

Bounded rational agents playing a public goods gamePrakhar Godara ^{*}, Tilman Diego Aléman [†] and Stephan Herminghaus [‡]*Max Planck Institute for Dynamics and Self-Organization (MPIDS), Am Faßberg 17, D-37077 Göttingen, Germany*

(Received 26 June 2021; revised 24 September 2021; accepted 17 December 2021; published 9 February 2022)

An agent-based model for human behavior in the well-known public goods game (PGG) is developed making use of bounded rationality, but without invoking mechanisms of learning. The underlying Markov decision process is driven by a path integral formulation of reward maximization. The parameters of the model can be related to human preferences accessible to measurement. Fitting simulated game trajectories to available experimental data, we demonstrate that our agents are capable of modeling human behavior in PGG quite well, including aspects of cooperation emerging from the game. We find that only two fitting parameters are relevant to account for the variations in playing behavior observed in 16 cities from all over the world. We thereby find that learning is not a necessary ingredient to account for empirical data.

DOI: [10.1103/PhysRevE.105.024114](https://doi.org/10.1103/PhysRevE.105.024114)**I. INTRODUCTION**

The arguably most important question of our time is how humankind can devise a sustainable management of its ecological niche on planet Earth [1]. Scientific problems concerning the many aspects of sustainability have thus attracted the interest of an increasingly active research community since about the turn of the millennium [2]. Aside from severe problems in dealing with limited planet resources and a changing global climate, a topic of major concern is the possible response of human societies to these stimuli. Having to deal with dire consequences of rapidly changing conditions, unwanted collective behavior may result, such as sedation or civil war. Hence an important goal of legislation and policy making is to have these systems evolve in a way which is as beneficial as possible for its agents.

Since legislation can only change the interaction rules which apply in human encounters, there is a need for theoretical modeling which is capable to predict the collective behavior in human societies on the basis of these interaction rules [3]. This bears close similarity to the physics of phase transitions and critical phenomena, where one seeks to predict the collective behavior of a large number of similar subsystems (such as molecules) solely from their known mutual interactions [4,5]. This paradigm has been applied successfully, e.g., in modeling the emergence of polarization in opinion dynamics [6–8], where collective behavior was found

to depend sensitively on details in the mutual interactions of agents. Hence in order to develop a predictive model of collective phenomena in societal dynamics, one has to model the interactions between individuals in a way sufficiently formal for access by theory, but still resembling human encounters as closely as possible.

While the interactions of opinions in topical space may be modeled by comparably simple mathematical structures [7,8], more general interactions between humans, including exchanges of resources and emotions, requires a much higher degree of modeling complexity. A classical paradigm for achieving such modeling is game theory, which has grown into a mature field of research, with extensions towards collective phenomena having emerged in recent years [9–14].

A frequently studied example is the so-called public goods game (PGG), in which players contribute resources to a common (public) pot, from which disbursements are paid back to all players equally [5,12,15,16]. This may be seen as modeling a wide variety of social interactions, since both contributions and disbursements may be monetary (e.g., taxes), goods (e.g., public infrastructure), activities (e.g., chores in a common household), or emotions (e.g., enjoying a tidy common household). In a society, every player is participating in many such games at the same time, such that society may, e.g., be viewed as a network of many PGG, coupled by the players they have mutually in common. The prediction of collective behavior in such a network rests, in the first place, on careful modeling of the agents and their interactions.

Despite extensive research in recent decades, the question how human behavior, as regards PGG, should be cast into a suitable model agent, must still be considered open. In PGG experiments played over a number of consecutive rounds among the same group of players, one widely observes that contributions to the common pot (referred to as cooperation) tend to decrease gradually from one round to the next [5,17]. This is not straightforward to account for through a simple Markovian agent model with a fixed transition matrix. Furthermore, it is commonly assumed that agents have full access

^{*}Corresponding author: prakhar.godara@ds.mpg.de[†]tilmandiego.aleman@stud.uni-goettingen.de[‡]stephan.herminghaus@ds.mpg.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

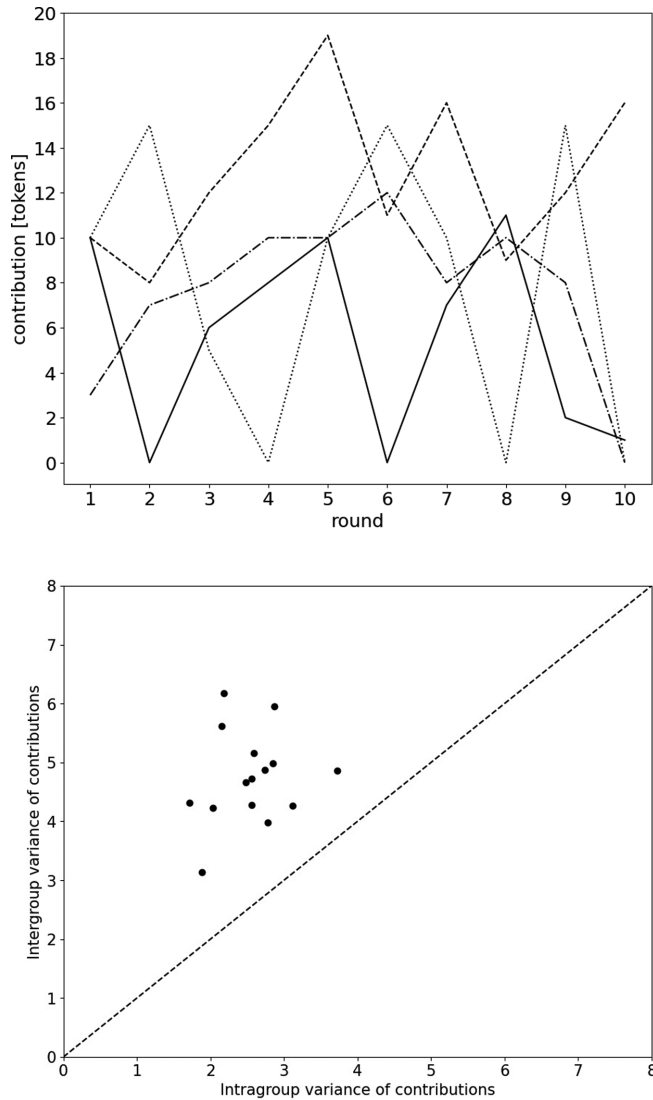


FIG. 1. Top: Trajectory of a single group game played among citizens of Athens [17]. Contributions of all four players are shown, in different styles. Bottom: Intragroup vs intergroup variance in the average contributions of players. Each data point corresponds to one city. The significant offset of the data above the first diagonal demonstrates the substantial coupling between players playing in the same group.

to all information necessary to perform optimal moves (“rational” agents), and are also designed to use this information exhaustively in search of a Nash equilibrium. As the latter can be shown to consist in defecting, i.e., zero contribution (assuming that the number of rounds are finite and known to all the players), this is clearly at variance with experimental data [18,19].

In search for less simplistic model agents which better account for experimental data, concepts of learning [20,21] have been put forward. In fact, the observed declining trend in average cooperation could be accounted for quite well [15,16]. However, a glance at typical game trajectories sheds some doubt on this to be the sole explanation. The top panel of Fig. 1 shows the trajectory of a single PGG (played among four citizens of Athens for ten subsequent rounds

[17]), showing the contributions of all players. It is not obvious what players should even try to “learn” from each other in such erratic trajectories. Nevertheless, the declining trend in contributions shows up once many trajectories are averaged [17]. It thus appears reasonable to investigate as well other possible extensions (other than learning) of agents with respect to the simple (fixed transition matrix rational) agent.

It is clear that real players are characterized by a certain lack of information as well as by limited resources to use the information they have access to. This observation is reflected in the concept of bounded rationality [22,23], which we focus on in the present paper. Our goal is to develop model agents whose parameters can be related to measurable individual preferences [24], and at the same time are capable to model experimental data on human playing behavior. As we will show, we are able to account for human playing behavior, both its “volatility” and the observed decline of cooperation, in 258 games [17], by merely assuming some foresight and bounded rationality, but without invoking any learning mechanism.

II. BOUNDED RATIONAL FORESIGHT

It appears reasonable to assume that, depending on the real-life situation modeled by the PGG, players may try to think a few steps ahead when deciding on their next move, or contribution. In fact, it is well accepted in the wider community that some foresight, usually cast in the notion of “agency” [25], is one of the key ingredients in human decision processes. This requires intellectual resources, as the course of the game needs to be anticipated somewhat into the future. As mentioned above, completely rational players (i.e., with infinite intellectual resources) will always defect in a finite PGG, as defection maximizes their gains irrespective of what other players do (Nash equilibrium). Experimental evidence about iterated PGG, however, shows that players rarely play the Nash equilibrium [18,19], but instead contribute substantially, with average contributions typically decreasing from one round to the next as the game proceeds, irrespective of the country or culture the players belong to [17].

Inspired from a formulation of bounded rationality [22,23], we use a path integral formulation to develop a type of model agent which tries to maximize its gain in future moves, but exhibits bounded rationality in a manner resembling some key aspects of a human player. Its parameters can be loosely attributed to certain traits of human character [24], as we will discuss in more detail further below. We then use these parameters to adjust the agent properties such as to fit data of public goods games which had been played by real human players [17]. As the public goods game can already be seen as a collective scenario in some sense, our approach is contrasting other work which assumes some agent behavior *a priori* and then immediately focuses on trying to predict collective effects [14,26,27].

A. Rules of the public goods game (PGG)

The public goods game (PGG) has become one of the major paradigms in game theory and is played by a finite set of N players. It can be summarized as follows:

(1) Each player is provided with the same number $\tau \in \mathbb{N}$ of tokens.

(2) Each player anonymously contributes an integer number of tokens to a “public” pot.

(3) Each player receives a return worth the total collection in the pot, multiplied by a number $\alpha \in \mathbb{R}$. This return, plus the tokens held back initially by the player, is called her reward.

This completes one round of the game. After that, the game continues for a finite (and previously known to all players) number T of rounds. Hence there is a chance for the players to “learn” over time how their companions are playing, and possibly to develop strategies to maximize their rewards accordingly. The PGG is completely characterized by the triple (N, T, α) . The number of tokens, τ , merely provides some kind of currency unit for the asset to be distributed. It is significant only to match real playing situations and does not affect the structure of the game.

Data have been made available from this game for a large number of players in different cities around the world [17]. In that study, each group consisted of four players. In each of the cities investigated, a number of groups of four players each were compiled from citizens in a random fashion. The number of groups per city varied from ten to 38. Each group played one game consisting of ten rounds. A maximum of 20 tokens could be invested in each round by a single player. It is well known from earlier experiments [18] that an increase in α leads to monotonously increasing contributions by the agents. In the study we refer to [17] it was chosen $\alpha = 0.4$. Wherever appropriate, we will choose the same conditions in our model in order to achieve maximum comparability, hence in the remainder of this paper we use $(N, T, \alpha) = (4, 10, 0.4)$.

While the average contributions were substantially different for different cities, it was quite generally found that the average (over all the trajectories in a city) contributions of the players tended to decrease gradually as the game proceeded. An example is shown in Fig. 1(a) for a single group from Athens. This as well as the rather erratic variations is clearly at variance with what one would expect for Nash equilibrium players.

The mutual interaction of the players within a group also becomes apparent when one considers the variance of the average contributions of the players. As Fig. 1(b) shows, the intragroup variance (abscissa) was generally smaller than the variance of average contributions among all groups of the same city (ordinate). Since players were picked from citizens in a random fashion, this suggests a certain degree of cooperativity, or peer pressure, between players within the same group as to the style of playing, either more parsimoniously or more generously. In the present paper, we will refer to this phenomenon using the neutral term *coupling*. The agent we want to develop should be capable of modeling as many of these traits as possible.

In the remainder of this section, we will develop our agent model in detail. In Sec. III we present the numerical implementation of the model and discuss some simulation results demonstrating its properties. Results of fitting the model to experimental data are shown in Sec. IV. Finally, we will suggest some future directions.

B. Developing the agent model

Exploiting the inherent symmetry in the game, we will state the model for only one agent with the index $k \in \{1, \dots, N\}$. From the perspective of the k th agent we shall often refer to the other agents as the *system*, and we call the k th agent the agent under consideration, or just the *agent*, for short. First we will introduce the model assuming full rationality and later introduce the concept of bounded rational agents. We use the following notations:

(1) $f_{k,t} \in \mathbb{N} :=$ the contribution (or “action”) of the k th agent at turn $t \in \{1, \dots, T\}$.

(2) $\vec{f}_t = (f_{1,t}, \dots, f_{N,t})$ is the state of the game at the end of turn t . The bar on the top represents a vector quantity.

(3) $\theta_t^T = (\vec{f}_t, \vec{f}_{t+1}, \dots, \vec{f}_T)$, is the total trajectory of the game from turn t to T .

(4) $G_k(\vec{f}_t) = \alpha \sum_i f_{i,t} + (\tau - f_{k,t})$ is the immediate gain of agent k at turn t . The last term represents the “gain” from what was not contributed.

(5) $\mathcal{G}_k[\theta_t^T] = \sum_{t'=t}^T G_k(\vec{f}_{t'})$ is the cumulative gain of agent k from turn t to T .

(6) $P(\theta_t^T) :=$ The probability of a trajectory from time t to T .

The aim of each agent is to maximize the cumulative gain achieved at the end of all turns. At each round t of the game, the k th agent chooses an action $f_{k,t}$ so as to maximize the expected cumulative gain. The latter, however, depends not only on the agent’s actions but also the other agent’s actions (i.e., the system). Hence due to the anonymity of the game and lack of information, the agent can have only a probabilistic model of the evolution of the system’s state.

For the discussion which follows, it will prove useful to introduce the following quantities:

(1) $\vec{f}_{-k,t} :=$ The contribution of all agents except the k th agent at turn t (we may identify $\vec{f}_t = (f_{k,t}, \vec{f}_{-k,t})$).

(2) $\pi_t^T = (\vec{f}_{-k,t}, \vec{f}_{-k,t+1}, \dots, \vec{f}_{-k,T})$, the system trajectory from turn t to T .

(3) $P(\pi_t^T) :=$ The probability of system trajectory from time t to T .

C. Rational agents

In order to play the game successfully, each agent has to calculate the likelihood of a particular trajectory of the game (system + agent) and hence be able to act so as to maximize the expected gain over the entire trajectory. In this model we assume that the agents are Markovian, i.e., their decisions in a given round are dependent only on the state of the game in the previous round and not on the states further back in the game history. While this may appear as a gross simplification, we will see that human playing behavior can be accounted for quite well by this assumption. The probability of realizing a trajectory $P(\theta_t^T)$ is then given by

$$P(\theta_t^T) = \prod_{t'=t}^T P(\vec{f}_{t'} | \vec{f}_{t'-1}) = \prod_{t'=t}^T P(f_{kt'}, \vec{f}_{-kt'} | \vec{f}_{t'-1}). \quad (1)$$

Note that there are two kinds of processes at play here. First, the stochastic process describing the system (from the perspective of the k th agent) which generates $\vec{f}_{-k,t}$. Second, there is the choice of the agent, $f_{k,t}$. Since the game is played

anonymously, both of these processes can be assumed to be independent. This can be used to write $P(f_{k,t}, \bar{f}_{-k,t} | \bar{f}_{t-1}) = P(f_{k,t} | \bar{f}_{t-1})P(\bar{f}_{-k,t} | \bar{f}_{t-1})$. We refer to $P(\bar{f}_{-k,t} | \bar{f}_{t-1})$ as the *transition function*. In order to avoid confusion, we rename it as $Q(\bar{f}_{-k,t} | \bar{f}_{t-1})$. This now allows us to write

$$\begin{aligned} P(\theta_t^T) &= \prod_{t'=t}^T P(f_{kt'} | \bar{f}_{t'-1}) Q(\bar{f}_{-kt'} | \bar{f}_{t'-1}) \\ &= P(\pi_t^T) P(f_t^T), \end{aligned} \quad (2)$$

where $P(f_t^T) = \prod_{t'=t}^T P(f_{kt'} | \bar{f}_{t'-1})$ is called the *policy* of the agent.

We now write the optimization problem the agent faces at turn t as

$$V_t[P(f_t^T)] \longrightarrow \max, \quad (3)$$

where

$$V_t[P(f_t^T)] = \sum_{\theta_t^T} P(\theta_t^T) \mathcal{G}[\theta_t^T] \quad (4)$$

is the expected cumulative gain (also called the value functional).

The maximum value of V_t will henceforth be called V_t^* . Writing $\mathcal{G}[\theta_t^T] = G(\bar{f}_t) + \mathcal{G}[\theta_{t+1}^T]$, the summation can be broken down into two parts, the first of which is the immediate expected gain, and the second is the future expected gain. By additionally using the normalization of the path probabilities we can write the equation in a recursive form as

$$V_t^*[P(f_t^T)] = \max_{f_t^T} \sum_{\bar{f}_t} P(\bar{f}_t | \bar{f}_{t-1}) [G(\bar{f}_t) + V_{t+1}[P(f_{t+1}^T)]]. \quad (5)$$

Now, making use of Eq. (2), we can write the above more explicitly as

$$\begin{aligned} V_t^* &= \max_{f_t^T} \sum_{\bar{f}_t} P(f_{k,t} | \bar{f}_{t-1}) \\ &\quad \times [Q(\bar{f}_{-k,t} | \bar{f}_{t-1}) G_k(\bar{f}_t) + Q(\bar{f}_{-k,t} | \bar{f}_{t-1}) V_{t+1}]. \end{aligned} \quad (6)$$

This equation is known in the literature as the *Bellman equation* [28], which was originally used in optimal control theory. In common applications, this equation also includes a *discount factor* $0 \leq \gamma \leq 1$, which is the factor by which the future gains are discounted in comparison to immediate rewards. One can then write

$$\begin{aligned} V_t^* &= \max_{f_t^T} \sum_{\bar{f}_t} P(f_{k,t} | \bar{f}_{t-1}) \\ &\quad \times [Q(\bar{f}_{-k,t} | \bar{f}_{t-1}) G_k(\bar{f}_t) + \gamma Q(\bar{f}_{-k,t} | \bar{f}_{t-1}) V_{t+1}], \end{aligned} \quad (7)$$

where $\gamma = 0$ would indicate an extremely myopic agent and $\gamma = 1$ would represent an extremely far-sighted agent. In terms of established preference dimensions, γ should be closely related to patience [24].

D. Rationally bounded agents

So far the agents have been completely rational. By rational we mean that the agent is able to perform the computations and solve the optimization problem mentioned above. The

reader must not confuse this with the notion of rationality used conventionally in game theory, which not only assumes infinite computational capabilities but also assumes that the agent has perfect information about the game and other players. In our model the agent doesn't have perfect information about other players, and this aspect is incorporated by the use of a transition function. Therefore our rational agents may not necessarily play the Nash equilibrium in the intermediate rounds as their contributions will depend on their respective transition functions.

In order to model real human players, we need to introduce a form of *bounded rationality*. It can be argued that human players playing the PGG do not quite maximize the functional as in Eq. (7), like a fully rational agent might do, due to either limited time or computational capabilities. Consider, for instance, a player who has maximally limited computational capabilities or has zero time to perform the optimization. Such a player is bound to contribute randomly and independently of other agents. An agent modeling such behavior would contribute random samples from a prior distribution $P_0(f_{k,t})$. This distribution represents the basal tendency of the agent, and any deviations from the basal play would involve some cost of performing computations. In this model we assume that the agents have a computational budget K , which represents the degree to which they can "deviate" from their basal tendency in search of an optimal strategy.

The functional form of the cost of computation is adopted from [22,23]. Making use of it, we constrain the optimization problem in Eq. (7) by the computational budget of the agent and write the optimization problem faced by the agent on turn t as

$$\begin{aligned} V_t^* &= \max_{f_t^T} \sum_{\bar{f}_t} P(f_{k,t} | \bar{f}_{t-1}) \left[Q(\bar{f}_{-k,t} | \bar{f}_{t-1}) G_k(\bar{f}_t) \right. \\ &\quad \left. + \gamma Q(\bar{f}_{-k,t} | \bar{f}_{t-1}) V_{t+1} \right], \\ &\text{with } D_{KL}[P(f_{k,t} | \bar{f}_{t-1}) | P_0(f_{k,t} | \bar{f}_{t-1})] \leq K. \end{aligned} \quad (8)$$

K is the computational budget of the agent in round t , $D_{KL}(\cdot | \cdot)$ is the Kullback-Leibler divergence, and $P_0(f_{k,t} | \bar{f}_{t-1})$ is the prior distribution. We can introduce a Lagrange parameter to write the optimization problem as

$$\begin{aligned} V_t^* &= \max_{f_t^T} \sum_{\bar{f}_t} P(f_{k,t} | \bar{f}_{t-1}) \left[Q(\bar{f}_{-k,t} | \bar{f}_{t-1}) G_k(\bar{f}_t) \right. \\ &\quad \left. - \frac{1}{\beta} \log \frac{P(f_{k,t} | \bar{f}_{t-1})}{P_0(f_{k,t} | \bar{f}_{t-1})} + \gamma Q(\bar{f}_{-k,t} | \bar{f}_{t-1}) V_{t+1} \right], \end{aligned} \quad (9)$$

where β is the inverse of the Lagrange parameter for the bounded optimization problem. Because we have an inequality constraint, an additional condition for optimality is given by the KKT condition [29], i.e., $\frac{1}{\beta} \{D_{KL}[P^*(f_{k,t} | \bar{f}_{t-1}) | P_0(f_{k,t} | \bar{f}_{t-1})] - K\} = 0$. This means that if the optimal action is within the bounds of the computational capabilities of the agent, the agent will act optimally, else β is chosen such that $D_{KL}[P^*(f_{k,t} | \bar{f}_{t-1}) | P_0(f_{k,t} | \bar{f}_{t-1})] = K$ and $P^*(f_{k,t} | \bar{f}_{t-1})$ is a solution of Eq. (9).

This concludes the model of the agent, which can be seen as defined by the quadruple $(Q(\bar{f}_{-k,t}|\bar{f}_{t-1}), P_0(f_{k,t}|\bar{f}_{t-1}), K, \gamma)$. The transition model encapsulates the agent's internal model of the system, the prior action represents the basal tendency of the agent, K expresses the computational limitations, and γ represents the degree of myopia of the agent. Although K could in principle vary from one round to another, we assume the computational constraint K to be the same for all the rounds, considering it as a trait of the agent.

We now turn to solving Eq. (9). The value function at turn t cannot be evaluated, because future actions are not known *ab initio* and yet they need to be considered in the optimization problem. This problem can be resolved in the same spirit as Bellman's, through what is called *backward induction*. Instead of starting from turn t , we can start from the last turn and obtain a series of nested functions which can then iteratively lead to turn t . Although simple in principle, an analytical solution can be obtained only in a few special cases. Numerical solution of this problem, however, is straightforward.

III. NUMERICAL SIMULATIONS

A. Model assumptions

First, we make the simplifying assumption that the prior of the agent $P_0(f_{kt}|\bar{f}_{t-1})$, is independent of the previous state, i.e., we replace $P_0(f_{kt}|\bar{f}_{t-1})$ with $P_0(f_{kt})$. This explicitly excludes any learning mechanism, as we want to explore to what extent we can account for observed behavior exclusively by bounded rational foresight.

Second, we assume that the agents have truncated Gaussian priors given by

$$TG(f; m, \sigma) = \begin{cases} \mathcal{N}e^{\frac{(f-m)^2}{2\sigma^2}}, & 0 \leq f \leq \tau, \\ 0, & \text{otherwise,} \end{cases}$$

where we set $\tau = 20$ in order to relate to the data we intend to compare our simulations with [17]. \mathcal{N} is the normalization constant, along with a fixed variance $\sigma^2 = 25$. By varying the peak $m \in (-\infty, \infty)$ of the distribution we can span the basal tendencies from being very greedy (small m), indifferent (intermediate m), or very benevolent (large m). The corresponding prior distributions are displayed in Fig. 2. The parameter m , which may serve as a fitting parameter determined individually for each agent, is constant over the full game. It can be seen as resembling a character trait of the respective player.

Third, because of the anonymity of the players in the game we can assume symmetry across the \bar{f}_{-kt} variables and separately across the \bar{f}_{-kt-1} variables in $Q(\bar{f}_{-k,t}|\bar{f}_{t-1})$. Additionally, we need only to consider the distribution of the means of \bar{f}_{-kt} and \bar{f}_{-kt-1} as these are the only relevant quantities in the game. Therefore, from the definition of the agent we replace the transition function $Q(\bar{f}_{-k,t}|\bar{f}_{t-1})$ with $Q(\mu_t|\mu_{t-1}, f_{kt-1})$, where $\mu_t = \frac{\sum_{i \neq k} f_{it}}{N-1}$.

Finally, we assume that the transition function $Q(\mu_t|\mu_{t-1}, f_{kt-1})$ is a truncated Gaussian with the most

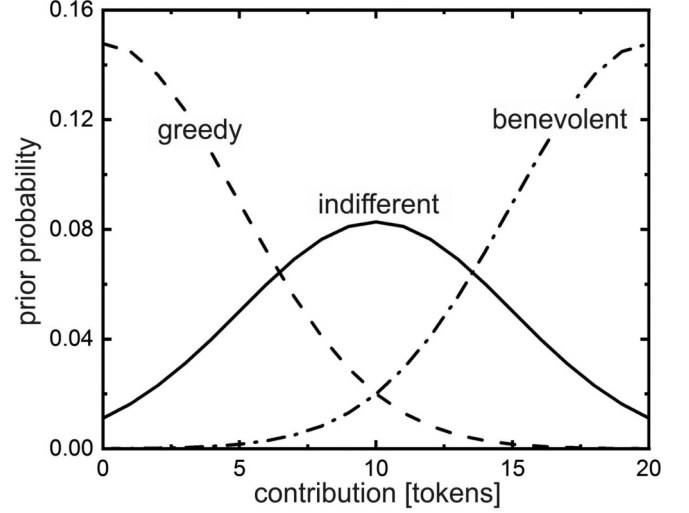


FIG. 2. Basal tendencies of the agents $P_0(f_{k,t})$ as given by truncated Gaussian prior distributions with variable m .

likely value of the Gaussian given by

$$\mu_t^{\text{peak}} = \begin{cases} \mu'_{t-1} + \xi_+ |\mu'_{t-1} - f_{kt-1}|, & \mu'_{t-1} - f_{kt-1} < 0 \\ \mu'_{t-1} - \xi_- |\mu'_{t-1} - f_{kt-1}|, & \mu'_{t-1} - f_{kt-1} > 0 \end{cases} \quad (10)$$

and some fixed variance ($\sigma_{\text{trans}} = 3$) [30]. Here μ_t^{peak} is the most likely value of μ_t , μ'_{t-1} is the observed value of μ_{t-1} and ξ_{\pm} are scalar parameters. This assumption is based upon the idea that an agent's contributions can either have an encouraging or a discouraging impact on other agents, and ξ_+ and ξ_- control the degree to which other agents are being encouraged or discouraged. In summary, the prior distributions are parameterized by m , while the transition functions are parameterized by ξ_{\pm} .

B. Relation to known human preferences

Accordingly, each agent is completely described by its tuple $(\xi_{\pm}, m, K, \gamma)$, which is considered constant over the game it plays. It is instructive to compare these parameters to known human preference parameters used, e.g., in the Global Preference Survey (GPS [24]). In that work, which provides a compilation of economic preferences all across the planet, six parameters affecting human choices are considered: patience (willingness to wait), risk taking (willingness to take risks in general), positive reciprocity (willingness to return a favor), negative reciprocity (willingness to take revenge), altruism (willingness to give to good causes), and trust (assuming people have good intentions).

There is a large body of literature on why these preferences are considered particularly important for economic decisions [31–34], but this shall not concern us here. What we immediately recognize is a relation between the parameter ξ_+ and ξ_- in our model on the one hand and positive and negative reciprocity on the other hand. Furthermore, patience, which is measured by the willingness to delay a reward if this increases the amount rewarded, obviously relates to the foresight expressed by the attempt to maximize the reward path integral (instead of focusing on the reward in a single

TABLE I. Backward Induction.

```

1: function POLICYM,  $t_{\max}$ 
2:    $t = t_{\max}$ 
3:    $\mathbf{E} = 0$ 
4:    $\mathbf{V} = 0$ 
5:   while  $t > 0$ 
6:     for  $0 \leq f_{1,t-1} \leq \tau$  do
7:       for  $0 \leq \mu_{t-1} \leq \tau$  do
8:         for  $0 \leq a \leq \tau$  do
9:            $\mathbf{E}[a] = \text{GAIN}(\mathbf{Q}, a, \mu_{t-1}, f_{1,t-1}, \mathbf{V})$ 
10:        end for
11:        $j = \arg \max \mathbf{E}$ 
12:        $\text{policy}[t, f_{1,t-1}, \mu_{t-1}] = (\delta_{kj})_{k \in \{0, \dots, (N-1)\tau\}}$ 
13:        $\mathbf{V} = \text{policy}[t, f_{1,t-1}, \mu_{t-1}] \cdot \mathbf{E}$ 
14:     end for
15:   end for
16:    $t = t - 1$ 
17: end while
18: RETURN policy
19: end function

```

$$\text{GAIN}(\mathbf{Q}, a, \mu_{t-1}, f_{1,t-1}, \mathbf{V}) = \mathbf{Q}(\cdot | \mu_{t-1}, f_{1,t-1}) \cdot \{\gamma \mathbf{V} + \alpha[(0, \dots, (N-1)\tau) + a] + \tau - a\},$$

where $\mathbf{Q}(\cdot | \mu_{t-1}, f_{1,t-1})$ refers to the probability distribution vector of having certain cumulative bets given μ_{t-1} and $f_{1,t-1}$.

For bounded rational agents, the method is more or less identical, except that instead of having delta distributions with their peak at the maximum expected value, we need to solve the constrained maximization problem. In terms of the algorithm, this means that instead of immediately taking the delta distribution, we first check if $D_{KL}(P_0 || \delta) \leq K$ for the agent's prior P_0 , the delta distribution as described in line 12 of the algorithm and his rationality parameter K . If this is the case, the computational cost of playing optimally in this situation is compatible with the computational budget of our agent, so his policy is exactly δ .

Otherwise, we find β such that

$$D_{KL}(P_0 || P^*) = K.$$

Here $P^*(f_{k,t}) = c P_0(f_{k,t}) \exp(\beta E[f_{k,t}])$, $f_{k,t} \in \{0, \dots, \tau\}$ where c is just a normalizing constant to get a probability distribution.

Note that for this algorithm one needs to know both μ_{t-1} and $f_{k,t-1}$ in order to evaluate the expected cumulative gain. This leads to a problem for the first round, as no history yet exists. Therefore in our implementation we manually initialize the group with an appropriate state \bar{f}_0 . For instance, when fitting simulations to experimental data, we initialize the group of agents with the initial contributions of the corresponding players. The code was written in Python using the just-in-time compiler Numba [35] and the numerical library Numpy [36].

D. Solution space of the model

In this section we will demonstrate the kind of behavior our model agents can exhibit. Since the main goal of the present work is to develop agents whose playing behavior is

round). Notions like risk taking, altruism, and trust will certainly reflect in the value m assigned to a player, and to some extent also affect ξ_{\pm} . Hence the ingredients of our model are by no means *ad hoc*, but are widely accepted to exist, to be relevant for decisions, and to vary considerably among different cultures across the globe [24].

C. Algorithm

Above we used backward induction to compute the (bounded rational) policy of an agent. For a more instructive description of the algorithm, we now describe the fully rational case in more detail (see Table I). The corresponding policy can be obtained again by backward induction, using the transition matrix for the system, with the transition matrix P .

We use the notation $\mu_{t-1} := \frac{1}{N-1} \sum_{k \neq 1} f_{k,t-1}$ for the cumulative bets of the other agents and assume for simplicity that the agent the policy of which we are interested in has index 1.

Here δ_{kj} is the Kronecker delta, bold notation refers to vector-valued variables, and the GAIN function is defined as

similar to that of the players in Ref. [17], we set $(N, T, \alpha, \tau) = (4, 10, 0.4, 20)$, as was chosen in that study, for the remainder of this paper. Due to the high dimensionality of the parameter space, we show game trajectories for four-player groups with only a few configurations given by $(\xi_{\pm}, m, K, \gamma)$.

I. Choice of ξ_{\pm}

In order to understand the significance of ξ_{\pm} , we consider fully rational agents for the sake of simplicity, with $(m, K, \gamma) = (10, \infty, 1)$, and consider the average (over rounds and agents) contribution of the group, $\langle A \rangle$, while the two components of $\bar{\xi}$ are varied. Writing $\xi_+ = r \cos \theta$ and $\xi_- = r \sin \theta$, it can be seen from Fig. 3 that $\langle A \rangle$ seems to depend only on r , while the polar angle θ has no effect on $\langle A \rangle$. The only region which is nongeneric is close to the origin ($r \leq 0.3$), where the dependence of $\langle A \rangle$ is steep, and levels off at minimal contributions for $r \leq 0.1$. $r = 0$ corresponds to the case when the agent decouples itself from the rest of the agents, i.e., the agent believes that its actions will have no impact on other agents' behaviors. In this case, the agent plays at Nash equilibrium, i.e., contributes nothing. This is known to be at variance with real player behavior, hence we should avoid small r when choosing ξ_{\pm} in the model.

Some caveat is in order concerning the polar angle, θ . When $\xi_+ > \xi_-$ ($\theta < \pi/4$), a strongly oscillating behavior is observed in the contributions for intermediate time (see inset in Fig. 3). These oscillations (which do not have a visible effect on the average $\langle A \rangle$) occur because agents believe, judging from their own ξ_+ , that it is easier to encourage people to contribute highly, and harder to discourage them. Therefore, the strategy agents adopt is to contribute highly once, so as to encourage all the other agents to contribute highly and then contribute nothing, reaping the benefits from the contribution

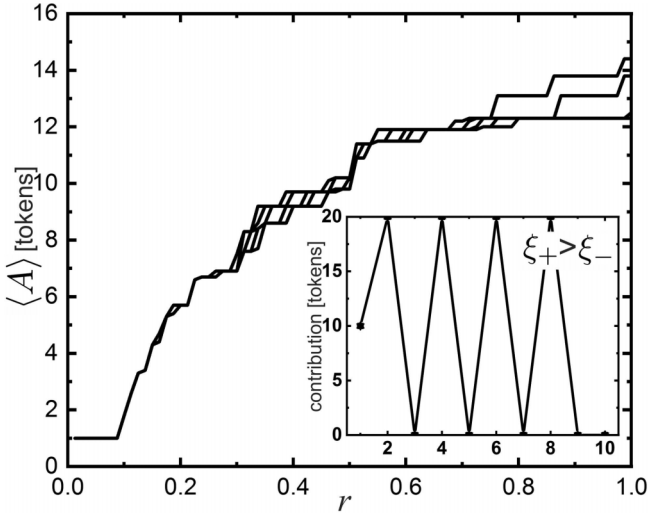


FIG. 3. Group average contribution, $\langle A \rangle$, as a function of $r = \sqrt{\xi_-^2 + \xi_+^2}$ for four different values of the polar angle, $\theta \in \{0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}\}$. The data collapse shows that the polar angle (hence the ratio ξ_+/ξ_-) is not relevant for the average contribution. The inset shows spurious oscillations for $\theta < \pi/4$, which are not observed in real games and should therefore be avoided by proper choice of ξ_+ and ξ_- .

of the other agents. The oscillations can then be observed because all the agents are employing the same strategy, and all of them are Markovian. Such oscillations are unnatural for human player groups, and can be considered an artifact due to the strictly Markovian character of the agents.

In order to model human players, it seems therefore reasonable to keep r sufficiently far away from zero and to assume $\xi_- > \xi_+$. The latter may as well be seen as reflecting a tendency to be risk-averse, which is characteristic of human players to a certain extent. Aside from the observations summarized above, we did not find our simulations to depend strongly on ξ_{\pm} . In our simulations, we therefore set $\xi_+ = 0.1$ and $\xi_- = 0.5$ and keep these values fixed for the remainder of this paper. Additionally, we initialize all following simulations with $f_0 = (10, 10, 10, 10)$, except when fitting experimental data.

2. Impact of K

It should be clear from the previous section that fully rational agents (i.e., $K = \infty$) act independently of their priors. Figure 4(a) shows simulations of a group of four identical rational agents with $(m, K, \gamma) = (5, \infty, 1)$. We see that identical rational agents play identically. Although fully rational ($K = \infty$), the agents do not play Nash equilibrium, but contribute substantially. This is because with our choice of ξ_{\pm} , $r = 0.26$ is sufficiently large to prevent players from “decoupling.” Notice also that rational agents always contribute zero tokens in the last round. This corresponds to the Nash equilibrium in the one-round PGG, as the sole purpose of contributing was to potentially encourage others to contribute in future rounds, which is expressed by the $\xi_p m$. In the inset of Fig. 4(a) we also see the impact of γ on rational agents. Here we have three agents (open circles)

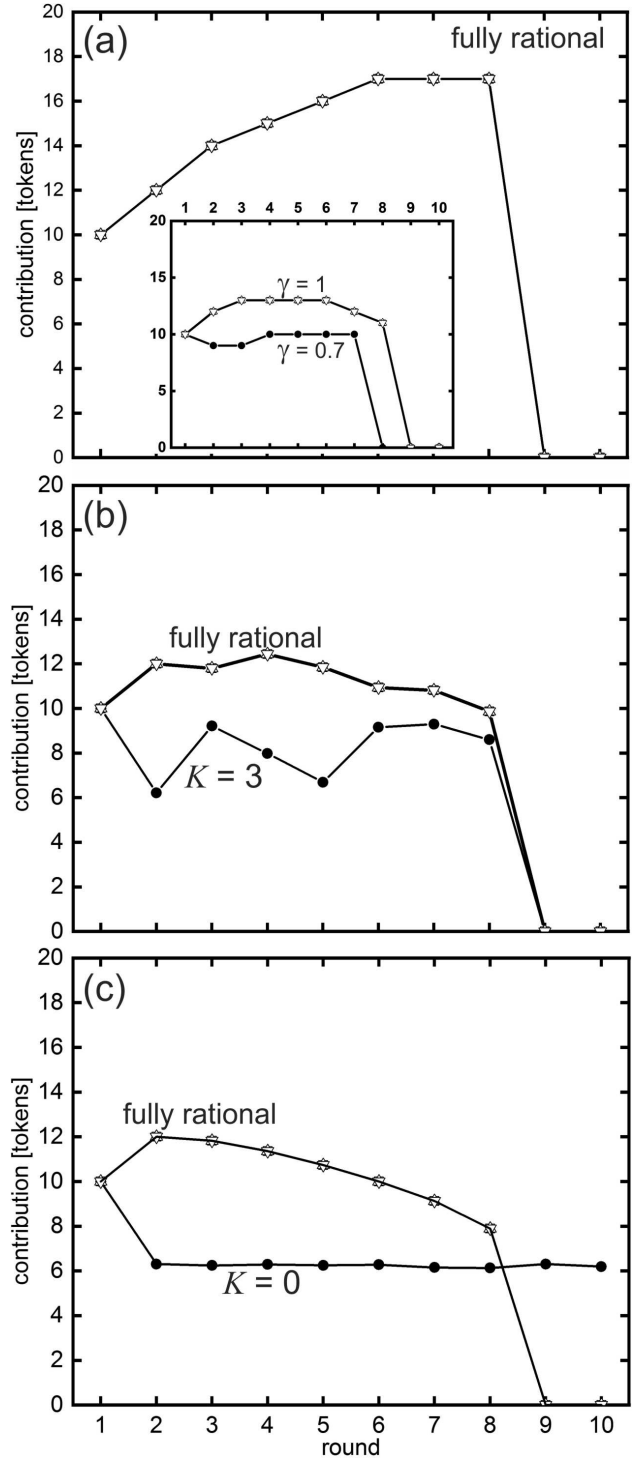


FIG. 4. Full game trajectories of a group of four fully rational agents. In (a) all agents are $(m, K, \gamma) = (5, \infty, 1)$. For the inset, one agent has different $\gamma = 0.7$ (solid circles). Panels (b) and (c) show ensemble average trajectories over 10 000 simulation runs with three agents (open circles) as before and one different agent (solid circles) with the same m and γ , but K reduced to 3 and 0, respectively.

with $(m, K, \gamma) = (5, \infty, 1)$ and one agent (solid circle) with $(m, K, \gamma) = (5, \infty, 0.7)$.

Computational limitations make the agent’s action random and dependent on the prior probabilities. Again we have a set

of three identical rational agents (circles) with $(m, K, \gamma) = (5, \infty, 1)$, and we show the impact of K by varying it for the fourth agent (solid circles). In Figs. 4(b) and 4(c), we show the ensemble average trajectory of the group and a different agent which has $K = 3$ and $K = 0$, respectively. Notice the dissimilarity from the case when all the agents were fully rational.

In Fig. 4(c) we see the effect of complete lack of any computational ability. The agent just acts according to its prior, unaffected by the play of other agents, as is seen from the flat average trajectory. The preferred contribution of the agent is given by the average of the truncated Gaussian prior with $m = 5$, which is ≈ 6.24 .

3. Impact of m

As mentioned previously, m only has an impact on agent behavior for finite K . In order to investigate its impact on agent behavior, we therefore construct a group of three identical rational agents with $(m, K, \gamma) = (5, \infty, 1)$ and a single agent with $K = 3$ and $\gamma = 1$, for which we vary m . As can be seen from the ensemble averaged trajectories shown in Fig. 5, m has a monotonous impact of the average contribution of all the agents. Notice that the agent with $m = 0$ plays like rational agents in the last rounds and the agent with $m = 20$ plays like the rational agents in the intermediate rounds. As mentioned before, this is because the optimal strategy is close to the agent’s basal tendency in these regimes. Bounded rational agents with higher values of m will not be able to play rationally in the last round, as can be seen in Figs. 5(b) and 5(c).

4. Mutual coupling of agents

Referring to the correlations displayed in Fig. 1(b), we now consider the intragroup coupling of agents. This can be investigated by composing a group of three identical agents with $K = 0$ as the “system” and one agent as the “probe.” K must vanish for the system in order to ensure that there are no repercussions of the probe agent’s behavior upon the system. We then vary m of the system and observe the ensuing changes on the contribution of the probe agent. The result is shown in Fig. 6. Here we have chosen a benevolent rational player as the probe. Clearly, its contributions are very much dominated by the contributions of the three system players, which demonstrates considerable coupling between the players within a group.

5. Groups of identical agents

So far we have focused on the impact of parameters on the behavior of individual agents. It is similarly instructive to study the behavior of groups of identical agents when their parameters are varied simultaneously. Results for the impact of m and K on $\langle A \rangle$ for groups of identical agents are summarized in Fig. 7. An initial contribution of 10 tokens is assumed for each agent. Obviously, m and γ have a monotonous impact on the average contributions. At high values of K , the average contribution of the group becomes more and more independent of the priors [convergence of all curves towards the right margin of Fig. 7(a)]. Also note that at very small K , the contribution is totally governed by the priors. This is not

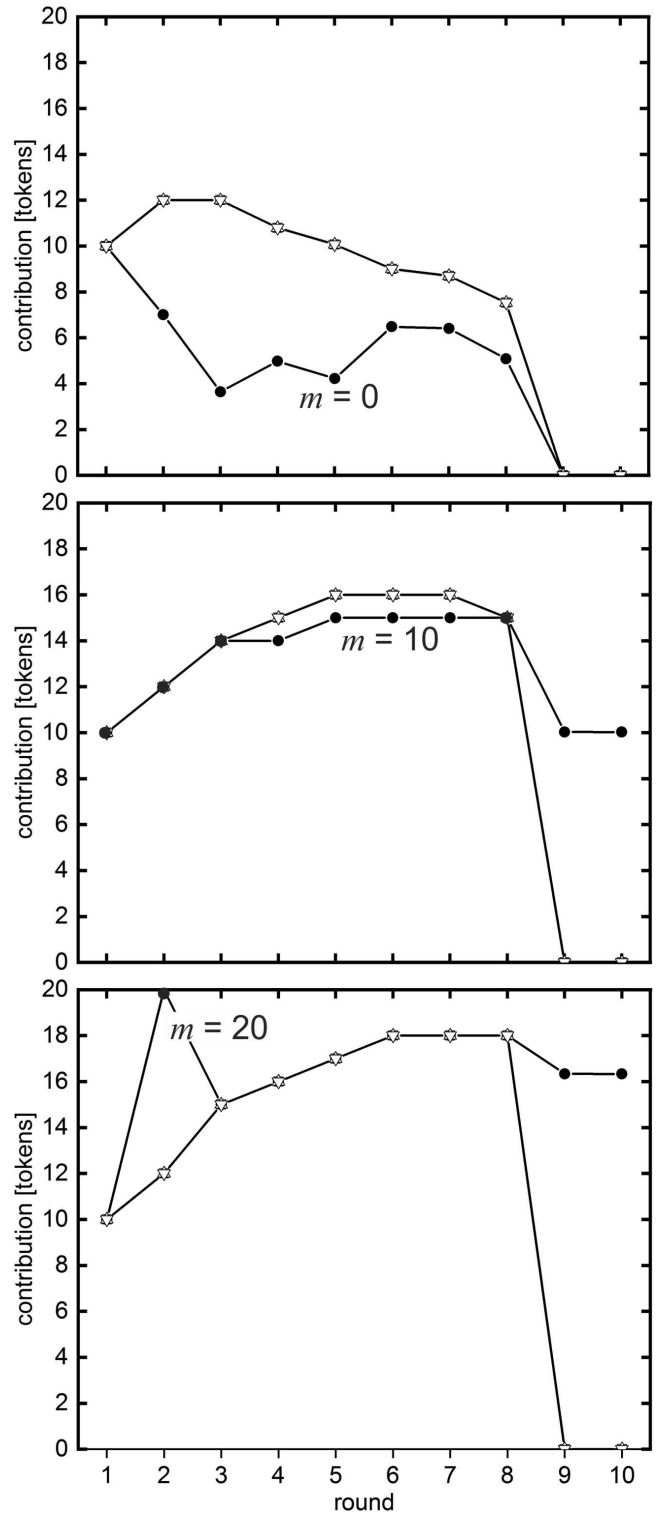


FIG. 5. Full game ensemble average trajectories of group of four agents. There are three identical agents (open circles) with $(m, K, \gamma) = (5, \infty, 1)$ and one different agent (solid circle) with (a) $(m, K, \gamma) = (0, 3, 1)$, (b) $(m, K, \gamma) = (10, 3, 1)$, and (c) $(m, K, \gamma) = (20, 3, 1)$.

the case with γ . At high values of K , γ has a large impact on $\langle A \rangle$ [Fig. 7(b)], while at low values of K , γ has little or no impact on the $\langle A \rangle$.

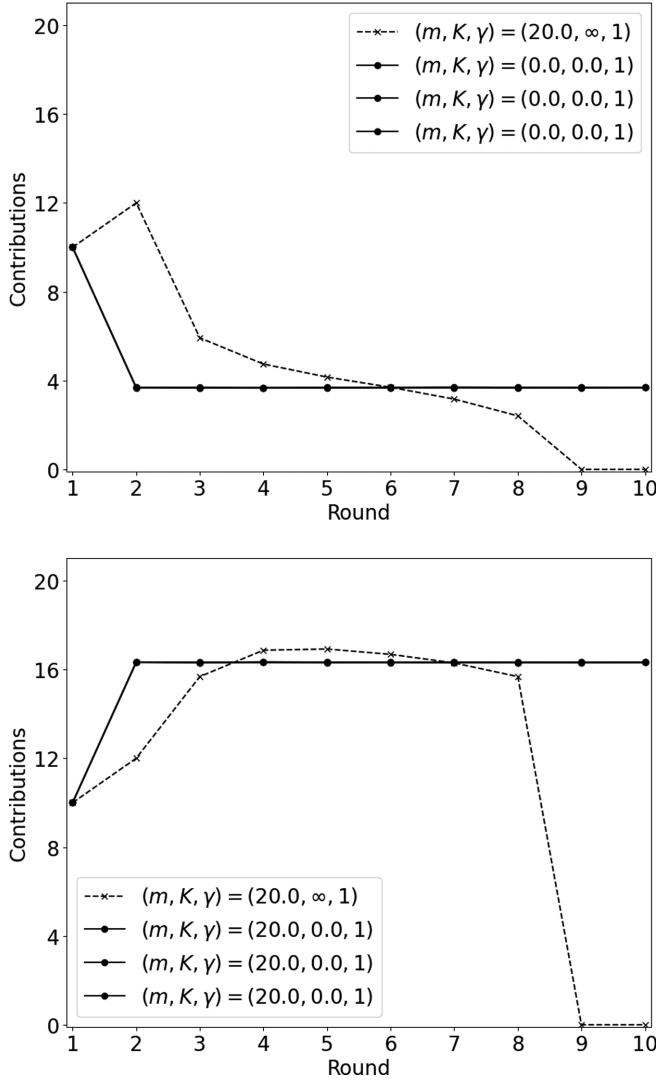


FIG. 6. Demonstration of coupling among agents within one group. A benevolent rational agent (dashed curve) with $(m, K, \gamma) = (20, \infty, 1)$ is made to play in two different systems, with either $m = 0$ (greedy, top, solid curves) or $m = 20$ (benevolent, bottom, solid curves). The system agents are all chosen with $K = 0$ in order to prevent repercussions of the agent under consideration (dashed) onto the system.

It is furthermore interesting to note that $\langle A \rangle$ varies with K appreciably only in an intermediate range of m . In the inset of Fig. 7(a), we plot the square of the derivative of $\langle A \rangle$ (suitably smoothed) with respect to K , averaged over the full range of m . We find a pronounced peak at $K \approx 2.5$. In this range, $\langle A \rangle$ is sensitive as well to m and γ . Hence we may say that the system has a particularly high susceptibility to parameter changes in this range. This is interesting in view of K being intimately related to the Lagrange parameter β , which can be viewed as an inverse generalized temperature [22,23]. A peak in susceptibility may be analogous to a phase transition, when thermal energy comes of the same order as the coupling energy between agents. This will be investigated in more detail in a forthcoming study.

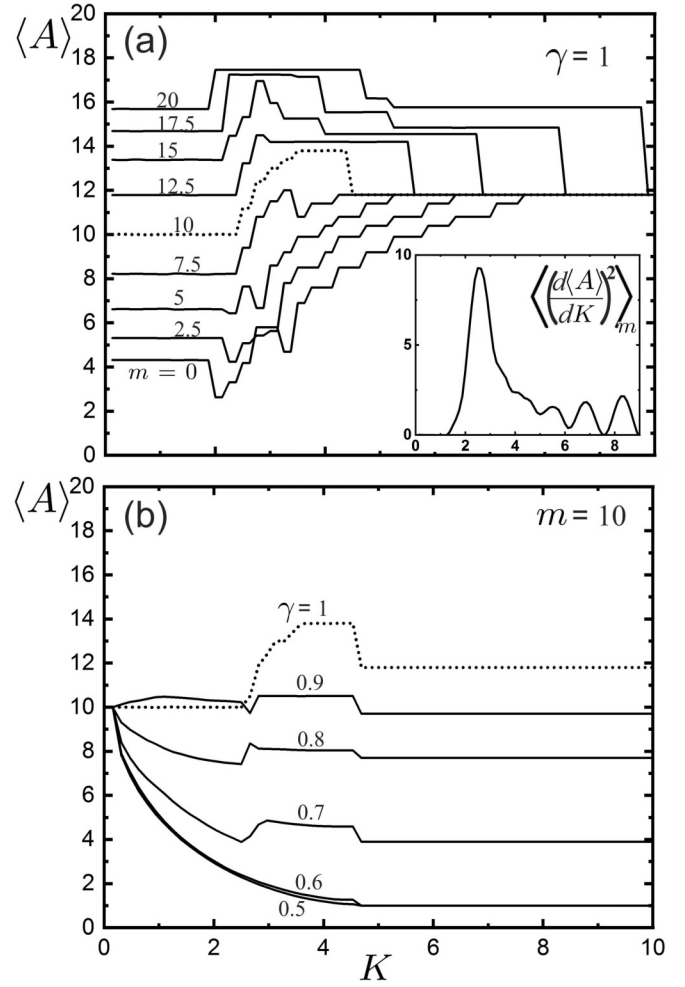


FIG. 7. The dependence of the average contribution $\langle A \rangle$ on agent parameters in a group of identical agents. The ensemble average was determined over 10 000 runs. In each panel, the dotted curve is for $(m, \gamma) = (10, 1)$.

IV. FITTING TO EXPERIMENTAL DATA

Let us now turn to fitting simulated game trajectories to the data obtained from experiments [17]. The data set includes full game trajectories of four-player games for ten rounds. The data spans over 15 different cities across the world. We fit our model agent to the actual players in the game. Note that all agent actions are correlated through their transition functions. Therefore we need to perform the fits on the whole group, rather than fitting individual players sequentially. This means we will have to fit the four quadruples $(Q(\bar{f}_{-k,t} | \bar{f}_{i-1}), K_k, m_k, \gamma_k)$, with $k \in \{1, 2, 3, 4\}$ (16 parameters), to 40 data points (four player contributions over ten rounds). This appears as rather sparse data, in particular as the data generated in simulations have a strong random contribution. We therefore have to seek some meaningful ways to reduce parameter space.

First, we fix some of the parameter values by adopting all assumptions from Sec. III A. Further suggestions emerge when fitting the eight parameters (K_k, m_k) , with $k \in \{1, 2, 3, 4\}$, to experimental group trajectories. This yields a two-dimensional histogram over the (K, m) -plane, which is

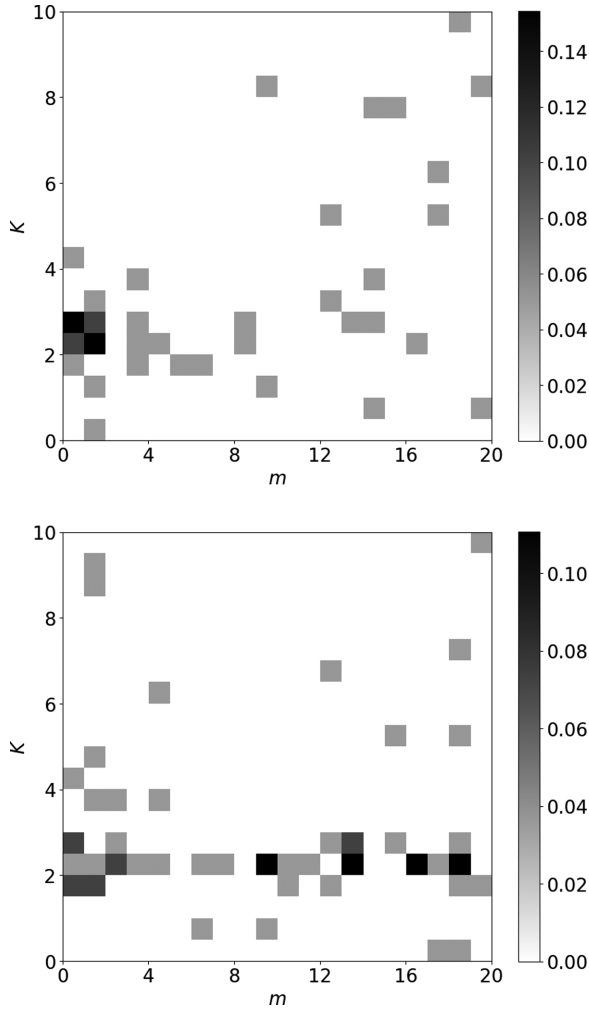


FIG. 8. Joint distributions of K and m for Melbourne (top) and Boston (bottom). There seems to be a clear preference for $K \approx 2.5$.

shown for two cities in Fig. 8, assuming $\gamma = 1$. While fitted values for m are scattered widely, there is a preference for $K \approx 2.5$ for both cities. This is in line with the susceptibility peak we identified in Fig. 7, where agents have access to a maximum range of game states. Hence we henceforth assume $K = 2.5$ for all agents in the fitting procedures. We furthermore assume that all the agents in a group have the same γ . We also choose the initial condition \tilde{f}_0 to be the same as that of the actual players, therefore effectively fitting only nine rounds. As a result, to each group from the experimental data we fit the quintuple $(\gamma, m_1, m_2, m_3, m_4)$.

We minimize the mean-squared deviation of the ensemble-averaged simulated trajectory from the experimental game trajectory. The quintuple mentioned above was numerically found using the Simulated Annealing algorithm in Scipy [37]. The optimization problem for the fitting procedure can be written as

$$\min_{(\gamma, m_1, \dots, m_4)} \sum_{t=2}^{10} \sum_{k=1}^4 (f_{k,t}^{\text{obs}} - \langle f_{k,t}^{\text{sim}} \rangle_{(\gamma, m_1, \dots, m_4)})^2, \quad (11)$$

where the $f_{k,t}^{\text{obs}}$ is the observed (from data) contribution of the k th agent in round t and $f_{k,t}^{\text{sim}}$ is the corresponding contribution

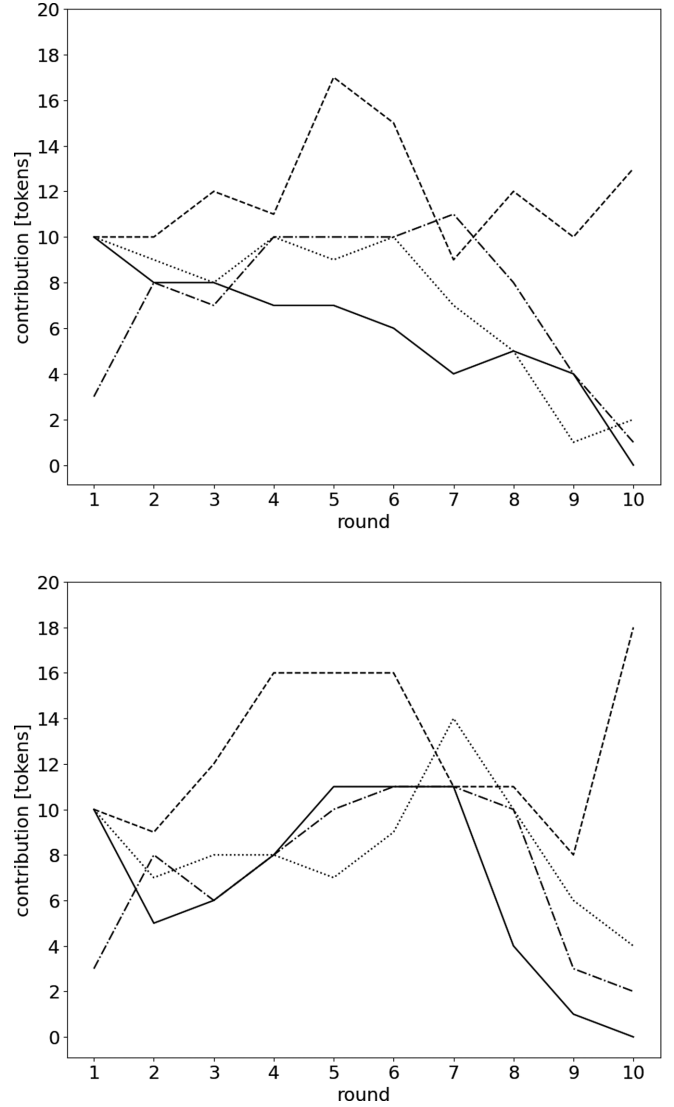


FIG. 9. Two trajectories simulated under identical conditions, with parameters obtained from fitting the group trajectory displayed in Fig. 1(a) (data from a Athens group). The same group of agents yields a different trajectory each time the simulation is run, due to the inherent randomness of the model. The fitting procedure minimizes the deviations of the average contribution at each round, as well as of the variance of these contributions from the observed variance of player contributions.

from the simulated agent. Furthermore, $\langle \cdot \rangle_{(\gamma, m_1, \dots, m_4)}$ denotes the average over multiple simulation runs of the group defined by the parameters $(\gamma, m_1, \dots, m_4)$.

The resulting parameter set found by fitting to a single group can be used to generate individual game trajectories of the so obtained group of agents, for comparison with the experimental trajectory. Figure 9 shows two examples from the agent group obtained by fitting to the trajectory from an Athens group, which we displayed in Fig. 1(a).

V. DISCUSSION

In Fig. 10 we compare the simulated city averages with the actual city averages. The error bars represent the city-wide

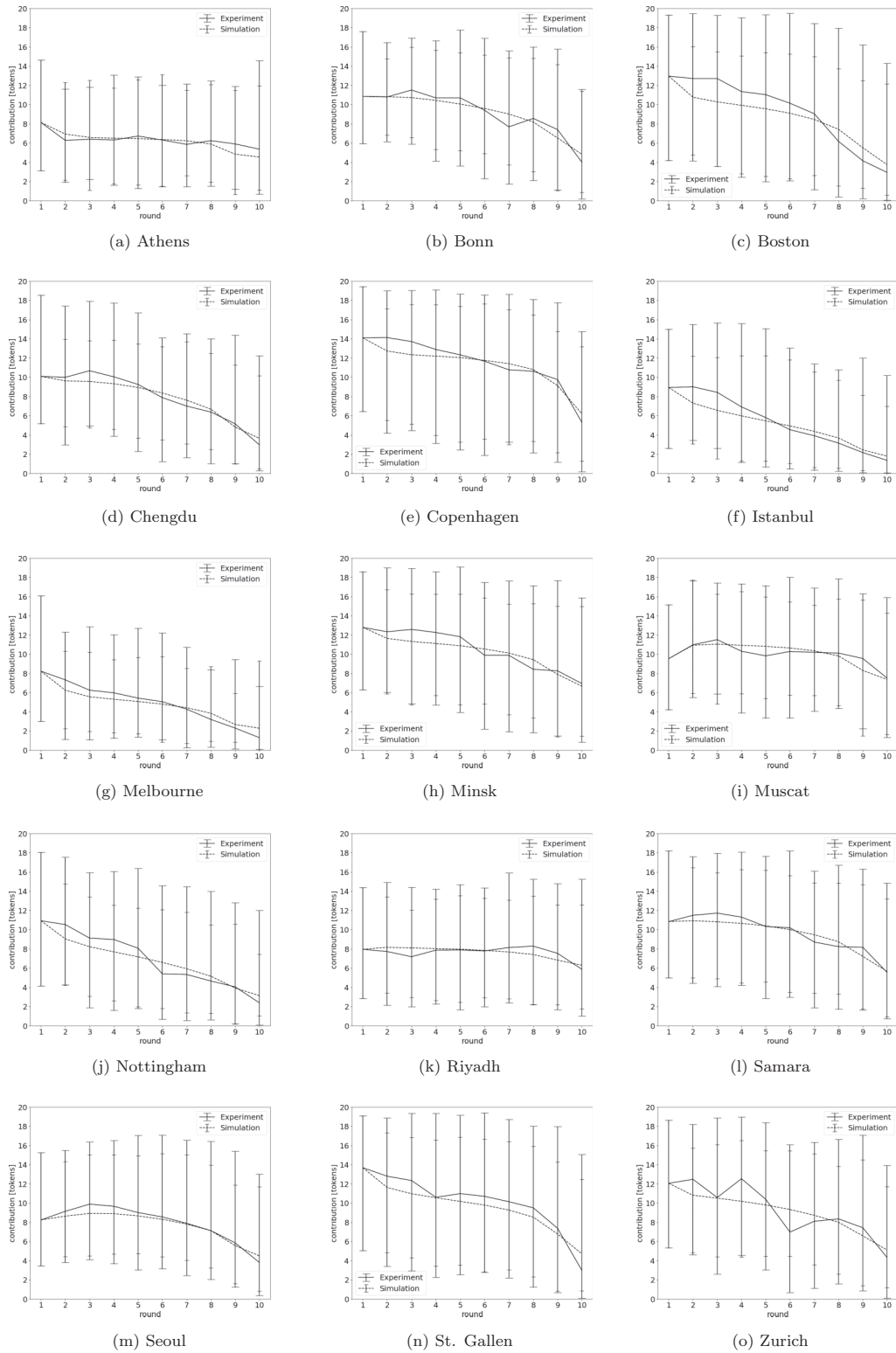


FIG. 10. Actual and simulated city-averaged contributions. The error bars indicate the variance of contributions.

standard deviation of contributions in that round for both the simulated city and the actual data. The simulated city averages were evaluated by averaging over multiple simulation runs of all the groups in the city while keeping the agent parameters to be the estimated parameters from the fits. Our model not

only accounts for the average contributions and their characteristic decline, but also for the variability of contributions. The slight underestimation of the variance can be attributed to the simplifying assumptions we have made in assigning the same γ to all the players within a group, and setting

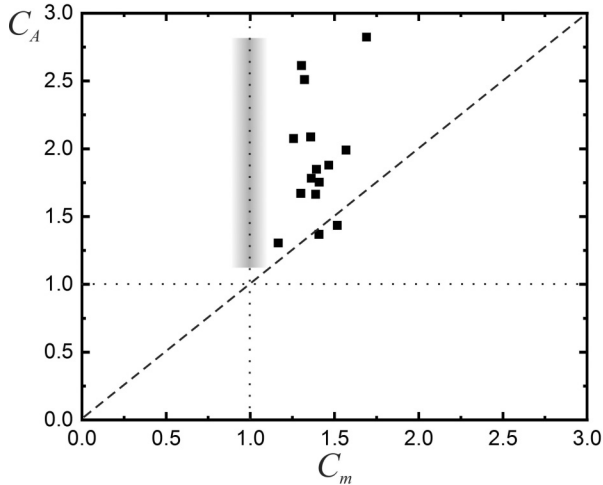


FIG. 11. Coupling in A vs coupling in m as obtained from the fitted simulations. The gray bar represents the data from Fig. 1(b).

$K = 2.5$ for all players in general. We see that even though we have limited our model severely in its scope, we are able to model the player behavior quite effectively. Note that the difference between the simulations and the data are much smaller than the variances of contributions throughout the data sets. Hence we find that γ and the individual values of m are sufficient as parameters for obtaining a very good agreement with experimental data for each city, although both the slope of the decline and the average contribution varies significantly. For the sake of completeness at this point, we show the distributions obtained for m and γ in the Appendix (see Fig. 12) for each city.

A. Game-induced interagent coupling

Let us now turn to the coupling of players within a group, as observed from the data in Fig. 1(b). We can quantify this coupling by taking the ratio of intergroup variance [ordinate in Fig. 1(b)] to intragroup variance [abscissa in Fig. 1(b)]. Specifically, for $\langle A \rangle$ and m , we write

$$C_m = \frac{\text{var}\{m|\text{intergroup}\}}{\langle \text{var}\{m_k|\text{intragroup}\} \rangle_l} \quad (12)$$

and

$$C_A = \frac{\text{var}\{\langle A \rangle|\text{intergroup}\}}{\langle \text{var}\{\langle A_k \rangle|\text{intragroup}\} \rangle_l}, \quad (13)$$

respectively, where “var” is the variance and the index l runs through all groups of a city.

Results for these couplings as obtained by fitting to the experimental game trajectories are presented in Fig. 11, where each data point corresponds to one city. The gray bar represents the data from Fig. 1(b), where C_A is found to range from 1.25 to 2.7. As m represents the preference inherent to a single player, the bar is placed at $C_m = 1$. This corresponds to the fact that players had been chosen randomly, such that the intergroup variance of personal preferences must *a priori* be equal to the intragroup variance, up to some statistical fluctuations which we indicate by the fuzzy boundaries of the gray bar.

The values obtained for C_A from the fitted simulations are found in the same range as that indicated by the gray bar. This is expected as we have fitted the contributions to those of the experiments. However, we see that most of the data points are well above the first diagonal, which shows that the individual m_k are less strongly coupled than the average individual contributions, $\langle A_k \rangle$. There is some coupling effect on the m since the fitting algorithm cannot distinguish to what extent a player contributes due to her own preference or due to entrainment by her fellow players. The offset above the first diagonal, however, clearly shows that the coupling effect is present among the model agents. This is another manifestation of the same phenomenon as demonstrated in Fig. 6.

B. Learning vs bounded rational foresight

While payoff-based learning has been suggested as *the* explanation of the commonly declining contributions in a PGG [15,16], we have shown that bounded rational foresight, as reflected by the model agent used in the present work, can perfectly well serve as an explanation of this phenomenon. Moreover, it is capable of accounting for the substantial in-game variance of contributions which is observed in real games.

Note that Fig. 10 shows that both the average contribution and the slope of its decline vary substantially among cities investigated. Our model suggests that this can be attributed to different human preference parameters, which are indeed well known to vary among different cultures. It is not straightforward to see why (or at least it is not known that) the parameters of payoff-based learning should vary in a similarly pronounced manner. This would be a necessary conclusion if one wanted to insist on payoff-based learning explaining all salient features of the data displayed in Fig. 10.

VI. CONCLUSIONS AND OUTLOOK

We have modeled the PGG game as an Markov decision process with bounded rational agents based on a path integral formulation of reward maximization, without invoking any learning mechanism. We found that at least in short games, for which experimental data are available, our bounded rational agent is able to account for human playing behavior in PGG (cf. Fig. 10).

One may argue that the bound on optimization in the form of relative entropy [Eq. (9)] is arbitrary for modeling human agents. In order to come up with a more “physical” bound on the computational abilities of humans, one would need to have a model of how humans perform computations through their neural network and evaluate the thermodynamic work done in order to perform those computations. Then from the energy limits of human agents we might be able to quantitatively derive an appropriate value for K . This is, however, well beyond what is currently known about computations in the brain.

Our results suggest that the most important parameters determining the playing behavior for our agents are m and γ . The question whether bounded rational foresight or payoff-based learning dominates human decision making in PGG-type situations must be considered open. Both mech-

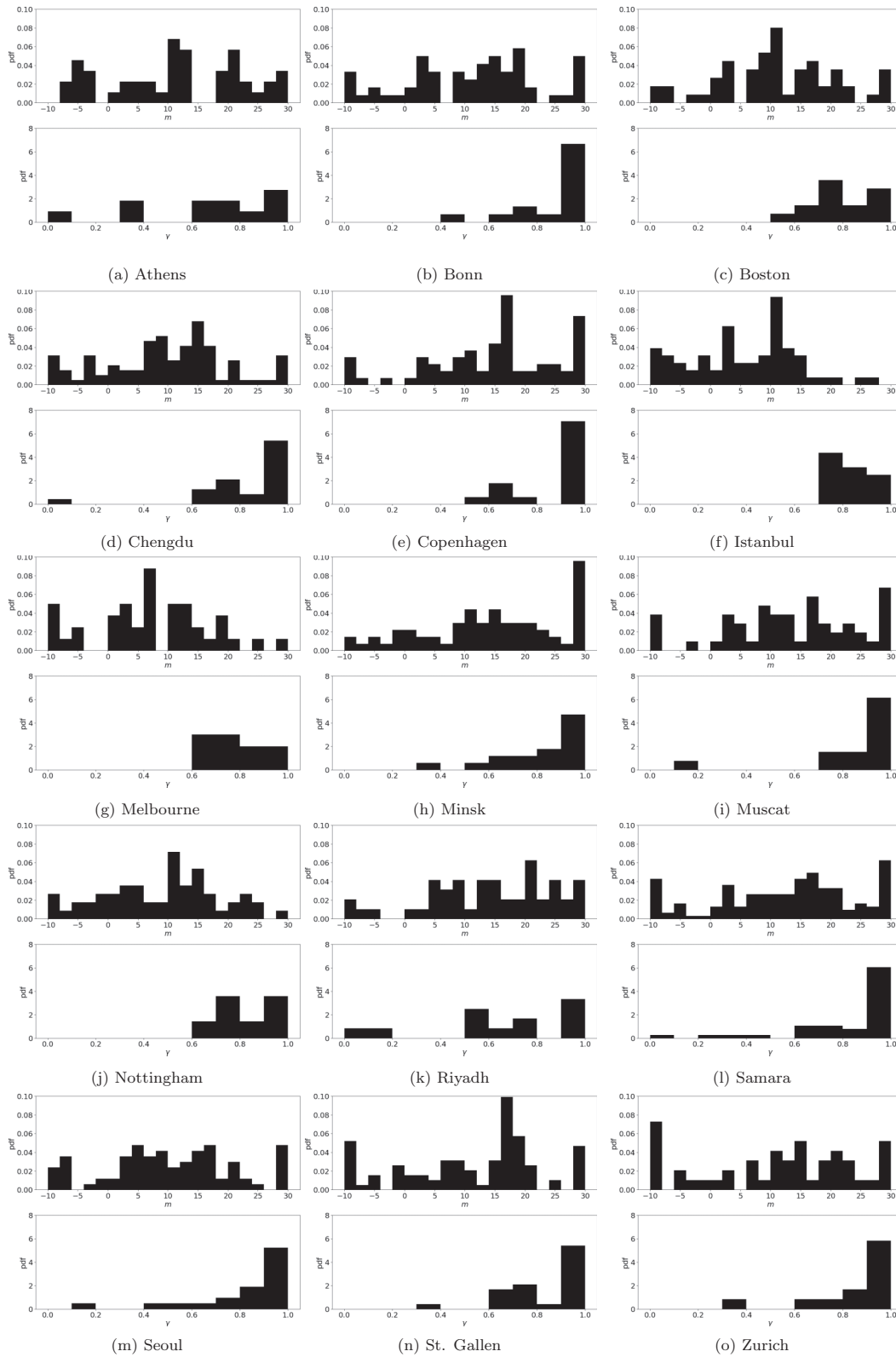


FIG. 12. Marginal distributions of γ and m as obtained by performing the fits for various cities.

anisms may play a role; actually, it would be surprising if one would be completely irrelevant. The relative importance of these mechanisms is a challenge. One might, for instance, perform PGG with players the preference parameters of which

have been determined beforehand. Correlating the results with the parameters m , γ , and ξ_{\pm} may then yield insight into the importance of bounded rational foresight in the respective game.

ACKNOWLEDGMENTS

The authors gratefully acknowledge useful discussions with K. Heidemann, V. Priesemann, S. Muehle, S. Gupta, and M. Khan. They are very grateful to the authors of Ref. [17] for sharing their raw data on public goods game experiments they had performed with human agents.

APPENDIX: MARGINAL DISTRIBUTIONS OF γ AND m FOR ALL THE CITIES

For the sake of completion, we provide the marginal distributions of γ and m for the various cities.

-
- [1] D. O'Neill, A. L. Fanning, W. F. Lamb, and J. K. Steinberger, A good life for all within planetary boundaries, *Nat. Sustain.* **1**, 88 (2018).
- [2] L. M. A. Bettencourt and J. Kaur, Evolution and structure of sustainability science, *Proc. Natl. Acad. Sci. USA* **108**, 19540 (2011).
- [3] L. J. Kotzé and R. E. Kim, Earth system law: The juridical dimensions of earth system governance, *Earth Syst. Govern.* **1**, 100003 (2019).
- [4] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, *Rev. Mod. Phys.* **81**, 591 (2009).
- [5] M. Perc, J. J. Jordan, D. G. Rand, and Z. Wang, Statistical physics of human cooperation, *Phys. Rep.* **687**, 1 (2017).
- [6] D. Garcia, A. Abisheva, S. Schweighofer, U. Serdült, and F. Schweitzer, Ideological and temporal components of network polarization in online political participatory media, *Policy Internet* **7**, 46 (2015).
- [7] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Modeling Echo Chambers and Polarization Dynamics in Social Networks, *Phys. Rev. Lett.* **124**, 048301 (2020).
- [8] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Emergence of Polarized Ideological Opinions in Multidimensional Topic Spaces, *Phys. Rev. X* **11**, 011012 (2021).
- [9] C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund, Via freedom to coercion: The emergence of costly punishment, *Science* **316**, 1905 (2007).
- [10] L.-L. Jiang, T. Zhou, M. Perc, and B.-H. Wang, Effects of competition on pattern formation in the rock-paper-scissors game, *Phys. Rev. E* **84**, 021912 (2011).
- [11] Q. Yu, D. Fang, X. Zhang, C. Jin, and Q. Ren, Stochastic evolution dynamics of the rock-scissors-paper game based on a quasi birth and death process, *Sci. Rep.* **6**, 28585 (2016).
- [12] T. Wu, F. Fu, and L. Wang, Coevolutionary dynamics of aspiration and strategy in spatial repeated public goods games, *New J. Phys.* **20**, 063007 (2018).
- [13] M. Tomassini and A. Antonioni, Computational behavioral models for public goods games on social networks, *Games* **10**, 35 (2019).
- [14] U. Alvarez-Rodriguez, F. Battiston, G. F. de Arruda, Y. Moreno, M. Perc, and V. Latora, Evolutionary dynamics of higher-order interactions in social networks, *Nat. Human Behav.* **5**, 586 (2021).
- [15] M. N. Burton-Chellew, H. H. Nax, and S. A. West, Payoff-based learning explains the decline in cooperation in public goods games, *Proc. R. Soc. B* **282**, 20142678 (2015).
- [16] M. Burton-Chellew and S. A. West, Payoff-based learning best explains the rate of decline in cooperation across 237 public-goods games, *Nat. Human Behav.* **5**, 1330 (2021).
- [17] B. Herrmann, C. Thöni, and S. Gächter, Antisocial punishment across societies, *Science* **319**, 1362 (2008).
- [18] R. M. Isaac and J. M. Walker, Group size effects in public goods provision: The voluntary contributions mechanism, *Q. J. Econ.* **103**, 179 (1988).
- [19] A. Gunthorsdottir, D. Houser, and K. McCabe, Disposition, history and contributions in public goods experiments, *J. Econ. Behav. Organ.* **62**, 304 (2007).
- [20] A. E. Roth and I. Erev, Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term, *Games Econ. Behav.* **8**, 164 (1995).
- [21] I. Erev and A. E. Roth, Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria, *Am. Econ. Rev.* **88**, 848 (1998).
- [22] D. A. Ortega and P. A. Braun, Information, utility and bounded rationality, in *Artificial General Intelligence*, edited by J. Schmidhuber, K. R. Thórisson, and M. Looks (Springer, Berlin, 2011), pp. 269–274.
- [23] D. H. Wolpert, in *Complex Engineered Systems: Science Meets Technology*, edited by D. Braha (Springer, Berlin, 2006), pp. 262–290.
- [24] A. Falk, A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde, Global evidence on economic preferences, *Q. J. Econ.* **133**, 1645 (2018).
- [25] R. Winkelmann, J. Donges, E. K. Smith, M. Milkoreit, C. Eder, J. Heitzig, A. Katsanidou, M. Wiedermann, N. Wunderling, and T. Lenton, Social tipping processes towards climate action: A conceptual framework, *Ecol. Econ.* **192**, 107242 (2022).
- [26] M. Tomassini and A. Antonioni, Public goods games on coevolving social network models, *Front. Phys.* **8**, 58 (2020).
- [27] Z. Xu, Z. Wang, and L. Zhang, Bounded rationality in volunteering public goods games, *J. Theor. Biol.* **264**, 19 (2010).
- [28] R. Bellman, On the theory of dynamic programming, *Proc. Natl. Acad. Sci. USA* **38**, 716 (1952).
- [29] H. W. Kuhn, Nonlinear programming: A historical view, *SIGMAP Bull.*, 6 (1982).
- [30] Varying σ_{trans} between 1 and 8 was found not to change the qualitative features of the group trajectories appreciably. We chose an intermediate value, which means that agents neither are entirely uninformed about their contemporaries nor have a very precise model of them.
- [31] K. J. Arrow, Gifts and exchanges, *Philos. Public Affairs* **1**, 343 (1972).
- [32] R. Boyd, H. Gintis, S. Bowles, and P. J. Richerson, The evolution of altruistic punishment, *Proc. Natl. Acad. Sci. USA* **100**, 3531 (2003).
- [33] T. Dohmen, A. Falk, D. Huffman, and U. Sunde, Representative trust and reciprocity: Prevalence and determinants, *Econ. Inquiry* **46**, 84 (2008).

- [34] C.-C. Chen, I.-M. Chiu, J. Smith, and T. Yamada, Too smart to be selfish? Measures of cognitive ability, social preferences, and consistency, *J. Econ. Behav. Organ.* **90**, 112 (2013).
- [35] S. K. Lam, A. Pitrou, and S. Seibert, Numba: A LLVM-based python JIT compiler, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, LLVM '15 (Association for Computing Machinery, New York, 2015), pp. 1–6.
- [36] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, et al., Array programming with NumPy, *Nature (London)* **585**, 357 (2020).
- [37] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, *Nat. Methods* **17**, 261 (2020).