Bounded learning and planning in public goods games

Prakhar Godara[*] and Stephan Herminghaus[†]
Max Planck Institute for Dynamics and Self-Organization (MPIDS), Am Faßberg 17, D-37077 Göttingen, Germany

(Received 6 October 2022; accepted 8 May 2023; published 30 May 2023)

A previously developed agent model, based on bounded rational planning, is extended by introducing learning, with bounds on the memory of the agents. The exclusive impact of learning, especially in longer games, is investigated. Based on our results, we provide testable predictions for experiments on repeated public goods games (PGG) with synchronized actions. We observe that noise in player contributions can have a positive impact of group cooperation in PGG. We theoretically explain the experimental results on the impact of group size as well as mean per capita return (MPCR) on cooperation.

DOI: 10.1103/PhysRevE.107.054140

I. INTRODUCTION

With the looming climate crisis, limited planetary resources, and the associated challenges to human societies, predicting human collective behavior in resource allocation is a problem of increasing importance [1–3]. Essential for such predictions is the development of models of human economic interactions which are both reliable and suitable for modeling entire societies.

A classical paradigm for achieving such modeling is game theory. It has lately grown into a mature field of research, with extensions toward collective behavior having emerged in recent years [4–9]. The capability of predicting human behavior in controlled environments such as games allows not only to test models of intelligence but also potentially allows policy makers to make more robust decisions [10,11] in situations of societal relevance. 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 [7,12–14].

However, human behavior in games lies in a much smaller dimensional space (game trajectories), than the physical system (agent + environment) that generates the behavior. This then leads to creation of a large number of ad-hoc models which account for human behavior in very limited settings only [15]. Such approaches may provide some predictive power in very specific scenarios but are likely to fail in predicting human behavior in different environments. Additionally, they do not have much potential in providing insight into the mechanisms of intelligent behavior.

In order to circumvent these problems, one needs to develop and systematically study models that are applicable in more general environments, with parameters which can be related to measurable human behavioral preferences [2]. To this end, we have demonstrated [16] that a general bounded rational planning agent is able to reproduce human behavior in public goods games (PGG). In particular, this was possible without needing to invoke any mechanisms of learning. This is not to say that humans do not learn when playing iterative PGG for 10 rounds. All this communicates is that learning is *not necessary* to reproduce human behavior in these games. Therefore as a next natural step we introduce learning in our model to see for what behaviors is it *necessary* to invoke mechanisms of learning. In other words, in this article we wish to observe the exclusive impact of learning on bounded rational agents, which couldn't have been generated by bounded rational agency alone. As in our previous work [16], we base our model on the specific case of playing PGG and compare the behavior of agents to known experimental results. Before we proceed with developing our model, we briefly describe the well known PGG.

The PGG is played with N players over a total of known T periods. In each period the players are given a fixed integer number of tokens τ, which they can anonymously invest into a public pot. Following a widely followed convention in the field [16,17], we use $\tau = 20$ throughout this article. In any period $t \leqslant T$ if the contribution of the kth player is $f_{k,t} \in [0, \tau]$ then the immediate reward of the player in that period is given by

$$G_{k,t} = \alpha(N-1)\mu_{k,t} - (1-\alpha)f_{k,t}, \tag{1}$$

where $\alpha < 1$ is multiplying factor which is known to all the players and $\mu_{k,t} = \frac{\sum_{i\neq k} f_{i,t}}{N-1}$ is the average contribution of other players. The total gain for the kth player can then be defined as $G_k = \sum_t^T G_{k,t}$.

It has been argued in artificial general intelligence (AGI) research that a minimal model of an intelligent agent embedded in an arbitrary environment (for instance, playing a game) has two main ingredients, learning and planning

[*] Corresponding author: prakhar.godara@ds.mpg.de
[†] stephan.herminghaus@ds.mpg.de

(AIXI [18]). At any point of time, an intelligent agent looks at the past trajectory of the environment (past game states and actions) to learn about the dynamics of the environment (modeling other players in the game). This knowledge of the dynamics is then used to simulate future trajectories of the environment (game), in order to choose the action which leads to the *best* trajectory, i.e., the trajectory maximizing a previously defined utility function. That is to say, learning is a mapping from observed behavior to mental models and planning is a mapping from mental models to actions. The readers should note that the notion of learning put forward is distinct from ”social learning” as is common in evolutionary game theory [19,20], where the agents learn *from* other agents by comparing their strategy's fitness with that of others in the population and then imitating the better strategy with a finite probability. In our approach the agents learn *of* the other player behavior by creating a model of other agents.

Also note that this distinction between learning and planning is not commonly made in most agent based models. Instead, learning is conceived to refer to figuring out which action leads to better immediate rewards, with the agent being oblivious to other agents (i.e., has no models of them) [12,21], i.e., to say that in these works learning is a mapping directly from observed behavior to actions. In contrast, by making a clear distinction between learning and planning, we can study, and potentially control, the distinct qualitative behaviors introduced by either of them.

The main problem in implementing AIXI to predict human behavior in games is that it is not computable [22]. Nonetheless, the idealized model can still be viewed as a guiding principle to generate models of human behavior in slightly less general environments by introducing specific approximations, whereby trading off the generality with computability of the model. Therefore, in this article our approach would be to introduce learning to our bounded rational agent model [16], while at the same time making use of context specific approximations that allow our model to be computable.

II. PLANNING AND LEARNING WITH BOUNDS

A. The planning mechanism

As aforementioned the planning mechanism is a mapping from mental models to actions (in this case, a policy). Therefore in this subsection we assume a mental model of the agent (given by the transition function) and seek to find the optimal policy of the agent. We describe the planning mechanism from the perspective of the kth agent and this extends to all k. We model the agent's decision making problem as a Markov decision process (MDP), with the transition function Q given by

$$Q(\mu_{k,t}|\mu_{k,t-1}, f_{k,t-1}) = TG(\mu_{k,t}; m, \sigma). \tag{2}$$

Here, TG is the truncated Gaussian distribution on the interval $[0, \tau]$ ($\tau = 20$), $f_{k,t}$ is the contribution of the kth agent in round t, and $\mu_{k,t} = \frac{\sum_{i\neq k} f_{i,t}}{N-1}$ is the average contribution of other players in round t. m is the peak position of the distribution given by

$$m = \begin{cases} \mu_{k,t-1} + \xi_+|\mu_{k,t-1} - f_{k,t-1}|, & \mu_{k,t-1} - f_{k,t-1} < 0 \\ \mu_{k,t-1} - \xi_-|\mu_{k,t-1} - f_{k,t-1}|, & \mu_{k,t-1} - f_{k,t-1} > 0. \end{cases} \tag{3}$$

As the kth agent can influence others actions through its contributions alone (because players play anonymously and do not interact otherwise), the parameters ξ_+ and ξ_- describe to which degree the agent believes to be able to encourage or discourage other agents to contribute. In that sense, $\xi_\pm$ is a model the agent has of its environment (i.e., the other agents) and represents its transition function. In so far as planning is concerned, we do not bother about how the agent comes up with a particular model (i.e., particular values of $\xi_\pm$), but rather what decisions (policy) does the agent come up with, given a model of its environment.

The bounded rational decision making problem in period $t \leqslant T$ as defined in Ref. [16] is described by a Bellman equation given by

$$V_t^* = \max_{P(f_t^T)} \sum_{f_{k,t},\mu_{k,t}} P(f_{k,t}|\bar{f}_{t-1})\bigg[Q\cdot G_{k,t}(f_{k,t},\mu_{k,t}) - \frac{1}{\beta}\log\frac{P(f_{k,t}|\bar{f}_{t-1})}{P_0(f_{k,t}|\bar{f}_{t-1})} + \gamma Q\cdot V_{t+1}\bigg], \tag{4}$$

where * is to indicate a maximized quantity, $\bar{f}_t = (f_{1,t}\dots f_{N,t})$ is the state of the game in period t, and β is a Lagrange parameter along with an additional constraint $\frac{1}{\beta}(D_{KL}(P^*(f_{k,t}|\bar{f}_{t-1})||P_0(f_{k,t}|\bar{f}_{t-1})) - K) = 0$, with K the computational capability of the agent. Intuitively speaking, K represents the maximum deviation (in policy space), from the prior policy, that the agent can afford in search of a better policy. For instance, setting $K = 0$ would mean that $P^* = P_0$, thereby the agent is maximally bounded and is going to play only according to its prior strategy P_0. On the other hand, if $K = \infty$, then one can see from the above constraint that $\frac{1}{\beta} = 0$, and hence Eq. (4) reduces to the completely rational case. All intermediate values of K span policies between the completely rational policy and the prior policy. Additionally, we consider another parameter $\gamma \in [0, 1]$ appearing in Eq. (4). It represents a foresight which exponentially “decays” into the future [16]. The solution of the optimization problem in Eq. (4) then provides us with the (bounded optimal) policy $P^*(f_t^T) = \prod_{t'=t}^{T} P^*(f_{k,t'}|\bar{f}_{t'-1})$ of the agent, which is the output of the planning mechanism.

B. The learning mechanism

1. *A subspace of all partial functions*

In AIXI [18], learning for an agent from past data happens through Solomonoff induction [23], which considers the space of all partial functions [24] on $\{0, 1\}^*$ [25], i.e., the space of all allowed “explanations” for the past trajectory. Although this form of learning guarantees convergence to the true distribution, it is not computable as a consequence of the halting problem [26]. In practice however, one might want to reduce the *search space* from the space of all partial functions on $\{0, 1\}^*$ to a smaller space.

In AIXI*tl* [22] it is proposed to consider only programs up to length l and computation time t. AIXI*tl* does this by running a brute force search over all the programs. Although this brute force search is computable, it still takes enormous computing power to compute. While this is not a problem for AIXI*tl*, which is focused on describing intelligence in an *arbitrary* environment, it seems unreasonable to model humans as brute force searchers which take enormous computing time in a *specific* environment such as PGG as they have context specific pre-play awareness of the game [17].

Another common way to reduce the search space is by creating a model class and then performing regression or maximum likelihood estimate to find the best model in the model class. The latter approach is not only easier to implement but also allows the opportunity to introduce easily interpretable parameters in the model as compared to AIXI*tl*. Therefore, it is the latter approach that we will take in this article.

As we intend to model human behavior in PGG, we exploit this context specificity and consider the model class introduced in Ref. [16], as it has been successful in explaining observed human behavior [17]. Therefore we only consider Markovian transition functions, given by truncated Gaussian distributions, parameterized via $\xi_\pm$.

2. The model

In the learning problem we are concerned with the agent learning the transition function Q from its past experiences in the game. As the transition function is parameterized by $\xi_\pm$, learning mechanism is then concerned with finding the $\xi_\pm$ values that are the most representative of the past experiences, i.e., those values of $\xi_\pm$ that have the highest likelihood of generating the past game trajectory.

Additionally, quite like the exponentially decaying foresight given by γ, we also introduce another parameter $\gamma_p \in [0, 1]$, which represents a hindsight decaying exponentially into the past (equivalently called "recency-bias" in Ref. [27]) of the agent. It signifies that when an agent evaluates the behavior of its environment, recent experiences guide its model more than earlier experiences. This is then achieved by weighting the maximum likelihood estimation with γ_p as below:

$$\xi^*_\pm(t) = \arg\max_{\xi_\pm}\Bigg[\sum_{i=t-1}^{2} \gamma_p^{t-i} \log Q(\mu_i|\mu_{i-1}, f_{k,i-1}) - (1-\gamma_p)(\xi_\pm(t) - \xi_\pm(t-1))^2\Bigg], \tag{5}$$

where the last summand captures the tendency of the agent to not update its model. Therefore $\gamma_p = 0$ would correspond to not updating the model given a past trajectory (no learning) and $\gamma_p = 1$ would correspond to learning from arbitrarily far back in the past.

C. An updated agent model

We now combine the planning and the learning mechanism into one agent which is described now by the tuple (m, K, γ, γ_p). In every period $2 < t \leqslant T$ the agent:

(i) plans: by considering the game state at $t-1$, making use of the current model $(\xi \pm (t))$ and solving Eq. (4) and evaluating the policy $P(f_t^T)$,

(ii) acts: by sampling a bet from the evaluated policy, and

(iii) learns: after observing the state of the game in the current period t and finding the $\xi_\pm(t)$ by making use of Eq. (5).

In period $t = 1$ the bets of the agent are sampled from its prior distribution $P_0(f_{k,t})$ and the agent is provided with a model $\xi_+, \xi_-(0) = (0.1, 0.5)$. In period $t = 2$ there is certainly planning and acting based on the model, but there is no learning, as the agent has not yet observed a transition.

III. BEHAVIOR SPACE OF THE MODEL

With the model being defined, in this section we explore the behavior space of the agent by considering two types of setups. Namely, considering contribution dynamics in groups of identical agents and groups of randomly chosen agents. In the former setup the agent parameters in a game are identical to each other. This setup is chosen to demonstrate the qualitative effects of the agent parameters on average contributions. In the latter setup the agent parameters are chosen randomly from a uniform distribution over the parameter space. This setup is chosen to observe the behavior of agents in a well-mixed population.

To the end of understanding the exclusive aspects of the dynamics introduced by learning and its interplay with planning, we only consider the computational bound K and the hindsight γ_p as the parameters of importance. For simplicity, the other parameters, namely, m and γ are fixed throughout the rest of the paper to 10 and 0.9 respectively [16,28]. Additionally throughout this section we consider games of length $T = 100$ and groups of size $N = 4$.

A. Groups of identical agents

In this subsection we explore cooperation in groups of identical agents playing a PGG for different values of K, γ_p. In Fig. 1 we show the average contribution $\langle A\rangle$ as a function γ_p for various values of K, where $A = \frac{\sum_k \sum_t f_{k,t}}{NT}$ and the $\langle\cdot\rangle$ is to denote an ensemble average over multiple simulation runs (1350 simulation runs for each datum).

Quite expectedly in groups of agents with $K = 0$, $\langle A\rangle$ is not impacted by learning. For $K > 0$ we see that introduction of learning monotonically decreases the contribution levels in groups of identical agents. Additionally, the rates at which $\langle A\rangle$ decreases with respect to γ_p depend on the value of K. Therefore we perform exponential fits on $\langle A\rangle$ with respect to γ_p in the small memory regime given by the interval [0,0.2]. I.e., we consider the ansatz $\langle A\rangle = \langle A\rangle_0 e^{d(K)\gamma_p}$ and find the coefficient $d(K)$ for different values of K. In the inset we plot $d(K)$ as function K and see the decay rate is linearly proportional to K. Here $\langle A\rangle_0$ is the average group contribution without learning (i.e., $\gamma_p = 0$).

$d(K)$ tells us the susceptibility of agents with a given computational budget K to learning. Note that the higher the value of K the faster the rate at which the learning mechanism bring you towards defection, which corresponds to the Nash equilibrium of the PGG. This observation highlights that rationality alone is not sufficient to produce Nash equilibrium

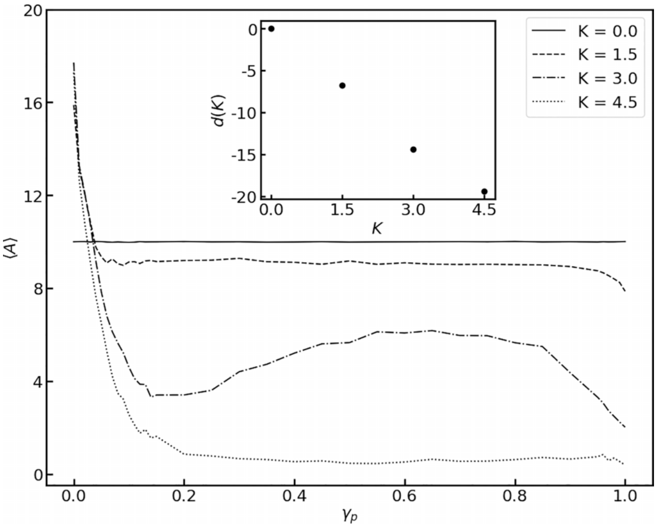

FIG. 1. Average contributions $\langle A\rangle$ as a function of learning strength γ_p for various values of K. Inset depicts the dependence of the decay rate of $\langle A\rangle$ with respect to K.

behavior. A rational agent also needs to develop predictive models of other rational agents to play the Nash equilibrium. This is reminiscent of the standard knowledge that rationality and mutual knowledge of rationality lead to Nash equilibrium [29] in games of more than two players. Our results in Figs. 1 and 2 then seem to indicate that through learning the behavior of other agents, some sort of a mutual knowledge of rationality is developed in a group of all rational agents.

Finally, in Fig. 2 one can see that the impact of K on $\langle A\rangle$ differs qualitatively for different values of γ_p. For lower γ_p, $\langle A\rangle$ increases with K and decreases for higher γ_p. This further highlights the exclusive impact that learning has on bounded rational agents. In so far as how such a qualitative difference is brought about in our model is concerned, we refer the reader to Sec. IV B 1, where the issue is explored in more detail.

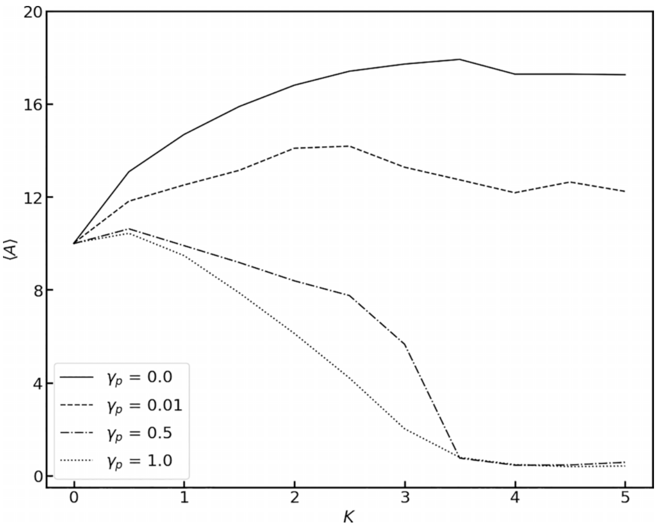

FIG. 2. Average contributions $\langle A\rangle$ as a function of computational budget K for various values of γ_p.

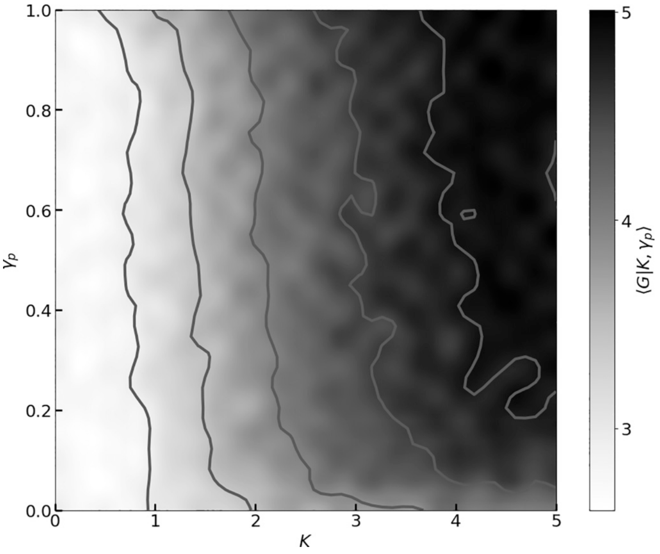

FIG. 3. Conditional expected gains $\langle G|K,\gamma_p\rangle$ (colorbar) and contours (solid grey curves) at $\langle G|K,\gamma_p\rangle = 3, 3.5, 4, 4.5, 4.8$.

B. Groups of random agents

In this section we consider groups of agents where the K, γ_p are i.i.d. (independent and identically distributed) with the uniform distribution $P(K,\gamma_p) = \mathcal{U}$ over the domain $\mathcal{D} = [0, 5] \times [0, 1]$. We are interested in the question: "In a random group of agents playing PGG, which agents gain the most?"

In order to do that, we create 5×10^5 groups and we consider the conditional expected reward $\langle G|K,\gamma_p\rangle$ defined as

$$\langle G|K,\gamma_p\rangle = \int_{\mathcal{D}} GP(G|K,\gamma_p)dG. \tag{6}$$

The gain of a particular agent is defined in Sec I. The conditional expected reward is as shown in Fig. 3. Much in line with our intuition, the conditional gain is maximized by higher values of K, γ_p, i.e., agents with higher computational budget and lesser recency bias earn the most reward when playing against a group of randomly chosen agents.

It is interesting to note that the data in Fig. 3 suggest that there is a trade-off between learning (γ_p) and planning (K). This shows up as a negative slope of the contours and a strong bend towards low γ_p (solid grey curves). Hence, in order to maintain a constant amount of gain, one can trade off the planning computational budget (K) with the learning memory (γ_p). Similar behavior has been observed before [30], although a different planning and learning algorithm was used. The authors defined a total computational budget that is to be allocated to learning and planning and find that optimal rewards are achieved at intermediate values of budget allocation toward planning (and consequently learning).

One can view γ_p also as a measure of computational resources allocated towards learning, as higher values of γ_p require the agent to have more memory and also perform a computationally intensive optimization over the $\xi_\pm$ space. Therefore, one can view the total computational budget of the agent as some linear combination of K and γ_p. In Fig. 3 this

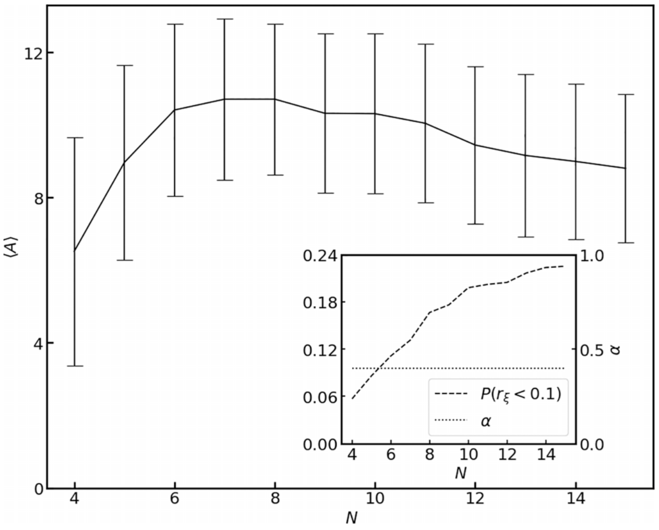

FIG. 4. Average contributions as a function of group size N and the variance (errorbars) for constant α. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

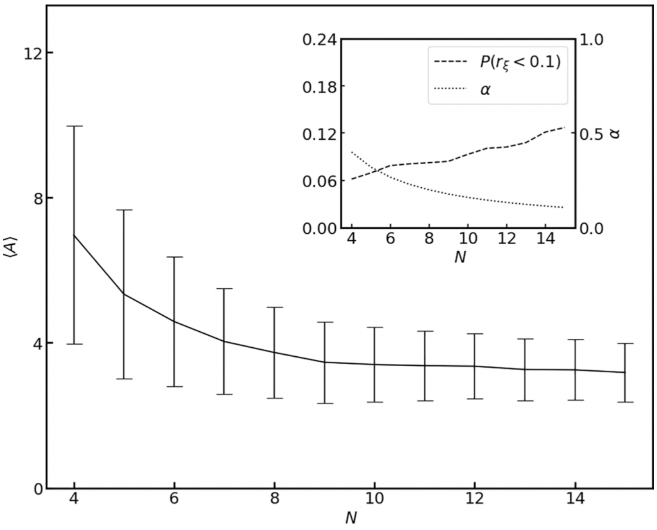

FIG. 5. Average contributions as a function of group size N and the variance (errorbars) for $\alpha \sim 1/N$. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

would be represented by straight lines with negative slope. Due to the bend in the contour lines of $\langle G|K, \gamma_p\rangle$, it can be anticipated that there is a maximum gain for some intermediate values of γ_p and K. This further indicates a potential universality in the trade-off between learning and planning and must be a direction for future research in so far as observing it in human players is concerned.

IV. COOPERATION AMONGST LEARNING AND PLANNING AGENTS

In this section we focus on certain computational experiments which are relevant to experimentally observed behavior in human players playing PGG. In Sec. IV A we observe the impact of group size on cooperation and in Sec. IV B we study how noise in game trajectories might impact the average contributions.

A. Impact of group size on cooperation

Experiments on PGG reveal different kinds of impacts that group size has on cooperation. Where some studies observe that group size positively impacts cooperation [31], some claim that cooperation is harder in larger groups whereas others claim a nonmonotonic impact of group size on cooperation [32,33].

In order to investigate the effect of group size on cooperation, we run simulations of randomly chosen bounded rational agents (i.e., K, γ_p are again i.i.d. with the uniform distribution as in Sec. III B), playing PGG for $T = 100$ periods [34]. Figures 4 and 5 show the ensemble average contribution $\langle A\rangle$ as a function of group size. In Fig. 4 we keep $\alpha = 0.4$ as a constant and we see that the cooperation is impacted nonmonotonically by group size. Cooperation peaks for intermediately sized groups. However, as can be seen from the size of the error bars, this is only the mean behavior of the ensemble, and the behavior of an individual group could vary substantially from the mean.

In order to explore reasons why cooperation may behave nonmonotonically, we first look at the values of $\xi_\pm$ for groups of each size, for all time. More specifically we look at the cumulative probability of having small values of r_ξ (taken to be less than 0.1 here), given by $P(r_\xi < 0.1)$ (see inset Fig. 4). Here $r_\xi = \sqrt{\xi_+^2 + \xi_-^2}$. We observe that $\langle r_\xi\rangle$ monotonically decreases with group size. Recall that $\xi_\pm$ is the degree to which we believe we can encourage or discourage other agents in their contributions. Lower values of r_ξ indicates that the agents are decoupled and this seems to be natural for larger groups, as an individual agent's action tends to have lesser impact on the group behavior as the group size increases. For a detailed calculation see Appendix A 1.

If this were the only process at play, one would be lead to believe that contributions monotonically decrease with group size. But there is a competing tendency. As we increase group size, cooperation is rewarded more steeply as the contributions in the pot are multiplied by αN [see Eq. (1)]. This increases linearly with N for constant α (see inset Fig. 4). Therefore the increase in αN with group size leads to cooperation being more beneficial in larger groups. Combining both these tendencies may lead to cooperation being maximized for intermediate sized groups.

To further verify this explanation we run simulations where we have $\alpha \propto \frac{1}{N}$ such that $\alpha N = const.$ (see Fig. 5). Now as expected, cooperation monotonically falls with group size N. This then seems to indicate that cooperation as a function of group size is influenced by two factors: the degree of control an agent thinks it has on the group contributions and the utility of cooperation. While the latter can be modulated by a parameter of the game (α) the former is a consequence of agent parameters. For instance, agents with smaller γ_p tend to not update their models as much, therefore they assume that they have similar control over larger groups as well. This then leads $\langle A\rangle$ to become monotonically increasing with group size (see Appendix A 3).

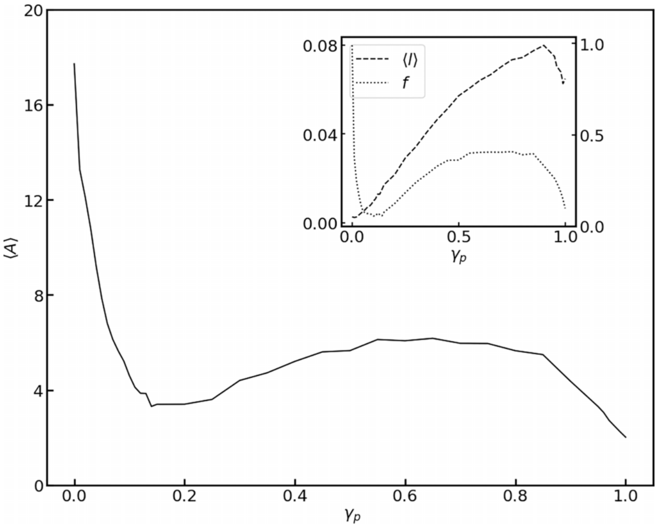

FIG. 6. Average contribution of groups of identical agents with $K = 3$. Inset shows the corresponding values of $\langle l \rangle$ and f as a function of γ_p.

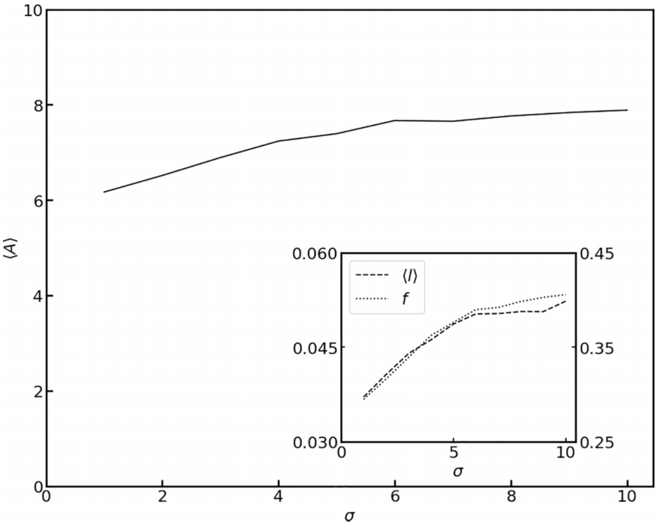

FIG. 7. Average contribution of groups of three randomly chosen agents and one noisy agent with variance of contributions given by σ. Inset shows the corresponding values of $\langle l \rangle$ and f as a function of σ.

B. Noise induced cooperation

1. Anomalous behavior of $K = 3$ agents

In Fig. 1 what is also interesting to note is that for bounded rational agents with $K \approx 3$, intermediate values of γ_p lead to an increase in cooperation, whereas for lower and higher values of K increasing γ_p beyond $\approx$0.2 is inconsequential to the average contribution. In the following we will explore why this is the case.

For agents that learn and plan, the contribution is not only impacted by their capability of choosing the best action (K), but also by their model of other agents ($\xi_\pm$). Certain models encourage the agent to contribute more than other models. More specifically, for the agent to contribute more than the group average contribution, one needs $\xi_+, \xi_- > 0$ (see Appendix A 2).

$\langle A \rangle$ then correlates with the occupation probability of the said quadrant of the $\xi_\pm$ space (see Fig. 6). This can be defined as $f = \frac{\sum_t \langle I_t \rangle}{T}$ where I_t is the indicator function given by

$$I_t = \begin{cases} 1, & \xi_\pm(t) \geqslant 0 \\ 0, & \text{else.} \end{cases} \tag{7}$$

For $\gamma_p = 0$ we start and stay in the aforementioned quadrant as the agent's model is not updated [Eq. (5)], as γ_p is increased the agent starts performing a random walk in the model space, with increasing mean step length l, thereby decreasing the occupation probability of the said quadrant and consequently decreasing the contribution (see Fig. 6). Here the mean step length l is defined as

$$l = \frac{1}{T-1} \sum_{t=2}^{T} \sqrt{(\boldsymbol{\xi}(t) - \boldsymbol{\xi}(t-1))^2}, \tag{8}$$

where $\boldsymbol{\xi}(t) = (\xi_+(t), \xi_-(t))$ and the corresponding ensemble average quantity is given by $\langle l \rangle$.

Upon further increasing γ_p and consequently the average step length $\langle l \rangle$, occupation probability of the said quadrant increases, similar to the manner in which increasing temperature leads to an increase in the probability density in the high potential energy regions. Finally when γ_p is close to 1, $\langle l \rangle$ reduces, because as the game length increases, every new observation has a decreasing impact on the $\xi_\pm$ value as obtained from Eq. (5). This then further reduces the occupation probability and also the contribution $\langle A \rangle$ as a consequence.

2. Adding a noisy agent to a group

Given the arguments above, it would be natural to expect that noise (i.e., greater $\langle l \rangle$) can enhance cooperation among bounded rational agents playing PGG. Apart from keeping γ_p in the intermediate region, $\langle l \rangle$ can be increased by adding a noisy agent to the group and increasing the variance of contributions of the noisy agent.

Hence in order to further explore the hypothesis above, we consider to add one noisy agent with $K = 0$, a fixed mean of contribution $m = 10$ and varying variances σ of the prior distribution $[P_0(f_{k,t})]$, to a group of three other randomly chosen agents, as done in Sec. III B. We then observe how the group average contributions $\langle A \rangle$ are impacted as we increase σ.

In Fig. 7 one can see that as the variance of the contributions of the noisy player σ is increased, the average contribution of the group increases. In the inset we also see the corresponding increase in $\langle l \rangle$ and also f. Thereby adding weight to the claim that cooperation can be induced by increasing noise in the game behavior. Whether this behavior is also observed in human players playing PGG, is yet to be tested experimentally.

V. CONCLUSIONS

We have demonstrated the exclusive impact of learning on the behavior of bounded rational agents in PGG. We explore the impact of noise on cooperation. Specifically, we find that the introduction of an agent that contributes in a noisy manner (i.e., with finite variance) to the public pot positively impacts the average contribution. It is found that this effect systematically increases as the variance is increased. This prediction remains to be tested via experiments.

We also provide a theoretical explanation to the observed impact of group size on cooperation, specifically we show that the shape of the curve of average contributions $\langle A\rangle$ vs group size N can be modulated by varying the MPCR and also the agent parameters. More specifically, there are qualitative differences in the contribution curves depending on whether the agents are learning or not. This provides us a quantifiable way of predicting cooperation in PGGs with varying number of players.

Our results not only justify the bounded rational model of human behavior but also show how rather simple assumptions on human behavior can lead to a large variety of behaviors that are also observed in experiments. This provides an alternative to the *ad hoc* cellular automata (CA) type models that are commonly found in literature. One criticism of this approach could be that it is rather cumbersome as opposed to CA based models. If there is any validity to the criticism then we suggest that this model be treated as a more fine-grained model of player behavior in games and one should then systematically find more coarse-grained CA type models which are effective descriptors of some coarse-grained observables.

While the presented model could be construed as a fine-grained model in comparison to CA based models, it is still an *effective* description of human decision making, as opposed to a *mechanistic* one. That is to say, our model makes statements such as "...humans behave in PGG *as if* they were solving Eqs. (4) and (5)...", as compared to a mechanistic model (such as DDM [35]) which makes statements such as "...humans play by *enacting* this procedure/algorithm...". We therefore do not make claims about how humans actually come up with their decisions. To check the veracity of either type of model of human decision making, it must first lend itself to experimental tests, as we are aiming at in the present study. Only then can a predictive simulation of human economic interactions, as alluded to in the introduction, be tackled successfully.

ACKNOWLEDGMENTS

The authors gratefully acknowledge useful discussions with S. Muehle and M. Khan. The authors thank P. Sharma for reviewing the manuscript and providing valuable feedback.

APPENDIX: EFFECT OF $\xi_{\pm}$ ON COOPERATION

1. The limit of $r_\xi \to 0$

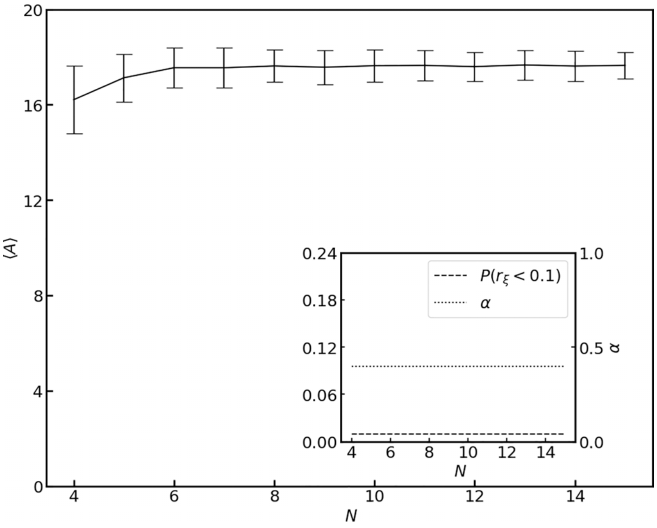

FIG. 8. Average contributions as a function of group size N and the variance (errorbars) for constant α and $\gamma_p = 0$. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

In the limit $r_\xi \to 0$ we can see that the transition function becomes independent of player contribution $f_{k,t}$ [see from Eq. (3) $\lim_{r_\xi \to 0} m = \mu_{k,t-1}$]. This essentially (from the perspective of the kth agent) decouples the agent from other players. We can see this effect more precisely in Eq. (4). For simplicity we ignore the bounded rationality term and consider that $T = t + 1$. Upon substituting for the value function and expanding we have

$$V_t^* = \max_{P(f_t^{t+1})} \sum_{\bar{f}_t} G_{k,t}\left[[PQ]_t + \gamma[PQ]_t \sum_{\bar{f}_{t+1}} G_{k,t+1}[PQ]_{t+1}\right], \tag{A1}$$

where $[PQ]_t$ is a short-hand notation for $P(f_{k,t})Q(\mu_{k,t}|\bar{f}_{t-1})$. We can perform the maximization over $P(f_{k,t+1})$ directly over the second summand as follows

$$\max_{P(f_{k,t+1})} \sum_{\bar{f}_{t+1}} \alpha(N-1)\mu_{t+1}[PQ]_{t+1} - \sum_{\bar{f}_{t+1}} (1-\alpha) f_{k,t+1}[PQ]_{t+1}, \tag{A2}$$

which further simplifies to

$$\max_{P(f_{k,t+1})} \sum_{\mu_{k,t+1}} \alpha(N-1)\mu_{t+1}Q_{t+1} - \sum_{f_{k,t+1}} (1-\alpha) f_{k,t+1}P_{t+1}. \tag{A3}$$

The first summand has no $f_{k,t+1}$ dependence, and the second summand can be seen to be maximized at $P(f_{k,t+1}) = \delta(f_{k,t+1})$ (because $\alpha < 1$) and therefore it vanishes upon maximization. Also, the first summand has no $f_{k,t}$ dependence, therefore it essentially reduces to $\alpha(N-1)\langle\mu_{k,t+1}|\mu_{k,t}\rangle$. Substituting this in Eq. (A1) we get

$$V_t^* = \max_{P(f_{k,t})} \sum_{\bar{f}_t} G_{k,t}[PQ]_t + \gamma[PQ]_t\langle\mu_{k,t+1}|\mu_{k,t}\rangle\alpha(N-1). \tag{A4}$$

The second summand in this equation when summed over $f_{k,t}$ gives a constant $\gamma Q_t\langle\mu_{k,t+1}|\mu_{k,t}\rangle\alpha(N-1)$ independent of $P(f_{k,t})$ and therefore it doesn't participate in the maximization. It then remains trivial to see that maximizing over $P(f_{k,t})$ gives $P(f_{k,t}) = \delta(f_{k,t})$. Therefore, it was optimal to defect in both the periods.

For $T > t + 1$ one can similarly see that at all periods the conditional expected contribution of other players will not depend on the player's play $(f_{k,t})$ and therefore the term will not participate in the maximization. Also, upon the introduction of the bounded rationality term for $K \approx 0$, the maximization will result in distributions similar to the prior and as K increases the mean contributions decrease, until at a critical threshold of computational budget K_{crit} where $D_{KL}(\delta(f_{k,t})||P_0(f_{k,t})) = K_{\text{crit}}$, it again resembles the solution for fully rational agents that we see above.

2. The cooperation quadrant

When r_ξ is not close to 0, the second summand in Eq. (A4) would be conditioned on $f_{k,t}$ as well, i.e., it would become $\gamma[PQ]_t\langle\mu_{k,t+1}|\mu_{k,t}, f_{k,t}\rangle\alpha(N-1)$. In order for the optimal action to not be defection it would be necessary (but not sufficient) that $\frac{\partial\langle\mu_{k,t+1}|\mu_{k,t},f_{k,t}\rangle}{\partial f_{k,t}} > 0$. From Eq. (3) one can see that this is the case when $\xi_\pm > 0$. Therefore when $\xi_\pm > 0$ the agents contribute the most.

3. Constant r_ξ and changing N

In Fig. 8 we show average contributions as a function of number of players in a group. Where agents in a group are described $\gamma_p = 0$ and K chosen uniformly randomly on the domain [0,5]. $P(r_\xi < 0.1) = 0$ for all values of N as can be seen in the inset. Note that as $P(r_\xi < 0.1)$ is constant w.r.t. N and αN is increasing linearly in N the average contributions increase with group size.

[1] L. M. A. Bettencourt and J. Kaur, Evolution, and structure of sustainability science, Proc. Natl. Acad. Sci. USA **108**, 19540 (2011).

[2] A. Falk, A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde, Global evidence on economic preferences, Q. J. Econ. **133**, 1645 (2018).

[3] L. J. Kotzé and R. E. Kim, Earth system law: the juridical dimensions of earth system, Earth Syst. Governance **1**, 100003 (2019).

[4] 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).

[5] 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).

[6] 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).

[7] 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).

[8] M. Tomassini and A. Antonioni, Computational behavioral models for public goods games on social networks, Games **10**, 35 (2019).

[9] 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. Hum. Behav. **5**, 586 (2021).

[10] J. R. Wright and K. Leyton-Brown, Predicting human behavior in unrepeated, simultaneous-move games, Games and Economic Behavior **106**, 16 (2017).

[11] A. Vazifedan and M. Izadi, Predicting human behavior in size-variant repeated games through deep convolutional neural networks, Prog. Artif. Intell. **11**, 15 (2022).

[12] 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).

[13] M. Perc, J. J. Jordan, D. G. Rand, and Z. Wang, Statistical physics of human cooperation, Phys. Rep. **687**, 1 (2017).

[14] M. Burton-Chellew and S. A. West, Payoff-based learning best explains the rate of decline in cooperation accross 237 public-goods games, Nat. Hum. Behav. **5**, 1330 (2021).

[15] E. Bonabeau, Agent-based modeling: Methods, and techniques for simulating human systems, Proc. Natl. Acad. Sci. USA **99**, 7280 (2002).

[16] P. Godara, T. D. Aléman, and S. Herminghaus, Bounded rational agents playing a public goods game, Phys. Rev. E **105**, 024114 (2022).

[17] B. Herrmann, C. Thöni, and S. Gächter, Antisocial punishment across societies, Science **319**, 1362 (2008).

[18] M. Hutter, A theory of universal artificial intelligence based on algorithmic complexity, arXiv:cs/0004001.

[19] J. Hofbauer and K. Sigmund, *Evolutionary Games, and Population Dynamics*, Cambridge University Press, 1998.

[20] K. Sigmund, H. D. Silva, A. Traulsen, and C. Hauert, Social learning promotes institutions for governing the commons, Nature **466**, 861 (2010).

[21] A. Amado, W. Huang, P. R. A. Campos, and F. F. Ferreira, Learning process in public goods games, Physica A **430**, 21 (2015).

[22] M. Hutter, Universal Algorithmic Intelligence: A Mathematical Top→Down Approach, *Artificial General Intelligence* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007), pp. 227–290.

[23] R. J. Solomonoff, A formal theory of inductive inference. part i, Inf. Control **7**, 1 (1964).

[24] Note 1, A partial function from $X \to Y$ is a function that is only defined on a subset $S \subset X$. If the function is defined on all of X, i.e., $S = X$ then the function is called a total function.

[25] Note 2, Here * refers to the space of finite strings in the alphabet $\{0, 1\}$, including the empty string.

[26] A. Church, An unsolvable problem of elementary number theory, Am. J. Math. **58**, 345 (1936).

[27] D. Fudenberg and David K. Levine, Recency, consistent learning, and nash equilibrium, Proc. Natl. Acad. Sci. USA **111**, 10826 (2014).

[28] Note 3, In Ref. [16] it has been shown that both m, and γ have a monotonic impact on $\langle A\rangle$. That is for a fixed k, and γ_p increasing m monotonically increases the contributions, and for a fixed m, and K, increasing γ monotonically decreases

the contributions. Therefore the qualitative features that are the focus of our attention in this article are unchanged by these parameters. Hence an arbitrary choice of these parameters was made.

[29] R. Aumann and A. Brandenburger, Epistemic conditions for nash equilibrium, Econometrica **63**, 1161 (1995).

[30] T. M. Moerland, A. Deichler, S. Baldi, J. Broekens, and C. M. Jonker, Think too fast nor too slow: The computational trade-off between planning, and reinforcement learning, arXiv:2005.07404.

[31] P. María, C. Valerio, and S. Angel, Group size effects, and critical mass in public goods games, Sci. Rep. **9**, 5503 (2019).

[32] R. M. Isaac and J. M. Walker, Group size effects in public goods provision: The voluntary contributions mechanism, Q. J. Econ. **103**, 179 (1988).

[33] W. Yang, W. Liu, A. Viña, M.-N. Tuanmu, G. He, T. Dietz, and J. Liu, Nonlinear effects of group size on collective action, and resource outcomes, Proc. Natl. Acad. Sci. USA **110**, 10916 (2013).

[34] Note 4, For each parameter configuration we run an ensemble of 1350 groups, and consider the average behavior.

[35] R. Ratcliff and G. McKoon, The diffusion decision model: theory, and data for two-choice decision tasks, Neural Comput. **20**, 873 (2008).