Balancing selfishness and norm conformity can explain human behavior in large-scale prisoner's dilemma games and can poise human groups near criticality

John Realpe-Gómez*
Quantum Artificial Intelligence Laboratory, NASA Ames Research Center, Moffett Field, California 94035, USA; Instituto de Matemáticas Aplicadas, Universidad de Cartagena, Cartagena de Indias, Bolívar 13001, Colombia; and SGT Inc., 7701 Greenbelt Road, Greenbelt, Maryland 20770, USA

Giulia Andrighetto
Institute of Cognitive Sciences and Technologies, National Research Council, Rome 00185, Italy; Mälardalen University, Högskoleplan 1, 721 23 Västerås, Sweden; and Institute for Futures Studies, Holländargatan 13, 101 31 Stockholm, Sweden

Luis Gustavo Nardin
Institute of Cognitive Sciences and Technologies, National Research Council, Rome 00185, Italy and Brandenburg University of Technology, 03046 Cottbus, Brandenburg, Germany

Javier Antonio Montoya
Grupo de Modelado Computacional, Universidad de Cartagena, Cartagena de Indias, Bolívar 13001, Colombia and The Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy

(Received 18 September 2017; published 30 April 2018)

Cooperation is central to the success of human societies as it is crucial for overcoming some of the most pressing social challenges of our time; still, how human cooperation is achieved and may persist is a main puzzle in the social and biological sciences. Recently, scholars have recognized the importance of social norms as solutions to major local and large-scale collective action problems, from the management of water resources to the reduction of smoking in public places to the change in fertility practices. Yet a well-founded model of the effect of social norms on human cooperation is still lacking. Using statistical-physics techniques and integrating findings from cognitive and behavioral sciences, we present an analytically tractable model in which individuals base their decisions to cooperate both on the economic rewards they obtain and on the degree to which their action complies with social norms. Results from this parsimonious model are in agreement with observations in recent large-scale experiments with humans. We also find the phase diagram of the model and show that the experimental human group is poised near a critical point, a regime where recent work suggests living systems respond to changing external conditions in an efficient and coordinated manner.

DOI: 10.1103/PhysRevE.97.042321

I. INTRODUCTION

Cooperation is crucial to human social life, from friendship and professional relationships to political participation and global level issues such as ecological conservation and international relations. Yet cooperation is often individually costly, making it inherently fragile. Many scholars have thus concentrated on understanding how to sustain it. Mechanisms such as reputation [1], communication and sanction [2], and social identity-related factors [3] have been found to play a key role in promoting human cooperative behavior.

Solid empirical and fieldwork evidence has been mounting to suggest that social norms are successful in the provision and maintenance of cooperation in everyday life [4–8]. Social norms are informal rules that prescribe what individuals ought or ought not to do and are typically enforced through informal sanctions, such as ostracism, negative gossip, shame, or disapproval [9–11]. They sustain behavior through shared beliefs and reciprocal expectations regarding the appropriate actions to perform in specific circumstances. Indispensable to social life, they are referred to as the cement [11] or the grammar of society [9].

Despite their importance, a rigorous and well-grounded model of how social norms affect human cooperative behavior is still lacking (see Sec. II). Using statistical-physics techniques and consistent with findings from the cognitive and behavioral sciences [5,9,12,13], we develop here an analytically tractable model in which the decision makers' utility is based on a balancing between the material rewards they obtain and the degree to which their action is in agreement with social norms. We explicitly incorporate the human ability to be sensitive to social norms, their so-called norm psychology [12], into the experience weighted attraction (EWA) [14] framework. Experience weighted attraction is a modeling approach that combines both reinforcement learning [15] and belief

*john.realpe@gmail.com

learning [16] that has been extensively explored in the field of behavioral economics and has been rather successful in explaining the interactive learning of humans in games [17,18].

Results from our cognitively inspired model are in agreement with observations from recent large-scale experiments with humans (625 subjects) playing simultaneously large-scale prisoner's dilemma (PD) games [19]. The model quantitatively reproduces both the global cooperation level (i.e., a decay from an initial value of 60% to around 35%) and the final distribution of agents according to their probability of cooperation. Other attempts we know of are reported in Refs. [20–24] but, except for Ref. [24], the focus of these works is on a qualitative rather than a quantitative understanding. The experiments studied in Ref. [24] have some differences from the type of experiments we analyze here, rendering a careful comparison more difficult (see Sec. II for a discussion of the main differences). Furthermore, the models presented in these works are not necessarily based on empirically grounded cognitively motivated assumptions, like the one that we introduce here.

Our model is also parsimonious enough to allow for a detailed characterization of its long-term dynamics. We identify three parameter regimes where the system can be monostable, bistable, or remain out of equilibrium. Such regimes are separated by surfaces that terminate on a line of critical points, where it is well known that systems can develop long-range correlations and become highly responsive to external stimuli [25–36].

Our findings suggest that groups of individuals who base their choice to cooperate on a balance between selfishness and compliance with social norms are poised near a critical point, where their capacity to respond efficiently to changing and widely diverse external conditions can be enhanced [26]. This experimental evidence that human cooperative groups may operate near criticality (see, e.g., Sec. IV of the recent review in Ref. [25] for a detailed description of relevant works) points to an unexplored feature of human cooperation that may suggest a way in which social norms, besides promoting cooperation, can also enhance the ability of human groups to adapt to external variability. Similar results have been found in experiments with ants [28].

This work is outlined as follows. In Sec. II we discuss previous research and provide an overview of the different components and assumptions of our agent-based model. In Sec. III we describe the learning component of the model. In Sec. IV we describe how agents make decisions on whether to cooperate or not by balancing individual and normative considerations. In Sec. V we make use of two further assumptions consistent with experiments, i.e., slow adaptation and the absence of network reciprocity. These assumptions allow us to turn the stochastic agent-based model on networks presented in Sec. IV into a four-parameter deterministic model of a single representative agent. In Sec. VI we determine the phase diagram of the effective single-agent model obtained in Sec. V and show that the model can display critical phenomena. In Sec. VII we extract the parameters of the effective single-agent model from experimental data and show that human groups playing in the experiments are posed near criticality. In Sec. VIII we summarize and present the conclusions of the work. In the Appendixes we present further technical details.

II. PREVIOUS WORK AND MODEL OVERVIEW

While there is a great deal of literature on physics-based models of human cooperation (see, e.g., Ref. [37] for a recent review), most of these models are theoretical works that do not take into account experimental evidence. Almost a decade ago a relevant review article [38] noted that the "contribution of physicists in establishing social dynamics as a sound discipline grounded on empirical evidence has been so far insufficient." In a recent perspective review [39], one of the leading researchers in the field remarked that "there are many relevant experimental results on cooperation on structured populations published in widely read journals while, unfortunately, many models are introduced in the literature without taking into account [such experimental] facts."

As summarized by Sánchez [39], some of the most relevant experimental findings are that (i) lattices or networks do not support cooperation; (ii) individuals display moody conditional cooperation (MCC), i.e., when deciding to cooperate subjects are responsive to the behavior of others, but only if they have cooperated themselves; (iii) individuals do not take into account the earnings of their neighbors; and (iv) cooperation can be sustained in dynamic networks. Indeed, as pointed out in the Introduction, we have identified only a few references [20–24] that have attempted to build empirically grounded models to explain the type of experiments we analyze here. However, except for Ref. [24], the focus of these works was on obtaining a qualitative understanding of the phenomena observed in this type of experiments.

In contrast, our work, as well as that by Horita *et al.* [24], is a quantitative study. Horita *et al.* [24] compare the explanatory power of models of conditional cooperation [40,41] and their moody variant (MCC) [19] to reinforcement learning models in explaining cooperation under multiplayer social dilemma games. They fit these models to empirical data obtained from behavioral experiments, namely, prisoner's dilemma and public goods games. However, because their experiments have some differences from the type of experiments we analyze here, a careful comparison is difficult. For instance, while we analyze experiments with 625 subjects interacting on a network during 52 rounds, Horita *et al.* study experiments in which 100 individuals interact during 20 rounds either within fixed groups of four people or with groups of four individuals chosen at random. The authors then aggregate the decisions made by individuals of all groups during all the rounds into a single data set [see, e.g., Eq. (13) in Ref. [24]]. It is not clear to us whether some relevant dynamical information is not lost in this aggregation process. In contrast, we extract our model parameters from relevant statistical features of individual large-scale experiments, using techniques that explicitly acknowledge the dynamical nature of our model [see, e.g., Eq. (E1)].

Horita *et al.* [24] provide evidence that (model-free) reinforcement learning algorithms where agents have no access to information about decisions made by their neighbors can account for the observed human behavior roughly as accurately as algorithms where agents can directly encode the MCC rule. This result is particularly evident in those treatments in which subjects interact with different people at every stage, i.e., where norms and expectations about the actions of others are more difficult to emerge. This finding is consistent with evidence

TABLE I. Summary of assumptions underlying the model presented here. The first block of assumptions corresponds to the EWA model introduced in Ref. [14], restricted to the special case discussed in the Supplemental Material of Ref. [44] (see Sec. I therein; cf. Ref. [45]). More specifically, (i) while EWA allows for belief and reinforcement learning to have different weights, here they have the same weight [see Eq. (2)]; (ii) while EWA allows for the interpolation between average and cumulative reinforcement learning, here the focus is exclusively on cumulative reinforcement learning [see Eq. (2)]. These lead to a model characterized by the drive [see Eq. (2)] and two parameters: (i) parameter α, which captures the exponential decrease of the relevance of past events, i.e., if $\alpha = 1$ agents only remember what happened in the previous round, while if $\alpha = 0$ agents have cumulative information of the full history of play, and (ii) parameter β, which captures the success of agents in choosing the optimal strategy, i.e., if $\beta \gg 1$ agents usually choose the optimal strategy, while if $\beta = 0$ agents choose strategies at random. The EWA model is based exclusively on self-regarding considerations. The second block of assumptions extends the EWA model to include norm-based considerations. The importance agents give to normative considerations is characterized by the parameter h; if $h = 0$ (if $h \gg 1$) only individual (normative) considerations matter. The normative component implements three processes characterized by the parameters w_C, w_O, and w_I. The more the norm is perceived as salient, i.e., relevant, to the agent the higher its impact on the agent's decision. These parameters determine how the norm salience is updated. The parameter w_C, however, can be absorbed in the parameter h, so we take $w_C = 1$. These two blocks of assumptions lead to a stochastic agent-based model where interactions take place on a given network. The third block of assumptions transforms the model into a deterministic analytically tractable model of a single representative agent characterized by the parameter α and three effective parameters [see Eqs. (12)–(17)]; these three effective parameters fully specify the long-term dynamics of the model (see Sec. IV).

Assumption	Description	Representation	Reference
	First block		
bounded rationality	agents do not always play the optimal strategy	β in Eq. (1)	[14,44,45]
belief learning	agents learn from what could have *potentially* happened	Eq. (2)	[14,44,45]
reinforcement learning	agents learn from what *actually* happened	Eq. (2)	[14,44,45]
memory decay	agents give more relevance to recent events	α in Eq. (2)	[14,44,45]
selfishness	agents base their decisions on self-regarding considerations	$\Delta I_C, \Delta I_D$, Eqs. (3) and (4)	[14,44,45]
	Second block		
norm conformity	agents base their decisions *also* on social norms	h in Eqs. (3) and (5)	[46,47]
—self-consistency	agents are consistent with own beliefs and self-ascribed norms	w_C in Eq. (5)	[48–50]
—social influence	norm compliance increases with the number of compliant peers	w_O in Eq. (5)	[40,51]
—moody conditional cooperation	social influence is stronger if aligned with self-consistency	w_I in Eq. (5)	[52]
	Third block		
slow adaptation	adaptation happens over several individual strategic choices	Eqs. (10) and (11)	[19,53]
no network reciprocity	interaction structure does not significantly influence behavior	Eqs. (12) and (13)	[19,51,53–55]

from the cognitive and behavioral sciences that inspired our model [5,40], showing that, although reinforcement learning plays an important role in governing human behavior, when involved in repeated and long-term interactions with the same people, individuals' choices are not independent of other people's behavior, but highly conditional on what they believe others will do.

In the experiments analyzed by Horita *et al.* and those we analyze here, the information about a neighbor's decisions necessary to compute the normative reasoning (see Sec. IV C) can in principle be extracted from the material payoffs. So it is not unreasonable to expect that in these experiments subjects can indirectly infer normative information from material payoffs only, as suggested by Horita *et al.* A possible way to resolve this ambiguity in the future could be to design experiments where this peculiar situation does not hold.

However, theoretical and empirical evidence suggests that human strategic behavior is based not only on model-free reinforcement learning, but also on model-based reinforcement learning (i.e., belief learning) [14,42,43]. These two types of algorithms are related to habits that subjects acquired from past experiences and goals that they expect to achieve in the future, respectively. In contrast to Horita *et al.*, our EWA-inspired model is a hybrid between these two learning algorithms. Although we assume an equal weight for both model-free and model-based reinforcement to simplify the analysis, the EWA component of our model can be easily generalized to incorporate the desired weight to each of these two algorithms. In future studies, this could be used to investigate which of the two approximations is more accurate, i.e., assuming all weight on model-free reinforcement learning, as done in Ref. [24], or assuming equal weight for both model-free and model-based reinforcement learning, as we do here.

Additionally, the model we present here encodes empirically grounded cognitive assumptions, as summarized in Table I. The first block of assumptions in Table I is specific to the EWA learning algorithm (see Sec. III). While Refs. [20–22] implement a heuristic evolutionary dynamics, none actually

implements the EWA learning dynamics [14], which is based on empirically sounder cognitive assumptions. Indeed, in Ref. [20] the authors recognize that "the original formulation of EWA cannot be trivially generalized to our MCC scenario" and aim to reproduce key features of the EWA updating by a linear combination of belief and reinforcement learning (see Supplemental Material of Ref. [20] under the section titled "SI EWA"). Experience weighted attraction, however, is known to be a better model than such a mixture (see, e.g., item 3 on p. 323 of Ref. [56]). Furthermore, in EWA agents learn solely from what they earned or could have earned, in agreement with experimental finding (iii) above.

The second block of assumptions in Table I is specific to the normative component. These assumptions rely on theoretical and empirical studies showing that human decisions are not only driven by selfish considerations but also influenced by social norms (i.e., informal social rules prescribing what individuals ought or ought not to do [12,13,46]). Moreover, those postulations are also aimed to account for the fact that the more salient, i.e., relevant, the norm is perceived to be, the stronger its impact on the individual's motivation to comply with it. Vilone *et al.* [22] point out that the interplay of social and strategic motivations in human interactions is a largely unexplored topic in collective social phenomena. They implement a heuristic algorithm where, at each iteration, agents choose with a certain probability either to update their strategy by imitating a neighbor picked at random or to update their strategy based on strategic considerations. In addition to being heuristic, i.e., not necessarily based on empirical evidence, in the strategic component of their update rule agents take into account the earning of their neighbors in contrast with the experimental finding (iii) above.

In addition to respecting experimental findings (iii), our model incorporates empirically grounded normative assumptions into the EWA framework while still conserving its general structure (see Sec. IV). Apart from being affected by the expectations and actions of their peers, individuals' decision to cooperate depends also on their mood. Consistently with experimental finding (ii) above, when deciding to cooperate the agents are responsive to the behavior of others, but only if they have cooperated themselves.

The first two blocks of assumptions lead to a stochastic agent-based model in which agents interact on a given (static) network and balance individual and normative considerations in their decision-making. It should not be difficult to extend this model to incorporate dynamical networks that can also take into account experimental finding (iv) above. However, we here restrict our analysis to static networks, which allows for further simplifications.

Finally, the third block of assumptions is also consistent with experimental evidence. Indeed, the nearly linear trend that usually characterizes the MCC rule in the type of experiments we analyze (see, e.g., Figs. 3 A and 3 B in Ref. [19]) is consistent with a relatively large randomness in agents' strategic choices (i.e., the parameter $\beta \ll 1$; see Table I). This implies that the timescale associated with individual strategic choices is smaller than the timescale on which adaptation happens, i.e., adaptation can be assumed slow; this assumption allows us to turn the stochastic agent-based model into a deterministic one (see Sec. V). The second assumption in this block exploits experimental finding (i) to turn the resulting deterministic model of agents interacting on a (static) network into an effective four-parameter model of a single representative agent (see Sec. V). This four-parameter model is parsimonious enough to allow for the analytical determination of its long-term dynamics (see Sec. VI). We emphasize once again that this effective model is restricted to the study of interactions on static networks.

A remark is in order: While the reinforcement learning algorithms studied by Horita *et al.* [24] conserve the identity of the individuals, our mean field model is based solely on a single representative agent. While our single-agent model depends on four parameters, the two models studied by Horita *et al.* depend only on two or three parameters. However, our mean field four-parameter model is parsimonious enough to allow for the analytical characterization of its different dynamical regimes. Avoiding the adiabatic and mean field approximations described above, as well as the equal weights between model-free and model-based reinforcement learning, we could turn our model into a stochastic agent-based model that includes as a special case the two-parameter model studied by Horita *et al.* In the future, such a more general model could be used, along with model selection techniques, to better compare ours with the work of Horita *et al.*

III. LEARNING ALGORITHM

Here we describe the learning component of the EWA model, which incorporates the first block of assumptions in Table I. In the next section we discuss how to extend this model to include normative considerations in the agents' decision-making.

Theoretical and empirical evidence shows that human social strategic behavior is based on a combination of model-free and model-based reinforcement learning algorithms [14,42,43]. These models are related to habits that subjects acquired from past experiences and goals that they expect to achieve in the future, respectively. Under some circumstances [44,45], carefully described in the first section of the Supplemental Material of Ref. [44], this can be captured by a simplified form of the EWA model [14]. Experience weighted attraction is a modeling approach that combines both reinforcement learning [15] and belief learning [16]. The former refers to reinforcing actions based on agents' past performance and the latter refers to reasoning about how actions that have not been chosen would have performed. One of the key insights provided by EWA is that belief learning can also be understood as a process in which actions are reinforced by forgone payoffs. In this sense, EWA is a combination of model-free and model-based reinforcement learning [57]. The simplified EWA model [44,45], which we are interested in here, can be described by the equations

$$x_i(t+1) = \frac{1}{1+e^{-\beta D_i(t+1)}}, \tag{1}$$

$$D_i(t+1) = (1-\alpha)D_i(t) + \Delta U_i(t). \tag{2}$$

Here $x_i(t+1)$ and $D_i(t+1)$ are, respectively, the probability and drive of agent i to cooperate at round $t+1$. When the parameter $\beta \geqslant 0$ is large, agent i tends to cooperate or defect

if the drive is positive or negative, respectively; if instead the drive or β is zero, the agent acts randomly. Intermediate values of β interpolate between these two extremes: rational optimization and random behavior. The term $\Delta U_i(t)$ in Eq. (2) is the difference in utilities resulting from the choice of either cooperate or defect. If $\Delta U_i(t) > 0$, agent i's drive to cooperate increases; if $\Delta U_i(t) < 0$, the drive decreases, while if $\Delta U_i(t) = 0$ it stays the same. Finally, the parameter α describes memory loss: If $\alpha = 1$, the agent remembers only the previous round t, while if $\alpha = 0$, the agent has cumulative information of the full history of play. The case of $0 < \alpha < 1$ amounts to an exponential discount of utility over time.

While the EWA model assumes that agents' motivation to cooperate $\Delta U_i(t)$ is specified exclusively by individual considerations, namely, material payoffs, in this work we extend the EWA formalism to incorporate normative considerations, as described in the next section.

IV. BALANCING INDIVIDUAL AND NORMATIVE CONSIDERATIONS

Here we discuss how we extend the EWA model (see Sec. III) to make agents balance between individual and normative considerations in their decision-making. The individual component is common to previous EWA models and implements the assumption of selfishness in Table I, while the normative component is introduced in this work and implements the second block of assumptions in Table I. We combine both considerations by defining agents' motivation to cooperate $\Delta U_i(t)$ as a weighted sum of the individual and normative components. This combination is also consistent with experimental observations suggesting that a common area in the brain correlates with the computation of both monetary and social rewards [58] (see also Ref. [59]). The idea that norms can be conceived as part of the utility function that individuals maximize has been receiving growing attention and empirical support [13,22,59–62]. In these works, however, social norms typically have an exogenously specified impact on individuals' behavior, while we assume that this impact is endogenously updated on the basis of how salient the specific norm is perceived to be within a certain group.

A. Decision rule to cooperate

Decades of theoretical and experimental work are nowadays putting on solid ground that when deciding whether to cooperate, humans do not always make choices that maximize their individual payoffs, but they also care about behaving in line with others in their group. Social norms accurately provide information about how members of a certain group will behave and more importantly about how they are prescribed to behave [7,9,63].

Consistent with this evidence and in analogy with Ref. [46], we develop a model in which the decision makers' utility $\Delta U_i(t)$ is based on the material rewards they obtain and on the degree to which their actions comply with social norms. Thus,

$$\Delta U_i(t) = \Delta I_i(t) + h\Delta N_i(t), \tag{3}$$

where the individual drive $\Delta I_i(t)$ models the motivation to maximize personal material payoffs and the normative drive $\Delta N_i(t)$ models the motivation to comply with social norms (see Sec. IV C below for more details). The parameter h weights the relative influence of selfish and norm-based motivations on cooperative decision-making: If $h = 0$, agents do not care about normative information, while if h is very large, agents' behavior is dominated by what the norm dictates. It also defines a relative timescale between selfish and prosocial reasoning, related to reflection and intuition [64].

In this way we have incorporated the ability to balance between normative and selfish considerations into the EWA model, keeping the standard EWA formalism almost intact. The impact of both types of considerations on individual decisions has been scarcely explored (see Refs. [13,22,59,62,64]).

B. Individual component

Here we describe the individual component of the model, which implements the assumption of selfishness in Table I. We also describe the specific case of the PD game because this is the game used in the experiment we analyze in Sec. VII. Evidently, other types of games can also be implemented by defining the payoffs accordingly.

We are interested in the situation where agents interact pairwise by playing a given two-player game with each neighbor in a social network [19,53]. In this case, we can write the individual motivation for agent i to cooperate in her interaction with her neighbor j at round t as $\Delta I_{ij}(t) = s_j(t)\Delta I_C + [1 - s_j(t)]\Delta I_D$. Here $s_j(t)$ refers to the strategy played by agent j at round t, i.e., whether she cooperated $s_j(t) = 1$ or defected $s_j(t) = 0$, while $\Delta I_C = R - T$ and $\Delta I_D = S - P$, where R is i's reward's payoff when both agents cooperate, P is i's punishment's payoff when both agents defect, S is i's sucker's payoff when i cooperates and j defects, and T is i's temptation's payoff when i defects and j cooperates.

The total payoff received by an agent i interacting with K neighbors is given by the average over the payoffs obtained on each of the K pairwise games the agent is involved in. So the individual drive of agent i to cooperate at round t is

$$\Delta I_i(t) = (\Delta I_C - \Delta I_D)\frac{1}{K}n_i(t) + \Delta I_D, \tag{4}$$

where $n_i = \sum_{j\in\partial i} s_j$ refers to the number of i's peers who cooperate, with ∂i standing for the set of neighbors of i in the social network.

If the payoffs satisfy $T > R > P > S$, we have the PD game; furthermore, $2R > T + S$ for iterated PD games. The structure of social dilemma is the following: Although the best individual choice for both is to defect, mutual cooperation yields a better payoff than mutual defection ($R > P$). The experiments that we analyze here correspond to a weak PD game where, in experimental currency units (ECUs), $R = 7$ ECUs, $T = 10$ ECUs, and $P = S = 0$ ECUs; so $\Delta I_C = -3$ ECUs and $\Delta I_D = 0$ ECUs [19].

C. Normative component

Here we describe the normative component of the model, which implements the second block of assumptions in Table I.

We also define the entire utility function [see Eq. (3)] for the specific case of the PD game.

The impact that social norms have on an agent's decisions is a function of how salient [46,47] the norm is perceived by agent i at round t within the social group. The higher the salience of the social norm, the stronger its impact on the motivation to comply with it. The norm salience is determined by two independent factors, weighted with the parameters $w_C, w_O \geqslant 0$, and their interaction, weighted with the parameter $w_I \geqslant 0$, i.e. (cf. [46]),

$$\Delta N_i(t) = w_C[2s_i(t) - 1] + w_O \frac{n_i(t)}{K} + w_I s_i(t) \frac{n_i(t)}{K}. \quad (5)$$

According to the first term, the salience of a norm is determined by the behavior at round t, namely, the agent's choice to comply with or violate the norm. If agent i complies with the norm, she will perceive it as more salient than if she violates it. This is justified by the fact that humans have a strong need to enhance their self-concepts by behaving consistently with their own beliefs and self-ascribed habits so that they can avoid ethical dissonance (self-consistency in Refs. [48–50]).

In line with theoretical and empirical findings on conditional cooperation [40,51], the second term containing $w_O > 0$ assumes that the salience of the norm is also affected by the share of peers that complied with it. The more peers comply with the norm, the more salient the norm becomes, and vice versa. The third term containing $w_I > 0$ disappears if the agent did not cooperate at round t [i.e., if $s_i(t) = 0$]; while not present in Ref. [46], this last term is introduced in this work to account for recent experimental observations that support the MCC rule, which assumes that in making decisions individuals are responsive to the behavior of others, but only if they have cooperated themselves [52]. It can be noticed that the second term relaxes the assumption behind the MCC rule by positing that when deciding whether to cooperate individuals are always sensitive to what others do and not just after having cooperated themselves (social influence). This relaxation is in line with recent findings reported in Ref. [24].

Comparing Eqs. (4) and (5), we can see that the information required about neighbors' action to estimate norm salience, i.e., $n_i(t)$, could in principle be inferred from the information on material payoffs. So it is not unreasonable to expect that agents can indirectly infer normative information from material payoffs only, as suggested by Horita *et al.* [24].

Now, introducing Eqs. (4) and (5) into Eq. (3), we get

$$\Delta U_i(s_i, n_i) = (as_i + b)n_i + 2hs_i - h, \quad (6)$$

where

$$a = hw_I/K, \quad (7)$$

$$b = (hw_O + \Delta I_C)/K \quad (8)$$

are effective parameters introduced to simplify the notation. We have dropped the index t to include explicitly the dependence of ΔU_i on the number n_i of agent i's peers who cooperated at round t. Furthermore, we took $w_C = 1$, as it can be absorbed in the parameter h, and $\Delta I_D = 0$ as we will focus our analysis on the weak PD game studied in Ref. [19].

V. SLOW ADAPTATION AND ABSENCE OF NETWORK RECIPROCITY

Here we describe how to implement the third block of assumptions presented in Table I and how to obtain an effective four-parameter model of a single representative agent. Our interest in such an effective model is that it allows for the complete analytical characterization of its long-term dynamics, indicating the existence of critical phenomena, while still quantitatively reproducing major features of large-scale experiments with human groups.

As already discussed in Sec. II, the large-scale experiments analyzed here are consistent with the assumption that adaptation is slow in comparison to the rate of change of individual strategic choices. Fluctuations around agents' average behavior induced by their stochastic nature can then be neglected (see Appendix A). This so-called adiabatic [44,45] approximation allows us to replace the stochastic variable s_i encoding the actual strategy chosen by each agent i for its mean value x_i, which is a deterministic quantity.

To see this, notice that by introducing Eq. (2) into Eq. (1) the system dynamics can be fully specified in terms of the cooperation probability as

$$x_i(t+1) = \frac{x_i(t)^{1-\alpha}}{x_i(t)^{1-\alpha} + [1 - x_i(t)]^{1-\alpha} e^{-\beta \Delta U_i(t)}}. \quad (9)$$

Replacing the stochastic term $\Delta U_i(t)$, which depends on the actual actions s_i and $s_{\partial i}$ of agent i and her neighbors ∂i [see Eq. (6)], by its average value $\overline{\Delta U_i}(t)$, which is obtained by changing each action s_j by its corresponding average x_j, we get the deterministic equation

$$x_i(t+1) = \frac{x_i(t)^{1-\alpha}}{x_i(t)^{1-\alpha} + [1 - x_i(t)]^{1-\alpha} e^{-\beta \overline{\Delta U_i}(t)}}. \quad (10)$$

More precisely,

$$\begin{aligned} \overline{\Delta U_i}(t) &= \sum_{s_i, s_{\partial i}} \Delta U_i(s_i, n_i) p_i(s_i, t) \prod_{j \in \partial i} p_j(s_j, t) \\ &= a x_i(t) \sum_{j \in \partial i} x_j(t) + b \sum_{j \in \partial i} x_j(t) + 2h x_i(t) - h, \end{aligned} \quad (11)$$

where $n_i = \sum_{j \in \partial i} s_j$, the term $s_{\partial i}$ denotes the set of strategies of i's peers, $p_j(s,t) = [x_j(t)]^s [1 - x_j(t)]^{1-s}$ is the probability that agent j plays strategy s, and we have used the expression for $\Delta U_i(s_i, n_i)$ in Eq. (6).

On the other hand, the interaction structure of a human group does not appear to significantly influence its cooperative behavior [19,51,53–55]; this is usually referred to as the absence of network reciprocity, which is the influence of network structure on cooperative behavior [65]. This implies that correlations between different agents can then be assumed to be weak, which leads to a mean field approximation [27], where $\sum_{j \in \partial i} x_j \approx xK$. Here x is the global mean value of x_i calculated over all agents i and K is the average number of neighbors of a generic agent i. This approximation allows us to describe the system in terms of a single representative agent that captures the typical behavior of a generic agent i. In this way, we obtain a deterministic learning dynamics of a single

representative agent given by the equation (see Appendix A and the first two sections in the supporting information of Ref. [44] for further details)

$$x(t+1) = \frac{x(t)^{1-\alpha}}{x(t)^{1-\alpha} + [1-x(t)]^{1-\alpha} e^{-\beta \overline{\Delta U}[x(t)]}}, \quad (12)$$

where x is the probability for the representative agent to cooperate and

$$\overline{\Delta U}[x] = aKx^2 + (bK + 2h)x - h \quad (13)$$

is obtained by replacing in Eq. (3) both s_i and n_i/K with the average value x. Equation (12) describes the relevant aspects of the dynamics of the global cooperation level and can reproduce the values observed in Ref. [19] with accuracy comparable to more complex models [20,66] (see Sec. VII).

VI. DYNAMICAL REGIMES AND PHASE DIAGRAM

Here we determine the phase diagram characterizing the long-term dynamics of the effective single-agent model described in detail in the previous sections; this diagram shows three regimes, monostability, bistability, and nonequilibrium, as well as a line of critical points.

To study the long-term dynamics of the model defined in Eq. (12), we look for fixed points, i.e., points x that satisfy $x(t+1) = x(t) = x$. The points at the boundary, i.e., $x = 0$ and $x = 1$, are fixed points of the mean field dynamics described by Eq. (12), but they are unstable since $\alpha, \beta > 0$. Only fixed points x^* satisfying $0 < x^* < 1$ can be stable. The condition that these points satisfy can be derived from Eq. (12) by setting $x(t+1) = x(t) = x$, which yields

$$x = f(x) \quad \text{with } f(x) = \tfrac{1}{2} + \tfrac{1}{2}\tanh[A(x - x_0)^2 + y_0], \quad (14)$$

where

$$A = \frac{aK}{2\gamma}, \quad (15)$$

$$x_0 = -\frac{bK + 2h}{2aK}, \quad (16)$$

$$y_0 = -\frac{(bK + 2h)^2}{8aK\gamma} - \frac{h}{2\gamma}, \quad (17)$$

and $\gamma = \alpha/\beta$. In Eq. (14) we have not made explicit the dependence of the function f on the effective parameters A, x_0, and y_0 to reduce clutter in the notation.

If the MCC assumption is dropped, i.e., $w_I = 0$ so $a = 0$, Eq. (14) becomes equivalent to the equation that determines the equilibrium magnetization, given by $m = 2x - 1$, of the Curie-Weiss model [67]. Indeed, when $a \to 0$ Eq. (14) can be written as $m = \tanh[\beta(J_{\text{eff}} m + H_{\text{eff}})]$, where $J_{\text{eff}} = (hw_O + \Delta I_C + 2h)/4\alpha$ would correspond to an effective ferromagnetic interaction (when $J_{\text{eff}} > 0$) and $H_{\text{eff}} = (hw_O + \Delta I_C)/4\alpha$ would correspond to an effective external field. As it is well known, the Curie-Weiss model can display two phases, paramagnetic and ferromagnetic, which are the magnetic analogs to the regimes of monostability and bistability of our model of human cooperative dynamics.

The MCC assumption ($w_I > 0$) introduces an additional nonlinearity, whose magnetic analog is an additive term of order m^2 in the argument of the hyperbolic tangent. Such term comes from the interaction between the agent's own cooperative behavior and that of her neighbors. Such additional nonlinearity renders the phase diagram of the model more complex and gives rise to a new nonequilibrium phase, where the cooperative dynamics never settles.

Indeed, as described in detail in Appendix B and shown in Fig. 1, depending on the values of the parameters A, x_0, and y_0, there can be zero, one, or two stable fixed points, corresponding to a nonequilibrium, monostable, or bistable long-term dynamics, respectively. This provides an analytical characterization of the system that helps to obtain insights into systems as complex as human groups that are typically difficult to obtain. In particular, this analytical characterization allows us to infer model parameters from experimental data and identify evidence that the human groups playing in the experiments of Ref. [19] are near criticality (see Sec. VII and Appendix E). The strategy we adopt here for the estimation of model parameters from experiments uses information about the dynamical regimes identified.

The regions of the phase diagram corresponding to the different dynamical regimes are separated by surfaces of discontinuous transitions that terminate on a line of critical points (see Appendix B). At these critical points, correlations are known to become long range [27] and systems have been shown to display a multitude of significant features, such as a large repertoire of dynamical responses, optimal transmission and storage of information, and extreme sensitivity to external perturbations [26–36].

Several mechanisms have been put forward in an attempt to explain how criticality could emerge in living systems [30]. A novel perspective posits that criticality is the evolutionary stable outcome of a group of individuals equipped with mechanisms aimed at representing each other with fidelity, wherein the best possible trade-off between accuracy and flexibility is achieved [26]. We here show evidence that mechanisms balancing between individuality and social conformity can underlie human cooperation and poise human groups near criticality.

Criticality is usually associated with the divergence of a properly defined susceptibility that quantifies the range of the correlations in the system and its response to external perturbations [27,67]. Here it can be defined in terms of the change in the global cooperation x when a certain model parameter θ varies, e.g., $\theta = h$. For illustration purposes, let us assume that the model parameters vary as $A(\theta) = A_c^* - \theta$, $x_0(\theta) = x_{0c}^*$, and $y_0(\theta) = y_{0c}^*$, for a generic parameter $\theta \gtrless 0$, where A_c^*, x_{0c}^*, and y_{0c}^* correspond to a point on the critical line. So, when approaching the critical point, i.e., when $\theta \to 0$ or $A \to A_c^*$, out of the nonequilibrium region the susceptibility is given by (see Appendix C)

$$\frac{\partial x}{\partial \theta} \propto \frac{1}{A_c^* - A}, \quad (18)$$

which clearly diverges. Notice that A varies with the original parameters of the model since it is defined in terms of them [see Eq. (15)]. In the Conclusion we discuss the implications that this feature may have for the adaptiveness of human groups.

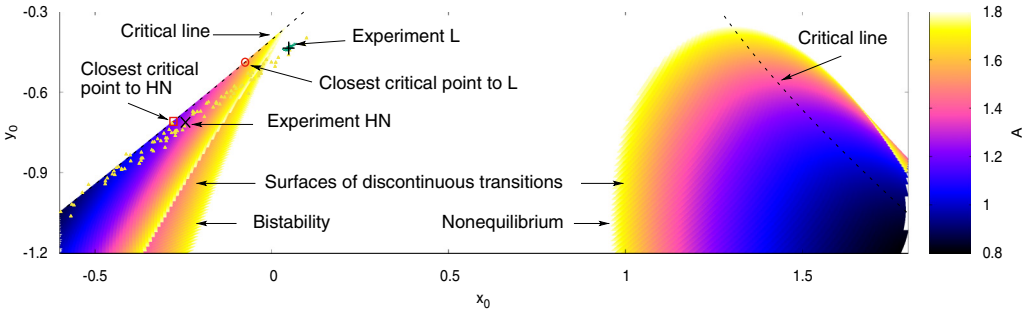

FIG. 1. Dynamical regimes of human cooperation. Colormap representation of the surfaces of discontinuous transitions defined by a function $A_d(x_0,y_0)$ that returns the transition value of A for each value of the effective parameters x_0 and y_0, defined in Eqs. (16) and (17), respectively. Each (color) gray level encodes a level curve $A_d(x_0,y_0) = A$, which partitions the x_0-y_0 plane into three regions corresponding to different long-term dynamical regimes: Inside the sharp triangularlike curve (left) the system is bistable; inside the paraboliclike region (right) the system never reaches equilibrium (see also Fig. 3 in Appendix B). Outside these two regions the system is monostable (see Appendix B). The cloud of densely packed (green) dots around the plus and the cloud of sparsely scattered (yellow) little triangles around the cross represent a projection on the x_0-y_0 plane of the posterior population of parameters α, A, x_0, and y_0 inferred here from the experiments performed in [19] on a heterogeneous network (HN) and on a lattice (L), respectively. We show the parameters estimated (HN, diagonal cross; L, vertical cross) and quantify their relative distance (HN, 3%; L, 11%) to the closest point (HN, square; L, circle) on the critical lines (dashed) (see Appendix E).

VII. ANALYSIS OF LARGE-SCALE EXPERIMENTS OF HUMANS PLAYING A PRISONER'S DILEMMA

Here we use experimental data from Ref. [19] to determine the parameters of the effective single-agent model described in Secs. II–V and locate the human group playing in the experiment into the phase diagram obtained in Sec. VI (see Fig. 1).

A. Brief review of experiments analyzed

To estimate where human groups may be located in the phase diagram of Fig. 1, we extracted the model parameters from two recent large-scale experiments in which more than 600 human participants play a prisoner's dilemma game simultaneously in two different network environments [19]. These experiments are aimed at testing the relative effect of homogeneous or heterogeneous network environments on cooperative behavior (for details see Appendix E). We build on these experiments because we expect them to offer more robust statistics than similar but smaller experiments.

In Ref. [19], one of the two experiments was conducted on a square lattice and the other on a heterogeneous network. However, their finding that network structure does not significantly affect behavior (i.e., the absence of network reciprocity) suggests that even though our mean field model neglects network structure, it can still provide a good description of the experiments, as shown below.

In these experiments, human subjects played a 2×2 multiplayer PD game with each of their K neighbors for 52 rounds. Players could take only one action, either to cooperate C or defect D, the action being the same against all the opponents. The experiment was simultaneously carried out on two different virtual networks: The first network consisted in a 25×25 lattice with a fixed number of 4 neighbors and periodic boundary conditions (625 subjects); the second network was a heterogeneous network with a fat-tailed degree distribution (604 subjects), where the number of neighbors varied between 2 and 16.

Subjects played a repeated (weak) prisoner's dilemma game with all their neighbors for an initially undetermined number of rounds. Payoffs were set to be 7 ECUs for mutual cooperation, 10 ECUs for a defector facing a cooperator, and 0 ECUs for any player facing a defector.

Participants received information about the actions and normalized payoffs of their neighbors in the previous round. Without knowledge of the duration of the game, participants had to make only one decision for all neighbors. Therefore, the situation becomes similar to a repeated public goods game. In public goods experiments, participants usually start cooperative, but in the absence of cooperation-enhancing mechanisms, such as punishment or reputation, their cooperation levels decrease over time. Information about the behavior of others allows participants to create expectations about how others will behave, namely, about the social norms ruling the group.

We focus here on two features observed in these experiments that can be reproduced by our model (Fig. 2). The first feature is the dynamics of the global cooperation level, which decays from an initial value of about 60% to a relatively constant value of about 35%, both on the heterogeneous network and on the square lattice [Figs. 2(a) and 2(b)]. The second feature is the probability $P(C|s,n)$ for a generic individual to cooperate C in a generic round, conditioned by her previous action s and the number n of neighbors who cooperated in the previous round [19,51,53–55]. Reference [19], for instance, reports a nearly linear dependence of $P(C|s,n)$ on n for both values of s [see Figs. 2(c) and 2(d)].

B. Inference of model parameters

To fit our model to the experimental data, we notice that the left-hand side of Eq. (1) for the representative agent can be interpreted as $x(t+1) = P(C,t+1|s,n,x,t)$, namely, the probability that the agent cooperates at round $t+1$ given that at round t the following three conditions are satisfied: (i) She played strategy s, (ii) n of her neighbors cooperated, and (iii) $x(t) = x$. To eliminate the explicit dependence on the history of the system, i.e., on x and t, we first assume that the system

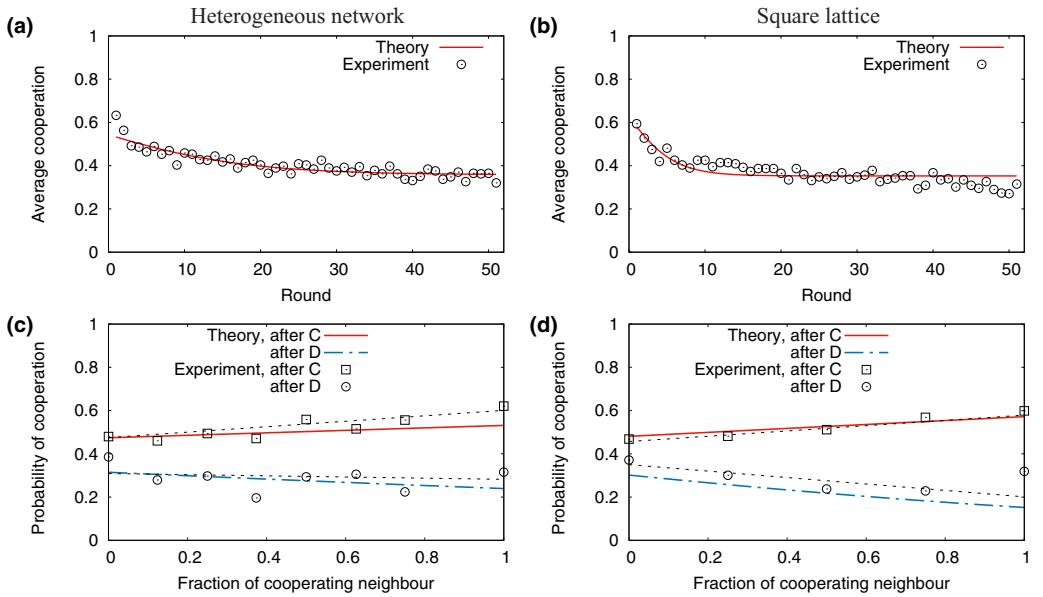

FIG. 2. (a) and (b) Balance of selfishness and conformity to social norms explains human behavior in large-scale prisoner's dilemma games. Shown is a comparison of the dynamics of the global cooperation level observed in the laboratory experiment conducted in [19] (circles) on (a) a heterogeneous network and (b) a square lattice with that predicted by Eq. (12) (line) with the corresponding parameters inferred from the same experiments (see the caption of Fig. 1). The model and experimental dynamics are in agreement even in the transient regime. Also shown is the empirical probability for a representative agent to cooperate in a generic round based on whether she cooperated ("after C," squares) or defected ("after D," circles) and on the number of neighbors who cooperated in the previous round obtained in the same experiments [19] on (c) a heterogeneous network and (d) a lattice. The experimental data are compared with the values predicted by Eq. (19) ("after C," red upper solid lines; "after D," blue lower dash-dotted lines) with the corresponding parameters inferred from the same experiments (see the caption of Fig. 1). We can see that the assumption of linearity is valid and that our model agrees with the experimental values to a large extent. We include the linear fits (dashed straight lines) directly obtained from experimental data [19] for comparison.

of interacting humans observed in the laboratory has reached a stationary state, so it can be accurately described by the long-term mean field dynamics [Eq. (12)]. We further assume that the system is essentially monostable and, accordingly, the model dynamics is dominated by a single fixed point. These assumptions considerably simplify the analysis and, as it turns out, are self-consistent with the results obtained (see Appendix D for a more general and detailed treatment).

Although the stationary state of a generic system can depend on its dynamical history, under the above assumptions, this is not the case. Thus, the right-hand side of Eq. (1) evaluated at the fixed point, i.e.,

$$P(C|s,n) = \frac{1}{1 + y_1^{1-\alpha} e^{-\beta \Delta U(s,n)}}, \tag{19}$$

should coincide with the experimental results, where $y_1 = (1 - x_1)/x_1$, with x_1 the global cooperation level at the dominant stable fixed point. The term $\Delta U(s,n)$ is the utility function of the representative agent in the mean field approximation, which is obtained by simply dropping the agent index i in Eq. (6). Similarly, $y_1 = (1 - x_1)/x_1$, with x_1 being the only stable fixed point of Eq. (12). This result indicates that when the system is deterministic and monostable, its long-term dynamics is independent of its history. When the system is bistable and we neglect fluctuations altogether, the probability $P(C|s,n)$ is given instead by a convex combination of terms like the one on the right-hand side of Eq. (19), one for each fixed point. More details can be found in Appendix D1.

To compare Eq. (19) with the nearly linear behavior [Figs. 2(c) and 2(d)] observed in [19,54] (but see Ref. [55]), we do a first-order approximation in β to obtain

$$P(C|s,n) = m_s n/K + r_s, \tag{20}$$

with

$$m_s = \beta K J(\alpha)(as + b), \tag{21}$$

$$r_s = I(\alpha) + \beta J(\alpha)[h(2s - 1)], \tag{22}$$

where

$$I(\alpha) \equiv \frac{1}{1 + y_1^{1-\alpha}}, \tag{23}$$

$$J(\alpha) \equiv \frac{y_1^{1-\alpha}}{\left(1 + y_1^{1-\alpha}\right)^2}. \tag{24}$$

This approximation is consistent with the results below (see Fig. 2).

The slopes m_s and intercepts r_s in Eq. (20) are better described in terms of the mean intercept r and the gap G between intercepts of the near linear trends that describe the MCC rule [19], i.e.,

$$r = \tfrac{1}{2}(r_C + r_D) = I(\alpha), \tag{25}$$

$$G = r_C - r_D = 2\beta h w_C J(\alpha), \tag{26}$$

TABLE II. Experimental data for the two experiments reported in Ref. [19], which were carried out on a square lattice and on a heterogeneous network. The first four rows are extracted from Table S2 in Ref. [19] and correspond to a linear fit of Figs. 3 A and 3 B in Ref. [19], while the last row is obtained from averaging the last ten rounds in Fig. 2 A in Ref. [19] (see Sec. E 2).

Quantity	Square lattice	Heterogeneous network
m_C	0.122 ± 0.034	0.126 ± 0.039
m_D	-0.149 ± 0.050	-0.0269 ± 0.035
r_C	0.457 ± 0.015	0.475 ± 0.016
r_D	0.350 ± 0.021	0.309 ± 0.069
$\bar{x}$	0.306 ± 0.024	0.355 ± 0.021

where, for convenience, we took $m_0 \equiv m_D$, $r_1 \equiv r_C$, etc. In experiments we have $m_D \neq 0$, which implies that $b \neq 0$ as $m_D \propto b$, so

$$m_C - m_D = \beta a K J(\alpha), \tag{27}$$

$$\frac{m_C}{m_D} = \frac{\beta a + \beta b}{\beta b}. \tag{28}$$

We are now in a better position to discuss the role played by the parameters w_C, w_O, and w_I encoding the normative assumptions. First, notice that we have reintroduced the parameter $w_C = 1$ in Eq. (26) to make explicit that if the assumption of self-consistency (see Table I) were dropped, i.e., if we set $w_C = 0$, the gap G would vanish, contradicting experimental observations (see Fig. 2). Analogously, we can see from Eq. (27) that if the MCC assumption (see Table I) were dropped, i.e., if we set $w_I = 0$, which implies $a = 0$, the two slopes $m_C = m_D$ are equal, contradicting experimental observations (see Fig. 2). In this sense, the parameters w_I and w_C play not only a quantitative but also a qualitative role. In contrast, the role of the parameter w_O is more quantitative than qualitative. Indeed, Eq. (21) implies that $m_D = \beta K J(\alpha) b$. If $w_O = 0$ we have $b = \Delta I_C / K$, which is fixed by the experimental conditions, and $m_D = \beta J(\alpha) \Delta I_C$ would be less than zero for the PD game, since $\Delta I_C < 0$. While this is consistent with experimental observations, with $w_O = 0$, the accuracy of the fit was rather poor and so we did not include it in our analysis. However, future analysis should study in further detail the relevance of this assumption.

As described in detail in Appendix D 2, there is a direct relationship between the parameters of the model and the experimental quantities defined above [see Eqs. (D27)–(D30)]. So the values of m_s and r_s, extracted from experimental data [19] (see Table II), constrain the values of the model parameters. There is a further constraint: The dynamics of the global cooperation level should be consistent with experimental results [Figs. 2(a) and 2(b)].

A population of parameters satisfying the resulting set of constraints was obtained via Bayesian inference by using the package POMP [68] and is illustrated in Fig. 1. Although the two-dimensional projection of the phase diagram in Fig. 1 may suggest otherwise, they all lie in the region of monostability. The technical details and the data obtained are provided in Appendix E.

C. Results

The parameters corresponding to the two experiments (see Table IV), inferred by the method described above (see also Appendix E) are at a relative Euclidean distance of 3% and 11% to the critical line (see Fig. 1 and Appendix E 5). Figures 2(a) and 2(b) compare the levels of global cooperation observed in the laboratory experiment [19] (circles) with the ones predicted by Eq. (12) (line), informed with values extracted from Ref. [19]. Results from both the heterogeneous [Fig. 2(a)] and homogeneous [Fig. 2(b)] networks are presented. Both figures show decay in cooperation over the 52 rounds from the initial value of 60% to around 35% in both treatments. Results show close agreement of the model dynamics with the laboratory experiments. Likewise, in Ref. [19], the network topology does not have any appreciable influence on the evolution of the level of cooperation.

Figures 2(c) and 2(d) show the probability for a representative agent to cooperate in a generic round based on whether she cooperated (C, squares) or defected (D, circles) and on the number of neighbors who cooperated in the previous round. Results obtained in both the heterogeneous network [Fig. 2(c)] and lattice [Fig. 2(d)] are shown. Again, both figures indicate that the probability defined in Eq. (19) is consistent with both the experiments and the linear approximation in Eq. (20). Our model reproduces human cooperative behavior observed in large-scale laboratory experiments more accurately than the MCC behavioral rule since, as shown in Refs. [19,52], the latter is not able to reproduce the slow decay of the cooperation level when the agents did not cooperate in the immediate past.

VIII. CONCLUSION

In this work we presented a statistical-physics-based model to account for human decision processes behind cooperative behavior. In this model, the decision makers' utility is based both on the material rewards they obtain and on the degree to which their actions comply with social norms. Results from this analytically tractable model are in agreement with observations from recent large-scale experiments with humans [19]. The model closely reproduces both the global cooperation level and the final distribution of agents according to their probability of cooperation. This provides support to our hypothesis that human cooperation is the outcome of the interaction between instrumental decision-making, aimed to maximize people's economic rewards and the norm psychology humans are endowed with. In doing so, we have provided experimental evidence of the effect of social norms in promoting cooperative behavior in large groups of humans facing a social dilemma situation.

The cognitively inspired model presented encapsulates important empirical knowledge on human cooperative behavior: (i) Humans' social strategic behavior operates with both model-free and model-based reinforcement learning [14,43] that are at the basis of the EWA framework adopted, (ii) population structure does not significantly influence the cooperative outcome [19,51,53–55] that in the model led to a mean field approximation, and (iii) adaptation is slow when compared with the timescale at which individual actions change, which allows us to neglect in the model stochastic fluctuations and obtain a deterministic dynamics [see Eq. (12) and Appendix A].

The presented model is parsimonious enough to allow for a detailed characterization of its long-term dynamics. By inferring the model's parameters from experimental data extracted from Ref. [19], we show that the cooperative system is located near criticality.

Recently, evidence has been mounting that living systems, such as the human brain, insect swarms, gene expression networks, bird flocks, and fish schools [27–31], operate near critical points and this might provide them functional advantages. Far from criticality, a system can be either too stable, which may favor maladaptive behaviors, or too uncoordinated with its members behaving essentially independently of each other. In both extremes the system as a whole is not very responsive to external changes, while around a critical point it is strongly correlated and highly sensitive to changes and its capacity to respond efficiently to varying external conditions can be maximized [26].

Even though still preliminary, our evidence of signatures of criticality in human cooperative groups is in agreement with recent findings on socioecological systems showing that social norms enhance the adaptiveness of cooperative systems to social and environmental variability [4,69]. These studies report that during times of institutional and ecological volatility, social norms facilitate the management of common resources, such as forests, water, and fisheries, more than the action of formal institutions. The long-range correlations between pairs of human subjects associated with a critical point could then help explain why norm-based cooperation may enhance the adaptiveness of human groups to external change. Social norms are then crucial mechanisms for both promoting cooperation and enhancing its resilience to external perturbation.

Clearly, more theoretical and empirical work is needed to reach solid conclusions. For example, machine learning techniques, such as the maximum entropy approach in Refs. [29,70–72], can be used to carry out a complementary data-driven analysis that does not rely on expert knowledge like the model we presented here. Moreover, experiments that vary some of the relevant parameters of the model, e.g., the payoff matrices, specifically targeted to more directly address our findings need to be performed.

However, the increasing amount of similar evidence [27–32,34–36] attesting criticality in living systems seems to support the plausibility of our results. Similarly to ants [28], human groups appear to reach optimal coordination at a suitable trade-off between individuality and social conformity [28] and this makes them poised at the critical point. Social conformity increases the ability of a group to coordinate to reach the desired collective outcome. However, behavioral conformism has also the disadvantages of increasing the stability of undesirable behaviors and of decreasing the ability of the system to react to external information [28]. Thus the optimal collective performance is achieved when group members are able to balance between social conformism and individuality so that they are able to achieve a high level of coordination within the group but also to maintain robust responsiveness to external perturbations.

How would humans tune to criticality? An intriguing possibility is that humans implicitly build a model of the external world and adjust its parameters accordingly, similar in a sense to what we did here. Model-based inference techniques apparently tend to produce parameter values that are close to a critical point [73]. Model-based learning mechanisms in humans [14,42] could then influence their behavior and drive human groups towards criticality [26]. Human subjects hardly possess global structural information about their group, which may explain why the mean field model developed here is accurate enough and ultimately why no significant impact of population structure on cooperative behavior has been observed [54]. An alternative idea [28,32] posits that biological groups can tune to criticality by growing until a suitable size. If so, it may be difficult to observe signatures of criticality in experimental setups with human groups of fixed size. Another interesting question that arises is whether there may be a connection between the signatures of criticality observed here in a group of decision makers and those that have been observed in the brains powering the decision-making itself [29,36,71].

ACKNOWLEDGMENTS

We thank Carlos Gersherson for facilitating this collaboration through the FuturICT Latin American Node. We also thank Alejandro Perdomo-Ortiz, Marcello Benedetti, Luca Tummolini, and Daniele Vilone for useful discussions. G.A. was partially supported by the Knut and Wallenberg Grant "How do human norms form and change?" 2016.0167 and the Horizon 2020 Framework Programme Project PROTON "Modelling the Processes leading to Organised crime and Terrorist Networks" under Grant Agreement No.: 699824.

All authors contributed to the design of the research, the development of the model, and the writing of the manuscript. J.R.-G. led the research and did the mean field analysis, the analytical calculations, and the Bayesian parameter estimation. J.R.-G. and G.A. wrote the first draft of the manuscript. J.A.M. contributed to algorithm development and data analysis.

APPENDIX A: SLOW ADAPTATION AND ADIABATIC APPROXIMATION

A way to justify the approach leading to Eqs. (10) and (11) in the main text is assuming that the cooperation probability, or equivalently the drive in Eq. (2), changes slowly during a batch of about T rounds [44,74,75]. For small values of T, a linear noise correction to the deterministic equation give good results even for a number of players as small as two [74].

Since we are interested here in games with hundreds of players and we are focusing exclusively on observed experimental features at the aggregate level, namely, the global level of cooperation and the MCC rule, we take $T = 1$ and neglect the noise altogether. This approach is expected to be better suited for games with a sufficiently large number of agents and is not expected to necessarily describe the initial transient regime in sufficient detail. As discussed in the main text, this approach can actually describe the major features of the largest experiment to date [19] with enough qualitative and quantitative detail.

If α and β both vanish, the probability of cooperation remains constant. We will thus assume that α and β are small so that changes in the drive during a few rounds are not appreciable. The accumulated changes will then only become noticeable after each batch of T rounds; these can be written as

$$D_i(t+T) = (1-\alpha)^T D_i(t) + \sum_{k=0}^{T-1} (1-\alpha)^k \Delta U_i(t+k). \quad \text{(A1)}$$

Here the sum is over the T consecutive rounds that start at round t. We can rewrite Eq. (A1) as

$$\frac{D_i(t+T)}{T} = (1-\alpha')\frac{D_i(t)}{T} + \frac{1}{T}\sum_{k=0}^{T-1}(1-\alpha')^{k/T}\Delta U_i(t+k), \tag{A2}$$

where $\alpha' = 1-(1-\alpha)^T \approx \alpha T$, since α is assumed small.

For large values of the batch size T, we can interpret the sum on the right-hand side of Eq. (A2) as a weighted time average. The weight is given by a discounting factor $(1-\alpha')^{k/T}$, which decreases from 1 to $1-\alpha'$ from the beginning ($k=0$) to end ($k=T-1$) of the batch, respectively. So, if we further assume that also $\alpha' \approx \alpha T$ is small, we can approximate such a sum by the ensemble average in Eq. (11) calculated with the corresponding mixed strategies [44,74,75].

Replacing the last term in Eq. (A2) with the term defined in Eq. (11) and writing everything in terms of a rescaled time $\tau \equiv t/T$, a rescaled drive $D_i'(\tau) = D_i(\tau T)/T$, and utility differences $\Delta U_i'(\tau) = \overline{\Delta U_i}(\tau T)/T$, we obtain

$$D_i'(\tau+1) = (1-\alpha')D_i'(\tau) + \Delta U_i'(\tau). \tag{A3}$$

Following Eqs. (1) and (A3) and the definitions $\tau \equiv t/T$ and $x'(\tau) \equiv x(\tau T)$, we can write

$$x(t+T) = x'(\tau+1) = \frac{1}{1+e^{-\beta D_i'(\tau+1)}}. \tag{A4}$$

In terms of a rescaled parameter $\beta' = \beta T$, we obtain an equation analogous to Eq. (9) but for updates on batches of T rounds, i.e.,

$$x_i'(\tau+1) = \frac{x_i'(\tau)^{1-\alpha'}}{x_i'(\tau)^{1-\alpha'} + [1-x_i'(\tau)]^{1-\alpha' e^{-\beta'\Delta U_i'(\tau)}}}, \tag{A5}$$

where introducing Eq. (6) into Eq. (11) for the case of the weak prisoner's dilemma that we are interested in, we have

$$\Delta U_i'(\tau) = a x_i'(\tau)\sum_{j\in\partial i} x_j'(\tau) + b\sum_{j\in\partial i} x_j'(\tau) + 2h x_i'(\tau) - h. \tag{A6}$$

Equation (A5) is a deterministic update rule obtained by neglecting the fluctuations in the last term in Eq. (A2), which is stochastic, and replacing it with the average in Eq. (11). Finally, notice that since we assumed α' is small then $\alpha \approx \alpha'/T$ should be even smaller. Notice also that in this case the ratio $\alpha'/\beta' = \alpha/\beta$ remains the same. If we take $T=1$ then $\tau = t$, so we can replace $x_i'(\tau) \to x_i(t)$ and $\Delta U'(t) \to \overline{\Delta U_i}(t)$ in Eq. (A6). Similarly, we can replace $\alpha' \to \alpha$ and $\beta' \to \beta$. This yields Eqs. (10) and (11).

APPENDIX B: CALCULATION OF THE PHASE DIAGRAM

Here we show that Eq. (12) indeed predicts three regimes with qualitatively different long-term dynamics: monostable, bistable, and nonequilibrium. Graphically, the solutions of Eq. (14) correspond to the intersections of the graphs of f and the identity function at points x that satisfy $0 < x < 1$. Their stability is determined by the magnitude of the derivative of f, i.e.,

$$f'(x) \equiv \frac{\partial f(x)}{\partial x} = \frac{A(x-x_0)}{\cosh^2[A(x-x_0)^2 + y_0]}, \tag{B1}$$

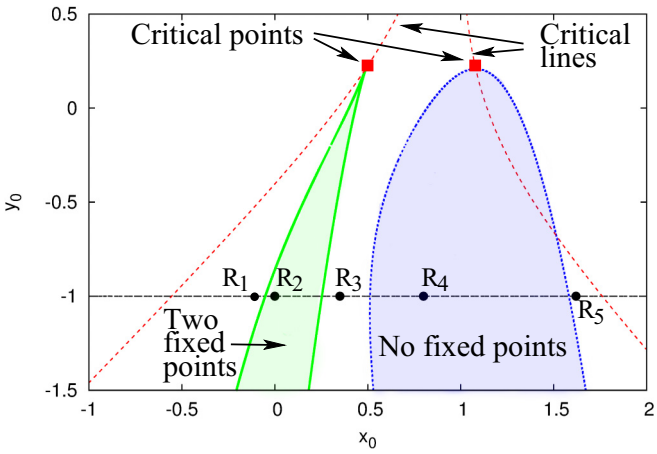

FIG. 3. Level curve of the surface $A_d(x_0, y_0)$ of discontinuous transitions [see Eqs. (B3) and (B4)], i.e., $A_d(x_0, y_0) = A$ (here $A = 5$). Inside the (green) shaded triangularlike region (left) there are two stable fixed points. Inside the (blue) shaded paraboliclike region (right) there are no stable fixed points. In all the remaining white area there is one stable fixed point. These regions terminate on critical points (red squares). If the value of A changes, these regions shift and so do the corresponding critical points along the critical lines (red dashed nonhorizontal lines crossing from the bottom left and bottom right to the top center). The points $R1, \ldots, R_5$ on the (black) dashed horizontal line show examples of the five regions described in the text (see Appendix B). These correspond to parameter values $A = 5$, $y_0 = -1$, and $x_0 = -0.90, 0.00, 0.35, 0.80, 1.62$, respectively.

evaluated at the corresponding intersection point x: If $|f'(x)| < 1$ [$|f'(x)| > 1$], then the fixed point is stable (unstable). We have used the partial rather than total derivative in Eq. (B1) to stress that f is also a function of A, x_0, and y_0.

We now proceed to derive the equations that define the surfaces separating the different regimes which, as we will see, are accompanied by a line of critical points. For clarity, we will first give a somewhat informal discussion before addressing the problem in more detail below. Notice that Eq. (14) is similar to the one yielding the equilibrium magnetization in the mean field Ising model on an external field. In analogy with the analysis of the Ising model and following the discussion in the preceding paragraph, the condition $|f'(x)| = 1$ plays a central role in determining the transition between different regimes. Using Eq. (B1), the condition $|f'(x)| = 1$ yields

$$\sqrt{A}|z| = \cosh^2(z^2 + y_0), \tag{B2}$$

where $z = \sqrt{A}(x - x_0)$. Using Eq. (B2), rewriting the definition of z as $x_0 = x - z/\sqrt{A}$, and using Eq. (14) to change x for $f(x)$, we can write

$$A_d(z, y_0) = \frac{\cosh^4(z^2 + y_0)}{z^2}, \tag{B3}$$

$$x_0(z, y_0) = \frac{1}{2} + \frac{1}{2}\tanh(z^2 + y_0) - \frac{z}{\sqrt{A_d(z, y_0)}}, \tag{B4}$$

from which we can obtain in parametric form the surface $A_d(x_0, y_0)$ that separates the three dynamical regimes, as a function of x_0 and y_0 with the parameter z [Fig. 1(a)]. Figure 3 shows a level curve of this surface, which is the set of points

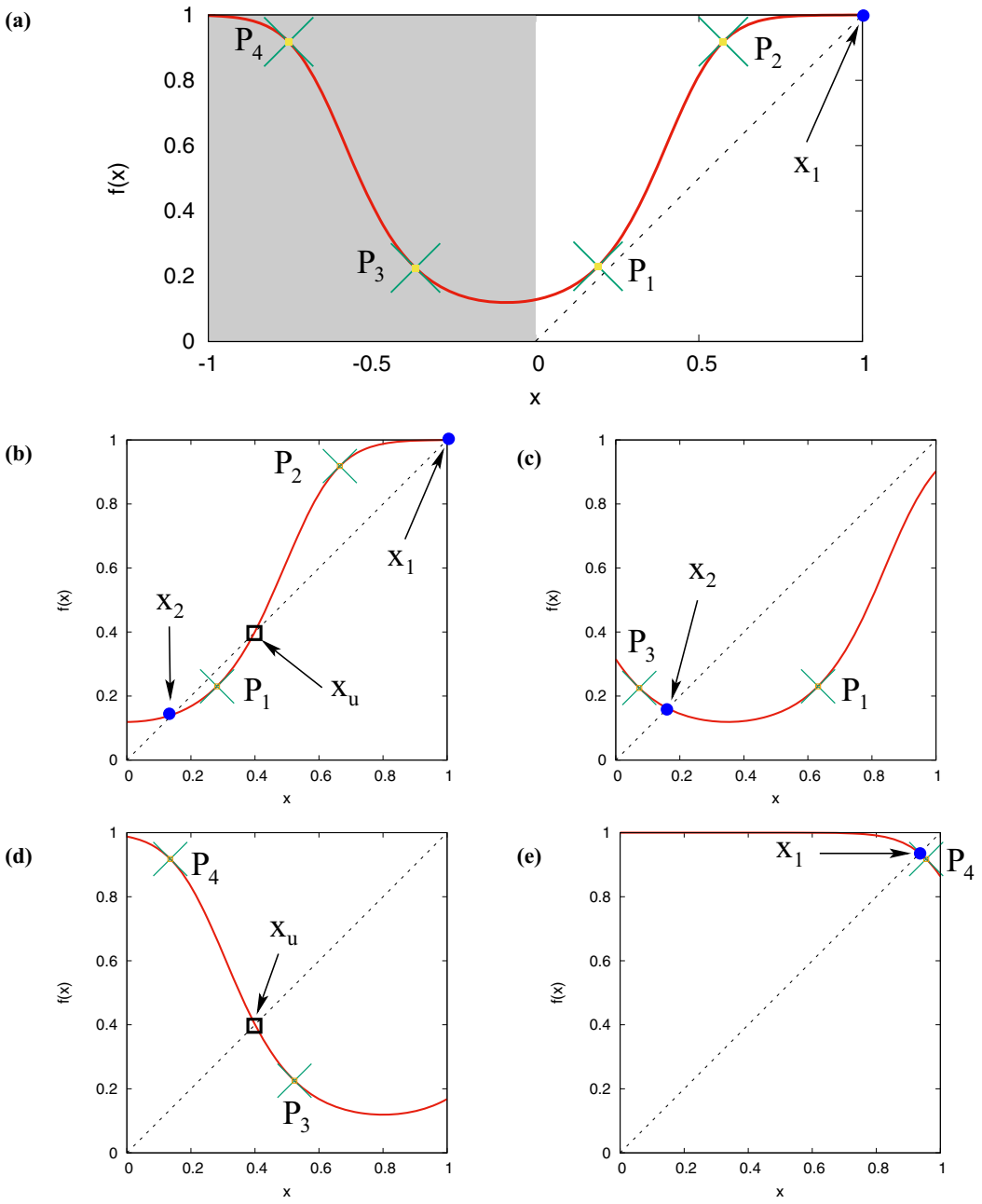

FIG. 4. Graph of f [see Eq. (14)] corresponding to points $R_1, \ldots, R_5$ in Fig. 3 in Appendix B. Notice that $f'(P_1) = f'(P_2) = 1$ and $f'(P_3) = f'(P_4) = -1$ (green crosses). The fixed points of f are its intersections with the identity function (dashed line); these are stable (blue circles) if $|f'(x)| < 1$ and unstable (black squares) if $|f'(x)| > 1$. (a) For small x_0 (e.g., point R_1 in Fig. 3), there is only one stable fixed point x_1. (b) Increasing x_0 (e.g., point R_2 in Fig. 3) until point P_1 touches the identity function, we enter region 2 where there are two stable fixed points x_1 and x_2 and an unstable one x_u in between. (c) Increasing x_0 (e.g., point R_3 in Fig. 3) until x_1 hits point P_2 to then disappear, we enter region 3 where only the stable fixed point x_2 survives. (d) Increasing x_0 (e.g., point R_4 in Fig. 3) until x_2 hits the point P_3 to become unstable, we enter region 4 where there are no stable fixed points. (e) Increasing x_0 (e.g., point R_5 in Fig. 3) until x_u hits point P_4 to become stable, we enter region 5 where there is only one stable fixed point x_1.

that satisfies $A_d(x_0, y_0) = A$, with $A = 5$. This value allows for a better visualization, while the discussion that follows remains qualitatively true for the case in Fig. 1 discussed in the main text, where the parameters inferred from the experiment [19] were used instead.

To fix ideas before we continue with a more formal description, we first show a more graphical discussion following Figs. 3 and 4. For this we fix parameters $A = 5$ and $y_0 = -1$ and vary x_0 moving from left to right along the horizontal dashed line in Fig. 3. This figure shows five different points labeled R_ℓ (with $\ell = 1, \ldots, 5$) on the said horizontal dashed line, which illustrate the five regions to be discussed next. Figure 4 depicts the respective functions $f(x)$ for $A = 5$ and $y_0 = -1$ at each of the five values of x_0 that correspond to the five points R_ℓ in Fig. 3. Notice that $f(x)$ always takes its minimum value at $x = x_0$. We have also identified four

points, labeled P_ℓ (with $\ell = 1, \ldots, 4$), where the magnitude of the slope of f is exactly one, i.e., $f'(P_1) = f'(P_2) = 1$ and $f'(P_3) = f'(P_4) = -1$ [see, for example, Fig. 4(a)]. Starting at R_1 [Fig. 4(a)], we can then shift $f(x)$ (red solid line) towards the right by increasing x_0, and when each of the four P_ℓ points hits the graph of the identity function (dashed line) they become fixed points x with $|f'(x)| = 1$ (i.e., marginally stable).

We now describe the different ways in which the identity function can intercept the graph of f. Referring to the sequence in Figs. 4(a)–4(e), we can imagine that we start from $x_0 < -0.1$ [Fig. 4(a)] and slowly increase its value so that the function f slowly moves from left to right traversing the conditions corresponding to the five points R_ℓ (with $\ell = 1, \ldots, 5$) in Fig. 3. In this process we traverse the following five regions.

Region 1. Initially x_0, where f takes its minimum value, is negative enough to cause the graph of the identity function to intersect f at a single point $x_1 \approx 1$ [see Fig. 4(a)]. Since $f'(x_1) \approx 0$, the fixed point x_1 is stable. Then, if we start increasing the value of x_0, the graph of f will move to the right and the value of x_1 will decrease until the graphs of f (red solid line) and the identity function (dashed line) intersect at point P_1 [see Fig. 4(a)] and that situation will mark the end of region 1. Point R_1 in Fig. 3 belongs to this region.

Region 2. From then on, a second stable fixed point x_2 emerges along with an unstable fixed point x_u, with $x_2 < x_u < x_1$ [see Fig. 4(b)]. Increasing the value of x_0 further, these fixed points shift to the left until x_1 hits point P_2 [see Fig. 4(b)] and then disappears. This can only happen if the curvature of the graph of f is not too large. Point R_2 in Fig. 3 belongs to this second region.

Region 3. Afterward, there is only one stable fixed point x_2 which shifts to the left while we keep increasing the value of x_0, until it hits point P_3 [see Fig. 4(c)], to then become unstable. This can only happen if the curvature of the graph of f is not too large. Point R_3 in Fig. 3 belongs to this region.

Region 4. After this, there is only one fixed point x_u, which is unstable, that shifts to the left when increasing the value of x_0, until it hits a point P_4, where $f'(P_4) = -1$ [see Fig. 4(d)]. Point R_4 in Fig. 3 belongs to this region.

Region 5. In this last regime, there is only one stable fixed point x_1 which keeps shifting to the left while we increase further the value of x_0 [see Fig. 4(e)]. Point R_5 in Fig. 3 belongs to this region.

Following the comments made in the description of regions 2 and 3 above, when the curvature of the graph of f is large enough the order in which points P_2 and P_3 in Fig. 4 meet gets inverted. However, the experimental results are not located in this regime and so we do not discuss this further.

In Fig. 3 the level curve $A_d(x_0, y_0) = A$ defines regions inside which there are zero (blue paraboliclike area) and two (green triangularlike area) stable fixed points. These regions terminate on a critical point (red squares), where a continuous transition takes place. By varying the value of A, we can change those regions and the corresponding critical points, which then gives rise to lines of critical points (red dashed nonhorizontal lines crossing from the bottom left and bottom right to the top center). One condition satisfied by a critical point is that there are only two points (instead of four) where the slope of f has magnitude one; one of those points is the reflection of the other around $x = x_0$. This is the case if Eq. (B2) has only one solution $z^* > 0$, since this implies a second solution $-z^*$ by the symmetry of Eq. (B2) under reflections $z \to -z$. The condition for Eq. (B2) to have a unique positive solution is that the slope of the function $g(z) = \cosh^2(z^2 + y_0)$ on its right-hand side equals $\sqrt{A}$, i.e.,

$$\sqrt{A} = g'(z) \equiv 4z \sinh(z^2 + y_0) \cosh(z^2 + y_0). \tag{B5}$$

We can safely assume that $A \neq 0$ and divide Eq. (B5) by Eq. (B2) for $z > 0$ to obtain the equation $4z^2 \tanh(z^2 + y_0) = 1$, from which we can obtain $y_{0c}(z)$, i.e., the critical value of y_0 as a function of z [see Eq. (B6) below]. Knowing this, we can use Eqs. (B3) and (B4) to obtain $A_c(z)$ and $x_{0c}(z)$, i.e., the corresponding critical values of A and x_0 as a function of z [see Eqs. (B7) and (B8) below]. More explicitly, the critical lines $(A_c(z), x_{0c}(z), y_{0c}(z))$ are described by

$$y_{0c}(z) = \tanh^{-1}\left(\frac{1}{4z^2}\right) - z^2, \tag{B6}$$

$$A_c(z) = \frac{\cosh^4[z^2 + y_{0c}(z)]}{z^2}, \tag{B7}$$

$$x_{0c}(z) = \frac{1}{2}\left(1 + \frac{1}{4z^2}\right) - \frac{z}{\sqrt{A_c(z)}}, \tag{B8}$$

where we have used the condition for criticality, i.e., $4z^2 \tanh(z^2 + y_0) = 1$, to obtain Eq. (B8).

In all the discussion so far the condition $|f'(x)| = 1$ has played a central role. Here we show in a more detailed way why this is the case. First, notice that the function f in Eq. (14) essentially contains a parabola given by the expression $A(x - x_0)^2 + y_0$ and transforms it by applying a hyperbolic tangent, a constant scaling, and a constant offset (both equal to $1/2$) to it. The parameter A defines the curvature of the parabola, while the parameters y_0 and x_0 define the minimum value it takes and where it takes it, respectively. These features remain qualitatively true for the graph of f, except that now y_0 also influences its curvature. Now note that the graph of the function f in Eq. (14) has the following properties (see Fig. 4).

Property 1. It is continuous and bounded, i.e., $0 \leqslant f(x) \leqslant 1$ for all x.

Property 2. It is symmetric around $x = x_0$, where it takes its minimum value, i.e., $f(x_0 - x) = f(x_0 + x)$ for $x_0 = \arg\min_x f(x)$.

Property 3. Starting from $x = x_0$ and moving towards $x > x_0$ ($x < x_0$), its slope monotonically increases (decreases) from zero up to a certain point, where its second derivative f'' vanishes, and then starts decreasing (increasing) until it asymptotically reaches zero again. In particular $\lim_{x \to \pm\infty} f'(x) = 0$.

Property 4. By varying x_0 it is translated horizontally, but otherwise its shape remains unchanged. In particular, this implies that $f'(x)$ depends only on the difference $x - x_0$, as observed in Eq. (B1).

According to properties 1 and 4, the graph of the function f in Eq. (14) can intersect with the graph of the identity function in any of its points, by choosing a proper value of x_0 (see Fig. 4). Furthermore, due to properties 3 and 4, we can always find a value of x_0 for which there is at least one stable fixed point,

since the function f always has points with slopes as close to zero as necessary.

Now, due to the continuity of f, if there are only two stable fixed points, say, x_1 and $x_2 < x_1$, then there is an unstable fixed point, say, x_u, such that $x_2 < x_u < x_1$ [see Fig. 4(b)]. In this case, because of the shape of f [see Fig. 4(b)], the unstable fixed point x_u and at least one of the two stable fixed points should be on the right side of x_0, i.e., $x_1 > x_u > x_0$. Following properties 1 and 3, if there is an unstable fixed point x_u such that $x_u > x_0$, there must also be two points where the slope of f is equal to one; this is due to the fact that for $x > x_0$ we have $f'(x) > 0$ and $f'(x)$ goes to zero for both $x = x_0$ and $x \to \infty$; furthermore, $f'(x_u) > 1$.

Hence, the existence of a point x with $f'(x) > 1$ signals also the existence of at least one value of the parameter x_0 for which there are two stable fixed points [see Fig. 4(b)]. Since f is symmetric around $x = x_0$, this also implies the existence of a point x' such that $f(x') < -1$. The first time this happens is when a point x with $|f'(x)| = 1$ emerges.

APPENDIX C: DIVERGING SUSCEPTIBILITY

The susceptibility χ of a system is related to its response to a small change in the external conditions. We could ask what the change δx is in the global level of cooperation when a generic parameter θ of the model is varied by a small amount $\delta\theta$. We have $\delta x \approx (\partial x/\partial\theta)\delta\theta$, so $\chi = \partial x/\partial\theta$. Arguably, the most natural parameters to consider in our model are ΔI_C introduced in Eq. (4), which is in principle under the influence of the experimenter, and perhaps also h [64]; these two parameters influence the effective parameters x_0 and y_0 [see Eqs. (16) and (17)]. Since f in Eq. (14) depends explicitly on the parameters A, x_0, and y_0 any change in a generic parameter θ that affects any of those three parameters would also affect f. To be more specific, let us assume that $A = A(\theta)$, $x_0 = x_0(\theta)$, and $y_0 = y_0(\theta)$ are well-behaved functions of θ.

Deriving both sides of Eq. (14) with respect to a generic parameter θ, we obtain

$$\chi = \frac{\partial x}{\partial\theta} = f'(x)\frac{\partial x}{\partial\theta} + \frac{\partial f(x)}{\partial\theta}, \tag{C1}$$

where $f'(x)$ is defined in Eq. (B1) and the additional term

$$\frac{\partial f(x)}{\partial\theta} = \frac{\partial f(x)}{\partial A}\frac{\partial A(\theta)}{\partial\theta} + \frac{\partial f(x)}{\partial x_0}\frac{\partial x_0(\theta)}{\partial\theta} + \frac{\partial f(x)}{\partial y_0}\frac{\partial y_0(\theta)}{\partial\theta} \tag{C2}$$

takes into account the explicit dependence of f on the parameters of the model, which vary when varying θ. The term defined in Eq. (C2) is smooth as long as we can assume, as we do here, that there are no spurious singularities in the definition of $A(\theta)$, $x_0(\theta)$, and $y_0(\theta)$.

Solving Eq. (C1) for the susceptibility $\chi = \partial x/\partial\theta$, we get

$$\chi = \frac{1}{1 - f'(x)}\frac{\partial f(x)}{\partial\theta}, \tag{C3}$$

which clearly diverges when, by varying θ, the fixed point x under consideration crosses continuously a point x_c where $f'(x_c) = 1$, i.e., a critical point. For illustration purposes, let us assume that the model parameters vary as $A(\theta) = A_c^* - \theta$, $x_0(\theta) = x_{0c}^*$, and $y_0(\theta) = y_{0c}^*$ for a generic parameter $\theta \geqslant 0$; here $A_c^* = A_c(z^*), x_{0c}^* = x_{0c}(z^*)$, and $y_{0c}^* = y_{0c}(z^*)$ correspond to a point on the critical line specified by a particular value of z^* through Eqs. (B6)–(B8). Using Eqs. (B1) and (B7), we can write $f'(x) = \sqrt{A(\theta)/A_c^*}$, which for small θ can be approximated as $f'(x) \approx 1 - \theta/2A_c^*$. So, following Eq. (C3), when approaching the critical point, i.e., $\theta \to 0$ or $A \to A_c^*$, the susceptibility diverges as

$$\chi \propto \frac{1}{\theta} = \frac{1}{A_c^* - A} \xrightarrow[A \to A_c^*]{} \infty, \tag{C4}$$

where in the last expression we have written $\theta = A_c^* - A$ in terms of A. The divergence of the susceptibility is one of the hallmarks of criticality [27].

APPENDIX D: CONNECTION TO EXPERIMENTS

1. Moody conditional cooperation

As mentioned in the main text, experiments show that the probability for a human to cooperate in a generic round of the game depends on whether she cooperated or not and how many of her peers cooperated. Here we explain how to connect this so-called moody conditional cooperation rule with the mean field model described by Eq. (12).

Indeed, the MCC rule can be expressed mathematically in terms of the conditional probability $P(C;t+1|s,n;t)$ for a generic agent to cooperate C at round $t+1$ given that she played strategy s and that n of her peers cooperated at round t. More precisely, the probability which the MCC rule refers to can be written as $P(C|s,n) = (1/T)\sum_t P(C;t+1|s,n;t)$, where T is the total number of rounds. We assume, however, that T is sufficiently large and that the system reaches a stationary state. In this case, the MCC rule is given by the conditional probability corresponding to the stationary state ($t \to \infty$), and we can drop the index t; we will keep the dependence on t for the most part to facilitate the discussion, though. We further assume that the stationary state can be described by the long-term dynamics of the mean field model.

Depending on the context, we will use interchangeably $s = C$ or $s = 1$ to refer to cooperation, and similarly we will use interchangeably $s = D$ or $s = 0$ to refer to defection. Now, writing $\Delta U_i(t) = \Delta U_i(s_i,n_i)$ [see Eq. (6)], the right-hand side of Eq. (9) gives the probability that an agent i cooperates at round $t+1$ given that, at round t, she played strategy s_i, that $n_i = \sum_{j\in\partial i} s_j$ of her peers cooperated, and that her cooperation probability was x_i. Indeed, this is a more detailed reading of Eq. (9). Since in the mean field approximation we are interested in a representative agent, we can drop the indices and write

$$P(C;t+1|s,n,x;t) = \frac{x^{1-\alpha}}{x^{1-\alpha} + (1-x)^{1-\alpha}e^{-\beta\Delta U(s,n)}}, \tag{D1}$$

where $\Delta U(s,n) = (as+b)n + 2hs - h$ [see Eq. (6)]. This conditional probability distribution depends on x, while the MCC rule does not. Informally, if we assume that the system is monostable we can get rid of the dependence on x to obtain

$$P(C|s,n) = \frac{1}{1 + y_1^{1-\alpha}e^{-\beta\Delta U(s,n)}}, \tag{D2}$$

where $y_1 = (1-x_1)/x_1$ and x_1 is the only stable fixed point; Eq. (D2) coincides with Eq. (11).

In the following we use the rules of probability theory to obtain a more general form of $P(C|s,n)$ from Eq. (D1) that reduces to Eq. (D2) if we assume that the system is monostable. Let us first write

$$P(C;t+1|s,n;t) = \frac{\int_0^1 P(\{C;t+1\},\{s,n,x;t\})dx}{\int_0^1 P(s,n,x;t)dx}, \quad \text{(D3)}$$

by definition of conditional probability. The term inside the integral in the numerator is the joint probability of all the variables involved. We have emphasized between curly brackets which round the variables refer to. On the other hand, the term inside the integral in the denominator is the joint probability of all the variables that refer only to round t.

Using the chain rule of probability theory, we can express such joint probabilities as

$$P(\{C;t+1\},\{s,n,x;t\}) = P(C;t+1|s,n,x;t) \times P(s,n;t|x;t)P(x;t), \quad \text{(D4)}$$

$$P(s,n,x;t) = P(s,n;t|x;t)P(x;t), \quad \text{(D5)}$$

Let N denote the set of neighbors of the representative agent and let $s_N = \{s_j | j \in N\}$ and $x_N = \{x_j | j \in N\}$ be the set of their strategies and cooperation probabilities, respectively. Then

$$P(s,n;t|x;t) = \sum_{s_N} P(s,s_N;t|x;t)\delta\left[n = \sum_{j\in N} s_j\right], \quad \text{(D6)}$$

where $\delta[p]$ is the indicator function, which is equal to one if proposition p is true and zero otherwise. Here $P(s,s_N;t|x;t)$ is the conditional probability that the representative agent play strategy s and her peers play strategies s_N, jointly, at round t given that the probability for the representative agent to cooperate at the same round is x. The definition of conditional probability and the chain rule allow us to write

$$P(s,s_N;t|x;t) = \frac{\int_0^1 P(s,s_N;t|x,x_N;t)P(x,x_N;t)\prod_{j\in N} dx_j}{P(x;t)}, \quad \text{(D7)}$$

where the term inside the integral in the numerator is the joint probability that at round t the representative agent and her neighbors play strategies s and s_N and their cooperation probabilities are x and x_N, respectively. The integral marginalizes this probability over x_N leaving the joint probability of the variables s, s_N, and x.

Although the next sentence may be redundant, its sole intention is to put everything in the formalism we are describing here: The probability that an agent plays strategy s at round t, given that the probability to cooperate at the same round t is x, can be written as $P(s;t|x;t) = x^s(1-x)^{1-s}$. Furthermore, at each round each agent picks her strategy independently of the rest. So

$$P(s,s_N;t|x,x_N;t) = P(s;t|x;t)\prod_{j\in N} P(s_j;t|x_j;t) = x^s(1-x)^{1-s}\prod_{j\in N} x_j^{s_j}(1-x_j)^{1-s_j}. \quad \text{(D8)}$$

This reflects the fact that correlations in the system come only through the correlations in the cooperation probability accumulated during the history of play. Equation (D7) can then be written as

$$P(s,s_N;t|x;t) = \frac{x^s(1-x)^{1-s}\int_0^1 P(x,x_N;t)\prod_{j\in N} x_j^{s_j}(1-x_j)^{1-s_j}dx_j}{P(x;t)}. \quad \text{(D9)}$$

Only now we resort to the mean field approximation which neglects correlations altogether to write $P(x,x_N;t) \approx P(x;t)\prod_{j\in N} P(x_j;t)$. Since we are interested in the stationary state, we can drop the round index t in the expressions that follow. So

$$P(s,s_N;t|x;t) = x^s(1-x)^{1-s}\rho^{s_j}(1-\rho)^{1-s_j}, \quad \text{(D10)}$$

where $\rho = \int_0^1 x_j P(x_j)dx_j$ is the average probability that a neighbor cooperates, which equals the average probability that the representative agent cooperates since we are working within a mean field approximation. Introducing Eq. (D10) into Eq. (D6), we get

$$P(s,n,t|x,t) = x^s(1-x)^{1-s}\binom{K}{n}\rho^n(1-\rho)^{K-n}, \quad \text{(D11)}$$

where $\binom{K}{n} = K!/n!(K-n)!$ is the binomial coefficient. Introducing Eqs. (D11) and (D1) into Eqs. (D3) and (D4), Eq. (D5) yields the desired result

$$P(C,t+1|s,n,t) = \frac{1}{\rho^s(1-\rho)^{1-s}}\int_0^1 \frac{x^{s+1-\alpha}(1-x)^{1-s}P(x)}{x^{1-\alpha}+(1-x)^{1-\alpha}e^{-\beta\Delta U(s,n)}}dx, \quad \text{(D12)}$$

where we have used $\int_0^1 x^s(1-x)^{1-s}dx = \rho^s(1-\rho)^{1-s}$. Using the change of variables $y = (1-x)/x$, which is monotonic for $x \in (0,1)$, we have

$$P(C,t+1|s,n,t) = \frac{1}{\rho^s(1-\rho)^{1-s}}\int_0^\infty \frac{y^{1-s}(1+y)^{-1}P_Y(y)}{1+y^{1-\alpha}e^{-\beta\Delta U(s,n)}}dy. \quad \text{(D13)}$$

Since we have assumed that β is small, we can expand the right-hand side of Eq. (D13) to first order in β to obtain

$$P(C,t+1|s,n,t) = I_s(\alpha) + \beta J_s(\alpha)\Delta U(s,n), \quad \text{(D14)}$$

where

$$I_s(\alpha) = \frac{1}{\rho^s(1-\rho)^{1-s}}\int_0^\infty \frac{y^{1-s}(1+y)^{-1}P_Y(y)}{1+y^{1-\alpha}}dy, \quad \text{(D15)}$$

$$J_s(\alpha) = \frac{1}{\rho^s(1-\rho)^{1-s}}\int_0^\infty \frac{y^{2-\alpha-s}(1+y)^{-1}P_Y(y)}{(1+y^{1-\alpha})^2}dy. \quad \text{(D16)}$$

Using Eq. (6), we can see that Eq. (D14) is of the form

$$P(C,t+1|s,n,t) = m_s n/K + r_s, \quad \text{(D17)}$$

where

$$m_s = \beta K J_s(\alpha)(as + b), \tag{D18}$$

$$r_s = I_s(\alpha) + \beta J_s(\alpha)[h(2s-1)]. \tag{D19}$$

This yields the general expression introduced in Eqs. (20)–(22); in the next section we make use of the assumption of monostability.

2. Regime of monostability

We now make use of the assumption that the long-term dynamics of the system is well described by the stable fixed points of the mean field dynamics. So $P(x) = (1-\mu)\delta(x-x_1) + \mu\delta(x-x_2)$, where x_1 and x_2 are the fixed points of Eq. (12) and μ yields their corresponding weights. In this case, the average probability for the representative agent to cooperate is $\rho = (1-\mu)x_1 + \mu x_2$. If there is only one fixed point we can take $\mu = 0$. If there are no fixed points, the analysis here does not apply. For simplicity and in agreement with the experiment we analyze [19], we assume that the dynamics is essentially monostable, say, $\mu \approx 0$. This implies that the experimental global average cooperation level in the stationary state $\bar{x}$ is close to the relevant fixed point, i.e., $\bar{x} \approx x_1$.

In this case we have $\rho = x_1$ and $P_Y(y) = \delta(y - y_1)$ with $y_1 = (1-x_1)/x_1$, which yields

$$P(C;t+1|s,n;t) = \frac{1}{1 + y_1^{1-\alpha} e^{-\beta \Delta U(s,n)}}; \tag{D20}$$

this is the expression in Eq. (19). In this case both terms in Eqs. (D15) and (D16) become

$$I_s(\alpha) = I(\alpha) \equiv \frac{1}{1 + y_1^{1-\alpha}}, \tag{D21}$$

$$J_s(\alpha) = J(\alpha) \equiv \frac{y_1^{1-\alpha}}{\left(1 + y_1^{1-\alpha}\right)^2}, \tag{D22}$$

independent of s. Equations (D18) and (D19) are better described in terms of the quantities

$$r = \tfrac{1}{2}(r_C + r_D) = I(\alpha), \tag{D23}$$

$$G = r_C - r_D = 2\beta h J(\alpha). \tag{D24}$$

Here r and G are the mean intercept and the gap between intercepts of the near linear trends that describe the MCC rule [19], respectively. We can safely assume that $b \neq 0$, which gives

$$m_C - m_D = \beta a K J(\alpha), \tag{D25}$$

$$\frac{m_C}{m_D} = \frac{\beta a + \beta b}{\beta b}. \tag{D26}$$

Now, Eq. (D23) can be readily inverted to obtain α in terms of the experimental quantity r [see Eq. (D27) below]. Similarly, using Eq. (D23) we can write $J(\alpha) = I(\alpha)[1 - I(\alpha)] = r(1-r)$ and so Eq. (D24) can be readily inverted to obtain βh in terms of the experimental quantities G and r [see Eq. (D28) below]. Finally, Eqs. (D25) and (D26) can be inverted to obtain βa and βb in terms of the experimental quantities m_C, m_D, r, and K [see Eqs. (D29) and (D30) below]. This yields

$$\alpha = 1 - \frac{\log[(1-r)/r]}{\log[(1-\bar{x})/\bar{x}]}, \tag{D27}$$

$$\beta h = \frac{G}{2r(1-r)}, \tag{D28}$$

$$\beta a = \frac{m_C - m_D}{Kr(1-r)}, \tag{D29}$$

$$\beta b = \frac{m_D}{Kr(1-r)}, \tag{D30}$$

where we have used the condition that the only stable fixed point should equal the experimental global cooperation level, i.e., $x_1 = \bar{x}$, to obtain Eq. (D27).

Although these equations leave the parameter β undetermined, this combination of parameters completely determines the coefficients that define the mean field dynamics through Eq. (12) and the parameters A, x_0 and y_0 that locate the system in the phase diagram. Indeed, multiplying by β in the numerator and denominator of Eqs. (15)–(17), using Eqs. (D27)–(D30), and doing some algebra we obtain the expressions

$$A = \frac{m_C - m_D}{2\alpha r(1-r)}, \tag{D31}$$

$$x_0 = -\frac{m_D + G}{2(m_C - m_D)}, \tag{D32}$$

$$y_0 = -\frac{m_D^2 + 2m_C G + G^2}{8\alpha r(1-r)(m_C - m_D)}, \tag{D33}$$

where the expression for α is given in Eq. (D27). Equations (D31)–(D33) in principle allow us to locate the system in the phase diagram of Fig. 1(a). However, we still need to check that these parameter values produce a dynamics through Eq. (12) that indeed agree with the dynamics observed in experiments, within the margin of error of the experimental results. Furthermore, we should check that indeed the assumption of monostability is indeed satisfied, i.e., that $x_1 \approx \bar{x}$.

Before finishing this section note that we can invert Eqs. (15)–(17) to recover the parameters defining the mean field dynamics in Eq. (12), which yields

$$\beta a = \frac{2\alpha A}{K}, \tag{D34}$$

$$\beta b = \frac{4\alpha}{K}[y_0 - x_0(1-x_0)A], \tag{D35}$$

$$\beta h = -2\alpha\left(y_0 + x_0^2 A\right); \tag{D36}$$

here we made use of the expression $y_0 = -(\tilde{a}x_0^2 + h)/2\gamma$.

3. Regime of bistability

In the case in which the system is bistable with a non-negligible value of μ, we have to deal with the whole expression $P(x) = (1-\mu)\delta(x-x_1) + \mu\delta(x-x_2)$ and Eqs. (D15)

and (D16) become

$$I_s(\alpha) = W_{1,s}\Upsilon_1(\alpha) + W_{2,s}\Upsilon_2(\alpha), \tag{D37}$$

$$J_s(\alpha) = W_{1,s}\Upsilon_1(\alpha)[1-\Upsilon_1(\alpha)] + W_{2,s}\Upsilon_2(\alpha)[1-\Upsilon_2(\alpha)], \tag{D38}$$

where we have defined the expressions

$$\Upsilon_\ell = \frac{1}{1+y_\ell^{1-\alpha}}, \tag{D39}$$

$$W_{\ell,s} = \mu^{i-1}(1-\mu)^{2-i}\left(\frac{x_i}{\rho}\right)^s\left(\frac{1-x_i}{1-\rho}\right)^{1-s}. \tag{D40}$$

Equations (D18) and (D19) defining the slopes and intercepts can still be inverted to obtain

$$\beta a K = \frac{m_C}{J_1(\alpha)} - \frac{m_D}{J_0(\alpha)}, \tag{D41}$$

$$\beta b K = \frac{m_D}{J_0(\alpha)}, \tag{D42}$$

$$\beta h = \frac{1}{2}\left[\frac{r_C}{J_1(\alpha)} - \frac{r_D}{J_0(\alpha)} + \frac{I_0(\alpha)}{J_0(\alpha)} - \frac{I_1(\alpha)}{J_1(\alpha)}\right]. \tag{D43}$$

However, the corresponding values of α are given implicitly by solutions to the equation

$$J_0(\alpha)[r_C - I_1(\alpha)] + J_1(\alpha)[r_D - I_0(\alpha)] = 0. \tag{D44}$$

In the bistable regime, we can parametrize the model in terms of α, the two stable fixed points x_1 and x_2, and the unstable fixed point x_u. Indeed, if we write the fixed point (14) as $A(x-x_0)^2 + y_0 = \tanh^{-1}(2x-1)$, evaluate it at two of the fixed points, say, x_1 and x_2, and subtract the two corresponding equations, we obtain the expression for A,

$$A = F_A(x_2,x_1|x_0) \equiv \frac{\Delta(x_2,x_1)}{(x_2-x_1)(x_2+x_1-2x_0)}, \tag{D45}$$

which defines the function $F_A(x_2,x_1|x_0)$; here

$$\Delta(x_2,x_1) \equiv \tanh^{-1}(2x_2-1) - \tanh^{-1}(2x_1-1). \tag{D46}$$

Any two fixed points that we choose would give us a different expression for A, which should be consistent. In particular, we should have $A = F_A(x_2,x_u|x_0) = F(x_u,x_1|x_0)$, from which we can obtain an expression for x_0 in terms of the three fixed points, i.e.,

$$x_0 = \frac{\left(x_2^2 - x_u^2\right)\Delta(x_u,x_1) - \left(x_u^2 - x_1^2\right)\Delta(x_2,x_u)}{(x_2-x_u)\Delta(x_u,x_1) - (x_u-x_1)\Delta(x_2,x_u)}. \tag{D47}$$

So we can introduce any feasible values for x_1, x_2, and x_u into Eq. (D47) to obtain the corresponding value for x_0; from Eq. (D45) we can obtain the corresponding value for A and from Eq. (14) the corresponding value for y_0.

APPENDIX E: FROM EXPERIMENTAL DATA TO MODEL PARAMETERS

1. General considerations

Here we depict how we estimated the parameters of the model described by Eq. (12). We follow the framework of partially observed Markov processes [68] and assume that the system is described by an underlying deterministic dynamics $x(t)$ satisfying Eq. (12). We further assume that the scientist in the laboratory observes a noisy version $x_{\rm obs}(t)$ of the underlying dynamics characterized by a probability distribution $\mathcal{P}_{\rm obs}[x_{\rm obs}(t)|x(t)]$. This collectively describes experimental uncertainty as well as intrinsic stochasticity [74,75] in the system dynamics that has been averaged out in the derivation of Eq. (12).

To each observed trajectory $\mathbf{x}_{\rm obs}(1:T) \equiv \{x_{\rm obs}(t) : t = 1,\ldots,T\}$ of T rounds and a trajectory of the underlying dynamics $\mathbf{x}(0:T) \equiv \{x(t) : t = 0,\ldots,T\}$, we can assign a joint probability distribution

$$\mathcal{P}[\mathbf{x}(0:T),\mathbf{x}_{\rm obs}(1:T)|\Theta] = \mathcal{P}_0[x(0)]\prod_{t=1}^{T}\mathcal{P}_{\rm obs}[x_{\rm obs}(t)|x(t)] \times \mathcal{P}_{\rm dyn}[x(t)|x(t-1)|\Theta], \tag{E1}$$

where Θ represents the parameters that define the system dynamics via Eq. (12), $\mathcal{P}_0[x(0)]$ represents the probability distribution over the initial condition $x(0)$ of the underlying dynamics, and $\mathcal{P}_{\rm dyn}[x(t)|x(t-1)|\Theta]$ is a Dirac delta function representing the deterministic dynamics described by Eq. (12). Finally, we assume that $\mathcal{P}_{\rm obs}[x_{\rm obs}(t)|x(t)] = \mathcal{N}[x_{\rm obs}(t);x(t),\sigma]$ is a Gaussian distribution of mean $x(t)$ and standard deviation σ to be determined later.

From the joint distribution defined in Eq. (E1) we can obtain the probability or likelihood of observing a particular realization of the observed dynamics $\mathcal{P}_{\rm lik}[\mathbf{x}_{\rm obs}(1:T)|\Theta]$ given the parameters Θ. Our aim is to determine the probability of the parameters given a particular realization of the observed dynamics which, following Bayes' rule, is given by $\mathcal{P}_\theta[\Theta|\mathbf{x}_{\rm obs}(1:T)] \propto \mathcal{P}_{\rm lik}[\mathbf{x}_{\rm obs}(1:T)|\Theta]\mathcal{P}_{\rm prior}[\Theta]$, where $\mathcal{P}_{\rm prior}[\Theta]$ contains the prior information on the parameters Θ. We estimate the parameter values by the average over the posterior, i.e., $\hat{\Theta} = \int \Theta\mathcal{P}_\theta[\Theta|\mathbf{x}_{\rm obs}(1:T)]d\Theta$. What follows depends on whether we are in a regime of monostability or bistability.

2. Parameter estimation in the regime of monostability

We build the prior $\mathcal{P}_{\rm prior}(\Theta)$ in an indirect way. First, we notice that Eqs. (D27)–(D30) yield a set of values for the parameters that, along with the initial condition $x(0)$, define the system dynamics through Eq. (12). So we can parametrize our model in Eq. (12) by the collection of experimental observables $\Theta = \mathcal{O} \equiv (m_C,m_D,r_C,r_D)$. To deal with the uncertainty in the experimental results, we assume that any observable O reported in the literature is described by a uniform probability distribution supported in the interval $[O^* - \zeta\delta O^*, O^* + \zeta\delta O^*]$. Here O^* and δO^* are, respectively, the value reported for O and its corresponding standard error (Table II); the parameter ζ is used to define a credible interval of the reported experimental results, e.g., if $\zeta \approx 1.28$ or $\zeta \approx 1.96$ we are dealing with a 90% or 97.5% credible interval, respectively. We rely on a uniform rather than a Gaussian distribution on experimental results to avoid the statistics being dominated by rare events [e.g., due to the logarithmic term in Eq. (D27)]. We will use this as a prior distribution $\mathcal{P}(\mathcal{O})$.

To compute α from $\mathcal{O}$ via Eq. (D27), we assume that the fixed point equals the average global cooperation over the last

TABLE III. Parameters inferred for the experiment in Ref. [19] on a heterogeneous network and a square lattice using $\zeta = 1.96$ and $\zeta = 1.28$, respectively (see Sec. VIII).

Parameter	Heterogeneous network	Square lattice
$\hat{m}_C$	0.0545	0.0868
$\hat{m}_D$	−0.0901	−0.2095
$\hat{r}$	0.3917	0.38765
$\hat{G}$	0.161	0.1809
$\hat{x}(0)$	0.5328	0.580

Covariance matrix (heterogeneous network)

1.3×10^{-5}	-7.3×10^{-6}	8.5×10^{-8}	4.3×10^{-6}	-3.8×10^{-6}
-7.3×10^{-6}	1.3×10^{-5}	-1.8×10^{-6}	6.1×10^{-6}	-3.6×10^{-6}
8.5×10^{-8}	-1.8×10^{-6}	7.1×10^{-5}	4.9×10^{-4}	-1.0×10^{-6}
4.3×10^{-6}	6.1×10^{-6}	4.9×10^{-4}	3.5×10^{-3}	-2.0×10^{-5}
-3.8×10^{-6}	-3.6×10^{-6}	-1.0×10^{-6}	-2.0×10^{-5}	2.2×10^{-5}

Covariance matrix (square lattice)

3.4×10^{-5}	-5.5×10^{-6}	-7.6×10^{-7}	3.4×10^{-6}	-3.0×10^{-5}
-5.5×10^{-6}	8.7×10^{-6}	-3.3×10^{-7}	1.1×10^{-6}	-5.4×10^{-6}
-7.6×10^{-7}	-3.3×10^{-7}	4.1×10^{-7}	9.1×10^{-8}	6.7×10^{-7}
3.4×10^{-6}	1.1×10^{-6}	9.1×10^{-8}	1.0×10^{-5}	-4.3×10^{-6}
-3.0×10^{-5}	-5.4×10^{-6}	6.7×10^{-7}	-4.3×10^{-6}	3.2×10^{-4}

TABLE IV. Parameters locating on the phase diagram of the model (Fig. 1 in the main text) the experiment performed in [19] on a heterogeneous network and a square lattice [19]. These data were obtained from Table III by using Eqs. (D27) and (D31)–(D33) to get the parameter values and first-order error propagation to get the corresponding covariance matrix.

Parameter	Heterogeneous network	Square lattice
$\hat{A}$	1.16	1.413
$\hat{x}_0$	−0.24	0.0483
$\hat{y}_0$	−0.71	−0.4346
$\hat{\alpha}$	0.263	0.4417

Covariance matrix (heterogeneous network)

9.3×10^{-2}	6.7×10^{-2}	5.9×10^{-2}	-2.0×10^{-2}
6.7×10^{-2}	5.1×10^{-2}	4.7×10^{-2}	-1.5×10^{-2}
5.9×10^{-2}	4.7×10^{-2}	4.4×10^{-2}	-1.3×10^{-2}
-2.0×10^{-2}	-1.5×10^{-2}	-1.3×10^{-2}	4.4×10^{-3}

Covariance matrix (square lattice)

1.6×10^{-3}	4.3×10^{-5}	-1.0×10^{-4}	-5.7×10^{-5}
4.3×10^{-5}	6.8×10^{-5}	5.4×10^{-5}	1.6×10^{-6}
-1.0×10^{-4}	5.4×10^{-5}	7.4×10^{-5}	1.2×10^{-5}
-5.7×10^{-5}	1.6×10^{-6}	1.2×10^{-5}	1.3×10^{-5}

ten rounds of the experiment, i.e., $x_1 \approx \bar{x}$. In this way we avoid the technical difficulty that the prior state would actually depend on the final state of the dynamics, i.e., the fixed point. We also estimate the standard deviation σ of the observation error to be equal to the standard deviation of the last ten points in the time series $\mathbf{x}_{\text{obs}}(1:T)$. We assign, however, a standard deviation three times larger 3σ to the first two points in the dynamics to take into account that the adiabatic approximation is expected to capture better the slower dynamics that follows the initial transient regime of rather fast decay.

3. Parameter estimation in the regime of bistability

This case is a bit more complex since now the relationships between parameters and experimental values, i.e., Eqs. (D41)–(D44), depend nontrivially on the fixed points of the underlying dynamics. Moreover, in contrast to the previous case, here we cannot disentangle this dependence as the observed long-term cooperation level is related to the underlying dynamics by $\bar{x} = (1-\mu)x_1^* + \mu x_2^*$. The prior here is also defined indirectly. As discussed in Appendix D 3, in this regime it is convenient to parametrize the model in Eq. (12) in terms of the two stable fixed points x_1^* and x_2^*, the only unstable fixed point x_u^*, and α, i.e., $\Theta = (\alpha, x_1^*, x_2^*, x_u^*)$.

To take into account the influence of the two stable fixed points on the underlying dynamics, we describe the observed dynamics as $x_{\text{obs}}(t) = (1-\mu)x_1(t) + \mu x_2(t)$, where $x_1(t)$ and $x_2(t)$ represent the dynamics given by Eq. (12) with two different initial conditions $x_1(0)$ and $x_2(0)$. For any given choice of the parameters Θ, we use Eqs. (D45) and (D47) and invert Eq. (14) to compute the corresponding values of A, x_0, and y_0. With these and the parameter α, we can use Eqs. (D34)–(D36) to compute the corresponding parameters $a_{\text{dyn}} = \beta K a$, $b_{\text{dyn}} = \beta K b$, and $h_{\text{dyn}} = \beta h$ that, along with the initial conditions $x_1(0)$ and $x_2(0)$, fully specify the underlying dynamics through Eq. (12). Furthermore, using Eqs. (D18) and (D19), we can estimate the corresponding values for the slopes and intercepts describing the MCC rule and then compare with the experimental values reported. If the values obtained happen to be outside the credible interval $[O^* - \zeta\delta O^*, O^* + \zeta\delta O^*]$ defined by the choice of parameter ζ, then such a specific value for the parameters Θ is rejected.

4. Implementation

We have used the package POMP [68] implemented in R to perform the Bayesian inference via a particle Markov chain Monte Carlo with an adaptive random walk as the proposed distribution. This is a package specifically designed for parameter inference of partially observed Markov processes.

5. Results

Table II summarizes the experimental results reported in Ref. [19]. The quantity $\bar{x}$ represents the global level of cooperation reached by the system of interacting humans in the laboratory. We estimate this quantity and its standard error by computing the average and standard deviation, respectively, of the global cooperation [Figs. 2(a) and 2(b)] over the last ten rounds of each of the two experiments performed in [19], on a heterogeneous network and on a square lattice. These are the two experiments that we analyze in this paper and to which we refer in this section.

Table III shows the parameters estimated for the two experiments and their corresponding covariance matrix. In the case of the experiment on a heterogeneous network (square lattice) we used $\zeta = 1.96$ ($\zeta = 1.28$) corresponding to a uniform distribution on the experimental quantities representing a 97.5% (90%) credible interval (Appendix E 2). Since these results were obtained in the regime of monostability (Appendix E 2),

the parameters over which we performed Bayesian inference were $\Theta_{\text{expt}} = (m_C, m_D, r, G, x(0))$. (In the regime of bistability we did not find satisfactory results.) The dynamical parameters are then determined through Eqs. (D27)–(D30). Figures 2(a) and 2(b) show the dynamics corresponding to these parameters in the case of the experiment on a heterogeneous network and on a square lattice, respectively. In contrast, Figs. 2(c) and 2(d) show the results of applying Eq. (11) using these parameter values.

Table IV shows the corresponding values of the parameters $\Theta_{\text{dyn}} = (\theta_{\text{phase}}, \alpha)$, where $\theta_{\text{phase}} = (A, x_0, y_0)$ directly locate the system in the phase diagram of the model (Fig. 1 in the main text) and, along with α, completely determine the dynamics of the system through Eq. (12). The covariance matrix reported is obtained by first-order error propagation of the results displayed in Table III. This was done to take into account that the constraints were enforced on Θ_{expt} during the inference process and produced the best visual results of Fig. 2. We also tried to first transform the posterior over the parameters Θ_{expt} into a posterior on the parameters Θ_{dyn} to then compute the average value over the latter, but the results were less satisfactory. In any case, Fig. 1 also shows the population of parameters Θ_{dyn} representing the corresponding posterior. We see that the values reported in Table IV, which are also shown in Fig. 1, indeed appear to be representative of the population.

Finally, we estimated the Euclidean distance $d(\theta_{\text{phase}}, \theta_c^*)$ of the parameters θ_{phase} to the closest point $\theta_c^* = (A_c^*, x_{0c}^*, y_{0c}^*)$ on the critical lines defined by Eqs. (B6)–(B8). Following the standard analysis of continuous phase transitions, we define a reduced or relative distance to the critical point as $\delta(\Theta_{\text{phase}}, \theta_c) = d(\theta_{\text{phase}}, \theta_c)/|\Theta_c|$, where $|\theta_c|$ stands for the Euclidean norm of the vector of parameters θ_c. Using the values in Table IV, we obtained the values of $\delta(\Theta_{\text{phase}}, \Theta_c) \approx 0.03$ and $\delta(\Theta_{\text{phase}}, \Theta_c) \approx 0.11$ for the experiments on a heterogeneous network and on a square lattice, respectively.

[1] M. Milinski, D. Semmann, and H.-J. Krambeck, Reputation helps solve the 'tragedy of the commons', Nature (London) **415**, 424 (2002).

[2] E. Ostrom, J. Walker, and R. Gardner, Covenants with and without a sword: Self-governance is possible, Am. Pol. Sci. Rev. **86**, 404 (1992).

[3] T. Yamagishi and T. Kiyonari, The group as the container of generalized reciprocity, Soc. Psychol. Q. **63**, 116 (2000).

[4] E. Ostrom, *Understanding Institutional Diversity* (Princeton University Press, Princeton, 2005).

[5] S. Bowles and H. Gintis, *A Cooperative Species: Human Reciprocity and its Evolution* (Princeton University Press, Princeton, 2011).

[6] M. A. Janssen, R. Holahan, A. Lee, and E. Ostrom, Lab experiments for the study of social-ecological systems, Science **328**, 613 (2010).

[7] E. Fehr and U. Fischbacher, Social norms and human cooperation, Trends Cognit. Sci. **8**, 185 (2004).

[8] K. Nyborg *et al.*, Social norms as solutions, Science **354**, 42 (2016).

[9] C. Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (Cambridge University Press, Cambridge, 2006).

[10] *Minding Norms: Mechanisms and Dynamics of Social Order in Agent Societies*, edited by R. Conte, G. Andrighetto, and M. Campennì, Oxford Series on Cognitive Models and Architectures (Oxford University Press, Oxford, 2013).

[11] J. Elster, *Nuts and Bolts for the Social Sciences* (Cambridge University Press, Cambridge, 1989).

[12] M. Chudek and J. Henrich, Culture-gene coevolution, norm-psychology and the emergence of human prosociality, Trends Cognit. Sci. **15**, 218 (2011).

[13] E. L. Krupka and R. A. Weber, Identifying social normal using coordination games: Why does dictator game sharing vary? J. Eur. Econ. Assoc. **11**, 495 (2013).

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

[15] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, 1998).

[16] N. Feltovich, *Encyclopedia of the Sciences of Learning* (Springer, New York, 2012), pp. 444–447.

[17] T. H. Ho, C. F. Camerer, and J.-K. Chong, Self-tuning experience weighted attraction learning in games, J. Econ. Theory **133**, 177 (2007).

[18] L. Zhu, K. E. Mathewson, and M. Hsu, Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning, Proc. Natl. Acad. Sci. U.S.A. **109**, 1419 (2012).

[19] C. Garcia-Lázaro, A. Ferrer, G. Ruiz, A. Tarancón, J. Cuesta, A. Sánchez, and Y. Moreno, Heterogeneous networks do not promote cooperation when humans play a prisoner's dilemma, Proc. Natl. Acad. Sci. U.S.A. **109**, 12922 (2012).

[20] G. Cimini and Sánchez, Learning dynamics explains human behavior in prisoner's dilemma on networks, J. R. Soc. Interface **11**, 20131186 (2014).

[21] G. Cimini and A. Sanchez, How evolutionary dynamics affects network reciprocity in prisoner's dilemma, JASSS **18**, 22 (2015).

[22] D. Vilone, J. J. Ramasco, A. Sánchez, and M. S. Miguel, Social imitation versus strategic choice, or consensus versus cooperation, in the networked prisoner's dilemma, Phys. Rev. E **90**, 022810 (2014).

[23] T. Ezaki, Y. Horita, M. Takezawa, and N. Masuda, Reinforcement learning explains conditional cooperation and its moody cousin, PLoS Comput. Biol. **12**, e1005034 (2016).

[24] Y. Horita, M. Takezawa, K. Inukai, T. Kita, and N. Masuda, Reinforcement learning accounts for moody conditional cooperation behavior: Experimental results, Sci. Rep. **7**, 39275 (2017).

[25] M. A. Munoz, Colloquium: Criticality and dynamical scaling in living systems, arXiv:1712.04499.

[26] J. Hidalgo, J. Grilli, S. Suweis, M. A. Muñoz, J. R. Banavar, and A. Maritan, Information-based fitness and the emergence of criticality in living systems, Proc. Nat. Acad. Sci. U.S.A. **111**, 10095 (2014).

[27] T. Mora and W. Bialek, Are biological systems poised at criticality? J. Stat. Phys. **144**, 268 (2011).
[28] A. Gelblum, I. Pinkoviezky, E. Fonio, A. Ghosh, N. Gov, and O. Feinerman, Ant groups optimally amplify the effect of transiently informed individuals, Nat. Commun. **6**, 7729 (2015).
[29] G. S. Tkačik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, and W. Bialek, Thermodynamics and signatures of criticality in a network of neurons, Proc. Natl. Acad. Sci. U.S.A. **112**, 11508 (2015).
[30] H. Chaté and M. A. Muñoz, Insect swarms go critical, Physics **7**, 120 (2014).
[31] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, and A. M. Walczak, Social interactions dominate speed control in poising natural flocks near criticality, Proc. Natl. Acad. Sci. U.S.A. **111**, 7212 (2014).
[32] A. Attanasi, A. Cavagna, L. Del Castello, I. Giardina, S. Melillo, L. Parisi, O. Pohl, B. Rossaro, E. Shen, E. Silvestri, and M. Viale, Finite-Size Scaling as a Way to Probe Near-Criticality in Natural Swarms, Phys. Rev. Lett. **113**, 238102 (2014).
[33] D. Krotov, J. O. Dubuis, T. Gregor, and W. Bialek, Morphogenesis at criticality, Proc. Natl. Acad. Sci. U.S.A. **111**, 3683 (2014).
[34] M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, Gene expression dynamics in the macrophage exhibit criticality, Proc. Natl. Acad. Sci. U.S.A. **105**, 1897 (2008).
[35] T. Mora, A. M. Walczak, W. Bialek, and C. G. Callan, Maximum entropy models for antibody diversity, Proc. Natl. Acad. Sci. U.S.A. **107**, 5405 (2010).
[36] J. M. Beggs, The criticality hypothesis: How local cortical networks might optimize information processing, Philos. Trans. R. Soc. A **366**, 329 (2008).
[37] M. Perc, J. J. Jordan, D. G. Rand, Z. Wang, S. Boccaletti, and A. Szolnoki, Statistical physics of human cooperation, Phys. Rep. **687**, 1 (2017).
[38] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. **81**, 591 (2009).
[39] A. Sánchez, Physics of cooperation: Experimental evidence and theoretical models, J. Stat. Mech. (2018) 024001.
[40] U. Fischbacher, S. Gächter, and E. Fehr, Are people conditionally cooperative? Evidence from a public goods experiment, Econ. Lett. **71**, 397 (2001).
[41] C. Keser and F. van Winden, Conditional cooperation and voluntary contributions to public goods, Scand. J. Econ. **102**, 23 (2000).
[42] D. Lee and H. Seo, Neural basis of strategic decision making, Trends Neurosci. **39**, 40 (2016).
[43] P. Glimcher and E. Fehr, *Neuroeconomics: Decision Making and the Brain* (Elsevier Science, Amsterdam, 2013).
[44] T. Galla and J. D. Farmer, Complex dynamics in learning complicated games, Proc. Natl. Acad. Sci. U.S.A. **110**, 1232 (2013).
[45] Y. Sato, E. Akiyama, and J. D. Farmer, Chaos in learning a simple two-person game, Proc. Natl. Acad. Sci. U.S.A. **99**, 4748 (2002).
[46] G. Andrighetto, J. Brandts, R. Conte, J. Sabater-Mir, H. Solaz, and D. Villatoro, Punish and voice: Punishment enhances cooperation when combined with norm signaling, PLoS One **8**, e64941 (2013).
[47] R. B. Cialdini and N. J. Goldstein, Social influence: Compliance and conformity, Annu. Rev. Psychol. **55**, 591 (2004).
[48] L. Festinger, *A Theory of Cognitive Dissonance* (Stanford University Press, Stanford, 1957).
[49] R. P. Abelson and A. Bernstein, A computer simulation model of community referendum controversies, Pub. Opin. Q. **27**, 93 (1963).
[50] S. Ayal and F. Gino, *The Social Psychology of Morality: Exploring the Causes of Good and Evil* (American Psychological Association, Washington, DC, 2011), pp. 149–166.
[51] A. Traulsen, D. Semmann, R. D. Sommerfeld, H.-J. Krambeck, and M. Milinski, Human strategy updating in evolutionary games, Proc. Natl. Acad. Sci. U.S.A. **107**, 2962 (2010).
[52] J. Grujić, C. Fosco, L. Araujo, J. A. Cuesta, and J. A. Sánchez, Social experiments in the mesoscale: Humans playing a spatial prisoner's dilemma, PLoS ONE **5**, e13749 (2010).
[53] J. Grujić, C. Garcia-Lázaro, M. Milinski, S. Semmann, A. Traulsen, J. A. Cuesta, Y. Moreno, and A. Sánchez, A comparative analysis of spatial prisoner's dilemma experiments: Conditional cooperation and payoff irrelevance, Sci. Rep. **4**, 4615 (2014).
[54] A. Sánchez, Theory must be informed by experiments (and back): Comment on "Universal scaling for the dilemma strength in evolutionary games", by Z. Wang *et al.*, Phys. Life Rev. **14**, 52 (2015).
[55] M. Gutiérrez-Roig, C. Gracia-Lázaro, J. Perelló, Y. Moreno, and A. Sánchez, Transition from reciprocal cooperation to persistent behavior in social dilemmas at the end of adolescence, Nat. Commun. **5**, 5362 (2014).
[56] C. Camerer and T.-H. Ho, Experience-weighted attraction learning in coordination games: Probability rules, heterogeneity, and time-variation, J. Math. Psychol. **42**, 305 (1998).
[57] D. Lee, Decision making: From neuroscience to psychiatry, Neuron **78**, 233 (2013).
[58] A. Lin, R. Adolphs, and A. Rangel, Social and monetary reward learning engage overlapping neural substrates, Soc. Cognit. Affect. Neurosci. **7**, 274 (2011).
[59] C. Declerck and C. Boone, *Neuroeconomics of Prosocial Behavior: The Compassionate Egoist* (Elsevier Science, Amsterdam, 2015).
[60] H. Gintis, The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms, J. Theor. Biol. **220**, 407 (2003).
[61] S. Gavrilets and P. J. Richerson, Collective action and the evolution of social norm internalization, Proc. Natl. Acad. Sci. U.S.A. **114**, 6068 (2017).
[62] S. Bowles, *The Moral Economy: Why Good Incentives are No Substitute for Good Citizens* (Yale University Press, New Haven, 2016).
[63] *Experimenting with Social Norms: Fairness and Punishment in Cross-Cultural Perspective*, edited by J. Ensminger and J. Henrich (Russel Sage Foundation, New York, 2014).
[64] D. G. Rand, J. D. Greene, and M. A. Nowak, Spontaneous giving and calculated greed, Nature (London) **489**, 427 (2012).
[65] M. A. Nowak and R. M. May, Evolutionary games and spatial chaos, Nature (London) **359**, 826 (1992).
[66] D. Villatoro, G. Andrighetto, J. Brandts, L. G. Nardin, J. Sabater-Mir, and R. Conte, The norm-signaling effects of group punishment: Combining agent-based simulation and laboratory experiments, Soc. Sci. Comput. Rev. **32**, 334 (2014).

[67] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters and Complexity*, Oxford Master Series in Physics (Oxford University Press, Oxford, 2006).

[68] A. A. King, D. Nguyen, and E. L. Ionides, Statistical inference for partially observed Markov processes via the R package POMP, J. Stat. Software **69**, 1 (2016).

[69] M. Schlüter, A. Tavoni, and S. Levin, Robustness of norm-driven cooperation in the commons, Proc. R. Soc. B **283**, 1 (2016).

[70] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. U.S.A. **113**, 3932 (2016).

[71] T. Mora, S. Deny, and O. Marre, Dynamical Criticality in the Collective Activity of a Population of Retinal Neurons, Phys. Rev. Lett. **114**, 078105 (2015).

[72] A. Cavagna, I. Giardina, F. Ginelli, T. Mora, D. Piovani, R. Tavarone, and A. M. Walczak, Dynamical maximum entropy approach to flocking, Phys. Rev. E **89**, 042707 (2014).

[73] I. Mastromatteo and M. Marsili, On the criticality of inferred models, J. Stat. Mech. (2011) P10012.

[74] T. Galla, Intrinsic Noise in Game Dynamical Learning, Phys. Rev. Lett. **103**, 198702 (2009).

[75] J. Realpe-Gomez, B. Szczesny, L. Dall'Asta, and T. Galla, Fixation and escape times in stochastic game learning, J. Stat. Mech. (2012) P10022.