such that $B(\varepsilon = 0) = 0$. Hence, one expects that, as N grows, fixed points or limit cycles will, in
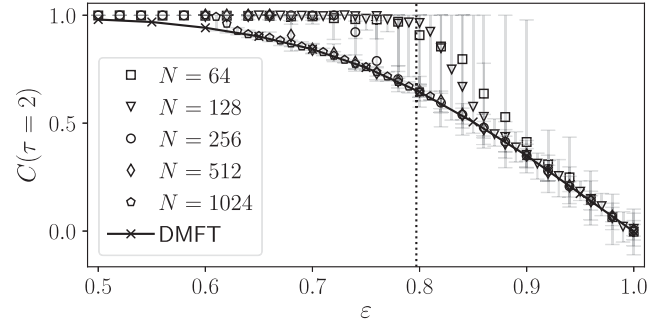

FIG. 11. Steady-state two-point correlation between configurations shifted by $\tau = 2$ time steps in the $\alpha = 1$, $\beta \to \infty$ limit from finite N numerical simulations averaged over 128 samples of disorder and initial conditions, error bars showing 95% confidence intervals. The $N = \{64, 128\}$ simulations are run for $t = 10^8$ time steps to illustrate taking the $t \to \infty$ limit before $N \to \infty$, whereas $N = \{256, 512, 1024\}$ are simulated for $t = 5 \times 10^6$ time steps to recover the $N \to \infty$ before $t \to \infty$ regime. The continuous line represents the $N \to \infty$ DMFT solution integrated numerically, while the vertical dotted line corresponds to the critical value ε_c found by Hwang *et al.* [73].

fact, never be reached, even if they are numerous. This is, in fact, what happens numerically; see Fig. 11.

What is then to be expected in the limit $N \to \infty$? In order to study the complicated learning dynamics that takes place, we resort to DMFT. In a nutshell, DMFT allows deterministic or stochastic dynamics in discrete or continuous time of a large number N of interacting degrees of freedom to be rewritten as a one-dimensional stochastic process with self-consistent conditions. While difficult to solve both analytically and numerically due to their self-consistent nature, DMFT equations have proved very effective at describing a very wide range of complex systems—see Ref. [104] for a recent review. Note, however, that such an approach is valid only when $N \to \infty$; as is clear later, strong finite size effects can appear and change the conclusions obtained using DMFT.

In our case, we write the DMFT for the evolution of the incentives $Q_i(t)$, which directly yield $m_i(t) = \tanh[\beta Q_i(t)]$. In order to do so, we rewrite our online learning process, which depends on the realized $S_i(t)$, as an expression solely in terms of $m_i(t)$ with additional fluctuations:

$$\sum_j J_{ij} S_j(t) = \sum_j J_{ij} m_j(t) + \eta_i(t), \quad \eta_i(t) = \sum_j J_{ij} \xi_i(t), \tag{27}$$

with $\xi_i(t) = S_i(t) - m_i(t)$ and, hence, $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t)\xi_i(s) \rangle = \{1 - [m_i(t)]^2\}\delta_{t,s}$. Now, assuming the central limit theorem holds, the random variables η_i become Gaussian for large N with

$$\langle \eta_i(t) \rangle = 0, \quad \langle \eta_i(t)\eta_j(s) \rangle = v(\varepsilon)[1 - q(t)]\delta_{t,s}\delta_{i,j}, \tag{28}$$

where $q(t) = C(t, t)$ as defined in Eq. (15). As required, in the noiseless limit $\beta \to \infty$ limit, one has $q(t) = 1 \ \forall \ t$ and the random variables η_i are identically zero.

Starting from the N equations

$$Q_i(t + 1) = (1 - \alpha)Q_i(t) + \alpha \sum_j J_{ij} m_j(t)$$
$$+ \alpha \eta_i(t) + \alpha h_i(t), \tag{29}$$

where we reintroduce an arbitrary external field $h_i(t)$ that eventually is set to 0, the DMFT can be derived using path integral techniques or the cavity method, the latter being detailed in Appendix E. Remaining in discrete time to explore the entire range of values of α, one finds, in the $N \to \infty$ limit,

$$Q(t + 1) = (1 - \alpha)Q(t) + \alpha^2(1 - \varepsilon) \sum_{s < t} G(t, s)m(s)$$
$$+ \alpha \phi(t) + \alpha h(t), \tag{30}$$

with $\langle \phi(t) \rangle = 0$, and

$$\langle \phi(t)\phi(s) \rangle = v(\varepsilon)[C(t, s) + [1 - q(t)]\delta_{t,s}]. \tag{31}$$

The memory kernel G and correlation function C are then to be determined self-consistently:

$$G(t, s) = \left\langle \left. \frac{\partial m(t)}{\partial h(s)} \right|_{h=0} \right\rangle, \qquad C(t, s) = \langle m(t)m(s) \rangle, \tag{32}$$

where the averages $\langle \cdots \rangle$ are over the realizations of the random variable ϕ. These discrete time dynamics can first be integrated numerically with an iterative scheme to update both the memory kernel and correlation function until convergence; see Refs. [105,106]. As detailed in the original work of Eissfeller and Opper [107,108], one can also make use of Novikov's theorem to compute the response function with correlations, avoiding the unpleasant task of taking finite differences on noisy trajectories, at the cost of the inversion of the correlation matrix. Note that this inversion, however, means that very long trajectories become difficult to integrate.

While we see that this numerical resolution can provide precious intuition to understand the role of finite N in the dynamics, a continuous description is much more convenient to obtain analytical insights. In the $\alpha \ll 1$, $t \gg 1$ regime, we can rescale the time as $t \to t/\alpha$. Interestingly, doing so requires expanding $Q(t + 1)$ to the second order if one is to keep an explicit dependence on α. The resulting continuous dynamics reads

$$\frac{\alpha}{2}\ddot{Q}(t) + \dot{Q}(t) = -Q(t) + (1 - \varepsilon)\int_0^t dsG(t, s)m(s)$$
$$+ \phi(t) + h(t) \tag{33}$$

with

$$\langle \phi(t)\phi(s) \rangle = v(\varepsilon)[C(t, s) + \alpha[1 - q(t)]\delta(t - s)], \tag{34}$$

and the memory kernel and correlation function are similarly defined self-consistently:

$$G(t, s) = \left\langle \left. \frac{\delta m(t)}{\delta h(s)} \right|_{h=0} \right\rangle, \qquad C(t, s) = \langle m(t)m(s) \rangle, \tag{35}$$

with, we recall, $q(t) = C(t, t) = \langle m^2(t) \rangle$. Very importantly, note that the rescaling in time introduces a prefactor α in the variance of the ϕ, which stems from the noise in the learning process. Since $1 - q(t) \sim \beta^{-2}$ for large β, this extra term is of the order of α/β^2, as anticipated above.

In the next sections, the DMFT equations are used to shed light on the dynamical behavior of the model in the limit $N \to \infty$.

VII. NOISELESS LEARNING

In this section, we use both the DMFT equations and the results on the complexity of fixed points and limit cycles to classify the different dynamical behaviors of the learning process in the noiseless case $\beta \to \infty$, where the realized and expected decisions are equal, $m_i(t) = S_i(t) = \text{sgn}[Q_i(t)]$.

A. The memoryless limit $\alpha = 1$

In this case, corresponding to Eq. (24), both approaches (DMFT and complexity of limit cycles) seem to agree on the overall picture: As ε increases from 0 to 1, the system transitions from $L = 2$ cycles to overdamped oscillations and chaos—see Fig. 3. However, upon scrutiny, one realizes that the perfect agreement between DMFT and direct numerical simulations of the dynamics for finite N is valid only in a region where ε is small and N large enough—see Fig. 11. In particular, when $0.5 \lesssim \varepsilon \lesssim 0.8$, $L = 2$ cycles do persist when N is smaller than approximately 200. For larger N, the lag 2 autocorrelation function $C(\tau = 2)$ is noticeably smaller than unity (consistent with [79]) and well predicted by DMFT as soon as $N \gtrsim 1000$.

What happens for $\varepsilon \lesssim 0.5$ when $N \to \infty$? The numerical solution of the DMFT equations suggest the following scenario: When $\varepsilon < \varepsilon_{\text{RM}} \approx 0.473$, the long-time value m_∞ of the correlation with the initial conditions $C(0, 2n)$ at even time steps is strictly positive, hence the subscript "RM" for remnant magnetization [109]. It is exactly equal to one only for $\varepsilon = 0$ (permanent oscillations) and decreases to reach zero when $\varepsilon \to \varepsilon_{\text{RM}}$ [108,110]. Below ε_{RM}, we can conclude that the system is not ergodic, which has important implications on the finite-temperature dynamics. For asymmetries greater than ε_{RM}, on the other hand, the decorrelation becomes exponential, and we enter a bona fide chaotic, ergodic regime.

Although memoryless learning is clearly unrealistic, these results are rather instructive. The system is indeed unable to display aggregate coordination when interactions are mutually independent (chaotic region around $\varepsilon = 1$). Placing ourselves in the socio-economic setting, it seems evident that the number of players vastly exceeds the number of iterations, and the results above indicate that the chaotic region is, in fact, quite large. Perhaps more importantly, in the absence of learning, agents see their decisions vary at a high frequency without ever reaching any static steady state, not only when the game is close to zero sum ($\varepsilon \to 2$), but even when it is fully reciprocal ($\varepsilon \to 0$). Clearly, this last point underlines the importance of introducing memory to recover realistic learning dynamics.

B. Memory helps convergence to fixed points

For N not too large, we observe numerically that the fraction of "frozen" agents for which $S_i(t+1) = S_i(t)$ quickly tends to 1 as α decreases from 1, as shown in Fig. 12. This is somewhat consistent with intuition, as the learning dynamics average rewards over a period $\tau_\alpha \sim 1/\alpha$, meaning that high-frequency cycles observed for $\alpha = 1$ are expected to be "washed out" when α is sufficiently small. Since fixed points exist in large numbers, it appears natural that they are eventually reached given their abundance at zero temperature. However, as we show in the previous section, $L = 2$ limit cycles are still much more numerous than fixed points for $\alpha \gtrsim 0.5$. The fact that $C(\tau = 1)$ approaches unity as α is reduced much faster than in Fig. 10(b) suggests that the basin of attraction of fixed points quickly expands, at the expense of $L = 2$ limit cycles [111].

Our numerical results, therefore, indicate that, for any finite size system which has enough time to reach a steady
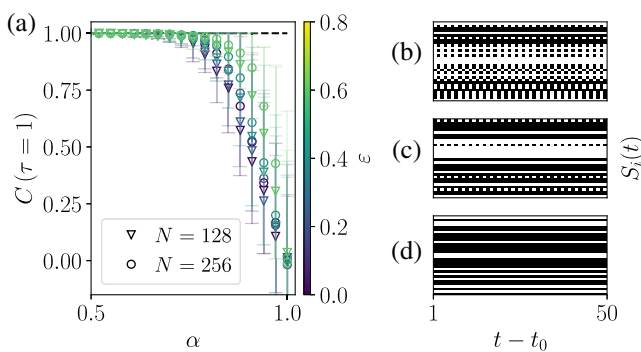

FIG. 12. Convergence to fixed points with decreasing α and for different values of $\varepsilon < 0.8$, $\beta \to \infty$, and finite N. (a) Steady-state two-point correlation function between successive configurations from finite N numerical simulations averaged over 200 samples of disorder and initial conditions, error bars showing 95% confidence intervals. (b)–(d) Sample trajectories of 32 randomly chosen sites among $N = 256$ for $\varepsilon = 0.4$, $t_0 = 10^5/\alpha$, for $\alpha = \{1.0, 0.88, 0.7\}$, respectively.
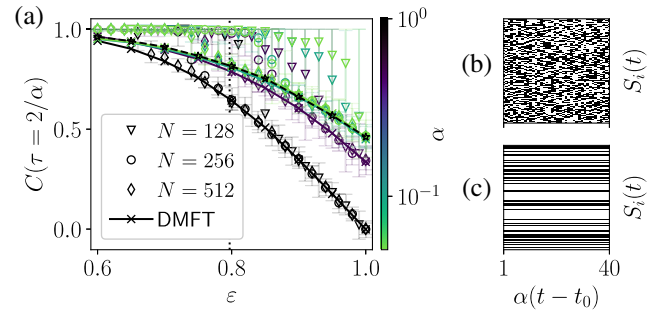

FIG. 13. Influence of finite size N, nonreciprocity ε, and memory span α on the learning dynamics. (a) Steady-state two-point correlation function shifted by $\tau = 2/\alpha$ in the $\beta \to \infty$ limit for different memory loss rates α, from light green to black (color map on the right axis). Symbols correspond to direct simulations and plain lines to the solutions of the DMFT equations, while the dashed line is the solution to the DMFT equations for $\alpha \to 0$. The $N = 128$ simulations are initialized with $t_0 = 10^8$ time steps, whereas $N = \{256, 512\}$ are simulated for $t_0 = 10^6$ iterations before taking measurements. Results are averaged over 32 samples, with error bars showing 95% confidence intervals. (b),(c) Sample trajectories for all $N = 128$ sites for $\varepsilon = 1$ and $\alpha = \{1.0, 0.1\}$, respectively, clearly reaching chaotic and fixed-point steady states.

state, the effect of α is effectively to help the system find fixed points—see Fig. 13. Focusing, for example, on the points corresponding to $N = 128$ and $\alpha = 0.1$, we indeed observe that the 95% quantile includes $C(2/\alpha) = 1$ even for $\varepsilon = 1$; i.e., fixed points can be reached even in the chaotic regime with modest simulation times, which would be an overwhelmingly improbable scenario in the memoryless case, as illustrated by Figs. 13(b) and 13(c).

As the number of agents N increases, we enter the DMFT regime shown as plain lines in Fig. 13. One finds that decreasing the value of α slows down the decorrelation of the system. However, for small α, the evolution becomes a function of $\alpha\tau$ only, as suggested by Eq. (33) when $\alpha \to 0$: The dynamical slowdown is dominated by the long memory of learning itself.

Figure 13 shows that, when $\varepsilon \gtrsim 0.5$, sufficiently large systems (described by DMFT) decorrelate with time for all α, and we expect $C(\tau \to \infty) \to 0$: Learning leads to chaos in such cases.

When $\varepsilon \lesssim 0.5$, on the other hand, we find that there is ergodicity breaking, in the sense that $C(\tau \to \infty) > 0$, as we find above for cycles when $\alpha = 1$. More precisely, a numerical analysis of the DMFT equations suggests that, when $\alpha \to 0$ and ε small, $1 - C(\tau \to \infty)$ is extremely small but nonzero. For example, when $\varepsilon = 0.4$ we find $1 - C(\tau \to \infty) \approx 0.005$. This is compatible with the numerical results in Ref. [108].

In other words, there seems to exist a critical value $\varepsilon_{\mathrm{RM}}(\alpha)$ separating the ergodic, chaotic phase for $\varepsilon > \varepsilon_{\mathrm{RM}}(\alpha)$ from the nonergodic, quasi-fixed-point behavior for $\varepsilon < \varepsilon_{\mathrm{RM}}(\alpha)$.

However, our numerical results are not precise enough to ascertain the dependence of ε_{RM} on α, which seems to hover around the value 0.473 found for $\alpha = 1$. More work on this specific point would be needed to understand such a weak dependence on the memory length.

The precise dynamical behavior of the autocorrelation function $C(\tau)$ can be ascertained in the continuous limit $\alpha \to 0$ when $\varepsilon = 1$. Indeed, the influence of the memory kernel vanishes in this case where interactions are exactly nonsymmetric, leaving us with

$$\dot{Q}(t) = -Q(t) + \phi(t) \quad (\alpha \to 0), \tag{36}$$

where we emphasize that the time variable has been rescaled as $t \to \alpha t$. From there, the classical solution method proposed by Crisanti and Sompolinsky [112,113] can be straightforwardly adapted with a small modification due to our parametrization of the interaction matrix that scales the variance of the entries by a factor $1/2$ for $\varepsilon = 1$; see Appendix F. The two-point autocorrelation function is found to be given by

$$C(\tau) = \frac{2}{\pi} \sin^{-1}\left(\frac{\Delta(\tau)}{\Delta(0)}\right), \tag{37}$$

where $\Delta(\tau) = \langle Q(t + \tau)Q(t)\rangle$ follows the second-order ordinary differential equation

$$\ddot{\Delta}(\tau) = \Delta(\tau) - \frac{1}{2}C(\tau), \tag{38}$$

with $\Delta(0) = 1 - (2/\pi)$ [114]. Very quickly, this means that the autocorrelation decays exponentially, $C(\tau) \propto e^{-\tau/\tau_1}$ with

$$\tau_1 = \sqrt{\frac{\pi - 2}{\pi - 3}} \approx 2.84. \tag{39}$$

Both the full solution, obtained by integrating the ordinary differential equation (ODE) numerically, as well as this exponential decay are shown in Fig. 14, displaying a very satisfactory match with numerical simulations.

C. Anomalous stretching of cycles

Having a better understanding of how the memory allows the system to find fixed points when they exist, a central question is what happens if the memory loss rate is reduced in the region of parameter space where there are only limit cycles. As previously stated, averaging over a period $\tau_\alpha \sim 1/\alpha$ clearly suggests that the occurrence of short cycles (starting at $L = 4$ for $\alpha = 1$) should gradually vanish.

Naively, one might expect a simple rescaling in time $t \to t/\alpha$, yielding cycles—when they exist—of period inversely proportional to α itself. Looking at the numerical
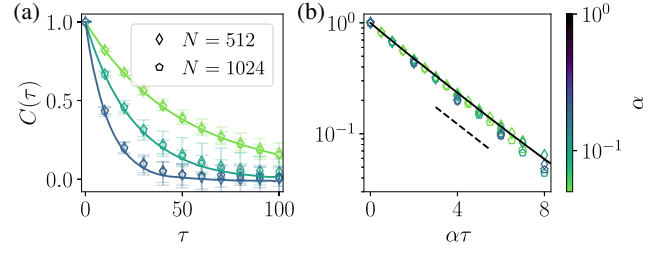

FIG. 14. Evolution of the time shifted autocorrelation function in the nonsymmetric case $\varepsilon = 1$, $\beta \to \infty$, for different system sizes and memory loss parameters averaged over 20 realizations. Left: lin-lin scale, error bars showing 95% confidence intervals and continuous lines representing the numerically integrated full DMFT equations [Eq. (36)]. Right: lin-log scale and rescaling of the time shift by α such that points collapse onto a single curve (error bars not shown), the black continuous line representing the analytical solution found by solving the Sompolinsky and Crisanti ODE Eq. (38) and the dashed line representing a pure exponential decay with characteristic time τ_1.

results from both the finite size game and the DMFT integrated numerically in Fig. 15(a), it quickly appears that such a simple rescaling in time does not provide the correct description. Indeed, the period of cycles is observed to be
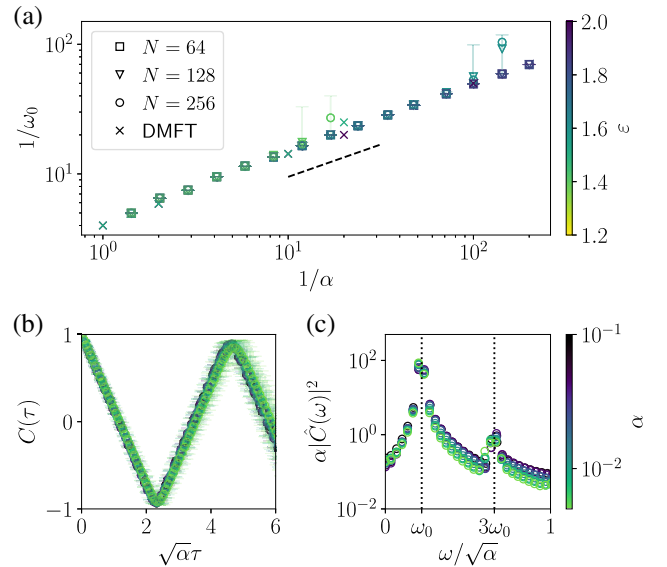

FIG. 15. Evolution of the oscillation frequency ω_0 with the memory loss rate α for $\beta \to \infty$, $\varepsilon > 2 - \varepsilon_c$ averaged over 96 samples of disorder and initial conditions, error bars showing 95% confidence interval. (a) Log-log plot highlighting the near square root dependency $\omega_0 \sim \sqrt{\alpha}$ the dashed line corresponding to an exponent $1/2$. (b) Two-point autocorrelation for $\varepsilon = 1.8$, $N = 256$ as a function of the rescaled time lag for different values of α, confirming the suitability of the scaling, in particular, at short timescales. (c) Power spectrum of the autocorrelation for the same parameters, displaying the maximum at ω_0 and secondary peaks at odd multiples of this fundamental frequency as expected from the triangular aspect of the autocorrelation.

proportional to $1/\sqrt{\alpha}$, i.e., much shorter than $1/\alpha$—see Figs. 15(b) and 15(c).

One important aspect to note is that there is some decorrelation, as the second peak of $C(\tau)$ does not quite reach unity [in Fig. 15(b)], meaning that we may see quasicycles and not exact limit cycles, complicating the analytical description of the phenomenon. Just as true fixed points in the $N \to \infty$ limit exist only when $\varepsilon = 0$, it appears that only the case $\varepsilon = 2$ does display true limit cycles.

Another subtle point to consider is that, similar to the $\varepsilon < 1$ cases discussed, we expect the time taken to reach these cycles to depend on the system size and the relative distance to the chaotic region. This is confirmed by the DMFT solved for fixed trajectory times for $\varepsilon = 1.5$ (light crosses), which progressively departs from the $\omega_0 \sim \sqrt{\alpha}$ regime around $\alpha = 0.1$.

To understand how such nontrivial stretching occurs, we go back to the continuous DMFT equation

$$\frac{\alpha}{2}\ddot{Q}(t) = -\dot{Q}(t) - Q(t) + (1-\varepsilon)\int_0^t ds\, G(t,s)m(s)$$
$$+ \phi(t) + h(t).$$

While the presence of the second-order derivative $\ddot{Q}(t)$ appears natural to recover limit cycles, it should be noted that this term, being prefactored by α, is superficially subdominant relative to the dissipation represented by $\dot{Q}(t)$. While we have seen that there is some decorrelation, the fact that robust oscillations are present, therefore, suggests that the complicated self-consistent forcing terms almost exactly compensate dissipation over a period, allowing the system to periodically revisit quasi-identical configurations. In fact, the shape of these oscillations is far from sinusoidal but rather of seesaw type; see Fig. 15(b). This suggests that, in the limit $\alpha \to 0$, $\ddot{Q}(t)$ diverges each time $\dot{C}(\tau)$ changes sign, such that $(\alpha/2)\ddot{Q}(t)$ cannot be neglected and, therefore, sets the relevant timescale to $\alpha^{-1/2}$. We have, however, not been able to perform a more precise singular perturbation analysis of this phenomenon.

Going beyond this rather loose argument and precisely characterizing such seesaw patterns appears very challenging and is left for future work. A possible approach would be to first take the $\varepsilon = 2$ case where true cycles should exist and to assume the correlation function is an exact triangular wave of frequency ω as suggested by Fig. 15(c). As a result, $m(s) = \text{sgn}[Q(s)]$ is an exact square wave, and the convolution with G can be written as a product in Fourier space. Enforcing the dissipation over a period to be zero, one could then perhaps find a closed equation for Q and ω if appropriate *Ansätze* for the response and forcing functions are taken.

VIII. NOISY LEARNING

While we have shown that the $\beta \to \infty$ deterministic limit can be relatively well understood with the analytical tools at our disposal, one of the key features of our model is the uncertainty in the decision occurring for boundedly rational agents. Besides, it is also in this situation that the online learning dynamics differ significantly from the more widely studied offline learning where the entire model can be understood in terms of deterministic mixed strategies parametrized by the coefficients $m_i(t)$ [compare Eqs. (5) and (7)].

When α is close to unity and β becomes small, the fluctuations are too large for coordination to occur. Taking, for instance, $\alpha = 1$, it is indeed clear that the iteration

$$m_i(t+1) = \tanh\left(\beta \sum_J J_{ij} S_j(t)\right)$$

will have an extremely large fluctuation in the argument on the right-hand side. As a result, we expect to lose the sharp transition as a function of T that can be observed for the NMFE (see Fig. 16). The order parameter q instead continuously tends to 0 with T, regardless of the asymmetry ε. This regime is shown in Fig. 16(a), representing the heat map of q for $\alpha = 0.5$. Clearly, the linear stability analysis of the paramagnetic fixed point presented in Sec. IV cannot hold when the thermal fluctuations are not averaged on large periods of time. To find a richer phenomenology, we, therefore, focus on the $\alpha \ll 1$ regime, where more complex dynamics can be observed.

A. Fixed points for intentions

In Sec. IV, we study the fixed points of the NMFE that the game may reach if the fluctuations from imperfect learning can be neglected, i.e., if $\alpha \to 0$. Now, we have seen that the DMFT equations proved effective in the zero-"temperature" limit and, importantly, established that true fixed points or two-cycles are reached only in finite time in the fully reciprocal case $\varepsilon = 0$ when the system size
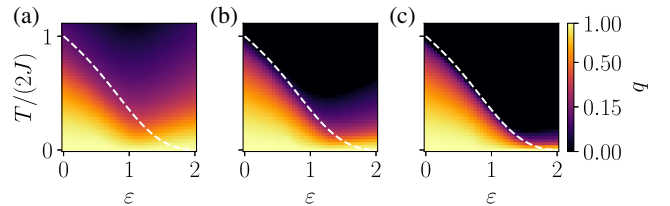

FIG. 16. Heat map of $q = C(0)$ in $(T = 1/\beta, \varepsilon)$ space from numerical simulations for $N = 256$, $t_0 = 10^6$ averaged over 32 samples of disorder and initial conditions and (a)–(c) corresponding to $\alpha = \{0.5, 0.1, 0.01\}$, respectively. The white dashed line represents the critical temperature $T_c(\varepsilon)$ where the paramagnetic solution ($q = 0$) becomes linearly unstable (Sec. IV).

diverges (see Fig. 11). The same equations can also be used to revisit this finite β regime.

Going back to Eq. (33) and neglecting the term in $\sqrt{\alpha}$ from the correlation function as we do in the static setup, fixed-point solutions require

$$Q = (1 - \varepsilon)m\chi + J\sqrt{qv(\varepsilon)}z, \tag{40}$$

where z is now a *static* white noise of unit variance, q is simply the now constant autocorrelation, and χ is the integrated response function that we assume to be time-translation invariant:

$$\chi = \int_0^\infty d\tau G(\tau). \tag{41}$$

The averages on the effective process can now be taken on z to self-consistently solve for q and χ (see, e.g., Ref. [61] for a more detailed description). The resulting set of equations are then

$$q = \langle m^2(z) \rangle_z, \tag{42}$$

to be solved simultaneously with

$$\chi = \left\langle \frac{\beta[1 - m^2(z)]}{1 - \beta(1 - \varepsilon)\chi[1 - m^2(z)]} \right\rangle_z, \tag{43}$$

where $m(z)$ is the solution to

$$m(z) = \tanh\left[\beta(1 - \varepsilon)\chi m(z) + \beta\sqrt{qv(\varepsilon)}z\right]. \tag{44}$$

Although our model is entirely built on a dynamical evolution equation and not on a notion of thermal equilibrium, this set of self-consistent equations coincides with the replica-symmetric solution of the NMFE model found by Bray, Sompolinsky, and Yu [92] for $\varepsilon = 0$. Since replica symmetry is broken in the whole low-temperature phase of the NMFE model, we expect that these static solutions of the DMFT cannot correctly describe the long-time limit of the dynamics, as we now show.

The numerical solutions for the DMFT fixed-point equations are shown in Fig. 17(a) and compared to numerical results of the game for small α and for $\varepsilon = 0.1$ (similar results, not shown, are obtained for other values of $\varepsilon \lesssim 0.8$) [115]. We find that the long-time behaviors of q for the direct simulation of the SK game (circles and squares) and for long-time dynamical solution of the DMFT equations match very well but differ from the value of q inferred from the set of self-consistent equations established above. This is expected, since with such solution the order parameter q approaches unity exponentially fast as $T \to 0$, whereas the fact that the probability of small local fields (i.e., rewards in the game analogy) vanish linearly (see Fig. 4) suggests that $q = 1 - \kappa T^2$, as for the full replica
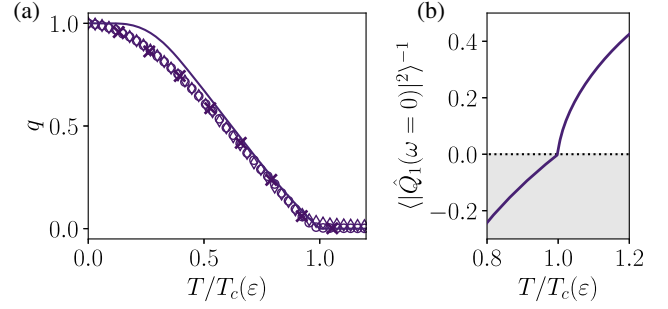

FIG. 17. (a) Order parameter $q = C(t, t)$ averaged in time in the (quasi)stationary regime versus rescaled temperature for $\varepsilon = 0.1$. Circular and diamond markers correspond direct, finite size simulations at $N = 256$ for $\alpha = 0.01$ and $\alpha = 0.1$, respectively, whereas crosses represent the (dynamical) numerical solution to the complete set of $N \to \infty$ DMFT equations for $\alpha = 0.1$. Continuous lines show the solution to the static DMFT fixed-point equations [Eqs. (42)–(44)] for $\varepsilon = 0.1$. (b) Spectral density of a small perturbation Q_1 to the fixed-point solution of the DMFT close to the critical temperature. As the quantity is necessarily positive for a valid solution, the gray region corresponds to instability.

symmetry breaking solution of [92] but with presumably a different value of κ.

To ascertain the range over which this nontrivial mean-field solution should be valid, we can study the stability of the DMFT fixed point close to the critical temperature $1/\beta_c$, following the procedure first detailed in Ref. [60]. Considering a random perturbation to the fixed point $\epsilon\xi(t)$, with $\xi(t)$ a Gaussian white noise and $\epsilon \ll 1$, we study the perturbed solution

$$Q(t) = Q_0 + \epsilon Q_1(t), \tag{45}$$

with Q_0 the fixed point given in Eq. (40), where the noise is no longer static but similarly given by $\phi(t) = \sqrt{qv(\varepsilon)}z + \epsilon\phi_1(t)$. Replacing in the DMFT continuous dynamics for $\alpha \to 0^+$ and collecting terms of the order of ϵ, we find that the perturbation evolves as

$$\dot{Q}_1(t) = -Q_1(t) + \beta(1 - \varepsilon)[1 - m^2(z)]\int_0^t G(t, s)Q_1(s)$$
$$+ \phi_1(t) + \xi(t), \tag{46}$$

where we use $\text{sech}^2(\beta Q_0) = 1 - m^2(z)$ from Eq. (44), giving in Fourier space

$$\hat{Q}_1(\omega) = \frac{\hat{\phi}_1(\omega) + \hat{\xi}(\omega)}{i\omega + 1 - \beta(1 - \varepsilon)[1 - m^2(z)]\hat{G}(\omega)}, \tag{47}$$

where we again assume that the memory kernel is time-translation invariant.

Now, in the limit $\beta Q_1 \ll 1$, i.e., close to the critical temperature, one can we linearize the hyperbolic tangent

tanh$[\beta Q_1(t)]$ and write a closed equation for the spectral density of Q_1 at the order of $\beta^2 Q_1^2$:

$$\frac{1}{\langle |\hat{Q}_1(\omega)|^2 \rangle} = |i\omega + 1 - \beta(1-\varepsilon)\hat{G}(\omega)|^2 - \beta^2 v(\varepsilon). \quad (48)$$

As a result, we have the criterion for the onset of instability for $\omega = 0$:

$$(1-\varepsilon)\chi = 1 - \beta\sqrt{v(\varepsilon)}, \quad (49)$$

where we notice $\hat{G}(\omega = 0) = \chi$, given, close to β_c, by

$$\chi = \frac{1}{2\beta(1-\varepsilon)}\left(1 - \sqrt{1 - 4\beta^2(1-\varepsilon)}\right). \quad (50)$$

Taking $\varepsilon = 0$, we recover the criterion found by Bray, Sompolinsky, and Yu [92] for the critical temperature, giving $T_c = 1/\beta_c = 2$ in their case.

For nonzero ε, we can also find the critical temperature by replacing Eq. (50) in Eq. (49), to recover yet again the critical value given by Eq. (20). The invalidity of the fixed-point solution is illustrated in Fig. 17(b), where the spectral density evaluated at $\omega = 0$ can be observed to become negative for $T < T_c(\varepsilon)$. While the linearization used to obtain the relation is expected to become invalid, we understand this negativity as a strong sign that the solution is unstable and, thus, likely invalid throughout the low-"temperature" phase, consistently with the discrepancy observed between the static prediction and the direct simulations and the dynamic solution of the DMFT shown in Fig. 17(a). In this finite β regime, we therefore effectively have only what we previously referred to as quasifixed points at best (including when $\varepsilon = 0$), even when neglecting the vanishing fluctuations caused by the online learning dynamics as $\alpha \to 0$.

B. Aging

The inadequacy of the static solution of the DMFT equations to describe the long-term dynamics of the system is a well-known symptom associated with the "aging" phenomenon [116], i.e., the fact that equilibrium is never reached and all correlation functions depend on the "age" of the system [117]:

$$C(t_w, t_w + t) = C_{\text{relax}}(t) + C_{\text{aging}}(t, t_w), \quad (51)$$

where t_w is the waiting time, or age of the system, $t_w = 0$ corresponding to a random initial condition.

Aging typically arises in complex systems in the low-temperature (low-noise) limit. Pictorially, the energy landscape of such systems (like spin glasses) are highly nonconvex and "rugged," with a very large number of local minima or quasistable saddle points in which the dynamics gets stuck for extended periods of time [117,118].

The consequence is then that the time required to exit a local minimum is a function of the age of the system, i.e., the time taken for the system to reach this configuration. Intuitively, the deeper in the energy landscape the solution is, the longer it takes for a sufficiently large random fluctuation to occur and allow the system to resume its exploration of the landscape. Such aging phenomena are known to occur in a wide range of complex systems with reciprocal interactions, such as glassy systems, populations dynamics [119], or neural networks that are described by very similar mean-field dynamics [120]. Aging dynamics was also recently found in a "habit formation" model; see Ref. [11].

Not surprisingly, in view of its similarity with usual spin glasses, the SK game displays aging for reciprocal interactions ($\varepsilon = 0$) and sufficiently low temperatures $\beta > \beta_c$; see Fig. 18(a), for which Eq. (51) accurately describes the data with the initial relaxation component $C_{\text{relax}}(t)$ well fitted by a power law t^{-x} and

$$C_{\text{aging}}(t, t_w) = \mathcal{C}\left(\frac{t}{t_w}\right), \quad (52)$$

corresponding to the aging behavior found in a wide range of glassy models [117,121–126]. Interestingly, this is not the behavior observed in the "physical" SK model [127,128], which is typically found to display "subaging," although the precise scaling of the aging correlation function remains unclear. It is, however, important to emphasize that the learning dynamics of the SK game is markedly different from the physical dynamics of the original SK model, so there is *a priori* no reason to expect their complicated out-of-equilibrium relaxation to be directly comparable.

What happens in our model when $\varepsilon > 0$? It is known from previous work (in somewhat different contexts) that aging is interrupted at long times in the presence of nonreciprocal interactions but survives for finite times provided the asymmetry is not too large [129–131]. If the asymmetry strength is further increased, we expect the amount of mixing in the system to eventually be large enough for the dynamics to no longer get stuck [132]. From our numerical simulations, shown in Figs. 18(b)–18(d), it appears that aging [as described by Eq. (51)] still holds when $\varepsilon < \varepsilon^\star$ but that the dynamics becomes time-translation invariant when $\varepsilon > \varepsilon^\star$. It is tempting to conjecture that aging disappears exactly when the dynamics becomes ergodic, i.e., when the correlation with a random initial condition decays to zero. This suggests that $\varepsilon^\star = \varepsilon_{\text{RM}}$, which is roughly in line with our numerical data. Note, however, that the transition between aging dynamics and time-translation-invariant correlations seems to occur somewhat progressively; hence, it may well be that we, in fact, observe interrupted aging beyond a time that decreases
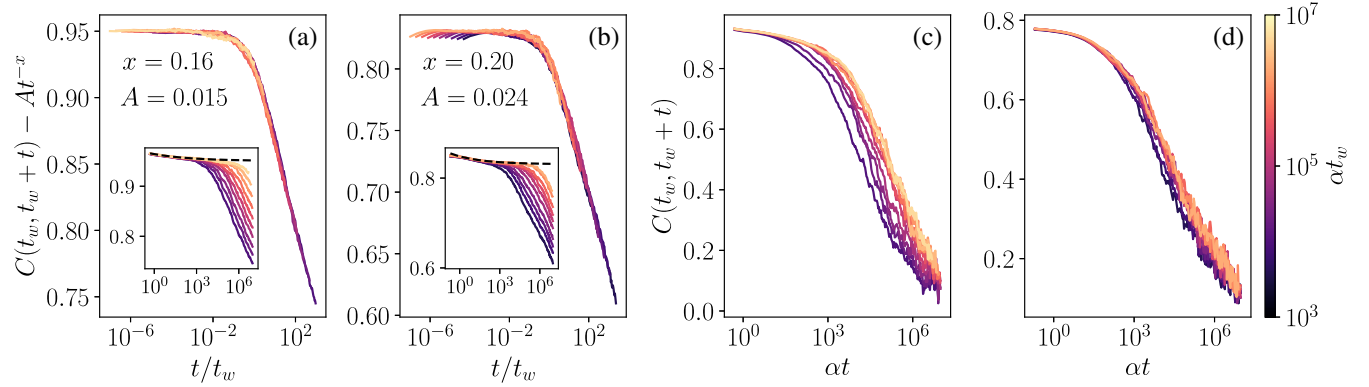
FIG. 18.   Aging behavior of the system for $N = 256$, averages performed over 576 realizations. Color indicates the value of αt_w; see the scale on the far right. (a),(b) Aging two-point correlation functions with the initial power law decay removed to isolate the aging component plotted as a function of t/t_w, the inset showing the entire correlation function as a function of αt and the dashed line representing the power law fit At^{-x} of the first relaxation. (a) $\alpha = 0.5$, $\beta = 4$, and $\varepsilon = 0$; (b) $\alpha = 0.2$, $\beta = 2$, and $\varepsilon = 0.25$. (c),(d) Partially aging two-point correlation functions plotted as a function of αt. (c) $\alpha = 0.5$, $\beta = 4$, and $\varepsilon = 0.5$; (d) $\alpha = 0.2$, $\beta = 2$, and $\varepsilon = 0.5$.

not only with ε, but also with temperature $1/\beta$. More analytical work is needed to clarify the situation.

A particularity of the aging dynamics of the SK game is related to its online learning dynamics. As clearly visible in both the DMFT equation and the derivation of the naive mean-field equation detailed in Appendix C, there will inevitably be some decorrelation in time of the *expected* decisions if $\alpha(1 - q)$ becomes significant. We, therefore, naturally expect the region where time-translation invariance breaks down to be dependent on all three parameters α, β, and ε.

The interpretation of aging in the socio-economic context is quite interesting and is discussed in Sec. III C. In a nutshell, it means that, as time goes on, agents get stuck in locally satisficing strategies for longer and longer, but, after a time proportional to the total time the game has already been played, the system eventually evolves and individual strategies m_i reach an altogether different configuration. This process goes on forever but becomes slower and slower with time: The notion of quasiequilibrium, therefore, makes sense at long times, for small enough noise and small enough nonreciprocity.

C. Chaos and (quasi)limit cycles

When the nonreciprocity of interactions is sufficiently small and quasifixed points exist, we have established that boundedly rational systems display complicated aging dynamics when learning noise, parametrized by the value of β, is present. The immediate question is now how such noise influences the complex dynamics, chaos, and (quasi) limit cycles that we have found in the $\beta \to \infty$, $\alpha \ll 1$ regime (see Sec. VII).

To qualitatively illustrate the effect of nonzero noise, we run simulations for $\alpha = 0.1$ and different values of ε and β. Figure 19 displays individual trajectories of the intentions

$m_i(t)$, as well as the distribution of the values of individual m_i over all agents, realizations, and time steps. We also show the autocorrelation function $C(\tau)$ (assumed to be time-translation invariant over the short timescales considered), which is also compared to the DMFT solved numerically. Note that, for the smallest value $\varepsilon = 0.1$, the trajectories illustrate the previously discussed quasifixed points emerging from the online dynamics. Clearly, while the correlation function remains close to constant, individual intentions are not exactly frozen (notice the wiggles in the top row in Fig. 19, especially for $\beta = 2$), explaining how the system as a whole eventually decorrelates and displays aging, as discussed in the previous section.

In the chaotic regime around $\varepsilon = 1$, it is clear that decreasing β (increasing noise) spreads the distribution of the m_i, which is less and less concentrated around ± 1 as an immediate consequence of the smoothed out hyperbolic tangent. As a result, the equal-time autocorrelation $C(0) = q$ naturally decreases when β decreases. It is furthermore interesting to note that its value significantly decreases as the asymmetry parameter ε increases. We expect a decrease in α to have a similar role, as suggested by the phase diagrams presented in Figs. 16(b) and 16(c).

Dynamically, the decay of the autocorrelation in this chaotic regime appears to be more or less independent of the strength of the noise, which can be seen by comparing the insets in the third row in Fig. 19. While it is known that external noise kills deterministic chaos in neural networks with uncorrelated couplings [133], what is interesting in our case is that both the nonlinearity of the hyperbolic tangent (governed by β) and the strength of the effective noise [which scales as $1 - q$; see Eq. (34)] are varied simultaneously and, in a sense, self-consistently.

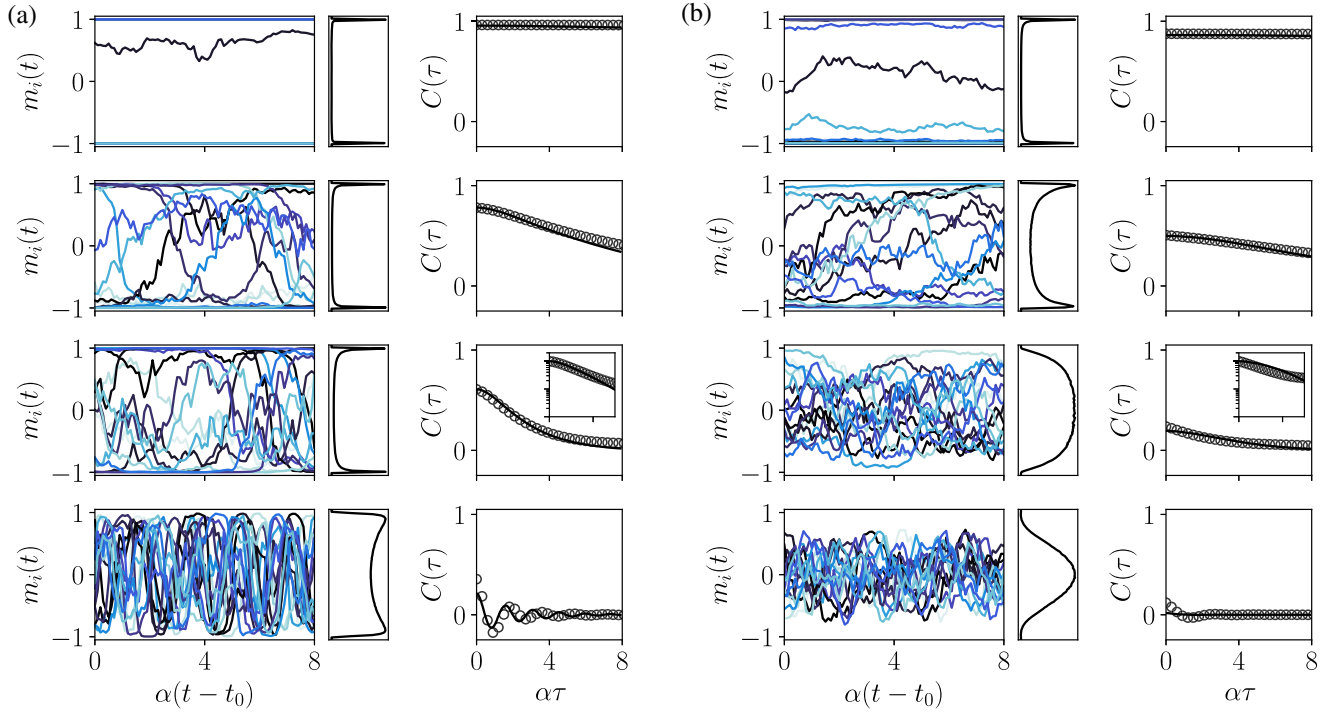FIG. 19. Finite β trajectories obtained from numerical simulations at $\alpha = 0.1$, $N = 256$, and $t_0 = 10^8$ averaged over 96 samples of disorder and initial conditions for (a) $\beta = 4$ and (b) $\beta = 2$ and from top to bottom $\varepsilon = \{0.1, 0.85, 1.05, 1.5\}$. Each line displays (from left to right) the evolution of 16 randomly selected agents, the associated histogram of m_i over both agents, time, and realizations and the autocorrelation function assumed to be time-translation invariant on short timescales. Third row: insets representing the evolution of the normalized autocorrelation $C(\tau)/C(0)$ with a logarithmic vertical scale.

Determining the way the decorrelation rate evolves with both β and ε is, therefore, quite nontrivial and would be a very interesting endeavor.

Where we previously had limit cycles, for $\varepsilon = 1.5$, for instance (last row in Fig. 19), it appears that oscillations survive for large values of β. Note, however, that in this case the value $q = C(0)$ appears to very quickly vanish when β decreases. This is consistent with the linear stability analysis of the paramagnetic solution that should be valid for vanishingly small α. As visible in Fig. 16(c), we indeed expect that the region in which the system displays any form of aggregate coordination becomes increasingly narrow as ε gets closer to its maximum value of 2. Precisely for $\varepsilon = 2$, the system likely becomes fully disordered ($q = 0$) for any finite values of β when $\alpha \to 0$. For small but finite values of α as those presented here, this is not quite the case, however, and large asymmetries $\varepsilon > 2 - \varepsilon_c$ do give rise to clear oscillations, in both individual trajectories and the correlation function (for instance, Fig. 19, bottom row, shows for $\varepsilon = 1.5$ that some oscillations can be somewhat sustained).

D. Role of the noise: Recap

In the presence of noise, the "conviction" of agents naturally goes down, in the sense that individual m_i's become smaller in absolute value, i.e., less polarized

around ± 1. For small α (long memory time), there exists a well-defined transition line in the plane ε (asymmetry), $1/\beta$ (amplitude of the noise) above which agents start playing randomly at each round (i.e., $m_i = 0$) but below which some instantaneous propensity to overplay $+1$ or -1 appears. The time evolution of this propensity depends on the value of ε: For small asymmetries, the fixed points that are reached in the absence of noise become quasifixed points around which the system settles for longer and longer periods of time, after eventually moving on to a completely different configuration (aging). For $\varepsilon \sim 1$ (uncorrelated influence from i to j and from j to i), deterministic chaos when noise is absent becomes noisy chaos, with not much change. In the highly competitive region $\varepsilon \to 2$, periodic cycles progressively become overdamped, as expected, since noise does not allow synchronization to survive at long times. In terms of individual rewards, not surprisingly, noise tends to be detrimental, except in the reciprocal region (ε small), where weak noise actually *helps* agents coordinating around mutually beneficial actions.

While the precise characterization of this very rich ecology of dynamical behaviors is left for future work, we emphasize that the decorrelation induced by the fact that incentives Q_i are themselves random variables is a key difference between the online learning presented here and

their offline counterpart that are often considered [13,53]. While the phenomenology between these two types of learning is somewhat similar for $\alpha \ll 1$, $\beta \gg 1$, there are clearly key differences whenever the ratio $\sqrt{\alpha}/\beta$ ceases to be negligible.

IX. SUMMARY AND DISCUSSION

A. Blind sided by complexity

Let us summarize our main conceptual assertions. As a schematic model of the complexity economic agents are confronted with, we introduced the SK game, a discrete time binary choice model with N interacting agents and three parameters: α (memory loss rate), β (inverse amplitude of noise in the learning process or *intensity of choice*), and ε (nonreciprocity of interactions).

We have shown that, even in a completely static environment where the payoff matrix does not evolve, agents are unable to learn collectively optimal strategies. This is either because the learning process gets trapped by a suboptimal fixed point (or remains around one for very long times) or because learning never converges and leads to a never-ending (chaotic or quasiperiodic) evolution of agent intentions.

Hence, contrarily to the hope that learning might save the "rational expectation" framework [8], which still holds the upper hand in macroeconomics textbooks, we argue that complex situations are generically unlearnable, as first suggested by Galla and J. Farmer [13]. Faced with complexity, therefore, agents must do with satisficing solutions—an idea proposed long ago by Simon [1] and made precise and tangible for the first time within our model specification.

Only a centralized, omniscient agent may be able to ascribe an optimal strategy to all agents—which incidentally raises the question of trust: Would agents even agree to follow the central planner advice? Would they even believe in her ability to solve complex problems, knowing that their solution sensitively depends on all parameters of the model? If a finite fraction of all agents fail to comply, the resulting average reward will drop precipitously below the optimal value and not be much better than the result obtained through individual learning.

As general ideas of interest in a socio-economic context, we have established that
  (1) long memory of past rewards is beneficial to learning, whereas overreaction to recent past is detrimental;
  (2) increased competition generically destabilizes fixed points and leads first to chaos and, in the high competition limit, to quasicycles;
  (3) some amount of noise in the learning process, quite paradoxically, allows the system to reach better collective decisions, in the sense that the average reward is increased;

  (4) nonergodic behavior spontaneously appears (in the form of aging) in a large swath of parameter space, when α, $1/\beta$, and ε are small.

On the positive side, we have shown that learning is far from useless: Instead of getting stuck among one of the most numerous fixed points with low average reward, the learning process does allow the system to coordinate around satisficing solutions with rather high (but not optimal) average reward. Numerically, the average reward at the end of the learning process is, for $\varepsilon = 0$, approximately 8% below the optimal value, when the majority of fixed points lead to a much worse average reward approximately 33% below the optimal value [72].

B. Technical results and conjectures

From a statistical mechanics perspective, our model is next of kin to but different from several well-studied models; a synthesis of our original results can be found in Figs. 1, 2, 9, and 16.

For example, when $\alpha = 1$, $\beta \to \infty$, the dynamics is equivalent to a Hopfield model of learning with nonsymmetric Gaussian synaptic couplings, for which many results are known, in particular, on the number of fixed points and L-cycles. Introducing some memory with $\alpha < 1$, we found that previously dynamically unattainable fixed points become typical solutions, replacing the short limit cycles in which the $\alpha = 1$ parallel dynamics get stuck. We also showed how the number of L-cycles can be calculated for all values of $\alpha < 1$.

The chaotic region that is known to exist when interactions are mostly nonsymmetric ($\varepsilon \approx 1$) also appears to be reduced by memory. When couplings are mostly nonreciprocal $\varepsilon \lesssim 2$, periodic oscillations survive but we found that decreasing α nontrivially increases the cycle length, as $\alpha^{-1/2}$.

When β is finite, the fixed-point solutions to the dynamics correspond to the so-called naive mean-field equation of spin glasses, another model that has been studied in detail [92]. One knows, in particular, that such solutions become exponentially abundant for small enough noise $1/\beta$ and for $\varepsilon = 0$, a result that we have extended to all $\varepsilon < 1$. When $\varepsilon > 1$, on the other hand, the only fixed point (or Nash equilibrium) is $m_i = 0$, $\forall\ i$, i.e., completely random decisions at each time step.

The DMFT is a tool of choice for investigating the dynamics of the model when $N \to \infty$. DMFT is, however, frustratingly difficult to exploit analytically in the general case, so we are left with numerical solutions of our DMFT equations that accurately match direct numerical simulations of the model when N is large but fail to capture some specific features arising when N is small. From our numerical results, we conjectured that quasifixed points (and, correspondingly, aging dynamics) persist for small noise and when $\varepsilon \leq \varepsilon^\star$, where the value of $\varepsilon^\star$ is difficult to

ascertain but could be as high as $\varepsilon_{RM} = 0.47$, perhaps related to the remnant magnetization transition found in Ref. [108].

For $\beta = \infty$, the long-time, zero-temperature autocorrelation $C(\infty)$ appears to drop extremely slowly with ε, but we have not been able to get an analytical result. DMFT equations also clearly lead to the anomalous $\alpha^{-1/2}$ stretching of the cycles mentioned above, but an analytic solution again eluded us.

One of the reasons analytical progress with DMFT is difficult is presumably related to the phenomenon of replica symmetry breaking and its avatar in the present context of dynamical learning. Indeed, any attempt to expand around a static solution of the DMFT equations leads to inconsistencies in the interesting situation $\beta > \beta_c(\varepsilon)$ when decisions are not purely random; see Fig. 17(b). In fact, as shown in Fig. 17(a), the value of the order parameter q predicted by such static solutions is substantially off the value found from the long-time, numerical solution of the DMFT equations, which itself coincides with direct simulations of the SK game. *En passant*, we noticed that the value of q seems to be given by a universal function of $\beta_c(\varepsilon)/\beta$, independently of the value of ε. Again, we have not been able to understand why this should be the case.

Finally, we have numerically established several interesting results concerning average and individual rewards that would deserve further investigations. For example, the average reward seems to converge toward its asymptotic value as $N^{-2/3}$, exactly as for the SK model, although, as already noted above, this asymptotic value is approximately 8% below the optimal SK value. Is it possible to characterize more precisely the ensemble of configurations reached after learning in the long memory limit $\alpha \to 0$? Can one, in particular, understand analytically the distribution of individual rewards shown in Fig. 4 and the corresponding asymptotic value of the average reward, as well as its nonmonotonic behavior as a function of the noise parameter β? These are, in our opinion, very interesting theoretical questions that one should solve in order to establish a *statistical mechanics of satisficing solutions*, which would be needed to understand the emergent properties of large assemblies of learning agents.

C. Extensions and final remarks

Many extensions of the very simple framework presented here can be imagined for the model to be more representative of real socio-economic systems. For example, by analogy with spin glasses, going beyond the fully connected interactions and toward a more realistic network structure should not change the overall phenomenology, although some subtle differences may show up. While analytical predictions become even more challenging, recent works on dynamical mean-field theories with finite connectivity, so far developed for Lotka-Volterra-type

systems, could perhaps be adapted to the learning dynamics of our model. Nonetheless, the preliminary simulations on sparse random graphs presented in Appendix B suggest that most of the salient results of our paper carry through to such situations. Having broadly distributed bonds J_{ij} could also be a valuable extension to our model (as mentioned in Refs. [49,77]), which we leave for future work.

Allowing the interaction network to evolve with time would, of course, also make the model more realistic, as in Ref. [12]; in this case, one would have to distinguish the case where the learning time is much longer or much shorter than the reshuffling time of the network. In the latter case, it interestingly appears feasible to devise an adaptive DMFT to describe the system on two separated timescales [134].

Other interesting additions to Ising games could also include the introduction of self-excitation [135] or of alternative decision rules that might be less statistical mechanics-friendly [42]. Extension to multinary decisions [33], beyond the binary case considered here, as well as higher-order interactions (i.e., a "p-spin game"; see also Ref. [136] for recent ideas), would obviously be interesting as well, especially as higher-order interactions are known to change the phenomenology of the SK model (see, e.g., Ref. [117] and references therein). In particular, we expect that in the p-spin case with $p \geq 3$ a much larger gap would develop between the optimal average reward and the one reached by learning.

Finally, whereas temperature in physics is the same for all spins, there is no reason to believe that the amount of noise β or the memory span α should be the same for all agents. Introducing such heterogeneities might be worth exploring, as some agents with longer memory may fare systematically better than others, like in minority games; see Ref. [40].

Beyond the socio-economic context that was our initial motivation for its design, we believe that the simplicity and generality of our model makes it a suitable candidate to describe a much wider range of complex systems. In the context of biological neural networks, the parameter α indeed allows one to interpolate between simple discrete-time Hopfield network [102] and continuous-time models where Q_i is an activation variable for the firing rates m_i [137–142]. Although in our case the influence of β introduces some perhaps unwanted stochasticity, these fluctuations can, in principle, be suppressed (at least partially) with sufficiently small α. The memory loss parameter could also represent an interesting way to tune the effective slowing down of the dynamics caused by symmetry and described in Ref. [120]. Here, the description of real neural networks would likely require much more sparse interactions but also perhaps the introduction of some dedicated dynamics for the interactions themselves; see, e.g., Ref. [143] for recent ideas.

Closer to our original motivation, more applied socio-economic problems might benefit from the introduction of this type of reinforcement learning. In macroeconomics, for instance, some form of "habit formation" could perhaps be relevant to extend existing descriptions of input-output networks [12,23], where client-supplier relationships are probably strongly affected by history (on this point, see Kirman's classic study on Marseilles' fish market [144]; see also Ref. [32]). Finally, while it is an aspect of our model we have not investigated here, previous works have reported that similar dynamics yield interesting volatility clusters and heavy tails, corresponding to sudden changes of quasifixed points. Such effects might be relevant to describe financial time series, as noted in Refs. [13,33].

In the context of financial markets, our model could also be used to challenge the idea that efficiency is reached by evolution. Indeed, a theory that has been proposed to explain empirical observations going against the efficient market hypothesis is the so-called "adaptive markets hypothesis," stating that inefficiencies stem from the (transient) evolution of a market toward true efficiency [145,146]. However, reinterpreting the Q_i in our model as some form of fitness measure, it appears unlikely that simple evolutionary dynamics (akin to the learning dynamics) could overcome the type of radical complexity discussed here. As a matter of fact, such an evolutionary twist to the SK game could also be used to conjecture that simple Darwinian dynamics is unlikely to lead to a global optimum in a complex biological setting [147].

Last but not least, we believe that the learning dynamics presented here may be useful from a purely algorithmic point of view in the study of spin glasses and so-called TAP states. Indeed, in the $\beta \to \infty$ limit, we have seen that our iteration relatively frequently finds fixed points in regions where their abundance is known to be subexponential (close to $\varepsilon = 1$, in particular) and this even for relatively large values of α. Interestingly, similar exponentially weighted moving averages have been employed in past numerical studies of TAP states for symmetric interactions [148,149], but on the magnetizations m_i themselves and not on the local fields Q_i like is the case above.

Using an offline version of our learning procedure could then be of use to effectively converge to fixed points of the TAP equations or naive mean-field equations and study their properties. Perhaps even more interestingly, the online dynamics and the resulting fluctuations of the m_i themselves could prove to be extremely valuable to probe hardly accessible regions of the solution space. In some sense, the fluctuations related to finite values of α and β could allow one to define "meta-TAP" states, in the sense of closely related TAP states mutually accessible thanks to such extra fluctuations, in the same spirit as standard Langevin dynamics in an energy landscape.

Finally, as mentioned in Sec. III G, in the zero-temperature limit and for $\varepsilon = 1$, it has recently been reported that for a certain range of self-interaction strengths J_d, there appears an exponential number of accessible solutions to the TAP equations that are seemingly not reachable with standard Hopfield dynamics [90]. Preliminary numerical experiments seem to suggest that our learning dynamics find such fixed points, as suggested by their effectiveness for $\varepsilon = 1$ without any form of self-interaction. Beyond existing interest around neural networks, such a self-reinforcement, habit formation term could also be, as stated above, interesting to study from the socio-economic perspective [9,11].

APPENDIX A: NORMAL-FORM REPRESENTATION OF TWO-PLAYER INTERACTIONS

In the main text, we define the rewards stemming from the interactions of player i with all other players as a single term. In order to make contact with the classical game-theoretic approach, one may also describe the SK game by writing the payoffs resulting from each two-player interaction in the "normal" or "strategic" form [150]. To do so, one simply decomposes the rewards $R_i^\pm$ as the sum of $N - 1$ terms corresponding to the outcome of each of the pairwise subgames being performed simultaneously by the agents. The result for $N = 2$ is shown in Table I.

TABLE I. Normal-form game representation of the payoffs in a two-player SK game. Note that the time dependency as well as the $1/2$ factor acting on all entries are omitted for readability.

	$S_2 = +1$	$S_2 = -1$
$S_1 = +1$	$(J_{12} + h_1, J_{21} + h_2)$	$(-J_{12} + h_1, -J_{21} - h_2)$
$S_1 = -1$	$(-J_{12} - h_1, -J_{21} + h_2)$	$(J_{12} - h_1, J_{21} - h_2)$

APPENDIX B: NUMERICAL EXPERIMENTS ON SPARSE RANDOM REGULAR GRAPHS

We test the validity of some of our results on sparser and, therefore, perhaps more realistic interaction matrices. We resort to random regular graphs with fixed connectivity $z = 4$, for which the strength of interactions is still sampled from a Gaussian distribution and where ε identically controls the nonsymmetry of the matrix. (Note that the standard deviation of the Gaussian now scales as $z^{-1/2}$ rather than $N^{-1/2}$.)

In such sparse graphs, the exact value of the optimal excess reward $\mathcal{R}_\infty$ is not known. However, it may be approximated using the zero-temperature cavity method developed by Mézard and Parisi [151] and applied to this specific problem in Ref. [152]. While this "belief propagation" (BP) algorithm is not strictly correct, in principle, as it corresponds to a replica-symmetric or one-step replica symmetry-breaking solution, it has been found to provide a very precise estimate for the ground state energy. In Ref. [152], the authors find $\mathcal{R}_\infty \approx 1.351$ for fully reciprocal interactions in the replica-symmetric approximation.

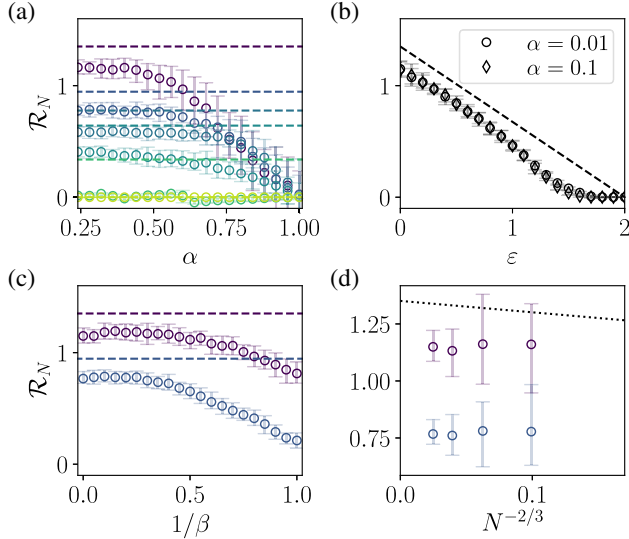Comparing this approximate $N \to \infty$ global optimum with the outcome of numerical simulations of the game as

FIG. 20. Reproduction of Fig. 2 for a random regular graph with fixed connectivity $z = 4$, dashed lines showing the approximate $N \to \infty$ solution obtained by belief propagation in Ref. [152]. Parameters: (a) $\varepsilon = \{0, 0.6, 0.85, 1.05, 1.5, 2\}$ from dark purple to light green, $\beta \to \infty$; (b) $\alpha = 0.01$ and $\alpha = 0.1$, $\beta \to \infty$; (c) $\alpha = 0.01$, $\varepsilon = \{0, 0.6\}$ (dark purple and blue, respectively); (d) $\alpha = 0.01$, $\beta \to \infty$, $\varepsilon = \{0, 0.6\}$ (dark purple and blue, respectively). In (a)–(c), $N = 256$. The dotted line in (d) is approximately reproduced from the numerical experiments in Ref. [152], while the finite size effects in our case are too faint to be fitted with reasonable confidence.
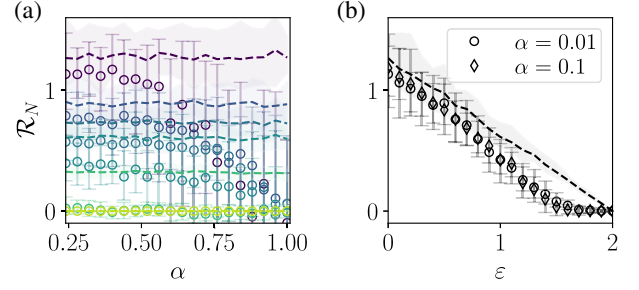

FIG. 21. Reproduction of Figs. 2(a) and 2(b) for a random regular graph with fixed connectivity $z = 4$ in small systems with $N = 24$, dashed lines showing the average reward at the global optimum computed by exact enumeration of all 2^N configurations (shaded areas showing 95% confidence intervals).

shown in Fig. 20, we recover the fact that agents may learn only satisficing outcomes. Similarly to the measurements in Ref. [152], we find minimal finite size corrections, meaning that the difference between the $N \to \infty$ ground state and the learned solutions may, in fact, be more sizable than in the fully connected case for diverging system sizes. To ensure that this may not be a result of the replica-symmetric assumption taken in the BP algorithm leading to an overestimate of the optimum, we also perform numerical experiments on small systems ($N = 24$), where the best outcome for each instance of the interaction matrix can be found by enumerating *all* 2^N solution vectors. The admittedly noisier results are shown in Fig. 21 and appear to be compatible with the same range of suboptimality, thereby also demonstrating the validity of our conclusions in relatively small systems.

As a side note, it is interesting to notice that learning the satisficing configurations in this sparser context requires a somewhat longer memory than in the fully connected case ($\alpha \lesssim 0.4$ in Fig. 20 versus $\alpha \lesssim 0.8$ in Fig. 2).

Albeit preliminary, these numerical experiments strongly support the idea that most of the results presented in the main text are highly universal and likely survive to a very wide range of adjustments to the model. We leave the perhaps more quantitatively dissimilar case of broadly distributed bonds for future work.

APPENDIX C: STATIC NMFE

We claim that for $\alpha \ll 1$

$$\tilde{m}_j^\alpha(t) \simeq \alpha \sum_{t'=1}^{t} (1-\alpha)^{t-t'} m_j(t'). \tag{C1}$$

Indeed, given the assumption of independence in time, i.e., $\langle (S_j(t') - m_j(t'))(S_j(t'') - m_j(t'')) \rangle = \delta_{t,t'}(1 - m_j(t)^2)$, we have

$$\left\langle \left(\tilde{m}_j^\alpha(t) - \alpha \sum_{t' \le t} (1-\alpha)^{t-t'} m_j(t') \right)^2 \right\rangle$$

$$= \alpha^2 \sum_{t' \le t} (1-\alpha)^{2(t-t')} \left\langle (S_j(t') - m_j(t'))^2 \right\rangle + \alpha^2 \sum_{t' \le t} \sum_{t'' \ne t'} (1-\alpha)^{t-t'} (1-\alpha)^{t-t''} \left\langle (S_j(t') - m_j(t'))(S_j(t'') - m_j(t'')) \right\rangle$$

$$= \alpha^2 \sum_{t' \le t} (1-\alpha)^{2(t-t')} (1 - (m_j(t'))^2)$$

$$\le \alpha^2 \sum_{t' \le t} (1-\alpha)^{2(t-t')} = \frac{\alpha}{2-\alpha} \xrightarrow[\alpha \to 0]{} 0. \tag{C2}$$

We can then make the *Ansatz* that the expected decision reaches a fixed point $m_j^\star$ after some time. For sufficiently large t and small but finite values of α, we therefore have

$$\langle \tilde{m}_j^\alpha(t) \rangle \simeq m_j^\star \tag{C3}$$

with fluctuations characterized by

$$\langle (\tilde{m}_j^\alpha(t) - m_j^\star)^2 \rangle = \frac{\alpha}{2} (1 - (m_j^\star)^2) + O(\alpha^2). \tag{C4}$$

APPENDIX D: LIMIT CYCLE COMPLEXITY WITH MEMORY

To study the influence of the memory loss rate $\alpha < 1$, we may adapt the method of Hwang *et al.*, although this requires the introduction of either a strong nonlinearity in the exponent and subsequent saddle equations or new variables. We opt for the latter and write the number of cycles of length L as

$$\mathcal{N}_L(N, \alpha, \varepsilon) = \sum_{\{S_i(t)\}} \int_{-\infty}^{\infty} \left(\prod_{i=1}^{N} \prod_{t=1}^{L} dQ_i(t) \right) |\det \mathcal{J}^\alpha|^N \prod_{i,t} \delta \left(Q_i(t+1) - (1-\alpha)Q_i(t) - \alpha \sum_j J_{ij} S_j(t) \right) \Theta[Q_i(t) S_i(t)], \tag{D1}$$

where the Dirac δ ensures that the dynamics are satisfied at each step while the second enforces $S_i(t) = \text{sgn}[Q_i(t)] \ \forall \ t$. The $L \times L$ matrix $\mathcal{J}^\alpha$ is the α-dependent Jacobian ensuring that the zeros of the δ function are correctly weighted, i.e., for $L > 1$,

$$\mathcal{J}_{ts}^\alpha = \begin{cases} -(1-\alpha), & \text{if } s = t \ (\text{mod } L), \\ 1, & \text{if } s = t+1 \ (\text{mod } L), \\ 0 & \text{otherwise.} \end{cases} \tag{D2}$$

The first step is, as usual, to perform the average over the disorder after introducing the integral representation of the Dirac δ:

$$\overline{\mathcal{N}_L}(N, \alpha, \varepsilon) = |\det \mathcal{J}^\alpha|^N \sum_{\{S_i(t)\}} \int_{-\infty}^{\infty} \left(\prod_{i=1}^{N} \prod_{t=1}^{L} dQ_i(t) \frac{d\lambda_i(t)}{2\pi} \right) \left(\prod_{i,t} \Theta[S_i(t) Q_i(t)] \right) \exp \left[-i \sum_{i,t} \lambda_i(t) Q_i(t+1) \right.$$

$$+ i(1-\alpha) \sum_{i,t} \lambda_i(t) Q_i(t) - \frac{1}{2N} \alpha^2 \left(1 - \frac{\varepsilon}{2} \right)^2 \sum_{i<j} \left(\sum_t [\lambda_i(t) S_j(t) + \lambda_j(t) S_i(t)] \right)^2$$

$$\left. - \frac{1}{2N} \alpha^2 \left(\frac{\varepsilon}{2} \right)^2 \sum_{i<j} \left(\sum_t [\lambda_i(t) S_j(t) - \lambda_j(t) S_i(t)] \right)^2 \right]. \tag{D3}$$

The last two terms, resulting from the average on disorder, may be rearranged to give

$$-\frac{\alpha^2}{2N} \sum_{t,s} \left(\sum_{i,j} [v(\varepsilon) \lambda_i(t) S_j(t) \lambda_i(s) S_j(s) + (1-\varepsilon) \lambda_i(t) S_j(t) \lambda_j(s) S_i(s)] - \frac{1}{2} (\varepsilon - 2)^2 \sum_i \lambda_i(t) S_i(t) \lambda_i(s) S_i(s) \right).$$

Similar to Ref. [73], we introduce a set of auxiliary functions:

$$U(t,s) = \frac{1}{N}\sum_i \lambda_i(t)S_i(t)\lambda_i(s)S_i(s), \qquad \text{(D4)}$$

$$V(t,s) = \frac{1}{N}\sum_i \lambda_i(t)S_i(s), \qquad \text{(D5)}$$

$$R(t,s) = \frac{1}{N}\sum_i \lambda_i(t)\lambda_i(s), \qquad \text{(D6)}$$

$$K(t,s) = \frac{1}{N}\sum_i S_i(t)S_i(s), \qquad \text{(D7)}$$

such that the last term gives

$$-\frac{1}{2}N\alpha^2 \sum_{t,s}\left[v(\varepsilon)R(t,s)K(t,s) + (1-\varepsilon)V(t,s)V(s,t) - \frac{(2-\varepsilon)^2}{2N}U(t,s) \right].$$

In the limit $N \to \infty$, the $O(N^{-1})$ term is insignificant and can, thus, be neglected. The complete expression is then

$$\overline{\mathcal{N}_L}(N,\alpha,\varepsilon) = |\det \mathcal{J}^\alpha|^N \sum_{\{S_i(t)\}} \int_{-\infty}^{\infty} \left(\prod_{i,t} dQ_i(t)\frac{d\lambda_i(t)}{2\pi} \right)\left(\prod_{i,t}\Theta[S_i(t)Q_i(t)] \right)$$

$$\times \left(\prod_{t,s} dK(t,s)\frac{d\hat{K}(t,s)}{2\pi}dR(t,s)\frac{d\hat{R}(t,s)}{2\pi}dV(t,s)\frac{d\hat{V}(t,s)}{2\pi} \right)$$

$$\times \exp\left[-i\sum_{i,t}\lambda_i(t)Q_i(t+1) + i(1-\alpha)\sum_{i,t}\lambda_i(t)Q_i(t) - \frac{1}{2}N\alpha^2\sum_{t,s}[v(\varepsilon)R(t,s)K(t,s) + (1-\varepsilon)V(t,s)V(s,t)] \right.$$

$$- i\sum_{t,s}\hat{K}(t,s)\left(NK(t,s) - \sum_i S_i(t)S_i(s) \right) - i\sum_{t,s}\hat{R}(t,s)\left(NR(t,s) - \sum_i \lambda_i(t)\lambda_i(s) \right)$$

$$\left. - i\sum_{t,s}\hat{V}(t,s)\left(NV(t,s) - \sum_i \lambda_i(t)S_i(s) \right) \right]. \qquad \text{(D8)}$$

Now, one can notice that

$$\int_{-\infty}^{\infty}\left(\prod_{t,s}\frac{dR(t,s)}{2\pi} \right)e^{-\sum_{t,s} iR(t,s)[N\hat{R}(t,s) - \frac{1}{2}iN\alpha^2 v(\varepsilon)K(t,s)]} = \prod_{t,s}\delta\left(N\hat{R}(t,s) - \frac{1}{2}iN\alpha^2 v(\varepsilon)K(t,s) \right), \qquad \text{(D9)}$$

which means the integral over $K(t,s)$ now reads

$$\int_{-\infty}^{\infty}\left(\prod_{t,s}\frac{dR(t,s)}{2\pi} \right)\prod_{t,s}e^{-iN\hat{R}(t,s)R(t,s)}\delta\left(N\hat{R}(t,s) - \frac{1}{2}iN\alpha^2 v(\varepsilon)K(t,s) \right) = e^{-\frac{2N}{\alpha^2 v(\varepsilon)}\sum_{t,s}\hat{R}(t,s)\hat{K}(t,s)}. \qquad \text{(D10)}$$

For $V(t,s)$, we have to be a bit more careful, as the expression is not linear for all terms, as the diagonal for $t = s$ gives a quadratic term that has to be treated separately. Noticing that the product $V(t,s)V(s,t)$ is symmetric, we start by considering the $t < s$:

$$\int_{-\infty}^{\infty}\left(\prod_{t<s}\frac{dV(t,s)}{2\pi} \right)e^{-\sum_{t<s} iV(t,s)[N\hat{V}(t,s) - iN\alpha^2(1-\varepsilon)V(s,t)]} = \prod_{t<s}\delta(N\hat{V}(t,s) - iN\alpha^2(1-\varepsilon)V(s,t)). \qquad \text{(D11)}$$

Now integrating over the $V(t,s)$ for $t > s$,

$$\int_{-\infty}^{\infty}\left(\prod_{t>s}\frac{dV(t,s)}{2\pi} \right)\prod_{t>s}e^{-iN\hat{V}(t,s)V(t,s)}\delta(-i\alpha^2(1-\varepsilon)V(s,t) + 2\hat{V}(t,s)) = e^{-\frac{N}{\alpha^2(1-\varepsilon)}\sum_{t>s}\hat{V}(t,s)\hat{V}(s,t)}. \qquad \text{(D12)}$$

Finally, the diagonal $V(t, t)$ for which the expression is quadratic can be computed with a Gaussian integral:

$$\int_{-\infty}^{\infty}\left(\prod_t dV(t,t)\right) e^{-\frac{1}{2}N\alpha^2(1-\varepsilon)\sum_t V(t,t)^2 - iN\sum_t V(t,t)\hat{V}(t,t)} \sim e^{-\frac{N}{2\alpha^2(1-\varepsilon)}\sum_t \hat{V}(t,t)^2}. \tag{D13}$$

Up to an $O(1)$ multiplicative constant, the complete expression now reads

$$\overline{\mathcal{N}_L}(N,\alpha,\varepsilon) \sim |\det \mathcal{J}^\alpha|^N \sum_{\{S_i(t)\}} \int_{-\infty}^{\infty}\left(\prod_{i,t} dQ_i(t)\frac{d\lambda_i(t)}{2\pi}\right)\left(\prod_{i,t}\Theta[S_i(t)Q_i(t)]\right)\left(\prod_{t,s} d\hat{K}(t,s)d\hat{R}(t,s)d\hat{V}(t,s)\right)$$

$$\times \exp\left(-i\sum_{i,t}\lambda_i(t)Q_i(t+1) + i(1-\alpha)\sum_{i,t}\lambda_i(t)Q_i(t) + i\sum_{i,t}\nu_i(t)Q_i(t)S_i(t)\right.$$

$$-\frac{2N}{\alpha^2 v(\varepsilon)}\sum_{t,s}\hat{R}(t,s)\hat{K}(t,s) - \frac{N}{2\alpha^2(1-\varepsilon)}\sum_{t,s}\hat{V}(t,s)\hat{V}(s,t)$$

$$+ i\sum_{t,s}\hat{K}(t,s)\sum_i S_i(t)S_i(s) + i\sum_{t,s}\hat{R}(t,s)\sum_i\lambda_i(t)\lambda_i(s) + i\sum_{t,s}\hat{V}(t,s)\sum_i\lambda_i(t)S_i(s)\right). \tag{D14}$$

Taking the changes of variable

$$\hat{R}(t,s) \to \frac{1}{2}\alpha^2 v(\varepsilon)\hat{R}(t,s),$$

$$\hat{K}(t,s) \to \frac{1}{2}\alpha^2 v(\varepsilon)\hat{K}(t,s),$$

$$\hat{V}(t,s) \to \alpha^2(1-\varepsilon)\hat{V}(t,s) \tag{D15}$$

for convenience, we can entirely factorize the problem in N, finally giving

$$\overline{\mathcal{N}_L}(N,\alpha,\varepsilon) \sim \int_{-\infty}^{\infty}\left(\prod_{t,s} d\hat{K}(t,s)d\hat{R}(t,s)d\hat{V}(t,s)\right)\exp\left(N\left[-\frac{1}{2}\alpha^2 v(\varepsilon)\sum_{t,s}\hat{R}(t,s)\hat{K}(t,s)\right.\right.$$

$$\left.\left. -\frac{1}{2}\alpha^2(1-\varepsilon)\sum_{t,s}\hat{V}(t,s)\hat{V}(s,t) + \log\mathcal{I}_L + \log|\det\mathcal{J}^\alpha|\right]\right), \tag{D16}$$

where

$$\mathcal{I}_L = \sum_{\{S(t)\}}\int_{-\infty}^{\infty}\left(\prod_t dQ(t)\frac{d\lambda(t)}{2\pi}\right)\left(\prod_t\Theta[S(t)Q(t)]\right)\exp\left[-i\sum_t\lambda(t)Q(t+1) + i(1-\alpha)\sum_t\lambda(t)Q(t)\right.$$

$$\left. +\frac{1}{2}\alpha^2 v(\varepsilon)\sum_{t,s}[S(t)S(s)i\hat{K}(t,s) + \lambda(t)\lambda(s)i\hat{R}(t,s)] + \alpha^2(1-\varepsilon)\sum_{t,s}\lambda(t)S(s)i\hat{V}(t,s)\right]. \tag{D17}$$

At this stage, we may notice that $S(t)^2 = 1$ $\forall$ t and as such that the diagonal part of the sum over $K(t,s)$ can be taken out of $\mathcal{I}_L$. Doing so and combining this contribution to the first term of the exponent in Eq. (D16), we have

$$\overline{\mathcal{N}_L}(N,\alpha,\varepsilon) \sim \int_{-\infty}^{\infty}\left(\prod_{t,s} d\hat{K}(t,s)d\hat{R}(t,s)d\hat{V}(t,s)\right)\exp\left(N\left[\frac{1}{2}\alpha^2 v(\varepsilon)\sum_t[i\hat{R}(t,t)+1]i\hat{K}(t,t)\right.\right.$$

$$\left.\left. +\frac{1}{2}\alpha^2\left(1-\varepsilon+\frac{\varepsilon^2}{2}\right)\sum_{s\neq t}i\hat{R}(t,s)i\hat{K}(t,s) - \frac{1}{2}\alpha^2(1-\varepsilon)\sum_{t,s}\hat{V}(t,s)\hat{V}(s,t) + \log\mathcal{I}_L + \log|\det\mathcal{J}^\alpha|\right]\right). \tag{D18}$$

Now, the first term in the exponent can be integrated over the $\hat{K}(t,t)$ exactly, yielding a product of δ functions fixing

$$i\hat{R}(t,t) = -1 \quad \forall \ t, \tag{D19}$$

as expected, and introducing only a subdominant correction $O(\log N/N)$ in the complexity.

In the $N \to \infty$ limit, the complexity is then finally given:

$$\Sigma_L(\alpha, \varepsilon) = \operatorname*{saddle}_{\hat{R},\hat{K},\hat{V}} \left\{ \frac{1}{2}\alpha^2 v(\varepsilon)\sum_{s\neq t} i\hat{R}(t,s)i\hat{K}(t,s) - \frac{1}{2}\alpha^2(1-\varepsilon)\sum_{t,s} \hat{V}(t,s)\hat{V}(s,t) + \log \mathcal{I}_L + \log|\det \mathcal{J}^\alpha| \right\}, \tag{D20}$$

with now

$$\mathcal{I}_L = \sum_{\{S(t)\}} \int_{-\infty}^{\infty} \left(\prod_t dQ(t)\frac{d\lambda(t)}{2\pi}\right)\prod_t \Theta[S(t)Q(t)]\exp\left[-i\sum_t \lambda(t)Q(t+1) + i(1-\alpha)\sum_t \lambda(t)Q(t)\right.$$
$$\left. - \frac{1}{2}\alpha^2 v(\varepsilon)\left[\sum_t \lambda(t)^2 - \sum_{s\neq t}[S(t)S(s)i\hat{K}(t,s) - \lambda(t)\lambda(s)i\hat{R}(t,s)]\right] + \alpha^2(1-\varepsilon)\sum_{t,s}\lambda(t)S(s)i\hat{V}(t,s)\right]. \tag{D21}$$

This integral may be rewritten as

$$\mathcal{I}_L = \sum_{\{S(t)\}} e^{\frac{1}{2}\alpha^2 v(\varepsilon)\sum_{s\neq t} S(t)S(s)i\hat{K}(t,s)} \int_{-\infty}^{\infty} \left(\prod_t \frac{dQ(t)}{\sqrt{2\pi}}\frac{d\lambda(t)}{\sqrt{2\pi}}\right)\prod_t \Theta[S(t)Q(t)]\exp\left(-\frac{1}{2}\boldsymbol{\lambda}^\top \mathbf{A}\boldsymbol{\lambda} - i\mathbf{b}^\top \boldsymbol{\lambda}\right) \tag{D22}$$

with the $L \times L$ matrix $\mathbf{A}$ constituted of

$$A(t,s) = \alpha^2 v(\varepsilon)[\delta_{ts} - (1-\delta_{ts})i\hat{R}(t,s)], \tag{D23}$$

and, $\mathbf{b} \in \mathbb{R}^L$,

$$b(t) = Q(t+1) - (1-\alpha)Q(t) - \alpha^2(1-\varepsilon)\sum_s \hat{V}(t,s)S(s), \tag{D24}$$

such that we, in fact, have

$$\mathbf{b} = \mathcal{J}^\alpha \mathbf{Q} + \mathbf{c}, \qquad c(t) = -\alpha^2(1-\varepsilon)\sum_s \hat{V}(t,s)S(s). \tag{D25}$$

Computing the Gaussian integral on λ,

$$\mathcal{I}_L = \sum_{\{S(t)\}} \frac{e^{\frac{1}{2}\alpha^2 v(\varepsilon)\sum_{s\neq t} S(t)S(s)i\hat{K}(t,s)}}{\sqrt{\det \mathbf{A}}} \int_{-\infty}^{\infty} \left(\prod_t \frac{dQ(t)}{\sqrt{2\pi}}\right)\prod_t \Theta[S(t)Q(t)]\exp\left(-\frac{1}{2}(\mathcal{J}^\alpha \mathbf{Q}+\mathbf{c})^\top \mathbf{A}^{-1}(\mathcal{J}^\alpha \mathbf{Q}+\mathbf{c})\right). \tag{D26}$$

Taking the change of variable $\mathbf{u} = \mathcal{J}^\alpha \mathbf{Q} + \mathbf{c}$, the above becomes

$$\mathcal{I}_L = \sum_{\{S(t)\}} \frac{e^{\frac{1}{2}\alpha^2 v(\varepsilon)\sum_{s\neq t} S(t)S(s)i\hat{K}(t,s)}}{|\det \mathcal{J}^\alpha|\sqrt{\det \tilde{\mathbf{A}}}} \int_{-\infty}^{\infty} \left(\prod_t \frac{dQ(t)}{\sqrt{2\pi}}\right)\prod_t \Theta\{S(t)[\mathbf{Q}-(\mathcal{J}^\alpha)^{-1}\mathbf{c}](t)\}\exp\left(-\frac{1}{2}\mathbf{Q}^\top \tilde{\mathbf{A}}^{-1}\mathbf{Q}\right), \tag{D27}$$

where $\tilde{\mathbf{A}} = (\mathcal{J}^\alpha)^{-1}\mathbf{A}(\mathcal{J}^\alpha)^{-1}$. As a result, the $|\det \mathcal{J}^\alpha|$ contributions in the complexity cancel out, and we find ourselves with

$$\Sigma_L(\alpha, \varepsilon) = \operatorname*{saddle}_{\hat{R},\hat{K},\hat{V}} \left\{ \frac{1}{2}\alpha^2 v(\varepsilon)\sum_{s\neq t} i\hat{R}(t,s)i\hat{K}(t,s) - \frac{1}{2}\alpha^2(1-\varepsilon)\sum_{t,s} \hat{V}(t,s)\hat{V}(s,t) + \log \mathcal{I}_L \right\}, \tag{D28}$$

with now

$$\mathcal{I}_L = \sum_{\{S(t)\}} \frac{e^{\frac{1}{2}\alpha^2 v(\varepsilon)\sum_{s\neq t} S(t)S(s)i\hat{K}(t,s)}}{\sqrt{\det \tilde{\mathbf{A}}}} \int_{-\infty}^{\infty} \left(\prod_t \frac{dQ(t)}{\sqrt{2\pi}}\right) \prod_t \Theta\{S(t)[\mathbf{Q} - (\mathcal{J}^\alpha)^{-1}\mathbf{c}](t)\} \exp\left(-\frac{1}{2}\mathbf{Q}^\top \tilde{\mathbf{A}}^{-1}\mathbf{Q}\right). \quad (D29)$$

The integral is challenging to study in generality, as the nondiagonal nature of $\mathcal{J}^\alpha$ and $\mathbf{A}$ means that we cannot factorize the integrand. We now move on to specific cases of interest. The equations can be further simplified through the rescalings $\alpha^2 v(\varepsilon)\hat{K}(t,s) \to \hat{K}(t,s)$, $\alpha\sqrt{v(\varepsilon)}\hat{V}(t,s) \to \hat{V}(t,s)$, and $Q(t) \to \alpha\sqrt{v(\varepsilon)}$, in which case we finally get

$$\Sigma_L(\alpha,\eta) = \underset{\hat{R},\hat{K},\hat{V}}{\text{saddle}}\left\{\sum_{s<t} i\hat{R}(t,s)i\hat{K}(t,s) - \frac{\eta}{2}\sum_{t,s} \hat{V}(t,s)\hat{V}(s,t) + \log \mathcal{I}_L\right\}, \quad (D30)$$

$$\mathcal{I}_L = \sum_{\{S(t)\}} e^{\sum_{s<t} S(t)S(s)i\hat{K}(t,s)} \Psi_L[\Gamma_1(\alpha,\eta),...,\Gamma_L(\alpha,\eta);\mathbf{C}], \quad (D31)$$

where $\Psi_L(x_1,...,x_L;\mathbf{C})$ is the cumulative distribution of an L-dimensional Gaussian with a zero mean vector and covariance matrix $\mathbf{C}$ evaluated up to x_t, $t\in\{1,...,L\}$. In our case, the upper bounds of integration are given by $\mathbf{\Gamma} = \mathbf{S}\circ(\mathcal{J}^\alpha)^{-1}\mathbf{c}$, with now

$$c(t) = \eta\sum_s \hat{V}(t,s)S(s), \quad (D32)$$

while the covariance matrix must be taken with care as the off-diagonal elements must be adapted to the presence of $S(t)$ in the Heaviside step function. As a result, off-diagonal elements must be multiplied by $S(t)S(s)$. As a result, we have

$$C(t,s) = \begin{cases} \tilde{A}(t,t) & \text{for } t = s, \\ S(t)S(s)\tilde{A}(t,s) & \text{for } t \neq s \end{cases} \quad (D33)$$

with $\tilde{\mathbf{A}} = (\mathcal{J}^\alpha)^{-1}\mathbf{A}(\mathcal{J}^\alpha)^{-1}$ and

$$A(t,s) = \delta_{ts} - (1-\delta_{ts})i\hat{R}(t,s). \quad (D34)$$

1. Fixed points

We can start by checking that we recover the known result for $L = 1$. In this case, $\mathcal{J}^\alpha = \alpha$, and we have only to solve for a scalar $\hat{V}(1,1) = x$. Here,

$$\mathcal{I}_1 = 2\Psi_1\left(\frac{\eta}{\alpha}x;\frac{1}{\alpha^2}\right) = 2\Phi(\eta x); \quad (D35)$$

therefore, the saddle point equation becomes

$$\Sigma_1(\alpha,\eta) := \Sigma_{\text{FP}}(\eta) = \max_x\left\{-\frac{1}{2}x^2 + \log 2 + \log\Phi(\eta x)\right\}, \quad (D36)$$

from which the expression in the main text can immediately be recovered.

2. Two-cycles

There are now six variables to solve for: $i\hat{R}$, $i\hat{K}$, $\hat{V}_1 = \hat{V}(1,1)$, $\hat{V}_2 = \hat{V}(2,2)$, $\hat{V}_{12} = \hat{V}(1,2)$, and $\hat{V}_{21} = \hat{V}(2,1)$. In this case, the Jacobian matrix is given by

$$\mathcal{J}^\alpha = \begin{bmatrix} -(1-\alpha) & 1 \\ 1 & -(1-\alpha) \end{bmatrix} \Rightarrow (\mathcal{J}^\alpha)^{-1} = \frac{1}{1-(1-\alpha)^2}\begin{bmatrix} 1-\alpha & 1 \\ 1 & 1-\alpha \end{bmatrix}, \quad (D37)$$

giving the bounds of integration

$$\mathbf{\Gamma} = \frac{\eta}{1-(1-\alpha)^2}\begin{bmatrix} \hat{V}_{21} + S_1 S_2 \hat{V}_2 + (1-\alpha)(\hat{V}_1 + S_1 S_2 \hat{V}_{12}) \\ \hat{V}_{12} + S_1 S_2 \hat{V}_1 + (1-\alpha)(\hat{V}_2 + S_1 S_2 \hat{V}_{21}) \end{bmatrix} \quad (D38)$$

and the covariance matrix

$$\mathbf{C} = \frac{1}{[1-(1-\alpha)^2]^2}\begin{bmatrix} 1+(1-\alpha)^2 - 2(1-\alpha)i\hat{R} & S_1 S_2\{2(1-\alpha) - [1+(1-\alpha)^2]i\hat{R}\} \\ S_1 S_2\{2(1-\alpha) - [1+(1-\alpha)^2]i\hat{R}\} & 1+(1-\alpha)^2 - 2(1-\alpha)i\hat{R} \end{bmatrix}. \quad (D39)$$

Now, it is immediately apparent that $\mathcal{I}_2$ is a function of the product $S_1 S_2 = \pm 1$ rather than S_1 and S_2. As a result, the complexity is given by

$$\Sigma_2(\alpha, \eta) = \underset{\hat{R}, \hat{K}, \hat{V}_1, \hat{V}_2, \hat{V}_{12}, \hat{V}_{21}}{\text{saddle}} \left\{ i\hat{R}i\hat{K} - \frac{\eta}{2}(\hat{V}_1^2 + \hat{V}_2^2 + 2\hat{V}_{12}\hat{V}_{21}) + \log 2 + \log \sum_{S=\pm 1} e^{Si\hat{K}} \Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) \right\}. \tag{D40}$$

The six saddle point equations read

$$i\hat{R} \sum_{S=\pm 1} e^{Si\hat{K}} \Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) + \sum_{S=\pm 1} S e^{Si\hat{K}} \Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) = 0, \tag{D41}$$

$$\left(i\hat{K} + \frac{i\hat{R}}{1 - (i\hat{R})^2} \right) \sum_{S=\pm 1} e^{Si\hat{K}} \Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) - \frac{1}{2} \sum_{S=\pm 1} e^{Si\hat{K}} \int_{-\infty}^{\Gamma_1} dQ_1 \int_{-\infty}^{\Gamma_2} dQ_2 \mathbf{Q}^\top \frac{\partial \mathbf{C}^{-1}}{\partial i\hat{R}} \mathbf{Q} \frac{e^{-\frac{1}{2}\mathbf{Q}^\top \mathbf{C}^{-1}\mathbf{Q}}}{2\pi\sqrt{\det \mathbf{C}}} = 0, \tag{D42}$$

$$-\eta \hat{V}_1 \sum_{S=\pm 1} e^{Si\hat{K}} \Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) + \frac{\eta(1-\alpha)}{1 - (1-\alpha)^2} \sum_{S=\pm 1} e^{Si\hat{K}} \int_{-\infty}^{\Gamma_2} dQ_2 \frac{e^{-\frac{1}{2}\mathbf{Q}_2^{\star\top} \mathbf{C}^{-1} \mathbf{Q}_2^\star}}{2\pi\sqrt{\det \mathbf{C}}}$$
$$+ \frac{\eta}{1 - (1-\alpha)^2} \sum_{S=\pm 1} S e^{Si\hat{K}} \int_{-\infty}^{\Gamma_1} dQ_1 \frac{e^{-\frac{1}{2}\mathbf{Q}_1^{\star\top} \mathbf{C}^{-1} \mathbf{Q}_1^\star}}{2\pi\sqrt{\det \mathbf{C}}} = 0, \tag{D43}$$

$$-\eta \hat{V}_2 \sum_{S=\pm 1} e^{Si\hat{K}} \Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) + \frac{\eta}{1 - (1-\alpha)^2} \sum_{S=\pm 1} S e^{Si\hat{K}} \int_{-\infty}^{\Gamma_2} dQ_2 \frac{e^{-\frac{1}{2}\mathbf{Q}_2^{\star\top} \mathbf{C}^{-1} \mathbf{Q}_2^\star}}{2\pi\sqrt{\det \mathbf{C}}}$$
$$+ \frac{\eta(1-\alpha)}{1 - (1-\alpha)^2} \sum_{S=\pm 1} e^{Si\hat{K}} \int_{-\infty}^{\Gamma_1} dQ_1 \frac{e^{-\frac{1}{2}\mathbf{Q}_1^{\star\top} \mathbf{C}^{-1} \mathbf{Q}_1^\star}}{2\pi\sqrt{\det \mathbf{C}}} = 0, \tag{D44}$$

$$-\eta \hat{V}_{21} \sum_{S=\pm 1} e^{Si\hat{K}} \Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) + \frac{\eta(1-\alpha)}{1 - (1-\alpha)^2} \sum_{S=\pm 1} S e^{Si\hat{K}} \int_{-\infty}^{\Gamma_2} dQ_2 \frac{e^{-\frac{1}{2}\mathbf{Q}_2^{\star\top} \mathbf{C}^{-1} \mathbf{Q}_2^\star}}{2\pi\sqrt{\det \mathbf{C}}}$$
$$+ \frac{\eta}{1 - (1-\alpha)^2} \sum_{S=\pm 1} e^{Si\hat{K}} \int_{-\infty}^{\Gamma_1} dQ_1 \frac{e^{-\frac{1}{2}\mathbf{Q}_1^{\star\top} \mathbf{C}^{-1} \mathbf{Q}_1^\star}}{2\pi\sqrt{\det \mathbf{C}}} = 0, \tag{D45}$$

$$-\eta \hat{V}_{12} \sum_{S=\pm 1} e^{Si\hat{K}} \Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) + \frac{\eta}{1 - (1-\alpha)^2} \sum_{S=\pm 1} e^{Si\hat{K}} \int_{-\infty}^{\Gamma_2} dQ_2 \frac{e^{-\frac{1}{2}\mathbf{Q}_2^{\star\top} \mathbf{C}^{-1} \mathbf{Q}_2^\star}}{2\pi\sqrt{\det \mathbf{C}}}$$
$$+ \frac{\eta(1-\alpha)}{1 - (1-\alpha)^2} \sum_{S=\pm 1} S e^{Si\hat{K}} \int_{-\infty}^{\Gamma_1} dQ_1 \frac{e^{-\frac{1}{2}\mathbf{Q}_1^{\star\top} \mathbf{C}^{-1} \mathbf{Q}_1^\star}}{2\pi\sqrt{\det \mathbf{C}}} = 0, \tag{D46}$$

with $\mathbf{Q}_1^* = [Q_1 \, \Gamma_2]^\top$ and similarly for $\mathbf{Q}_2^*$, and

$$\frac{\partial \mathbf{C}^{-1}}{\partial i\hat{R}} = \frac{1}{[1-(i\hat{R})^2]^2} \begin{bmatrix} 2i\hat{R}[1+(1-\alpha)^2] - 2(1-\alpha)[(i\hat{R})^2+1] & S\{-4i\hat{R}(1-\alpha) + [1+(1-\alpha)^2][(i\hat{R})^2+1]\} \\ S\{-4i\hat{R}(1-\alpha) + [1+(1-\alpha)^2][(i\hat{R})^2+1]\} & 2i\hat{R}[1+(1-\alpha)^2] - 2(1-\alpha)[(i\hat{R})^2+1] \end{bmatrix}. \tag{D47}$$

We now take the particular case $\alpha = 1$. In this case, we expect two-cycles with a zero overlap in between consecutive steps, meaning $i\hat{R} = 0$. As a result, the matrix $\mathbf{C}$ is the identity matrix, and, thus,

$$\Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) = \Phi(\eta(\hat{V}_{21} + S\hat{V}_2))\Phi(\eta(\hat{V}_{12} + S\hat{V}_1)). \tag{D48}$$

In order for $i\hat{R} = 0$ to satisfy Eq. (D41), we then require $i\hat{K} = 0$ and $\hat{V}_1 = \hat{V}_2 = 0$ such that $\sum_{S=\pm 1} Se^{i\hat{K}}\Psi_2(\Gamma_1, \Gamma_2; \mathbf{C}) = 0$. Given

$$\frac{\partial \mathbf{C}^{-1}}{\partial i\hat{R}}\bigg|_{i\hat{R}=0} = \begin{bmatrix} 0 & S \\ S & 0 \end{bmatrix} \tag{D49}$$

for $\alpha = 1$, the solution $i\hat{K} = 0$ satisfies the saddle point equation (D42), while the independence of the integrands on S ensures that $\hat{V}_1 = \hat{V}_2 = 0$ are compatible with Eqs. (D43) and (D44). Equations (D45) and (D46) then suggest the *Ansatz* $\hat{V}_{12} = \hat{V}_{21} = x$, in which case the problem is finally reduced to

$$\Sigma_2(\alpha = 1, \eta) = \max_x \{-\eta x^2 + 2\log 2 + 2\log \Phi(\eta x)\}$$

$$= 2\Sigma_{\text{FP}}(\eta), \tag{D50}$$

recovering the known solution of Refs. [73,100].

APPENDIX E: DERIVATION OF THE DMFT EQUATIONS

We start from the $N \gg 1$ discrete difference equations to which we have added an external field $h_i(t)$:

$$Q_i(t+1) = (1-\alpha)Q_i(t) + \alpha\sum_j J_{ij}m_j(t) + \alpha\eta_i(t) + h_i(t),$$

$$\langle \eta_i(t)\eta_j(s) \rangle \approx v(\varepsilon)[1 - q(t)]\delta_{t,s}\delta_{i,j} \tag{E1}$$

and introduce a new agent at index $i = 0$. The influence of this newly introduced agent on the dynamic equation of agents $i \neq 0$ is then

$$\sum_j J_{ij}m_j(t) + h_i(t) \rightarrow \sum_j J_{ij}m_j(t) + J_{i0}m_0(t) + h_i(t). \tag{E2}$$

Given N is large, we consider the response at the linear order, meaning that the expected decision of all agents $i > 0$ becomes

$$m_i(t) \rightarrow m_i(t) + \sum_{s<t}\sum_j \underbrace{\frac{\partial m_i(t)}{\partial h_j(s)}\bigg|_{h=0}}_{\chi_{ij}(t,s)} J_{j0}m_0(s), \tag{E3}$$

as $J_{j0}m_0(t)$ can be seen as the modification of the effective field "felt" by all agents j, and $\chi_{ij}(t,s)$ is simply the linear response function to a small external field. The dynamics that the newly introduced agent follows is then given by

$$Q_0(t+1) = (1-\alpha)Q_0(t) + \alpha\sum_{i>0} J_{0i}m_i(t)$$

$$+ \sum_{s<t}\left(\sum_{ij} J_{0i}\chi_{ij}(t,s)J_{j0}\right)m_0(s)$$

$$+ \alpha\eta_0(t) + h_0(t). \tag{E4}$$

The sum on i and j can be split into its diagonal and off-diagonal parts. On the diagonal, we assume the central limit theorem holds, yielding

$$\sum_i J_{0i}\chi_{ii}(t,s)J_{i0} = N\left[\overline{\langle J_{0i}\chi_{ii}(t,s)J_{i0}\rangle} + O\left(\frac{1}{\sqrt{N}}\right)\right]$$

$$\approx (1-\varepsilon)\langle \chi_{ii}(t,s)\rangle, \tag{E5}$$

while the off-diagonal contribution will be subdominant as its mean will be zero, given nonopposing entries in the interaction matrix are uncorrelated. We can also assume the central limit theorem is valid for the sum on indices i at the leading order, in which case

$$\sum_i J_{0i}m_i(t) \approx 0 + v(\varepsilon)\xi_0(t), \tag{E6}$$

where ξ_0 is a Gaussian of zero mean and correlated in time as $\langle \xi_0(t)\xi_0(s)\rangle = C_0(t,s) = \langle m_0(t)m_0(s)\rangle$. Bringing everything together, one realizes that there are no cross contributions between agents at the leading order and, thus, that we can drop the index 0 and recover the equation in the main text:

$$Q(t+1) = (1-\alpha)Q(t) + \alpha^2(1-\varepsilon)\sum_{s<t} G(t,s)m(s)$$

$$+ \alpha\phi(t) + \alpha h(t), \tag{E7}$$

with a new noise term combining the original thermal-like fluctuations and the effective contribution from the disorder averaging, with $\langle \phi(t)\rangle = 0$, and

$$\langle \phi(t)\phi(s)\rangle = v(\varepsilon)\{C(t,s) + [1-q(t)]\delta_{t,s}\}, \tag{E8}$$

and where the memory kernel and correlation function are to be determined self-consistently:

$$G(t,s) = \langle \chi_{ii}(t,s)\rangle = \left\langle \frac{\partial m(t)}{\partial h(s)}\bigg|_{h=0}\right\rangle,$$

$$C(t,s) = \langle m(t)m(s)\rangle. \tag{E9}$$

Note that, in order to eliminate the external field, we express the susceptibility with the noise term, resulting in a rescaling $G(t, s) \to \alpha G(t, s)$. From this expression, the continuous limit can be taken, changing the sum in time weighted by α into an integral:

$$\frac{\alpha}{2} \ddot{Q}(t) + \dot{Q}(t) = -Q(t) + (1 - \varepsilon) \int_0^t ds\, G(t, s) m(s) + \phi(t) + h(t). \tag{E10}$$

APPENDIX F: ADAPTING THE SOMPOLINSKY AND CRISANTI RESULT

We start from the much simplified DMFT equation:

$$\dot{Q}(t) = -Q(t) + \phi(t), \qquad \langle \phi(t)\phi(s) \rangle = \frac{1}{2} C(t, s) \tag{F1}$$

with $C(t,s) = \langle m(t)m(s) \rangle$, $m(t) = \text{sgn}(Q(t))$ in the $\beta \to \infty$ limit. As shown by Sompolinsky and Crisanti [112], we can write a second-order differential equation for the two-point autocorrelation of $Q(t)$:

$$\ddot{\Delta}(\tau) = \Delta(\tau) - \frac{1}{2} C(\tau), \tag{F2}$$

with, therefore, $\Delta(\tau) = \langle Q(t + \tau)Q(t) \rangle$, and where thanks to the Gaussian nature of the fluctuations we have

$$
\begin{aligned}
C(\tau) &= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \left[\int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \text{sgn}(\sqrt{\Delta(0) - |\Delta(\tau)|}x + \sqrt{|\Delta(\tau)|}z) \right]^2 \\
&= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \text{erf}\left(\sqrt{\frac{|\Delta(\tau)|}{\Delta(0) - |\Delta(\tau)|}} \frac{z}{\sqrt{2}} \right)^2 \\
&= \frac{2}{\pi} \sin^{-1}\left(\frac{\Delta(\tau)}{\Delta(0)} \right)
\end{aligned}
\tag{F3}
$$

(see Ref. [153] for the last step above). Now, by identifying that the only difference with the original problem is a factor 2 in Δ, we can directly use the result $\Delta(0) = 1 - (2/\pi)$, found by enforcing the condition required for $\Delta(\tau)$ to decay monotonically [114]. We may finally expand the inverse sine to recover:

$$\ddot{\Delta}(\tau) \underset{\tau \gg 1}{\sim} \left(1 - \frac{1}{\pi\Delta(0)} \right) \Delta(\tau) \tag{F4}$$

and, therefore,

$$C(\tau) \underset{\tau \gg 1}{\sim} \frac{2}{\pi} e^{-\frac{\tau}{\tau_1}}, \qquad \tau_1 = \sqrt{\frac{\pi - 2}{\pi - 3}} \approx 2.84. \tag{F5}$$

We notice that the value of the characteristic time is identical in the standard Sompolinsky and Crisanti case [the only difference being a factor 2 in the magnitude of $\Delta(\tau)$], which is, therefore, *not* equal to the value $\sqrt{(\pi/\pi - 2)} \approx 1.66$ originally given in Refs. [112,113].

[1] H. A. Simon, *A behavioral model of rational choice*, Q. J. Econ. **69**, 99 (1955).

[2] J.-M. Grandmont, *Expectations formation and stability of large socioeconomic systems*, Econometrica **66**, 741 (1998).

[3] A. Kirman, *Complex Economics: Individual and Collective Rationality* (Routledge, New York, 2010).

[4] C. Hommes, *Behavioral and experimental macroeconomics and policy analysis: A complex systems approach*, J. Econ. Lit. **59**, 149 (2021).

[5] G. Dosi and A. Roventini, *More is different... and complex! The case for agent-based macroeconomics*, J. Evol. Econ. **29**, 1 (2019).

[6] M. King and J. Kay, *Radical Uncertainty: Decision-Making for an Unknowable Future* (Hachette, London, 2020).

[7] A. Marcet and T. J. Sargent, *Convergence of least squares learning mechanisms in self-referential linear stochastic models*, J. Econ. Theory **48**, 337 (1989).

[8] G. W. Evans and S. Honkapohja, *Learning and macroeconomics*, Annu. Rev. Econ. **1**, 421 (2009).

[9] R. Pemantle, *A survey of random processes with reinforcement*, Probab. Surv. **4**, 1 (2007).

[10] D. Lamberton, G. Pagès, and P. Tarrès, *When can the two-armed bandit algorithm be trusted?*, Ann. Appl. Probab. **14,** 1424 (2004).

[11] J. Moran, A. Fosset, D. Luzzati, J.-P. Bouchaud, and M. Benzaquen, *By force of habit: Self-trapping in a dynamical utility landscape*, Chaos **30,** 053123 (2020).

[12] C. Colon and J.-P. Bouchaud, *The radical complexity of rewiring supplier-buyer networks*, SSRN 4300311, 10.2139/ssrn.4300311.

[13] T. Galla and J. D. Farmer, *Complex dynamics in learning complicated games*, Proc. Natl. Acad. Sci. U.S.A. **110,** 1232 (2013).

[14] J.-P. Bouchaud and R. E. Farmer, *Self-fulfilling prophecies, quasi nonergodicity, and wealth inequality*, J. Polit. Econ. **131,** 000 (2023).

[15] G. Parisi, *Complex systems: A physicist's viewpoint*, Physica (Amsterdam) **263A,** 557 (1999).

[16] G. Parisi, *Physics complexity and biology*, Adv. Complex Syst. **10,** 223 (2007).

[17] J.-P. Bouchaud, *Radical complexity*, Entropy **23,** 1676 (2021).

[18] "Common knowledge" means that "you know what I know and I know that you know what I know."

[19] S. Galluccio, J.-P. Bouchaud, and M. Potters, *Rational decisions, random matrices and spin glasses*, Physica (Amsterdam) **259A,** 449 (1998).

[20] J. Garnier-Brun, M. Benzaquen, S. Ciliberti, and J.-P. Bouchaud, *A new spin on optimal portfolios and ecological equilibria*, J. Stat. Mech. (2021) 093408.

[21] D. Sharma, J.-P. Bouchaud, M. Tarzia, and F. Zamponi, *Good speciation and endogenous business cycles in a constraint satisfaction macroeconomic model*, J. Stat. Mech. (2021) 063403.

[22] W. Brian Arthur, S. N. Durlauf, and D. A. Lane, *Process and emergence in the economy*, Complexity and the Economy, edited by W. Brian Arthur (2015), pp. 89–102.

[23] T. Dessertaine, J. Moran, M. Benzaquen, and J.-P. Bouchaud, *Out-of-equilibrium dynamics and excess volatility in firm networks*, J. Econ. Dyn. Control **138,** 104362 (2022).

[24] P. Bak, *How Nature Works: The Science of Self-Organized Criticality* (Springer, New York, 2013).

[25] P. W. Anderson, *The Economy as an Evolving Complex System* (CRC Press, Boca Raton, 2018).

[26] J. B. Rosser, Jr., *On the complexities of complex economic dynamics*, J. Econ. Perspect. **13,** 169 (1999).

[27] P. Bak, K. Chen, J. Scheinkman, and M. Woodford, *Aggregate fluctuations from independent sectoral shocks: Self-organized criticality in a model of production and inventory dynamics*, Ric. Econ. **47,** 3 (1993).

[28] R. Bookstaber, *The End of Theory: Financial Crises, the Failure of Economics, and the Sweep of Human Interaction* (Princeton University Press, Princeton, NJ, 2017).

[29] M. Shiino and T. Fukai, *Replica-symmetric theory of nonlinear analogue neural networks*, J. Phys. A **23,** L1009 (1990).

[30] J. Moran and J.-P. Bouchaud, *May's instability in large economies*, Phys. Rev. E **100,** 032307 (2019).

[31] W. B. Arthur, *Foundations of complexity economics*, Nat. Rev. Phys. **3,** 136 (2021).

[32] G. Dosi, *The Foundations of Complex Evolving Economies: Part One: Innovation, Organization, and Industrial Dynamics* (Oxford University, New York, 2023).

[33] J. B. Sanders, J. D. Farmer, and T. Galla, *The prevalence of chaotic dynamics in games with many players*, Sci. Rep. **8,** 1 (2018).

[34] For other examples of models where learning leads to chaotic dynamics, see, e.g., Refs. [35,36].

[35] W. A. Brock and C. H. Hommes, *A rational route to randomness*, Econometrica **65,** 1059 (1997).

[36] W. A. Brock and C. H. Hommes, *Heterogeneous beliefs and routes to chaos in a simple asset pricing model*, J. Econ. Dyn. Control **22,** 1235 (1998).

[37] T. Hirano and J. E. Stiglitz, *Land speculation and wobbly dynamics with endogenous phase transitions*, technical report, National Bureau of Economic Research, 2022.

[38] G. Dosi, M. Napoletano, A. Roventini, J. E. Stiglitz, and T. Treibich, *Rational heuristics? Expectations and behaviors in evolving economies with heterogeneous interacting agents*, Economic inquiry **58,** 1487 (2020).

[39] W. A. Brock and S. N. Durlauf, *Discrete choice with social interactions*, Rev. Econ. Stud. **68,** 235 (2001).

[40] D. Challet, M. Marsili, and Y.-C. Zhang, *Minority Games: Interacting Agents in Financial Markets* (Oxford, New York, 2004).

[41] M. B. Gordon, J.-P. Nadal, D. Phan, and V. Semeshenko, *Discrete choices under social influence: Generic properties*, Math. Models Methods Appl. Sci. **19,** 1441 (2009).

[42] J.-P. Bouchaud, *Crises and collective socio-economic phenomena: Simple models and challenges*, J. Stat. Phys. **151,** 567 (2013).

[43] S. P. Anderson, A. De Palma, and J.-F. Thisse, *Discrete Choice Theory of Product Differentiation* (MIT Press, Cambridge, MA, 1992).

[44] For a complete review of the motivations behind this rule, see Refs. [42,43], and references therein.

[45] C. J. Watkins and P. Dayan, *Q-learning*, Mach. Learn. **8,** 279 (1992).

[46] C. Camerer and T. Hua Ho, *Experience-weighted attraction learning in normal form games*, Econometrica **67,** 827 (1999).

[47] D. Sherrington and S. Kirkpatrick, *Solvable model of a spin glass*, Phys. Rev. Lett. **35,** 1792 (1975).

[48] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and beyond: An Introduction to the Replica Method and Its Applications* (World Scientific, Singapore, 1987), Vol. 9.

[49] Note that these models break the $\mathbf{S}(t) \to -\mathbf{S}(t)$ symmetry for all odd values of p.

[50] D. Panchenko, *Introduction to the SK model*, Current Developments in Mathematics (2014), Vol. 2014, pp. 231–291, .

[51] Y. Sato, E. Akiyama, and J. P. Crutchfield, *Stability and diversity in collective adaptation*, Physica (Amsterdam) **210D,** 21 (2005).

[52] T. Galla, *Intrinsic noise in game dynamical learning*, Phys. Rev. Lett. **103,** 198702 (2009).

[53] T. Galla, *Cycles of cooperation and defection in imperfect learning*, J. Stat. Mech. (2011) P08007.

[54] D. Vilone, A. Robledo, and A. Sánchez, *Chaos and unpredictability in evolutionary dynamics in discrete time*, Phys. Rev. Lett. **107**, 038101 (2011).

[55] A. Kianercy and A. Galstyan, *Dynamics of Boltzmann Q learning in two-player two-action games*, Phys. Rev. E **85**, 041145 (2012).

[56] J. Burridge, *Limit cycles and the benefits of a short memory in rock-paper-scissors games*, Phys. Rev. E **92**, 042111 (2015).

[57] M. Pangallo, T. Heinrich, and J. Doyne Farmer, *Best reply structure and equilibrium convergence in generic games*, Sci. Adv. **5**, eaat1328 (2019).

[58] V. Semeshenko, M. B. Gordon, J.-P. Nadal, and D. Phan, *Choice under social influence: Effects of learning behaviours on the collective dynamics*, Contrib. Econ. Anal. **280**, 177 (2006).

[59] V. Semeshenko, M. B. Gordon, and J.-P. Nadal, *Collective states in social systems with interacting learning agents*, Physica (Amsterdam) **387A**, 4903 (2008).

[60] M. Opper and S. Diederich, *Phase transition and $1/f$ noise in a game dynamical model*, Phys. Rev. Lett. **69**, 1616 (1992).

[61] T. Galla, *Random replicators with asymmetric couplings*, J. Phys. A **39**, 3853 (2006).

[62] A. Leonidov, A. Savvateev, and A. G. Semenov, *Ising game on graphs*, Chaos Solitons Fractals **180**, 114540 (2024).

[63] A. Leonidov and E. Vasilyeva, *Strategic stiffening/cooling in the Ising game*, Chaos Solitons Fractals **160**, 112279 (2022).

[64] D. Smug, D. Sornette, and P. Ashwin, *A generalized 2d-dynamical mean-field Ising model with a rich set of bifurcations (inspired and applied to financial crises)*, Int. J. Bifurcation Chaos Appl. Sci. Eng. **28**, 1830010 (2018).

[65] Note that the precise distribution of the interactions is irrelevant; as long as we have $J_{ij} \sim 1/\sqrt{N}$, the first two moments are finite [48]. We could, for example, take signed interactions $J_{ij} = \pm 1/\sqrt{N}$, which could importantly be much easier to measure from empirical data. Having broadly distributed J_{ij} would be an interesting extension to our model, which we leave for future work.

[66] The excess reward is the gain obtained by the agent from their current choice relative to that that they would be rewarded making the other choice, which in this binary case is simply twice the reward, i.e., $R_i^{S_i} - R_i^{-S_i} = 2R_i^{S_i}$.

[67] G. Parisi, *A sequence of approximated solutions to the SK model for spin glasses*, J. Phys. A **13**, L115 (1980).

[68] In fact, it was recently shown that one can devise a smart algorithm with run-time growing like $K(\epsilon)N^2$ to find configurations $\{S_i\}$ that reach a value of at least $1 - \epsilon$ times the optimum (13), with $\epsilon > 0$ independent of N [69]. However, such a smart algorithm is of no use for our learning agents.

[69] A. Montanari, *Optimization of the Sherrington-Kirkpatrick Hamiltonian*, SIAM J. Comput. FOCS19-1 (2021). 10.1137/20M132016X

[70] Note that there are finite N corrections that must be taken into account for such a comparison, which read [71] $\mathcal{R}_N \approx \mathcal{R}_\infty - A/(N^{2/3})$, $A \approx 0.75 \times (2 - \epsilon)$.

[71] S. Boettcher, *Simulations of ground state fluctuations in mean-field Ising spin glasses*, J. Stat. Mech. (2010) P07002.

[72] A. J. Bray and M. A. Moore, *Metastable states in spin glasses*, J. Phys. C **13**, L469 (1980).

[73] S. Hwang, V. Folli, E. Lanza, G. Parisi, G. Ruocco, and F. Zamponi, *On the number of limit cycles in asymmetric neural networks*, J. Stat. Mech. (2019) 053402.

[74] J. Sterman, *Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment*, Manage. Sci. **35**, 321 (1989).

[75] Note that the role of ϵ is somewhat similar to that of parameter Γ in Ref. [13]: Increasing competition leads to chaos.

[76] Note that this is strictly speaking true only when β is large but not infinite. For infinite β and finite N, one always gets oscillations of length $L = 4$.

[77] Of course, in the $\beta \to \infty$ limit discussed in this subsection, one can replace $m_i(t)$ by the actual decision $S_i(t)$ in the definition of $C(t, t + \tau)$. More generally, the spin-spin correlation function is given by $[1 - q(t)]\delta(\tau) + C(t, t+\tau)$ with $q(t) = \overline{\langle m_i^2(t) \rangle}$.

[78] Note the rather large error bar on $C(2n + 1)$, meaning that there are actually substantial fluctuations of the number of idle spins around the value $N/2$ when N is finite.

[79] K. Nutzel and U. Krey, *Subtle dynamic behaviour of finite-size Sherrington-Kirkpatrick spin glasses with nonsymmetric couplings*, J. Phys. A **26**, L591 (1993).

[80] H.-J. Sommers and W. Dupont, *Distribution of frozen fields in the mean-field theory of spin glasses*, J. Phys. C **17**, 5785 (1984).

[81] S. Pankov, *Low-temperature solution of the Sherrington-Kirkpatrick model*, Phys. Rev. Lett. **96**, 197204 (2006).

[82] M. Müller and M. Wyart, *Marginal stability in structural, spin, and electron glasses*, Annu. Rev. Condens. Matter Phys. **6**, 177 (2015).

[83] One would expect a different behavior in the possibly relevant case of a fat-tailed distribution of the J_{ij}; see Refs. [84–86]. We leave this question for further investigations.

[84] P. Cizeau and J.-P. Bouchaud, *Mean field theory of dilute spin-glasses with power-law interactions*, J. Phys. A **26**, L187 (1993).

[85] K. Janzen, A. Engel, and M. Mézard, *Thermodynamics of the Lévy spin glass*, Phys. Rev. E **82**, 021127 (2010).

[86] I. Neri, F. Metz, and D. Bollé, *The phase diagram of Lévy spin glasses*, J. Stat. Mech. (2010) P01010.

[87] F. Krzakala and J.-P. Bouchaud, *Disorder chaos in spin glasses*, Europhys. Lett. **72**, 472 (2005).

[88] T. Aspelmeier, *Bond chaos in the Sherrington-Kirkpatrick model*, J. Phys. A **41**, 205005 (2008).

[89] G. Toulouse, *On the mean field theory of mixed spin glass-ferromagnetic phases*, J. Phys. (Paris), Lett. **41**, 447 (1980).

[90] R. Zecchina, *Liquid computing in neural networks* (unpublished).

[91] D. J. Thouless, P. W. Anderson, and R. G. Palmer, *Solution of "solvable model of a spin glass"*, Philos. Mag. **35**, 593 (1977).

[92] A. J. Bray, H. Sompolinsky, and C. Yu, *On the 'naive' mean-field equations for spin glasses*, J. Phys. C **19,** 6389 (1986).

[93] H. Takayama and K. Nemoto, *Spin glass properties of a class of mean-field models*, J. Phys. Condens. Matter **2,** 1997 (1990).

[94] K. Nishimura, K. Nemoto, and H. Takayama, *Metastable states of the naive mean-field model for spin glasses at finite temperatures*, J. Phys. A **23,** 5915 (1990).

[95] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein, *Spectrum of large random asymmetric matrices*, Phys. Rev. Lett. **60,** 1895 (1988).

[96] V. Ros, F. Roy, G. Biroli, G. Bunin, and A. M. Turner, *Generalized Lotka-Volterra equations with random, non-reciprocal interactions: The typical number of equilibria*, Phys. Rev. Lett. **130,** 257401 (2023).

[97] T. Aspelmeier, A. J. Bray, and M. A. Moore, *Complexity of Ising spin glasses*, Phys. Rev. Lett. **92,** 087203 (2004).

[98] A. Bovier, F. Dunlop, A. Van Enter, F. Den Hollander, and J. Dalibard, *Mathematical Statistical Physics: Lecture Notes of the Les Houches Summer School 2005* (Elsevier, New York, 2006).

[99] F. R. Waugh, C. M. Marcus, and R. M. Westervelt, *Fixed-point attractors in analog neural computation*, Phys. Rev. Lett. **64,** 1986 (1990).

[100] H. Gutfreund, J. Reger, and A. Young, *The nature of attractors in an asymmetric spin glass with deterministic dynamics*, J. Phys. A **21,** 2775 (1988).

[101] F. Tanaka and S. Edwards, *Analytic theory of the ground state properties of a spin glass. I. Ising spin glass*, J. Phys. F **10,** 2769 (1980).

[102] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. U.S.A. **79,** 2554 (1982).

[103] U. Bastolla and G. Parisi, *Relaxation, closing probabilities and transition from oscillatory to chaotic attractors in asymmetric neural networks*, J. Phys. A **31,** 4583 (1998).

[104] L. F. Cugliandolo, *Recent applications of dynamical mean-field methods*, Annu. Rev. Condens. Matter Phys. **15,** 177 (2023).

[105] 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 **52,** 484001 (2019).

[106] F. Mignacco, F. Krzakala, P. Urbani, and L. Zdeborová, *Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification*, Adv. Neural Inf. Process. Syst. **33,** 9540 (2020), https://proceedings.neurips.cc/paper_files/paper/2020/file/6c81c83c4bd0b58850495f603ab45a93-Paper.pdf.

[107] H. Eissfeller and M. Opper, *New method for studying the dynamics of disordered spin systems without finite-size effects*, Phys. Rev. Lett. **68,** 2094 (1992).

[108] H. Eissfeller and M. Opper, *Mean-field Monte Carlo approach to the Sherrington-Kirkpatrick model with asymmetric couplings*, Phys. Rev. E **50,** 709 (1994).

[109] The convergence to m_∞ is as slow a power law of τ, which makes difficult its numerical determination. Obtaining the precise value of ε_{RM} is, therefore, challenging [108].

[110] A. Scharnagl, M. Opper, and W. Kinzel, *On the relaxation of infinite-range spin glasses*, J. Phys. A **28,** 5721 (1995).

[111] Note that $C(\tau = 2)$ remains equal to 1 in the whole range of parameters shown in Fig. 12; i.e., we observe only fixed points or 2-cycles.

[112] H. Sompolinsky, A. Crisanti, and H.-J. Sommers, *Chaos in random neural networks*, Phys. Rev. Lett. **61,** 259 (1988).

[113] A. Crisanti and H. Sompolinsky, *Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics*, Phys. Rev. A **37,** 4865 (1988).

[114] A. Crisanti and H. Sompolinsky, *Path integral approach to random neural networks*, Phys. Rev. E **98,** 062120 (2018).

[115] We, in fact, note that in this regime all results seem to collapse when plotted versus $T/T_c(\varepsilon)$, as in Fig. 17(a).

[116] L. F. Cugliandolo and J. Kurchan, *On the out-of-equilibrium relaxation of the Sherrington-Kirkpatrick model*, J. Phys. A **27,** 5749 (1994).

[117] J.-P. Bouchaud, L. F. Cugliandolo, J. Kurchan, and M. Mézard, *Out of equilibrium dynamics in spin-glasses and other glassy systems*, Spin Glasses Random Fields **12,** 161 (1998).

[118] E. Vincent, J. Hammann, M. Ocio, J.-P. Bouchaud, and L. F. Cugliandolo, *Slow dynamics and aging in spin glasses*, in *Complex Behaviour of Glassy Systems: Proceedings of the XIV Sitges Conference Sitges, Barcelona, Spain, 1996* (Springer, New York, 2007), pp. 184–219.

[119] A. Altieri, F. Roy, C. Cammarota, and G. Biroli, *Properties of equilibria and glassy phases of the random Lotka-Volterra model with demographic noise*, Phys. Rev. Lett. **126,** 258301 (2021).

[120] D. Martí, N. Brunel, and S. Ostojic, *Correlations between synapses in pairs of neurons slow down dynamics in randomly connected neural networks*, Phys. Rev. E **97,** 062314 (2018).

[121] J.-P. Bouchaud, *Weak ergodicity breaking and aging in disordered systems*, J. Phys. I (France) **2,** 1705 (1992).

[122] H. Rieger, *Nonequilibrium dynamics and aging in the three-dimensional Ising spin-glass model*, J. Phys. A **26,** L615 (1993).

[123] L. F. Cugliandolo, J. Kurchan, and F. Ritort, *Evidence of aging in spin-glass mean-field models*, Phys. Rev. B **49,** 6331 (1994).

[124] H. Yoshino, *Off-equilibrium dynamics of a $(1 + 1)$-dimensional directed polymer in random media*, J. Phys. A **29,** 1421 (1996).

[125] L. Berthier and J.-P. Bouchaud, *Geometrical aspects of aging and rejuvenation in the Ising spin glass: A numerical study*, Phys. Rev. B **66,** 054404 (2002).

[126] T. Arnoulx de Pirey and G. Bunin, *Aging by near-extinctions in many-variable interacting populations*, Phys. Rev. Lett. **130,** 098401 (2023).

[127] E. Marinari, G. Parisi, and D. Rossetti, *Numerical simulations of the dynamical behavior of the SK model*, Eur. Phys. J. B **2,** 495 (1998).

[128] A. Baldassarri, *Numerical study of the out-of-equilibrium phase space of a mean-field spin glass model*, Phys. Rev. E **58,** 7047 (1998).

[129] L. F. Cugliandolo, J. Kurchan, P. Le Doussal, and L. Peliti, *Glassy behaviour in disordered systems with nonrelaxational dynamics*, Phys. Rev. Lett. **78,** 350 (1997).

[130] E. Marinari and D. Stariolo, *Off-equilibrium dynamics of a four-dimensional spin glass with asymmetric couplings*, J. Phys. A **31,** 5021 (1998).

[131] L. Berthier and J. Kurchan, *Non-equilibrium glass transitions in driven and active matter*, Nat. Phys. **9,** 310 (2013).

[132] G. Iori and E. Marinari, *On the stability of the mean-field spin glass broken phase under non-Hamiltonian perturbations*, J. Phys. A **30,** 4489 (1997).

[133] L. Molgedey, J. Schuchhardt, and H. G. Schuster, *Suppressing chaos in neural networks by noise*, Phys. Rev. Lett. **69,** 3717 (1992).

[134] T. M. Pham and K. Kaneko, *Theory for adaptive systems: Collective robustness of genotype-phenotype evolution*, arXiv:2306.01403.

[135] A. Antonov, A. Leonidov, and A. Semenov, *Self-excited Ising game*, Physica (Amsterdam) **561A,** 125305 (2021).

[136] A. Civilini, O. Sadekar, F. Battiston, J. Gómez-Gardeñes, and V. Latora, *Explosive cooperation in social dilemmas on higher-order networks*, Phys. Rev. Lett. **132,** 167401 (2024).

[137] C. Van Vreeswijk and H. Sompolinsky, *Chaos in neuronal networks with balanced excitatory and inhibitory activity*, Science **274,** 1724 (1996).

[138] N. Brunel, *Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons*, J. Comput. Neurosci. **8,** 183 (2000).

[139] K. Rajan, L. F. Abbott, and H. Sompolinsky, *Stimulus-dependent suppression of chaos in recurrent neural networks*, Phys. Rev. E **82,** 011903 (2010).

[140] M. Stern, H. Sompolinsky, and L. F. Abbott, *Dynamics of random neural networks with bistable units*, Phys. Rev. E **90,** 062710 (2014).

[141] J. Kadmon and H. Sompolinsky, *Transition to chaos in random neuronal networks*, Phys. Rev. X **5,** 041030 (2015).

[142] D. G. Clark and L. F. Abbott, *Theory of coupled neuronal-synaptic dynamics*, Phys. Rev. X **14,** 021001 (2024).

[143] U. Pereira-Obilinovic, J. Aljadeff, and N. Brunel, *Forgetting leads to chaos in attractor networks*, Phys. Rev. X **13,** 011009 (2023).

[144] A. P. Kirman and N. J. Vriend, *Learning to be loyal. A study of the Marseille fish market*, in *Interaction and Market Structure: Essays on Heterogeneity in Economics* (Springer, New York, 2000), pp. 33–56.

[145] A. W. Lo, *The adaptive markets hypothesis*, J. Portfolio Manage. **30,** 15 (2004).

[146] A. W. Lo, *Reconciling efficient markets with behavioral finance: The adaptive markets hypothesis*, J. Invest. Consult. **7,** 21 (2005), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1702447#.

[147] T. Brotto, G. Bunin, and J. Kurchan, *A model with Darwinian dynamics on a rugged landscape*, J. Stat. Phys. **166,** 1065 (2017).

[148] T. Aspelmeier, R. A. Blythe, A. J. Bray, and M. A. Moore, *Free-energy landscapes, dynamics, and the edge of chaos in mean-field models of spin glasses*, Phys. Rev. B **74,** 184411 (2006).

[149] T. Aspelmeier and M. A. Moore, *Realizable solutions of the Thouless-Anderson-Palmer equations*, Phys. Rev. E **100,** 032127 (2019).

[150] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton University, Princeton, NJ, 1947).

[151] M. Mézard and G. Parisi, *The cavity method at zero temperature*, J. Stat. Phys. **111,** 1 (2003).

[152] F. Liers, M. Palassini, A. K. Hartmann, and M. Jünger, *Ground state of the Bethe lattice spin glass and running time of an exact optimization algorithm*, Phys. Rev. B **68,** 094406 (2003).

[153] A. Prudnikov, Y. A. Brychkov, and O. Marichev, *More Special Functions*, Integrals and Series Vol. 3 (Gordon and Breach, New York, 1990).