Resource bounded agency in public goods games

Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm/Promotionsstudiengang

der Georg-August University School of Science (GAUSS)

vorgelegt von:

Prakhar Godara

aus:

Rajasthan, India

July 19, 2023

Betreuungsausschuss

1. **Prof. Dr. Stephan Herminghaus**
   Dynamic of Complex Fluids
   Max Planck Institute for Dynamics and Self-organisation
2. **Prof. Dr. Peter Sollich**
   Non-equilibrium statistical physics
   Institute for Theoretical Physics, Göttingen university
3. **Prof. Dr. Stefan Klumpp**
   Theoretical biophysics
   Institut für Dynamik komplexer Systeme, Göttingen university

Mitglieder der Prüfungskommission

1. Referee: **Prof. Dr. Stephan Herminghaus**
   Dynamic of Complex Fluids
   Max Planck Institute for Dynamics and Self-organisation
2. Referee: **Prof. Dr. Stefan Klumpp**
   Theoretical biophysics
   Institut für Dynamik komplexer Systeme, Göttingen university
3. **Prof. Dr. Peter Sollich**
   Non-equilibrium statistical physics
   Institute for Theoretical Physics, Göttingen university
4. **Prof. Dr. Viola Priesemann**
   Theory of Complex Systems
   Max Planck Institute for Dynamics and Self-organisation
5. **Dr. Igor Kagan**
   Cognitive Neuroscience Laboratory
   German Primate Center
6. **Prof. Dr. Michael Waldmann**
   Georg-Elias-Müller-Institut für Psychologie
   Göttingen university

Tag der mündlichen Prüfung: July 6, 2023

"I want to claim that AI is better viewed as sharing with traditional epistemology the status of being a most general, most abstract asking of the top-down question: how is knowledge possible?"

Daniel Dennett

Abstract

Prakhar Godara

Resource bounded agency in public goods games

The new and emerging field of sociophysics aims to explain and understand collective behavior demonstrated by humans in a variety of domains. This usually requires to make some assumptions about the microscopic inter-human rules of interaction. Often these interactions rules are inspired from physical systems, and are therefore criticized to be too simplistic, to give meaningful predictions of human behavior in novel domains. In this thesis, I take inspiration from general intelligence research and game theory to develop a model of human agency in the well known public goods game (PGG), in order to respond to the criticism. The model agent has two aspects called *learning* and *planning*, both of which, are bounded by constraints. While the former is bounded by the amount of memory the agent holds about it's past (recency bias), the latter is bound by an information theoretic constraint inspired from information thermodynamics. The thesis presents a route from microscopic behavior to collective behavior of human players in PGG. I start with demonstrating that a bounded planning agent is sufficient to explain human behavior in short games. Following that, I also include the bounded learning mechanism and demonstrate the exclusive effects of the learning mechanism on agent behavior. Finally I make use of the model to explore collective effects of these humanized agents, the behavior of which, as opposed to contemporary models, is justified by comparing to experimental data on human behavior.

Acknowledgements

I thank the following people, for the mentioned reasons.

1. Stephan Herminghaus-
   (a) For providing me with academic freedom to explore my own ideas and helping to develop them. I cannot stress how important this has been in shaping me.
   (b) For instilling in me, explicitly, to find the "simplest" problem first.
   (c) For guiding through extra-academic concerns, such as discussions on quality of science, traits one should build as a good scientist, and life in general.
   (d) For letting me unrestrictedly use the resources of the institute. For instance, loaning me a blackboard (which I should return right after I submit my thesis) and letting me occupy the office space in the old MPIDS office in Bunsenstrasse.
2. Steffen Muehle - for engaging in a wide set of discussions with me. A lot of which ended up shaping my research work presented in the thesis. Also for discussions that didn't pan out.
3. Marcus Hutter and Joscha Bach - despite not having met them ever, their ideas have inspired a lot of my own.
4. Monika Teuteberg - for helping me navigate the bureaucratic, language and organisational concerns, from day 1 in Germany, to my last.
5. Kris Hantke and Thomas Eggers - for tolerating my computer/cluster related illiteracy. Also for guiding me through the use of the cluster, VPN, etc..
6. Mariyam Khan, Devarshi Mukherjee, Simran Kaur, Joscha Tabet, Marco Mazza, Sidhi Gupta, Hiu Fai Yik, Pang Shiang Tay, Knut Heidemann, Puneet Sharma and Benoit Mahault - for intentionally or unintentionally helping me formalize my thoughts on a wide variety of subjects. In the downstream of a lot of those discussions, are some ideas, that found their way into the thesis.

7. Mariyam Khan, Rishbha Godara, Rameshwar Godara and Alpana Godara - for rejoicing in my victories, and not, in my failures.

Contents

List of Figures

Dedicated to my goal of
understanding intelligent
behavior.

Chapter 1

Introduction

"I can calculate the motion of heavenly bodies, but not the madness of people." - such was the sentiment of Sir Isaac Newton in the year 1720 after having lost approximately 2×10^4 pounds in an economic collapse.

1.1 Physics of the madness of people

The main goal of social physics is to describe and understand collective social phenomena that emerge from interactions between individual humans. While studying the collective behavior of humans is not the same as studying the collective behavior of (the more *sane* counterparts, for instance) molecules of water, there still are quite a few macroscopic regularities in human behavior (Castellano, Fortunato, and Loreto, 2009), thereby making social systems susceptible to mathematical tools of statistical physics.

As it turns out, statistical physics of social systems seems to "work" for quite a few classes of social systems and has led social physics to become a mature field of research, with domains of interest to social physicists ranging from traffic flows, economic time series modelling, opinion dynamics, cooperation in social games, migration and epidemiology, to name a few (Jusup et al., 2022 provides a comprehensive overview of the field).

1.1.1 How is it done?

Statistical physics is driven by the idea that observables of collective behavior (say melting point of water) lie in a smaller dimensional space as compared to the state space of the microscopic degrees of freedom of the system (position and momenta of all the molecules) and therefore a lot of information regarding the microscopic dynamics can be ignored when concerned with the collective effects (hence universality classes). A dominant modelling paradigm in statistical physics of social systems is *agent based modelling* (ABM). In principle, every ABM approach is carried out in the following manner. In each step we state the corresponding quantity in the Ising model to aid understanding.

1. **Macroscopic**: choose a domain, therefore the coarse grained observables of interest, like magnetism.

2. **Microscopic**:

   (a) define the microscopic entities (*agents* in ABM and spins in the Ising model) and

   (b) choose an interaction rule for the entities, like the Hamiltonian in the Ising model.

3. **Emergence**: from the microscopic dynamics, use the tools of statistical physics to evaluate the macroscopic observables of interest.

If the outcome of the model doesn't match the experimental observation then step 2 is redone.

The above blueprint of ABM will become apparent with the following two examples from social physics. The first example comes from modelling opinion dynamics of agents interacting via a network as described in Baumann et al., 2020.

1. The objective of the study is to explore the emergence of *polarization* on the opinion space. It will become clear in the following, what their definition of polarization is.

2. (a) The opinion space is given by $X = [-\infty, +\infty]$ where $x \in X$ represents an opinion. One also considers N agents interacting via a network given by the adjacency matrix A. The opinion of the ith agent at time t is given by $x_i(t)$.

   (b) The evolution of the opinions of agents is given by the dynamical law
$$\frac{dx_i(t)}{dt} = -x_i(t) + K \sum_{j=1}^{N} A_{ij} \tanh(\alpha x_j). \tag{1.1}$$

3. The dynamical law is then used to evaluate the distribution $P(x, t)$ of the opinions of the ensemble of agents. Polarization is then evaluated from the distribution. At any time t the population is considered to be polarized if there are two well-separated peaks in the opinion distribution around the neutral consensus ($x = 0$).

The second example comes from modelling player behavior in the well-known public goods game (PGG). The PGG is played between N players where each player is given some initial endowment (say 20 tokens) and they are allowed to contribute all or a part of their endowment into a public pot anonymously. Once all the contributions are collected, the total contribution is multiplied by a factor greater than 1 and then divided among all the players equally. The payoff of a player at the end of the game is the sum of the tokens not contributed and the rewards reaped from the common pot. In an iterated PGG this game is repeatedly played for a known number of periods. In the following, we describe the agent based model approach generally taken in this field (for instance the work of Szabó and Hauert, 2002 serves as an example).

1. A common goal is to explain the cooperative and non-Nash behavior shown by human players playing iterated PGG.
2. (a) Each player playing the game is given an identity, in the simplest case, from the set $\{C, D\}$, where C stands for cooperator and D for defector. Both cooperators and defectors behave in a prescribed manner (for instance, cooperator always contributes all the tokens, defector always doesn't contribute).
   (b) The agents play based on the rules above and the payoffs are observed. The identities of the agents are then updated through some transition probability distribution (also depending on the payoffs) before the next period begins and the game continues.
3. The dynamical rule described above is then used to evaluate the emergence of (or the lack of) cooperation, which is measured via the total average contribution of the players in a given period of the game.

1.1.2 Problems with ABM

Essentially, the burden of a physicist (or any other modeller) comes down to performing a "good" selection of the microscopic interaction rule for the ABM. The true test of the goodness of the rule is whether the macroscopic observable of interest is faithfully replicated. But the space of all allowed rules is much larger than the space of allowed values of the macroscopic observable. This essentially creates a degeneracy in the rule space and allows the modeller to make an arbitrary choice of the microscopic rules to replicate a simple behavior of the macroscopic observable.

The reason why this is not a big concern for traditional domains of physics is that in physical systems, we exploit various symmetries and conservation laws that restrict the space of allowed interaction rules. This is not a luxury one has when modelling microscopic entities as complex as humans. Therefore, especially in social physics, such ad hoc ABM provides little insight into what is actually happening and is also bound to fail in novel situations, when suddenly a new observable becomes of interest. The author claims that such is the state of the majority of publications in the social physics literature.

1.1.3 What can be done?

Essentially the above problem can be reduced to that of non-invertibility of macroscopic dynamics to microscopic dynamics. There could be two ways to resolve this issue.

1. One approach to deal with this non-invertibility could be to increase the number of macroscopic observables that need to be replicated, as this would put further restriction on the space of microscopic rules. The limiting process of increasing the number of observables might just come down to making the macroscopic observable a microscopic one. This would necessitate a more detailed inspection into individual agency.

2. Another route could be to use the same interaction rule across different domains of application. This would require that the agent based models follow a general enough framework that allows for it to be applicable in various settings, i.e., one needs to develop a view of agency that is independent of the domain in which the agent operates.

Essentially the above indicate the need for a more **accurate** and **general** framework of human agency. As it turns out there are ideas in other academic disciplines that fill this void. More specifically, economists have pondered the nature of human agency, which has led to Game theory, which has become a dominant modelling paradigm for economic agents. On the other hand, computer scientists working on general intelligence have pondered on what general models of agency can be built that allow the agents to perform in a wide array of environments. The author believes that physicists should make use of the progress made in these disciplines to guide their models of agency in social physics.

1.2 Modelling individual agency

As discussed in the previous section, we will bring our attention to modelling individual agency. We will do this by exploring what the literature in other disciplines has to say about human agency and agency in general. This will also serve as a motivation for the line of work presented in this thesis. Before we do any of this, we need to make some matters precise.

1.2.1 What even is an agent?

A notion that is central to this line of research is the notion of an *agent*. In the most abstract sense, an agent can be viewed as a function f (Russell and Subramanian, 1994) whose domain is the space of *observations* O and co-domain

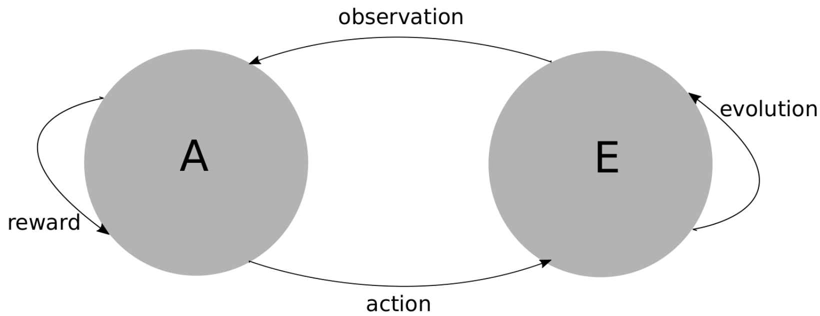


FIGURE 1.1: Relationship between the agent A and environment E.

is a space of *actions* A, i.e. $f : O \to A$. Through these observations and actions, the agent is in constant interaction with, what can be called the *environment*. The environment is essentially viewed as a controllable dynamical system (stochastic or otherwise) and the control variable is the action of the agent, equivalently it can be viewed as a function $g : X \times A \to X$ where X is the state space of the environment. The observations $o \in O$ on the other hand are some functions of the state of the environment i.e. $h : X \to O$.

In addition to the agent, environment and observation functions there is also a utility function $U : X \times A \to$ R, which describes the goals of the agent. A utility function essentially induces an order relation on the space of actions i.e. some actions are better than others. The relationship between the agent and the environment has been provided a schematic representation as shown in Fig. 1.1. The agent-environment interactions can be viewed as a cycle that starts with the environment in a given state, the agent observes the state of the environment and outputs an action, which then feeds into the evolution law of the environment, consequently generating a new environment state and also a reward for the agent.

There are a couple of things to note already.

1. The notion of an agent in consideration is oblivious to how the agent function is implemented physically i.e. a machine for instance. This is only a functional description of agency as opposed to an *architectural* one.

2. The description of the environment as a function is sufficiently general as it also allows for the possibility for the environment to include other agents.

1.2.2 Searching for "good" agent functions

The general notion of agency defined above allows us to cast the pursuits of agent based modelling across different fields as a search problem for different agent functions. In the following, we explore the relevant explorations in the fields of general intelligence research and game theory.

Artificial general intelligence (AGI)

Very broadly speaking it can be said that the search for general intelligence is a search for those f that perform well/intelligently (with respect to the utility function) for a wide range of g. A formalization of this order relation on agent functions has been provided by Legg and Hutter, 2007. The authors not only define the order relation but also describe (one of) the maximal element(s) of the set of agents with respect to the order relation and it corresponds to the AIXI[1] agent function (Hutter, 2007).

The AIXI agent function is essentially an amalgamation of the Solomonoff's theory of induction (Solomonoff, 1964) and Bellman's theory of optimal control (Bellman, 1954). The AIXI agent function can be viewed as a composition of two functions called the *learning* and the *planning* functions. Where the learning function is a mapping from observations to models of the environment and the planning function is a mapping from models of the environment to actions. It is then shown that this learning and planning agent is Pareto optimal with respect to the aforementioned order relation (although this result comes with its own set of nuances (Leike and Hutter, 2015)).

Game theory

Even before its formal inception with the work of Neumann, 1928, game theory was being developed with the motivation of describing optimal behavior in economic interactions for instance Cournot, Bacon, and Fisher, 1897 discussed what is now known as the *Cournot duopoly*. Economic interactions are usually abstracted into games such as prisoner's dilemma, public goods game (which will be the focus of our attention in this thesis), dictator game, etc. and agent behavior in these specific environments is studied.

Broadly speaking there are two sub branches to game theory - *normative* game theory and *descriptive* game theory (or behavioral game theory) and they can be characterized by different searches for agent functions. Normative game theorists concern themselves with how the agent **ought** to act in economic interactions. Von Neumann in his aforementioned work states a set of axioms that any "good" economic agent must act as, thereby restricting the set of good agent functions. For instance *perfect rationality* and *common knowledge* of rationality are commonly assumed by normative game theorists (we will discuss this in a little more detail in Chapter 2).

[1]This is short for Artificial Intelligence - ξ.

On the other hand behavioral game theorists have the concern of describing how agents (humans) **actually** interact in economic interactions as observed by experiments of humans playing economic games. Here the hope is to develop a theory that finds agent functions that most closely represent human behavior in the environments related to economic interactions. Camerer, 2011 serves as a good overview of the subject.

1.3 Returning to social physics: Our approach

In this thesis we combine the insights from AGI research (learning and planning agent) which provides generality (i.e. applicability to a wide range of environments) to our agent model and also add some specific insights from behavioral game theory (cognitive bounds on learning and planning) to give our agent model the human touch, thereby alleviating the two main worries that face the ABM approach to study social systems. We do this by developing our agent model and employing it in a very well studied paradigm afforded by the public goods game (PGG). We consider PGG simply for our results to be comparable to contemporary literature in sociophysics. Finally. we end up with an agent model that has parameters which can be interpreted as individual human preferences and can be inferred from experimental data.

Before we move on to describe our agent model, in the next chapter we describe the PGG and review some of the work that has already been done to the end of explaining human behavior in PGG and provide further motivation for the modelling route we choose.

Chapter 2

The public goods game

In this chapter, we will set the stage for the work done in the thesis. We start by defining the public goods game (PGG) and then providing a brief overview of some game theoretic concepts associated with "optimal" behavior in PGG. We will also describe the experimental evidence on human behavior in PGG, which our models in the subsequent chapters will aim to explain. To contrast our model with the current standards in the field we will also explore contemporary models that explain human behavior in PGG.

2.1 Game definition

A public goods game (PGG)[1] is defined by the tuple (N, T, τ, α), where N in the number of players, T is the number of iterations (also called periods) for which the game is played, τ is in initial endowment provided to each player in each period and α is the return factor (also called MPCR - mean per capita return). In each period of the game, the following happens (Fig. 2.1 shows a schematic of one period of the game, and references to it will be in brackets.).

1. Each player is provided with the same number $\tau \in \mathbb{N}$ of tokens ($\tau = 20$ in the figure).
2. Each player anonymously contributes an integer number of tokens $(20, 0, 20, 20)$ to a "public" pot (represented by the cyan box in the middle).
3. Each player receives a return worth the total collection in the pot, multiplied with a number $\alpha \in \mathbb{R}$ ($\alpha = 0.4$). This return, plus the tokens held back initially by the player, is called her reward $(24, 44, 24, 24)$.

This completes one period of the game. After that, the game continues for a finite (and previously known to all players) number T of periods. The total gain of an individual is then the sum of the rewards in each period.

[1] Technically it must be called the iterated public goods game. Regardless, we continue this misnomer throughout the thesis for convenience. Whenever referring to the actual PGG we may refer to it as a "single shot PGG" or a "single period PGG".

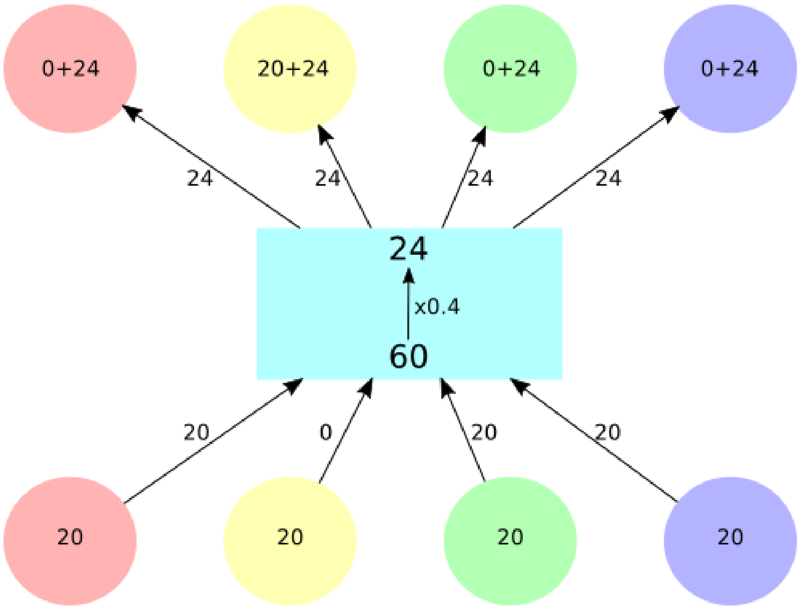


FIGURE 2.1: One period of a PGG with $N = 4$ players and $\alpha = 0.4$.

Giving notations to the above, we can represent the contribution of a player $i \leq N$ in period $t \leq T$ as $f_{i,t} \in [0, \tau]$. Then the gain of player i in period t (also called the immediate gain) is given by

$$G_{i,t} = \alpha \sum_{j=1}^{N} f_{j,t} - f_{i,t}. \tag{2.1}$$

This can also be written as

$$G_{i,t} = \alpha(N-1)\mu_{i,t} - (1-\alpha) f_{i,t}, \tag{2.2}$$

where $\mu_{i,t} = \frac{\sum_{k \neq i} f_{k,t}}{N-1}$ is the average contribution of other players. The total gain for the ith player can then be defined as $G_i = \sum_t^T G_{i,t}$.

2.2 Why focus on PGG?

The main reason to focus on PGG is due to the wide scope it has. It is a classic example of a "social dilemma", i.e. a situation where the individual and collective goals clash. This will become clear in the following. As can be seen from the definition of the game, if $\alpha \geq 1$ then it is trivial to see that every individual gains by contributing all of her initial endowment into the public pot. On the other hand, if $\alpha < 1$ things are not as clear. We can certainly see from Eq. 2.2 that individual gains are maximized when the individual "defects" (i.e. doesn't contribute) and others do not (like the yellow player in Fig. 2.1). Although if everyone follows this strategy and defects, no one ends up with more reward than the initial endowment τ. On the other hand, collective gain (i.e. the sum of gains of all the players) gets maximized when the collection in the public pot is maximized, i.e. when everyone contributes. In such a scenario everyone gets a better reward as compared to the case where everyone defects. But then again, this goes against the self-interest of the players.

This aspect of a social dilemma that PGG captures allows it to be a good metaphor for a variety of situations we encounter. For instance, let's consider mask-wearing to prevent the spread of the Covid-19 virus. Clearly, there is a personal cost to wearing a mask, as evidenced by plenty of us wearing it underneath the nose. Those of us who wore a mask essentially contributed to the public pot, where the return on our investment was a lowered probability of being infected. Social gains would be maximized by everyone wearing masks. Although, the best outcome for an individual is when everyone wears it and the individual doesn't, much like a PGG! Other domains where the PGG metaphor applies are for example state taxes, reducing the carbon footprint for climate, cleanliness in the house with multiple roommates, etc.. Such a wide scope of applicability of PGG has made it a fundamental paradigm in experimental economics, which also means that there is plenty of experimental data available that theoreticians can aim to explain. Therefore, developing good models for human behavior in PGG enhances our understanding of a wide variety of socio-economic environments that we are embedded in.

2.3 How we ought to act in games

2.3.1 Rationality

Before we pursue how we ought to act in a multiplayer game such as a PGG, let's consider how we ought to act in a single player game $\mathcal{G}_1$. The game is defined by a tuple $\mathcal{G}_1 = (X, Y, C, U)$, where

- X is an action space, denoting the set of all actions available to the player,
- Y a consequence space, denoting the consequences of each action,
- $C : X \to Y$ the causal map from actions to consequences and
- $U : Y \to \mathbb{R}$ is the utility function denoting the reward of the player.

In this game if a player takes an action $x \in X$, then it earns a reward $U(C(x))$. In order for the player to be optimal or **rational**[2], the player **must** be able to choose an action x^* such that

$$U(C(x^*)) \geq U(C(x)) \forall x \in X. \tag{2.3}$$

This is a deterministic game, i.e. the causal map from actions to consequences is given by a function C. In a more general setting, where we allow for actions to have uncertain consequences the causal map C transforms to $C' : X \to P(Y)$, where $P(Y)$ is a probability measure on the space Y. In such a case a rational agent **must** be capable of maximizing the *expected utility*, i.e. the agent must be able to choose an action x^* such that

$$\int_Y C(x^*)U(y)dy \geq \int_Y C(x)U(y)dy \forall x \in X. \tag{2.4}$$

So far, playing the game $\mathcal{G}_1$ optimally or rationally means that the agent must be able to maximize either the utility or the expected utility function. $\mathcal{G}_1$ is a trivial game in the sense that there is no *strategic* interaction between players. This simplicity vanishes the moment we have a multiplayer interaction where the utility of each player depends on the actions of others. Let us explore this possibility by considering a two player game $\mathcal{G}_2 = (X_1, X_2, Y, C, U_1, U_2)$. Here

- X_1 and X_2 are the action spaces of the respective players,
- Y the consequence space,
- $C : X_1 \times X_2 \to Y$ a causal map and
- $U_1 : Y \to \mathbb{R}$ and $U_2 : Y \to \mathbb{R}$ are the respective utility functions.

If we now proceed as before then from the perspective of player 1, the player is rational if she can choose an action $x_1^* \in X_1$ such that U_1 gets maximized. But U_1 depends on $y \in Y$ and y depends on the actions of both the players. Therefore if player 1 wants to be able to maximize her payoff, she needs some information regarding what player 2 is going to do, but this information is not provided to player 1.

2.3.2 Common knowledge of rationality

It seems that mere **rationality** (i.e. the capability to maximize (expected) payoff) is insufficient in guiding us to act in a multiplayer setting and we need to

[2]Here we have tacitly introduced the notion of "rationality" which is formally defined via the VNM axioms. For pedagogical reasons, we will not get into the formal description, but rather just carry on with these operational notions. Any interested reader is encouraged to go through Von Neumann and Morgenstern, 2004.

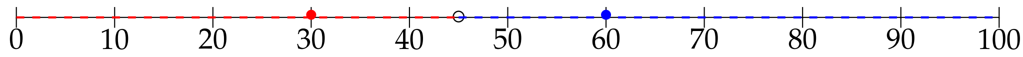


FIGURE 2.2: Two player location game with the potential locations of the red and blue players are as shown at 30 and 60 respectively. The dashed lines represent the part of the city that flocks to the respectively colored shop.

make some assumptions about the kind of knowledge each player has about other players. It may be better to proceed by an example. Consider the two player location game. In this game we have two players named red and blue , who both want to open an ice cream shop in a one dimensional city given by the interval $[0, 100]$. The population in this city is uniformly distributed on this interval but the shops can only be opened at x coordinates that are multiples of 10 as marked in Fig. 2.2. The goal of the players is to open the shop at such a place that most people from the city buy ice-creams from their shop. The city however is famous for its lazy people and it is known by both players that the customers are only going to visit the shop that is closest to them. Let us go through some observations sequentially.

- End points ($x = 0$ and $x = 100$) are the worst locations to open the shop, therefore any player that is capable of maximizing her utility (i.e. is rational) will not open shop at the end points of the city.
- If red knows that blue is rational, then (according to red) blue will not open shop at either $x = 0$ or $x = 100$. So red will now consider $x = 10$ and $x = 90$ as viable shop locations (for blue).
- If red knows that blue knows that red is rational then blue will not expect red to play $x = 0$ or $x = 100$ as well. This will mean that the new end points of the game are $x = 10$ and $x = 90$.
- If this process continues *ad infinitum*, we say that the players have **common knowledge of rationality**.
- So if the players are **rational** and if they have **common knowledge of rationality** then in the 2 player location game they will open shop at $x = 50$. This is also the **Nash equilibrium** of the game.

Coming back to $\mathcal{G}_2$, the issue of not being able to maximize personal utility due to lack of sufficient knowledge can now be resolved by invoking the assumption of common knowledge of rationality. As it turns out, in *normative game theory* the properties of rationality and common knowledge of rationality are always assumed on the part of the agents. When both these conditions are met, these rational and all-knowing agents act according to the

Nash equilibrium. For $\mathcal{G}_2$, the Nash equilibrium is a set of strategies (x_1^*, x_2^*) such that

$$\begin{aligned} U_1(C(x_1^*, x_2^*)) &\geq U_1(C(x_1, x_2^*)) \forall x_1 \in X_1, \\ U_2(C(x_1^*, x_2^*)) &\geq U_2(C(x_1^*, x_2)) \forall x_2 \in X_2. \end{aligned} \tag{2.5}$$

Essentially, the Nash equilibrium is a collection of optimal actions. The actions are optimal in the sense that for all the players, these actions are the best response to the actions of the rest of the players.

2.4 How we ought to act in a PGG

We are now ready to answer what the Nash equilibrium of PGG is going to be. This is what normative game theorists would say, one ought to act like. Let us start with the single shot PGG that the players play in period t of the actual PGG and consider this from the perspective of player $i \leq N$. In order to find the Nash equilibrium action for player i we will have to find $f_{i,t}^*$ such that

$$G_{i,t}(f_{i,t}^*, \bar{f}_{-i,t}^*) \geq G_{i,t}(f_{i,t}, \bar{f}_{-i,t}^*) \forall f_{i,t} \in [0, \tau]. \tag{2.6}$$

Here $\bar{f}_{-i,t}^*$ represents the vector of contributions from all the players except the one with index i and * is used to represent optimal quantities as usual. Using Eq. 2.2 we can see that irrespective of what $\bar{f}_{-i,t}^*$ is, the optimal action for player i is given by

$$f_{i,t}^* = \begin{cases} 0 & \alpha < 1, \\ \tau & \alpha \geq 1. \end{cases} \tag{2.7}$$

Because the chosen index i is arbitrary, the Nash equilibrium for PGG is complete defection or total contribution depending on the value of α. As this is true for an arbitrary period in a PGG, defection in all periods is the Nash equilibrium in PGG. Therefore if we are rational and our rationality is in common knowledge, then when playing PGG we should either defect or contribute everything. But are we rational? And if so, is our rationality in common knowledge?

2.5 How we actually act in a PGG: experimental evidence

Herrmann, Thöni, and Gächter, 2008 conduct experiments with human players across sixteen different cities all across the world. In the experiments, the players were made to play PGG with $\tau = 20$, $N = 4$ and $\alpha = 0.4$. Fig. 2.3 shows city-averaged contributions as a function of game period for all sixteen cities. As can be seen, human behavior in PGG can be characterized by a **declining trend** of the average contributions as the game progresses. While

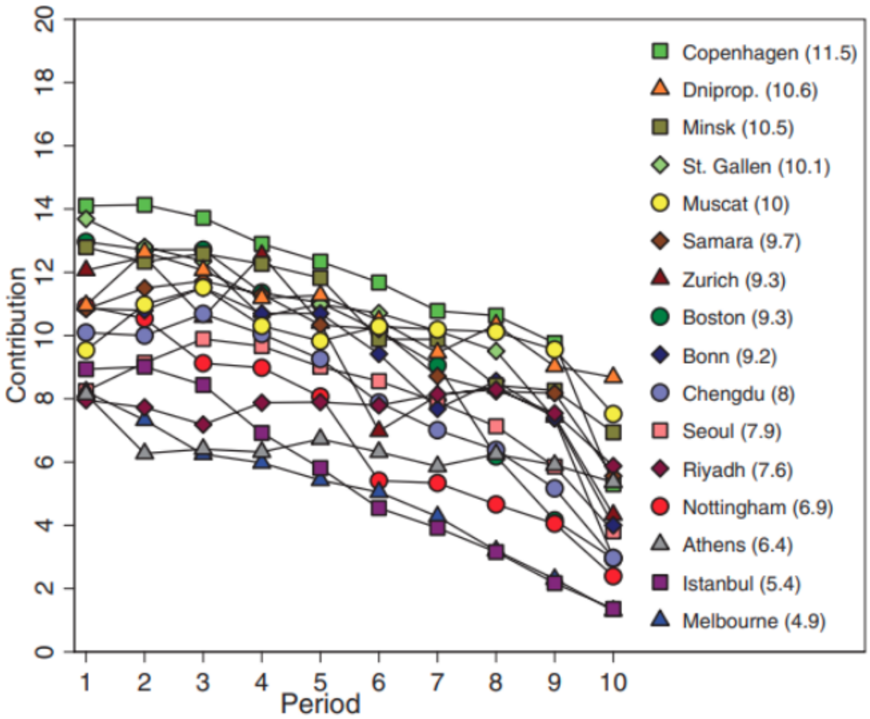


FIGURE 2.3: Group average contributions as a function of game period for various cities. Figure taken directly from Herrmann, Thöni, and Gächter, 2008.

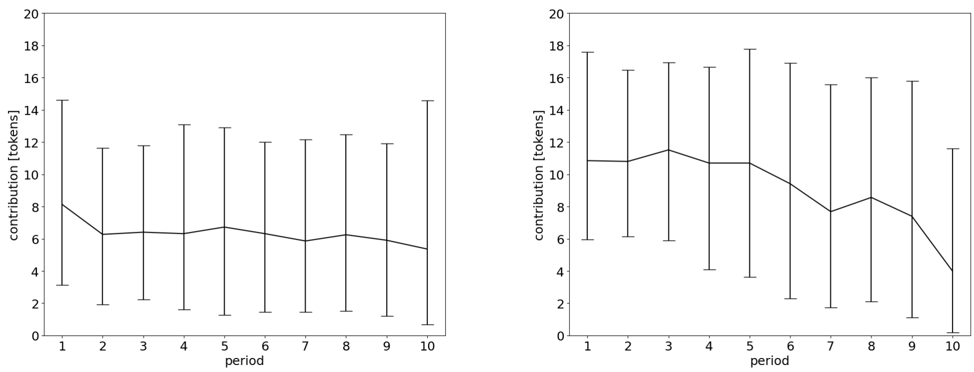


FIGURE 2.4: City average contributions and the corresponding variances represented as error bars for two cities- Athens (left panel) and Bonn (right panel). The plots are generated using the data from Herrmann, Thöni, and Gächter, 2008.

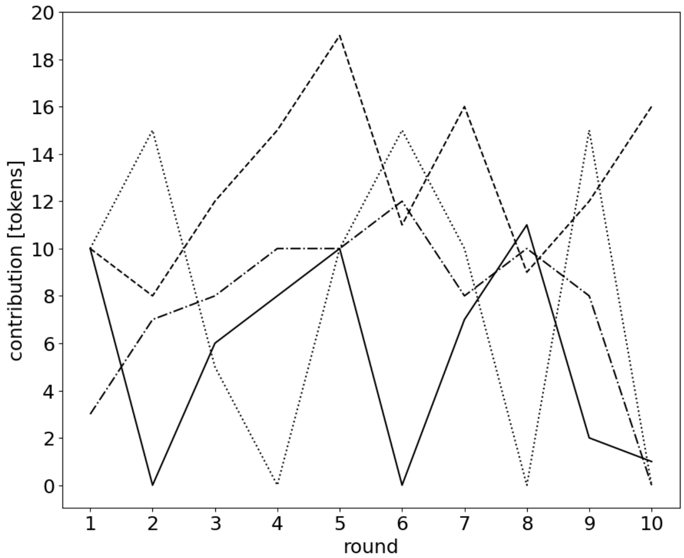


FIGURE 2.5: Trajectory of a single group game played among citizens of Athens. Contributions of all four players are shown, in different styles.

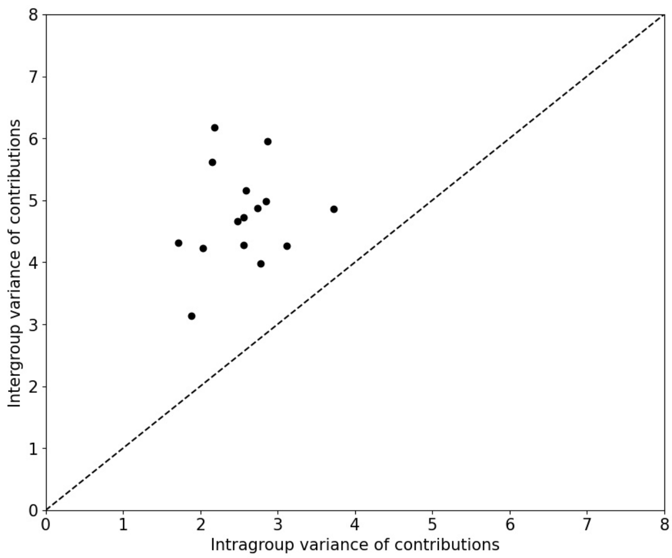


FIGURE 2.6: Intragroup vs. intergroup variance in the average contributions of players. Each data point corresponds to one city. The significant offset of the data above the fist diagonal demonstrates the substantial coupling between players playing in the same group.

it is true that the declining trend in average contribution has been observed in various other studies (for instance see the works of Janssen and Ahn, 2003; Croson, 2007) and is therefore considered characteristic of human behavior in PGG, the period to period variance in the contributions is also very high (as can be seen in Fig. 2.4). This can also be seen by taking a look at the game trajectory of an individual group. Fig. 2.5 demonstrates the very **erratic** nature of human play in PGG.

Beyond the erraticity, human players also showed evidence of **systematic interactions** amongst themselves. The mutual interaction of the players within a group becomes apparent when one considers the variance of the average contributions of the players. As Fig. 2.6 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 completely random fashion as a part of the experiment, this suggests a certain degree of co-operativity, or peer pressure, between players within the same group as to the style of playing, either more parsimoniously or more generously.

It seems obvious that the prescriptions of normative game theory are incapable of reproducing the **declining trend** of average contributions, the **erraticity** of human play (far from the Nash equilibrium) or the **mutual interactions**. Naturally, it leads one to question the prescriptions of normative game theory, namely, rationality and common knowledge of rationality, which are the prerequisites of playing Nash equilibrium. This is the main motivation behind *behavioral game theory* (a good overview can be found in Camerer, 2011). Arising from these motivations is the notion of *bounded rationality* as first discussed by Simon, 1955 and this will be explored in more detail in Chapter 3 as it will be central to our modelling approach. Before we get into our model, let us look at the current state of the art, when it comes to modelling human behavior in PGG.

2.6 Existing models

A broad look over the existing literature on human behavior in PGG shows that most research primarily focuses on explaining the declining trend of contributions in PGG (in some cases also intra-group correlations) and disregarding both the erratic nature of human play and the mutual interactions they demonstrate. The explanations for the declining average contributions that come forward are *conditional cooperation*, *pay off based learning*, and *inequity aversion*.

2.6.1 Conditional cooperation

Conditional cooperators are those players whose contributions to the public good positively correlate with their beliefs about how the rest of the group is going to contribute. For instance, in the work of Dufwenberg and Kirchsteiger, 2004, they consider model agents that hold beliefs about how other players are going to contribute, given a history of the game. The manner in which these beliefs enter the dynamics of the agents is through the utility function. The authors add two components to the utility function- first is the (what they call) "material payoff" which would be given by Eq. 2.2 in our case, and second is the "reciprocity payoff" which encodes the agent's beliefs about other players' actions. As the game progresses the agents update their beliefs about other players via a Bayesian update rule and this changes their preferences and consequently their behavior. A very similar approach is also employed by Falk and Fischbacher, 2006 but only initial beliefs enter the utility function. In these studies, the declining trend is explained via these conditionally cooperating agents realizing the existence of free-riders (players who always defect) and therefore leading to also a decrease in their contributions. Additionally, it has been reported that a small self-serving bias to conditional cooperation is also responsible for the declining trends (Fischbacher, Gächter, and Fehr, 2001).

2.6.2 Inequity aversion

Inequity aversion is the tendency to avert extreme outcomes in rewards in a group. The work of Fehr and Schmidt, 1999 suggests that the declining trends can be attributed to the interaction between fair (inequity averse) and selfish agents. The model of inequity aversion proposed includes a modification to the utility function. A fair agent's utility includes not only the material payoffs but also a negative payoff for either earning more (guilt) or less (envy) material payoff than other players. Therefore the declining trend can yet again be seen as a response of the fair players to the presence of selfish players that always defect. In addition to the declining trend, inequity aversion also holds the potential to explain the correlations in intra-group dynamics.

2.6.3 Payoff based learning

Yet another approach to explain the declining trends in average contributions is the payoff based learning approach as adopted by Burton-Chellew, Nax, and West, 2015a. The authors make the case that the declining trends in average contributions are a consequence of the players learning how to play the game. They hypothesize that the players start with an incomplete understanding of playing the game and as the game progresses, they learn which actions give them more utility (which in the case of the concerned work only includes the material payoff) and just by definition (Eq. 2.2) the players decrease their contributions as the game progresses.

2.7 How do we proceed?

Our work starts with the objective to explain the non-Nash behavior that human players, playing iterated PGG for 10 periods, show as reported in experimental findings by Herrmann, Thöni, and Gächter, 2008. Our first contribution to this line of work is challenging the explanations of payoff based learning and inequity aversion and presenting an alternative (discussed in Chapter 3) based on the notion of bounded rationality (without invoking a learning mechanism) and a slightly different manner of incorporating beliefs about other agents. More specifically we show that just bounded rational planning is not only sufficient to explain observed human behavior in PGG but also does a better job than all existing models. We not only replicate the mean decreasing trends but also the period to period variances (erraticity) and the intra-group correlations.

We do not imply that human players do not learn when playing iterated PGG for 10 periods. All our work does is point out the fact that there are multiple explanations for human behavior in iterated PGG and one should increase the complexity of the explanation only when faced with observations that cannot be explained by the previous one. With this motivation in mind, in our second work (Chapter 4) we introduce the complete model with both learning and planning and bounds on them and focus our attention on what **exclusive** impact does learning have of agent behavior. We go on to show that results from studies on human players playing PGG with varying number of players cannot be replicated by planning alone and learning mechanisms need to be invoked. In addition to this, we provide experimentally testable predictions for human behavior in PGG.

Only after having reasonably justified the use of our microscopic agent based model, we move to collective effects and provide our predictions. We consider our bounded learning and planning agent playing the spatially extended variant of the PGG, also called spatial PGG or SPGG (Chapter 5). Quite surprisingly we find that the model agents in an SPGG show a very simple collective behavior that is mostly characterized by local topological descriptors of the underlying hypergraph.

Chapter 3

Bounded rational agency in PGG

Following the observation from experiments that human behavior is not congruent with the prescriptions of normative game theory, in this chapter we will develop an agent model that relaxes the two assumptions of rationality and common knowledge of rationality (CKR from now on). We will start with building up a model of fully rational agents but without CKR and post that we will introduce bounds on their rationality and complete the agent model. With this updated bounded rational agent, through this chapter, we will explore how it explains the behavior of human players in PGG.

3.1 Notation

We start the discussion by exploring the notation that we employ. In the following, we consider a PGG given by the tuple (N, T, τ, α).

1. $f_{i,t} \in \mathbb{N} :=$ the contribution (or 'action') of the i^{th} agent in period $t \in \{1, ..., T\}$.

2. $\bar{f}_t = (f_{1,t}, ...f_{N,t})$ is the state of the game at the end of turn t. The bar on the top will be used to represent a vector quantity as before.

3. $G_{i,t}(\bar{f}_t) = \alpha \sum_k f_{k,t} + (\tau - f_{i,t})$ is the immediate gain of agent i in period t. The last summand represents the 'gain' from what was not contributed.

4. $\theta_t^T = (\bar{f}_t, \bar{f}_{t+1}, ..., \bar{f}_T)$, is the trajectory of all the players in the game from period t to T.

5. $G_i(\theta_t^T) = \sum_{t'=t}^T G_i(\bar{f}_{t'})$ is the cumulative gain of agent i from period t to T.

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

7. $\bar{f}_{-i,t} :=$ The contribution of all agents except the i^{th} agent in period t (we may identify $\bar{f}_t = (f_{i,t}, \bar{f}_{-i,t})$).

8. $\pi_t^T = (\bar{f}_{-i,t}, \bar{f}_{-i,t+1}, ..., \bar{f}_{-i,T})$, is the trajectory of all players except the i^{th} player, from period t to T.

9. $P(\pi_t^T)$:= The probability of the trajectory of all players except the i^{th} player, from period t to T.

3.2 Rational agency over finite time horizons

We start the discussion by modelling a completely rational agent (and as we will see, without common knowledge of rationality) with rationality meaning the capability to find the best action by maximizing for an objective (utility) as already discussed in Sec. 2.3.1.

Exploiting the inherent symmetry in the game, we will state the model for only one agent with the index $i \in \{1, ...N\}$. From the perspective of the i^{th} agent, we shall often refer to the other agents as the *system* and we call the i^{th} agent the *focal* agent, or just *agent*, for short.

The aim of agent i is to maximize the cumulative gain G_i achieved at the end of all iterations i.e. to say that in each period t of the game, agent i chooses an action $f_{i,t}$ in service of maximizing the cumulative gain G_i. The latter, however, depends not only on the agent's actions but also the other agent's actions (i.e., the system) like $\mathcal{G}_2$ in Sec. 2.3.1 and therefore we need to make some comment about what information agent i has about the system. Recall from Chapter 2 that in PGG the contributions of the players are **anonymous** i.e. there is no communication between the players. It is then safe to assume that agent i can only have a probabilistic model of other agents, and this is how we proceed[1].

In order for the agent to be rational, agent i has to calculate the likelihood of a particular trajectory of the game (i.e. 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 period are dependent only on the state of the game in the previous period and not on the states further back in the game history[2]. The probability of realizing a trajectory $P(\theta_t^T)$ is then given by

$$P(\theta_t^T) = \prod_{t'=t}^{T} P(\bar{f}_{t'}|\bar{f}_{t'-1}) = \prod_{t'=t}^{T} P(f_{it'}, \bar{f}_{-it'}|\bar{f}_{t'-1}). \tag{3.1}$$

From the perspective of agent i, note that there are two kinds of processes at play here. First, the stochastic process describing the system which generates

[1] Notice that we have already relaxed the CKR assumption by doing this.

[2] While this may appear as a gross simplification, we will see that human playing behavior can be accounted for quite well by this assumption.

$\bar{f}_{-i,t}$. Second, there is the choice of the agent, $f_{i,t}$. Since the game is played anonymously, both of these processes can be assumed to be statistically independent[3]. This can be used to write $P(f_{i,t}, \bar{f}_{-i,t}|\bar{f}_{t-1}) = P(f_{i,t}|\bar{f}_{t-1})P(\bar{f}_{-i,t}|\bar{f}_{t-1})$. We refer to $P(\bar{f}_{-k,t}|\bar{f}_{t-1})$ as the *transition function* as it denotes the likelihood of the system responding with $\bar{f}_{-i,t}$ to the previous state $\bar{f}_{t-1}$ of the game. In order to avoid confusion with other Ps, we rename it as $Q(\bar{f}_{-i,t}|\bar{f}_{t-1})$. This now allows us to write the total game trajectory as

$$\begin{aligned} P(\theta_t^T) &= \prod_{t'=t}^{T} P(f_{it'}|\bar{f}_{t'-1})Q(\bar{f}_{-it'}|\bar{f}_{t'-1}) \\ &= P(\pi_t^T)P(f_t^T), \end{aligned} \tag{3.2}$$

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

So far we have broken down the path probability into two parts - one that the agent controls i.e. its *policy* and the other holds the information of how the system transitions into new states. With each of these paths, there is also an associated gain for the agent. If we combine the path probability with the path gain, we can write the optimization problem agent i faces in period t as,

$$V_t[P(f_t^T)] \longrightarrow \max, \tag{3.3}$$

where

$$V_t[P(f_t^T)] = \sum_{\theta_t^T} P(\theta_t^T)G_i(\theta_t^T), \tag{3.4}$$

is the expected cumulative gain (also called the value functional). The maximum value of V_t will henceforth be called V_t^*.

Noticing that $G_i(\theta_t^T)$ can be written as $G_i(\theta_t^T) = G_{i,t}(\bar{f}_t) + G_i(\theta_{t+1}^T)$, we can break down the summation in Eq. 3.4 into two parts, the first of which is the immediate expected gain, the second is the future expected gain.

$$V_t[P(f_t^T)] = \sum_{\theta_t^T} \Big[P(\bar{f}_t|\bar{f}_{t-1})P(\theta_{t+1}^T)G_{i,t}(\bar{f}_t) + P(\bar{f}_t|\bar{f}_{t-1})P(\theta_{t+1}^T)G_i(\theta_{t+1}^T) \Big] \tag{3.5}$$

Now we can sum over θ_{t+1}^T. In the first summand, the path probability is going integrate to 1 and the second summand will lead to the value function at time $t+1$. Finally, we get a recursive equation of the form

$$V_t[P(f_t^T)] = \sum_{\bar{f}_t} P(\bar{f}_t|\bar{f}_{t-1}) \Big[G_{i,t}(\bar{f}_t) + V_{t+1}[P(f_{t+1}^T)] \Big]. \tag{3.6}$$

[3]Notice that the independence is of the variables at the same period t. Agent i could still influence future system states through its actions.

Now, we can make use of Eq. 3.2 and write the above even more explicitly as

$$V_t = \sum_{\bar{f}_t} P(f_{i,t}|\bar{f}_{t-1})\Big[Q(\bar{f}_{-i,t}|\bar{f}_{t-1})G_{i,t}(\bar{f}_t) + Q(\bar{f}_{-i,t}|\bar{f}_{t-1})V_{t+1}\Big]. \tag{3.7}$$

Finally the optimization problem that agent i faces in period t can be written as

$$V_t^* = \max_{f_t^T} \sum_{\bar{f}_t} P(f_{i,t}|\bar{f}_{t-1})\Big[Q(\bar{f}_{-i,t}|\bar{f}_{t-1})G_{i,t}(\bar{f}_t) + Q(\bar{f}_{-i,t}|\bar{f}_{t-1})V_{t+1}\Big]. \tag{3.8}$$

This equation is known in literature as the *Bellman equation* as developed by Richard Bellman, 1952, 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. This is written as

$$V_t^* = \max_{f_t^T} \sum_{\bar{f}_t} P(f_{i,t}|\bar{f}_{t-1})\Big[Q(\bar{f}_{-i,t}|\bar{f}_{t-1})G_{i,t}(\bar{f}_t) + \gamma Q(\bar{f}_{-i,t}|\bar{f}_{t-1})V_{t+1}\Big], \tag{3.9}$$

where $\gamma = 0$ would indicate an extremely myopic agent as it is only trying to maximize the immediate gains and $\gamma = 1$ would represent an extremely far-sighted agent as it values all gains, no matter how far away in the future, equally.

This concludes the description of our agent i which is rational as it finds the optimal policy $P^*(f_t^T)$ by solving Eq. 3.9, but there is no CKR as the agent's model of other agents is a probabilistic model given by the Markovian transition function Q. In the next section, we will explore one particular manner of bounding the rationality of the agent. More specifically we will briefly go over some information thermodynamics to evaluate the cost of performing a computation.

3.3 The cost of computing- Szilard engine

Szilard, 1929 created a thought experiment through which he described the thermodynamic consequences of possessing information. The main idea is rather simple. Consider a single particle in a box as in Fig. 3.1 that is currently in equilibrium with a heat bath at temperature T. If one were to ask you what the probability of the particle was for being in the left half of the box (without you observing it), you would have said that it was 0.5. Let us say you now make an observation and find that the particle is indeed in the left half of the box with unity probability. Now you can do the following.

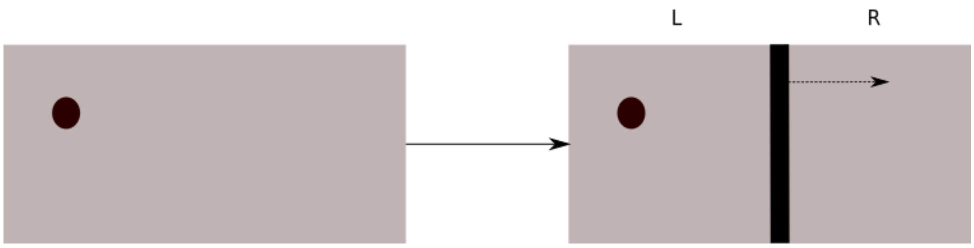


FIGURE 3.1: A particle in a box and a massless partition being introduced into it, dividing it into two halves- left (L) and right (R).

1. You are now allowed to introduce a massless partition in the middle of the box (thick black line in the figure) and thereby allowing you to extract work from the system[4]. The amount of work you would be able to extract by doing this would be

$$W = \int_{V/2}^{V} \frac{k_B T}{V} dV = k_B T \ln 2 \tag{3.10}$$

2. Once the partition reaches the right end of the box, you lift it out and wait for the particle to again equilibriate with the heat bath.

3. Once it is in equilibrium, repeat step 1.

Essentially what you have done here, is that you have converted 1 bit of information (L and R can be represented as a binary variable) into $k_B T \ln 2$ of work. Does this mean that we can keep freely extracting work out of a single heat bath at constant temperature? Well, no. While it is true, that the entropy of the physical system decreases when we possess information about the system, the global entropy (system + observer) doesn't. Loosely speaking we are ignoring the work done by the observer in figuring out where the particle was in the first place! That is, the amount of work the observer has to do to update its knowledge about the system through observation is at least as much as the amount of energy that can be extracted out of the system as a consequence of the information.

In a more general case, we can say the following. The observer starts with an *a priori* knowledge of the system state represented by the distribution P_0 where $P_0(L) = P_0(R) = 0.5$. This would be the guess of the observer without performing any observation at all. Let us now say that the observer has made a (potentially flawed) observation and the new knowledge state can be represented by P where $P_0(L) = p$ and $P_0(R) = 1 - p$. The expected

[4] The particle is going to hit the massless partition and push it in the direction of the dashed arrow. The partition could as well have been tied to a mass which could have been lifted using the pressure from the particle.

work this observer can extract out of a Szilard engine can be written as

$$\langle W \rangle = k_B T[p \log(\frac{p}{1/2}) + (1-p)\log(\frac{1-p}{1/2})] = k_B T D_{KL}(P||P_0). \tag{3.11}$$

3.4 The bounded planning agent

You may ask, what does the previous section have to do with agents playing a game? The connection is rather simple. One may identify the Szilard engine as a game where the expected work extracted can be identified as the expected utility of the game. The agent in the case of a PGG then has an *a priori* policy of playing the game called P_0 and if the agent doesn't engage in any computation for instance searching through the box to find the particle or searching through the policy space to find a better policy, then the agent is likely to act in the game according to P_0. On the other hand, a more computationally capable agent will expend some effort through deliberation and find a better policy P as a consequence. What we do know is that for the agent to deviate from its basal tendency P_0 to a new policy P it has to expend effort that is proportional to the KL divergence of the two policies (for more details the reader is encouraged to go through Ortega and Braun, 2011).

Making use of the above analogy we allocate a computational budget K that represents the maximal deviation from its basal tendency P_0 that the agent is allowed in search of a better policy, to agent i. Consequently, we constrain (hence the name "bounded") the optimization problem in Eq. 3.9 by the computational budget of the agent and write the optimization problem faced by the agent in period t as,

$$\begin{aligned} V_t^* = \max_{f_t^T} \sum_{\bar{f}_t} P(f_{i,t}|\bar{f}_{t-1}) \Big[Q(\bar{f}_{-i,t}|\bar{f}_{t-1}) G_{i,t}(\bar{f}_t) + \gamma Q(\bar{f}_{-i,t}|\bar{f}_{t-1}) V_{t+1} \Big], \\ \text{with} \quad D_{KL}(P(f_{i,t}|\bar{f}_{t-1})||P_0(f_{i,t}|\bar{f}_{t-1})) \leq K. \end{aligned} \tag{3.12}$$

Here $D_{KL}(.||.)$ is the Kullback-Leibler divergence and $P_0(f_{k,t}|\bar{f}_{t-1})$ is the prior policy for this period. The constrained optimization problem can be written as an unconstrained one by introducing a Lagrange parameter to write the optimization problem as,

$$\begin{aligned} V_t^* = \max_{f_t^T} \sum_{\bar{f}_t} P(f_{i,t}|\bar{f}_{t-1}) \Big[Q(\bar{f}_{-i,t}|\bar{f}_{t-1}) G_{i,t}(\bar{f}_t) - \frac{1}{\beta} \log \frac{P(f_{i,t}|\bar{f}_{t-1})}{P_0(f_{i,t}|\bar{f}_{t-1})} \\ + \gamma Q(\bar{f}_{-i,t}|\bar{f}_{t-1}) V_{t+1} \Big], \end{aligned} \tag{3.13}$$

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 Karush–Kuhn–Tucker (KKT) condition (see Kuhn, 1982), i.e.,

$$\frac{1}{\beta}(D_{KL}(P^*(f_{k,t}|\bar{f}_{t-1})||P_0(f_{k,t}|\bar{f}_{t-1})) - K) = 0. \tag{3.14}$$

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. 3.13[5].

This concludes the description of the bounded rational agent. The bounded planning agent can be seen as defined by the quadruple $(Q(\bar{f}_{-k,t}|\bar{f}_{t-1}), P_0(f_{k,t}|\bar{f}_{t-1}), K, \gamma)$[6]. The transition model encapsulates the agent's internal model of the system (or other players), the prior policy 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 period to another, in this work we assume the computational constraint K to be the same for all the periods, considering it as a trait of the agent.

We now turn to solving Eq. 3.13. The value function at period 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 Richard Bellman's, through what is called *backward induction*. Instead of starting from period t, we can start from the last period and obtain a series of nested functions which can then iteratively lead to turn t. Although simple in principle, an analytical solution can only be obtained in a few special cases. The numerical solution to this problem, however, is straightforward.

[5]One may argue that the bound on optimization in the form of relative entropy in Eq. 3.13 is arbitrary for modelling human agents as it follows from the physical system of a particle in a box. 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 that physical system 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.

[6]By describing the agent as such, we have deviated from the bounded rationality framework of Ortega and Braun, 2011. In that work, the authors require the game to specify the Lagrange parameter β for each period of the game, whereas, in our model, β is determined from K after performing the constrained optimization.

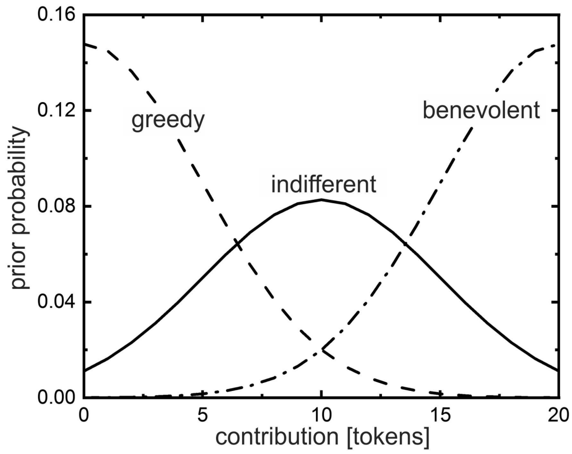


FIGURE 3.2: Basal tendencies of the agents $P_0(f_{k,t})$ as given by truncated Gaussian prior distributions with variable m.

3.5 Numerical simulations

3.5.1 Model assumptions

First, we make the simplifying assumption that the prior of the agent $P_0(f_{it}|\bar{f}_{t-1})$, is independent of the previous state, i.e. we replace $P_0(f_{i,t}|\bar{f}_{t-1})$ with $P_0(f_{i,t})$. This simplification essentially implies that the agents don't have a very complex prior strategy that takes into account specific game states and the agent's response to each of them. Rather the agent has an apriori preference toward some specific actions and it doesn't change as the game evolves.

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

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

where we set $\tau = 20$ in order to relate to the data we intend to compare our simulations with (Herrmann, Thöni, and Gächter, 2008). $\mathcal{N}$ 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. 3.2. 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 $\bar{f}_{-i,t}$ variables and separately across the $\bar{f}_{-i,t-1}$

variables in $Q(\bar{f}_{-i,t}|\bar{f}_{t-1})$. Additionally, we only need to consider the distribution of the means of $\bar{f}_{-i,t}$'s and $\bar{f}_{-i,t-1}$'s as these are the only relevant quantities in the game from the perspective of agent i (see Eq. 2.2). Therefore, from the definition of the agent we replace the transition function $Q(\bar{f}_{-i,t}|\bar{f}_{t-1})$ with $Q(\mu_t|\mu_{t-1}, f_{i,t-1})$. Where $\mu_t = \frac{\sum_{k\neq i} f_{k,t}}{N-1}$.

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

$$\mu_t^{\text{peak}} = \begin{cases} \mu'_{t-1} + \xi_+|\mu'_{t-1} - f_{i,t-1}|, & \mu'_{t-1} - f_{i,t-1} < 0 \\ \mu'_{t-1} - \xi_-|\mu'_{t-1} - f_{i,t-1}|, & \mu'_{t-1} - f_{i,t-1} > 0 \end{cases} \tag{3.15}$$

and some fixed variance ($\sigma_{\text{trans}} = 3$) [7]. Here μ_t^{peak} is the most likely value of μ_t, μ'_{t-1} is the observed value of μ_{t-1} and $\xi_\pm$ 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 as the game is being played anonymously and the agent can only influence the game trajectory via its contributions. Therefore, 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 $\xi_\pm$.

3.5.2 The agent model and relation to known human preferences

With the above assumptions in mind the agent model described at the end of Sec. 3.4 gets modified and the simplified agent is completely described by its tuple $(\xi_\pm, m, K, \gamma)$, which is considered constant over the game it plays. As a side note, it is instructive to compare these parameters to known human preference parameters used, e.g., in the Global Preference Survey (GPS Falk et al., 2018). 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 (Arrow, 1972; Boyd et al., 2003; Dohmen et al., 2008; Chen et al., 2013), 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

[7] 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 are neither entirely uninformed about their contemporaries nor have a very precise model of them.

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 period). Notions like risk taking, altruism, and trust will certainly reflect in the value m assigned to a player, and to some extent also affect $\xi_{\pm}$. 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 (Falk et al., 2018).

3.5.3 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 (i.e. Sec. 3.2). 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 $\mu_{t-1} := \frac{1}{N-1}\sum_{i\neq 1} f_{i,t-1}$ 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.

Algorithm 1 Backward Induction

```
function POLICY(M, t_max)
    t = t_max
    E = 0
    V = 0
    while t > 0 do
        for 0 ≤ f_{1,t-1} ≤ τ do
            for 0 ≤ μ_{t-1} ≤ do
                for 0 ≤ a ≤ τ do
                    E[a] = GAIN(Q, a, μ_{t-1}, f_{1,t-1}, V)
                end for
                j = arg max E
                policy[t, f_{1,t-1}, μ_{t-1}] = (δ_{ij})_{i∈{0,...,(N-1)τ}}
                V = policy[t, f_{1,t-1}, μ_{t-1}] · E
            end for
        end for
        t = t - 1
    end while
    RETURN policy
end function
```

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

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

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. For the algorithm-wise, 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, $P^*(f_{i,t}) = c \cdot P_0(f_{i,t}) \cdot \exp(\beta E[f_{i,t}])$, $f_{i,t} \in \{0, \ldots, \tau\}$ 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 $f_{i,t-1}$ in order to evaluate the expected cumulative gain. This leads to a problem for the first period, as no history yet exists. Therefore in our implementation, we manually initialize the group with an appropriate state $\bar{f}_0$[8]. 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 (Lam, Pitrou, and Seibert, 2015) and the numerical library Numpy (Harris et al., 2020).

3.5.4 Solution space of the model

Now we will demonstrate the kind of behavior our model agents can exhibit, to the end of developing some intuition about the agent parameters. Since we aim to develop agents whose playing behaviour is similar to that of human players in ref. Herrmann, Thöni, and Gächter, 2008, we set $(N, T, \alpha, \tau) = (4, 10, 0.4, 20)$, as was chosen in that study. Due to the high dimensionality of the parameter space, we show game trajectories for 4-player groups with only a few configurations given by $(\xi_\pm, m, K, \gamma)$.

[8] As the contribution in the first period has to depend on the pre-play awareness of the human player, we can just treat it as an initial condition for the purposes of our model.

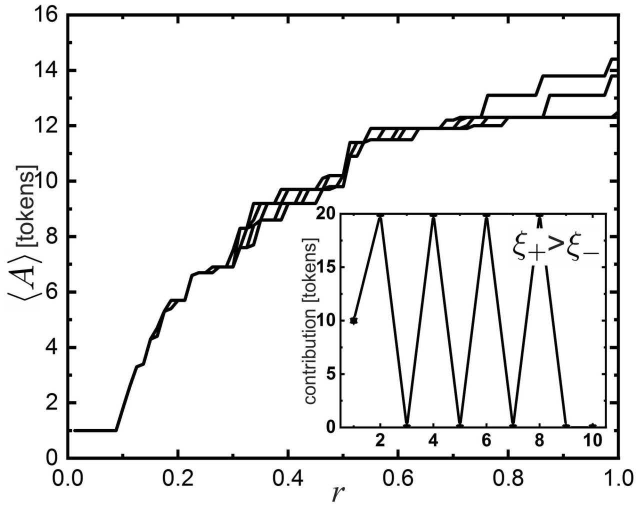


FIGURE 3.3: group average contribution, $\langle A \rangle$, as a function of $r = \sqrt{\xi_-^2 + \xi_+^2}$ for four different values of the polar angle, $\theta \in \{0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}\}$. The data collapse shows that the polar angle (hence the ratio ξ_+/ξ_-) is not relevant 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_-.

Choice of $\xi_\pm$

In order to understand the significance of $\xi_\pm$, we consider fully rational agents for the sake of simplicity, with $(m, K, \gamma) = (10, \infty, 1)$, and consider the average (over periods and agents) contribution of the group, $\langle A \rangle$[9], while the two components of $\overline{\xi}$ are varied. Writing $\xi_+ = r\cos\theta$ and $\xi_- = r\sin\theta$, it can be seen from Fig. 3.3 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 non-generic 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 behaviour, hence we should avoid small r when choosing $\xi_\pm$ 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. 3.3). These oscillations (which do not have a visible effect on the average $\langle A \rangle$) occur because agents believe, judging

[9]Here $\langle \cdot \rangle$ is used to denote an ensemble average over multiple simulation runs.

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 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 $\xi_\pm$. 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 $\bar{f}_0 = (10, 10, 10, 10)$, except when fitting experimental data.

Impact of K

It should be clear from the previous section that fully rational agents (i.e. $K = \infty$) act independently of their priors. Fig. 3.4 (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 $\xi_\pm$, $r = 0.26$ is sufficiently large to prevent players from "decoupling". Notice also that rational agents always contribute zero tokens in the last period. This corresponds to the Nash equilibrium in the one-period PGG, as the sole purpose of contributing was to potentially encourage others to contribute in future periods, which is expressed by the $\xi_\pm$. In the inset of Fig. 3.4 (a) we also see the impact of γ on rational agents. Here we have 3 agents (open circles) 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 3 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 Fig. 3.4 (b) and (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. 3.4 (c) we see the effect of a complete lack of any computational ability. The agent just acts according to its prior, unaffected by the play of

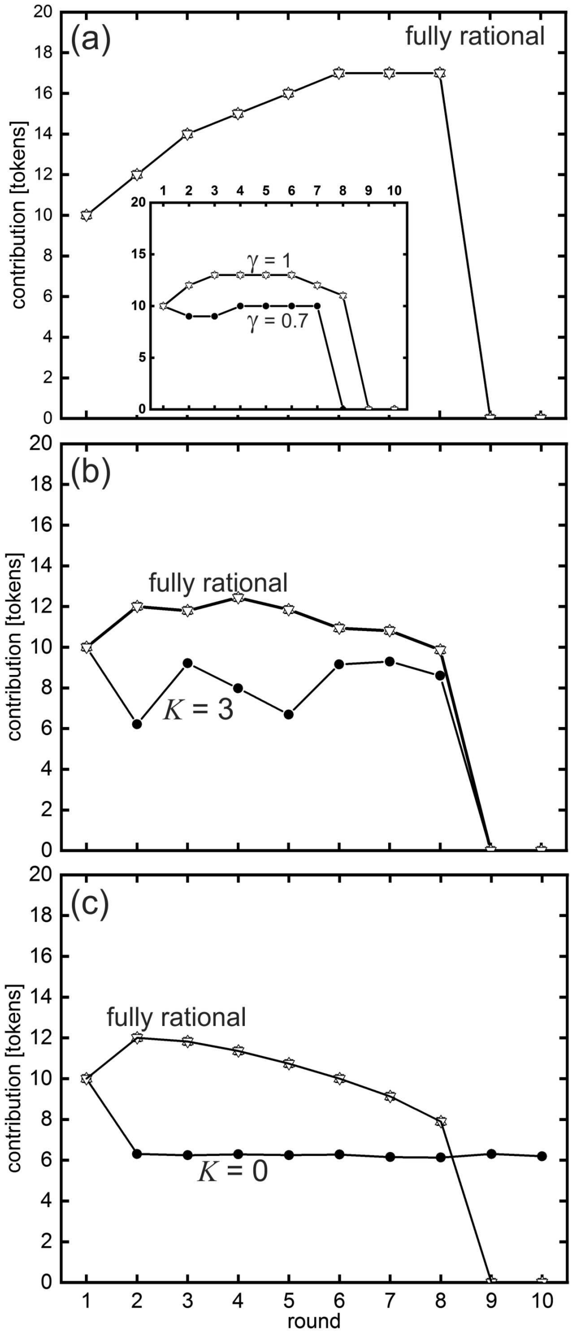


FIGURE 3.4: Full game trajectories of a group of four fully rational agents. In (a) all agents are $(m, K, \gamma) = (5, \infty, 1)$. For the inset, one agent has different $\gamma = 0.7$ (solid circles). (b) and (c) show ensemble average trajectories over 10000 simulation runs with three agents (open circles) as before and one different agent (solid circles) with the same m and γ, but K reduced to 3 and 0, respectively.

other agents, as it 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.

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 behaviour, 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. 3.5, m has a monotonous impact on the average contribution of all the agents. Notice that the agent with $m = 0$ plays like rational agents in the last periods and the agent with $m = 20$ plays like the rational agents in the intermediate periods. 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 period, as can be seen in Fig. 3.5 (b) and (c).

Mutual coupling of agents

Referring to the correlations displayed in Fig. 2.6, 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 behaviour 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. 3.7. 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.

Groups of identical agents

So far we have focused on the impact of parameters on the behaviour of individual agents. It is similarly instructive to study the behaviour 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. 3.6. 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

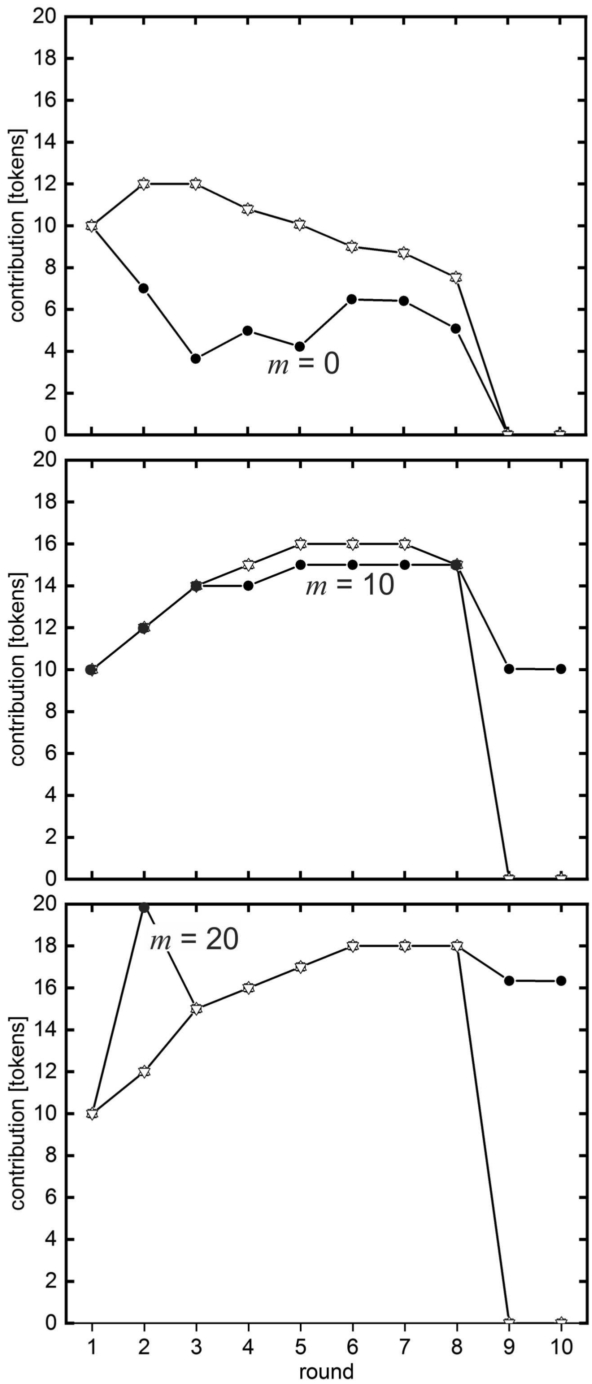


FIGURE 3.5: Full game ensemble average trajectories of group of four agents. There are three identical agents (open circles) with $(m, K, \gamma) = (5, \infty, 1)$ and one different agent (solid circle) with (a) $(m, K, \gamma) = (0, 3, 1)$, (b) $(m, K, \gamma) = (10, 3, 1)$ and (c)$(m, K, \gamma) = (20, 3, 1)$.

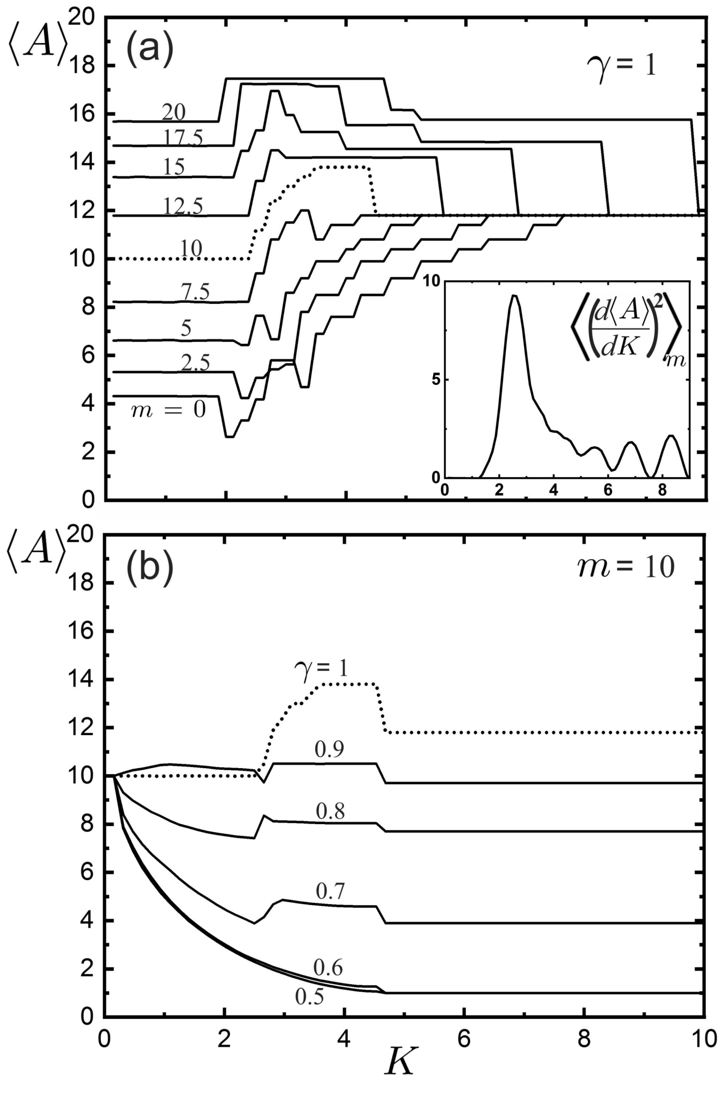


FIGURE 3.6: The dependence of the average contribution $\langle A \rangle$ on agent parameters in a group of identical agents. The ensemble average was determined over 10000 runs. In each panel, the dotted curve is for $(m, \gamma) = (10, 1)$.

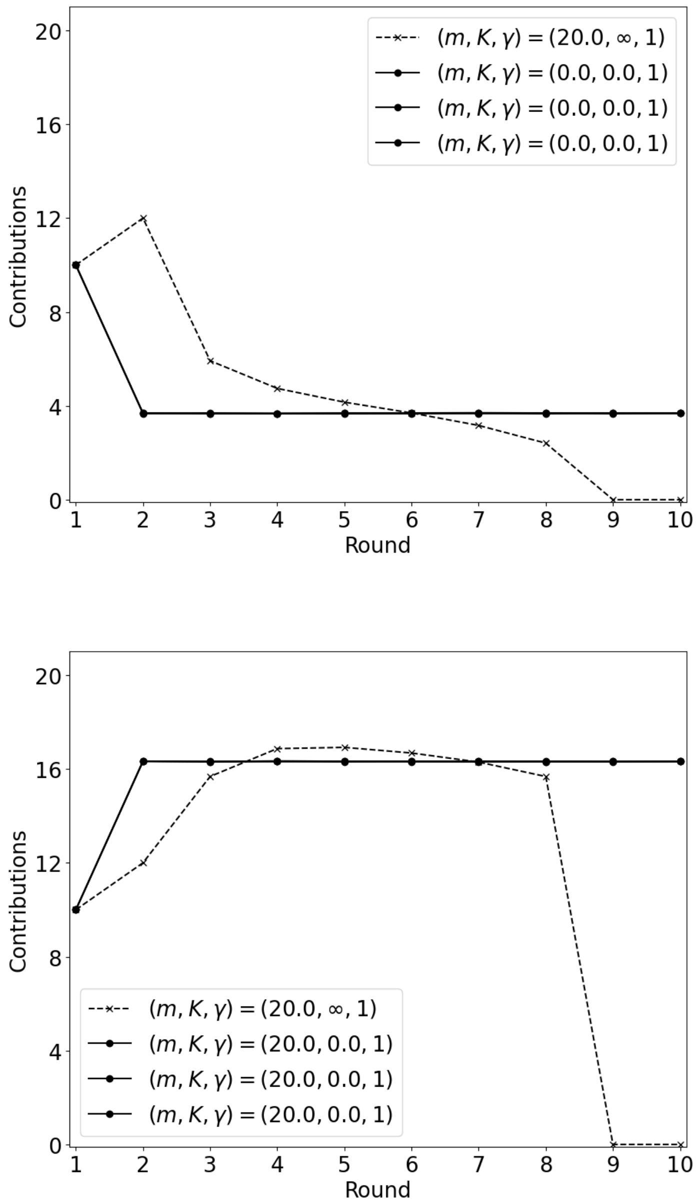


FIGURE 3.7: Demonstration of coupling among agents within one group. A benevolent rational agent (dashed curve) with $(m, K, \gamma) = (20, \infty, 1)$ is made to play in two different systems, with either $m = 0$ (greedy, top, solid curves) or $m = 20$ (benevolent, bottom, solid curves). The system agents are all chosen with $K = 0$ in order to prevent repercussions of the agent under consideration (dashed) onto the system.

the right margin of Fig. 3.6 (a)). Also note that at very small K, the contribution is entirely governed by the priors. This is not the case with γ. At high values of K, γ has a large impact on $\langle A \rangle$ (Fig. 3.6 (b)), while at low values of K, γ has little or no impact on the $\langle A \rangle$.

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. 3.6 (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 (Ortega and Braun, 2011). 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.

3.6 Fitting to experimental data

We are finally ready to fit the simulated game trajectories to the data obtained from experiments done by Herrmann, Thöni, and Gächter, 2008. 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 $(\xi^i_\pm, K_i, m_i, \gamma_i)$, with $i \in 1, 2, 3, 4$ (16 parameters), to 40 data points (4 player contributions over 10 periods). 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 of all, we fix some of the parameter values by adopting all assumptions from section 3.5.1. Further suggestions emerge when fitting the eight parameters (K_i, m_i), with $i \in \{1, 2, 3, 4\}$, to experimental group trajectories. This yields a two-dimensional histogram over the (K, m)-plane, which is shown for two cities in Fig. 3.8, 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. 3.6, 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 $\bar{f}_0$ to be the same as that of the actual players, therefore effectively fitting only 9 periods. As a result, to each group from the experimental data we fit the quintuple $(\gamma, m_1, m_2, m_3, m_4)$.

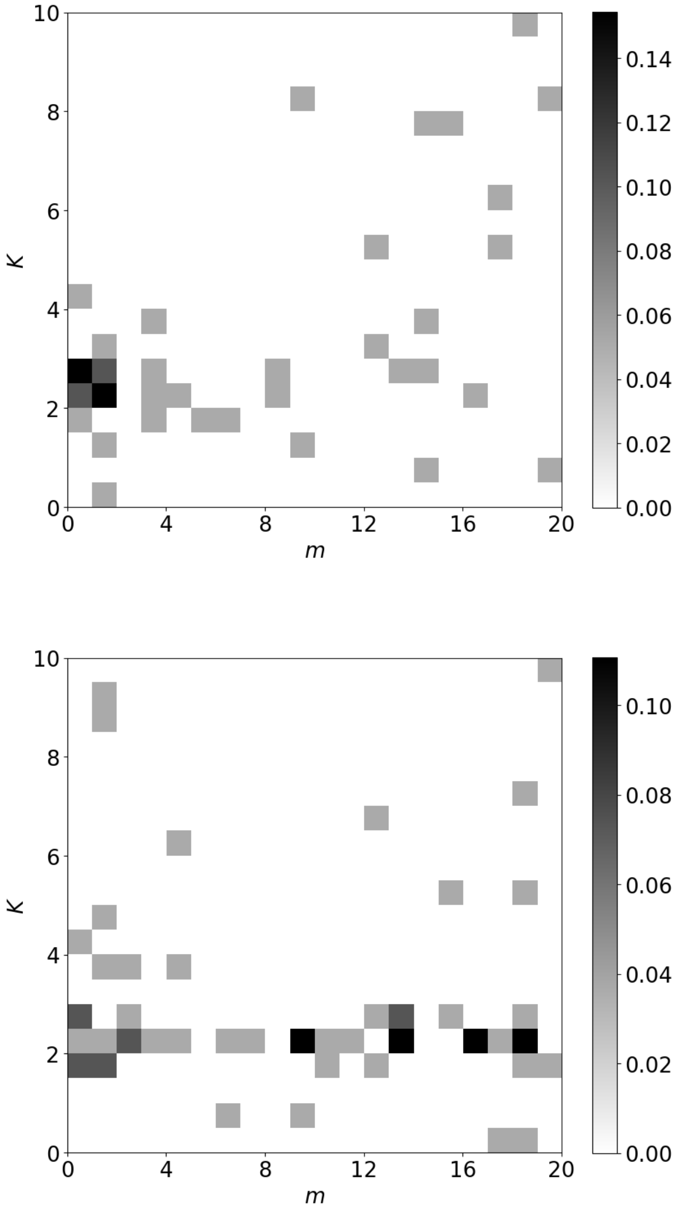


FIGURE 3.8: Joint distributions of K and m for Melbourne (top) and Boston (bottom). There seems to be a clear preference for $K \approx 2.5$.

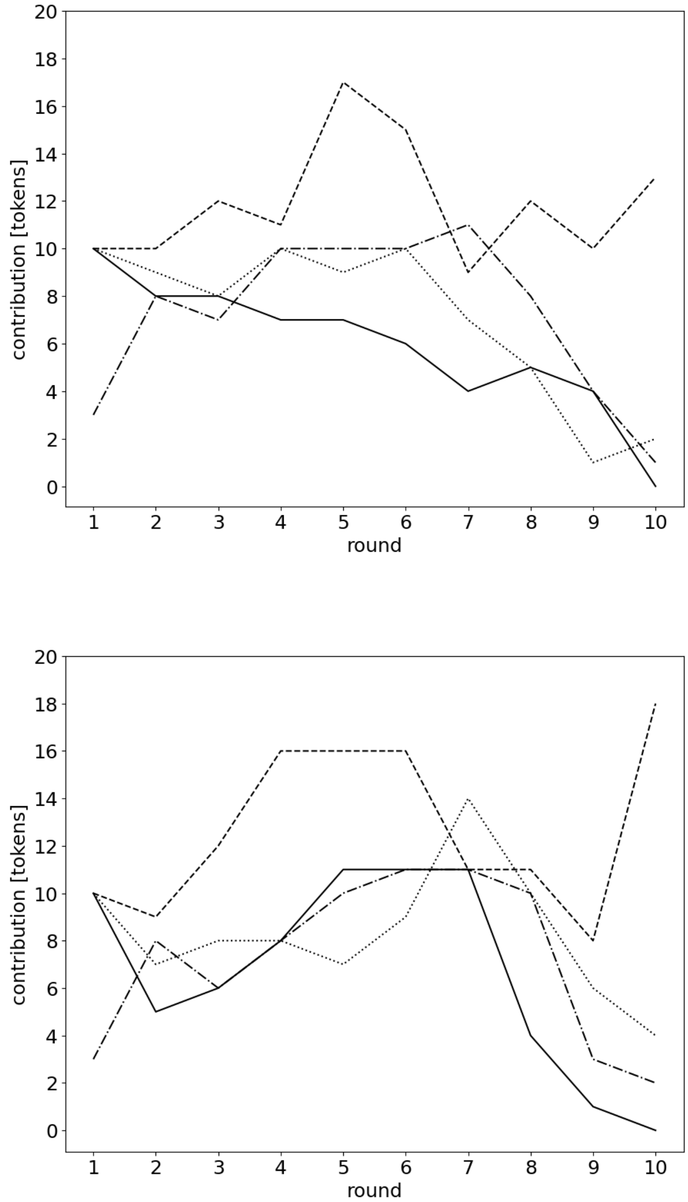


FIGURE 3.9: Two trajectories simulated under identical conditions, with parameters obtained from fitting the group trajectory displayed in Fig. 2.5 (data from an Athens group). The same group of agents yields a different trajectory each time the simulation is run, due to the inherent randomness of the model. The fitting procedure minimizes the deviations of the average contribution at each period, as well as of the variance of these contributions from the observed variance of player contributions.

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* Virtanen et al., 2020. The optimization problem for the fitting procedure can be written as

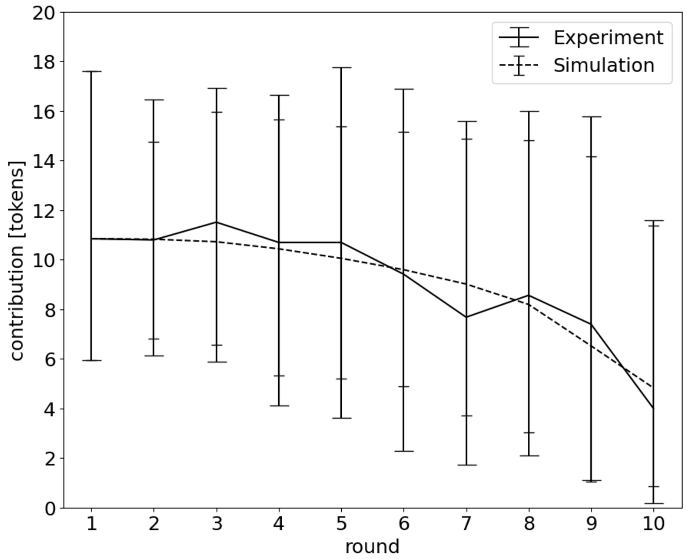
$$\min_{(\gamma,m_1,\ldots,m_4)} \sum_{t=2}^{10} \sum_{k=1}^{4} \left(f_{k,t}^{\mathrm{obs}} - \langle f_{k,t}^{\mathrm{sim}} \rangle_{(\gamma,m_1,\ldots,m_4)} \right)^2, \tag{3.16}$$

where the $f_{k,t}^{\mathrm{obs}}$ is the observed (from data) contribution of the k'th agent in period t and $f_{k,t}^{\mathrm{sim}}$ is the corresponding contribution from the simulated agent. Furthermore, $\langle \cdot \rangle_{(\gamma,m_1,\ldots,m_4)}$ denotes the average over multiple simulation runs of the group defined by the parameters $(\gamma, m_1, \ldots, m_4)$.

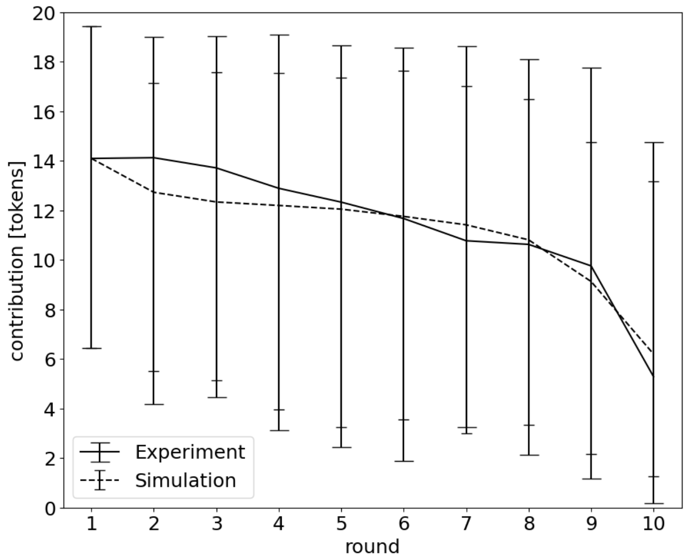
The resulting parameter set found by fitting to a single group can be used to generate individual game trajectories of the so obtained group of agents, for comparison with the experimental trajectory. Fig. 3.9 shows two examples from the agent group obtained by fitting to the trajectory from an Athens group, which we displayed in Fig. 2.5.

3.7 Discussion

In Fig. 3.10 we compare the simulated city averages with the actual city averages for Bonn and Copenhagen as two examples (for comparisons in other cities, please refer to the publication in the appendix A). The error bars represent the city-wide standard deviation of contributions in that period 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 $K = 2.5$ 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 is 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. We show the distributions obtained for m and γ in appendix A for each city.



(A) Bonn



(B) Copenhagen

FIGURE 3.10: Actual and simulated city-averaged contributions. The error bars indicate the variance of contributions.

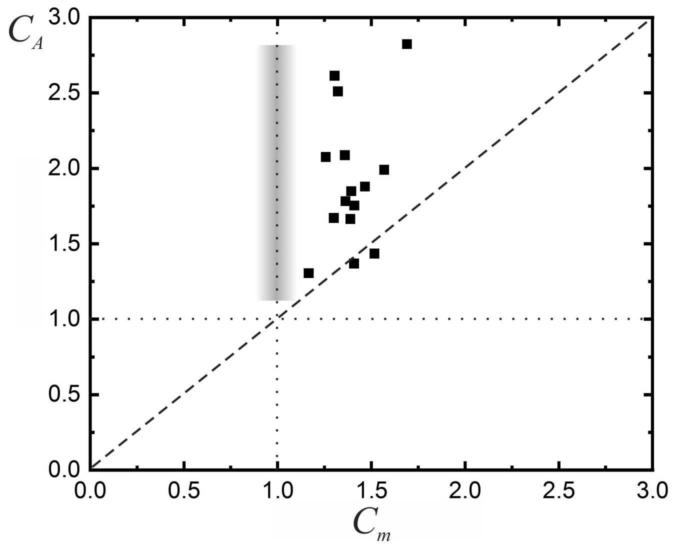


FIGURE 3.11: Coupling in A vs. coupling in m as obtained from the fitted simulations. The grey bar represents the data from Fig. ??b.

3.7.1 Game-induced inter-agent coupling

Let us now turn to the coupling of players within a group, as observed from the data in Fig. 2.6. We can quantify this coupling by taking the ratio of intergroup variance (ordinate in Fig. 2.6) to intragroup variance (abscissa in Fig. 2.6). Specifically, for $\langle A \rangle$ and m, we write

$$C_m = \frac{var\{m \mid \text{intergroup}\}}{\langle var\{m_i \mid \text{intragroup}\}\rangle_l} \tag{3.17}$$

and

$$C_A = \frac{var\{\langle A \rangle \mid \text{intergroup}\}}{\langle var\{\langle A_i \rangle \mid \text{intragroup}\}\rangle_l}, \tag{3.18}$$

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. 3.11, where each data point corresponds to one city. The grey bar represents the data from Fig. 2.6, 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 grey bar.

The values obtained for C_A from the fitted simulations are found in the same range as that indicated by the grey 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's 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. 3.7.

3.7.2 Comparison to other approaches

While payoff-based learning has been suggested as *the* explanation of the commonly declining contributions in a PGG (For instance the works of Burton-Chellew, Nax, and West, 2015a; Burton-Chellew and West, 2021), we have shown that bounded rational foresight, as reflected by the bounded rational model agent, can perfectly well serve as an explanation of this phenomenon. Moreover, it is capable to account for the substantial in-game variance of contributions that is observed in real games.

Note that Fig. 3.10 shows that both the average contribution and the slope of its decline vary substantially among the 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. 3.10.

Additionally, we note that the conditional cooperator model (Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) which modelled agents by modifying the utility function to also include the agent's model of other agents, is similar to our model in spirit but not exactly in how conditional cooperation is implemented. The main difference is that, in our approach, we do not modify the utility function (only consider the material payoffs), and the mental model of the agents is incorporated as a transition function instead. Finally, we also see that the need for inequity aversion based explanations (Fehr and Schmidt, 1999) is eliminated to explain the intra-group correlations, as in our model they are mediated by the transition function the agent holds.

3.8 Conclusions

In this chapter we have developed agents by deviating from the prescriptions of normative game theory. Specifically, we have created a bounded rational agent which is described by the tuple $(\xi_{\pm}, m, K, \gamma)$ which is considered constant over the game and shown that they replicate human behavior in PGG (for games of length 10 periods) quite well (cf. Fig. 3.10). Our results also demonstrate how the bounded rational foresight explains the three main observables highlighted in Chapter 2, namely, declining trends, erraticity (cf. Fig. 3.10) and intra-group correlations (cf. Sec. 3.7.1).

One may argue that the constancy of $\xi_{\pm}$ (the model of other agents the focal agent has) over the game is too strict a restriction and is probably not representative of humans who probably also learn and update their models as the game progresses. While it is true that human players indeed learn, the objective of this chapter was to demonstrate that learning is not *necessary* to replicate human behavior. Our results then indicate that the space of models that generates the behavior is much larger than the space in which the behavior exists (game trajectories). Therefore any attempts to make realistic models of human agency must increase the complexity of the model only incrementally.

This is exactly how we proceed. In the next chapter, we will, in addition to the bounded planning mechanism, also introduce a learning mechanism and therefore explore the *exclusive* effects of learning on human behavior in PGG.

Chapter 4

Exploring the exclusive effects of learning

In the previous chapter, we explored how bounded rational agency was sufficient to explain human behavior in PGG given by $(N, T, \tau, \alpha) = (4, 10, 20, 0.4)$. 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 chapter, we wish to observe the exclusive impact of learning on bounded rational agents, which couldn't have been generated by bounded rational agency alone. Like before we will compare the behavior of agents to known experimental results and also provide experimentally testable findings where possible.

4.1 Motivating the learning mechanism

As briefly discussed in Chapter 1, 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 (AIXI Hutter, 2007). 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 (modelling other players in the game). This knowledge of the dynamics is then used to simulate future trajectories[1] 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. Employing the notion from Sec. 1.2.1, learning is a mapping from observed behavior to mental models given by $f_l : O \to M$, and planning is a mapping from mental models to actions

[1]This is essentially equivalent to the path integral formulation of reward maximization as introduced in Chapter 3.

given by $f_p : M \to A$. Here O is the observation space, A the action space, and M the space of all possible models[2].

4.1.1 Relation to other notions of learning

Social learning

The readers should note that the notion of learning put forward is distinct from "social learning" as is common in evolutionary game theory (for instance refer to the works of Hofbauer and Sigmund, 1998; Sigmund et al., 2010), 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. This is in contrast with our approach, as (in our approach) the agents learn *of* the other players' behaviors by creating a model of them.

Payoff based learning

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), for example, the notion of payoff based learning in the works of Amado et al., 2015; Burton-Chellew, Nax, and West, 2015b. In these works learning is a mapping directly from observed behavior to actions. Again, making use of the functional notation, we can write the learning function as $f_l^{\text{payoff}} : O \to A$, which is in contrast with our approach. By making a clear distinction between learning and planning, we can study, and potentially control, the distinct qualitative behaviors introduced by either of them, which is not a luxury payoff based learning provides.

So why not just implement AIXI? The main problem in implementing AIXI to predict human behavior in games is that it is not computable (Legg and Hutter, 2007) (although a variant of it is, and this is discussed in the next section). 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 the computability of the model. Therefore, in this chapter we will introduce learning, in the aforementioned sense, to our bounded rational agent model of Chapter 3, while at the same time making use of context specific approximations that allow our model to be computable.

[2] In the previous chapter we assumed agents to have a fixed point in the model space represented by the transition function with $\xi_{\pm} = (0.1, 0.5)$.

4.2 The learning mechanism

4.2.1 A sub-space of all partial functions

In AIXI Hutter, 2007, learning for an agent from past data happens through Solomonoff induction Solomonoff, 1964, which considers the space of all partial functions[3] on $\{0,1\}^*$ [4] i.e. the space of all allowed "explanations" for the past trajectory. Although this form of learning guarantees convergence to the true distribution, is not computable as a consequence of the halting problem (Church, 1936). 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* (a variant of AIXI described by Legg and Hutter, 2007) 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 (c.f. supplementary article by Herrmann, Thöni, and Gächter, 2008).

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 chapter. We will exploit the context specificity afforded by PGG and consider the model class M introduced in Chapter 3. Therefore we only consider Markovian transition functions Q, given by truncated Gaussian distributions, parameterized via $\xi_{\pm}$.

4.2.2 The model

The bounded rational agent of the previous chapter was defined by the tuple $(\xi_{\pm}, m, K, \gamma)$ which was considered to be constant throughout the game. More specifically the planning mechanism required as its input a transition function Q which represented the agent's model of other agents. In so far as the planning mechanism was concerned, it was oblivious to how this transition function was obtained by the agent.

[3] 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.

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

In the learning problem, however, we will be concerned with how the agent comes up with the transition function Q. As the transition function is parameterized by $\xi_\pm$, the 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 hindsight decaying exponentially into the past (equivalently called "recency-bias" in Fudenberg and Levine, 2014) 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 also equivalent to the agent having a bounded memory of its past. This is then achieved by weighting the maximum likelihood estimation with γ_p as below,

$$\xi^*_\pm(t) = \arg\max_{\xi_\pm} \left[\sum_{i=t-1}^{2} \gamma_p^{t-i} \log Q(\mu_i|\mu_{i-1}, f_{ki-1}) - (1-\gamma_p)(\xi_\pm(t) - \xi_\pm(t-1))^2 \right]. \tag{4.1}$$

Here 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, therefore the agent reduces to that of Chapter 3), and $\gamma_p = 1$ would correspond to learning from arbitrarily far back in the past.

4.3 An updated agent model

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

1. **plans**: by considering the game state at $t-1$, making use of the current model ($\xi_\pm(t)$) and solving Eq. 3.13 and evaluating the policy $P(f_t^T)$,
2. **acts**: by sampling a bet from the evaluated policy and
3. **learns**: after observing the state of the game in the current period t and finding the $\xi_\pm(t)$ by making use of Eq. 4.1.

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 $\xi_+, \xi_-(0) = (0.1, 0.5)$[5]. In

[5] These are the values from Chapter 3.

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.

4.4 Behavior space of the updated model

Now that we have defined the learning mechanism, 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 chapter to 10 and 0.9 respectively[6]. Additionally, as we intend to qualitatively exaggerate the effects of learning on the agent, throughout this section we consider longer games of length $T = 100$ and keep the group size to be $N = 4$.

4.4.1 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. 4.1 we show the average contribution $\langle A \rangle$ as a function γ_p for various values of K, where $A = \frac{\sum_k \sum_t f_{k,t}}{4T}$ 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 the 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

[6] In the previous chapter in Sec. 3.5.4 it was 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 chapter are unchanged by these parameters. Hence we make an arbitrary choice of these parameters.

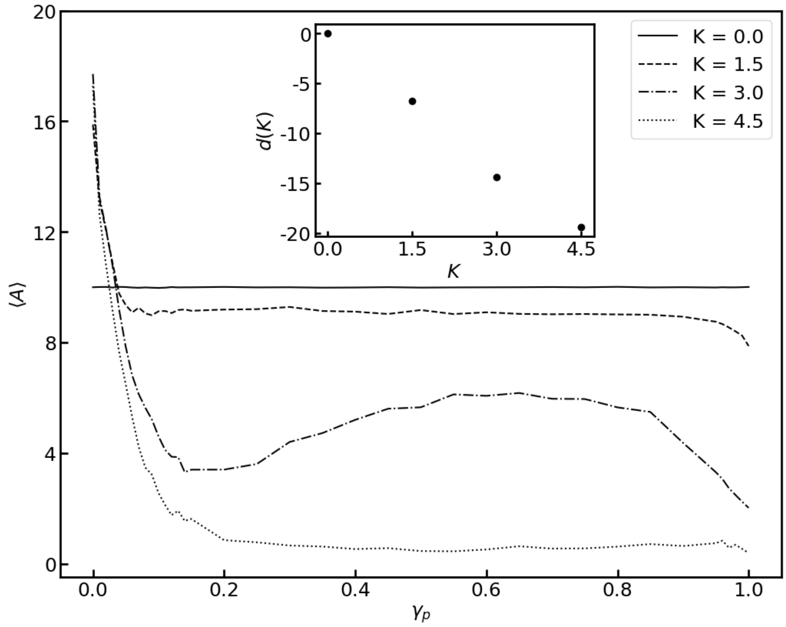


FIGURE 4.1: Average contributions $\langle A \rangle$ as a function of learning strength γ_p for various values of K. Inset depicts the dependence of the decay rate of $\langle A \rangle$ with respect to K.

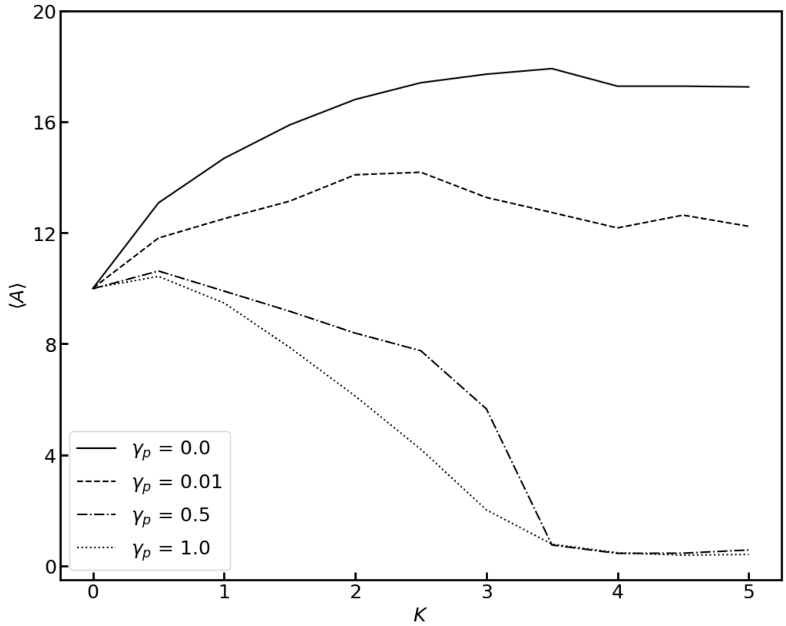


FIGURE 4.2: Average contributions $\langle A \rangle$ as a function of computational budget K for various values of γ_p.

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 brings you towards defection, which corresponds to the Nash equilibrium of the PGG. This observation reiterates that rationality alone is not sufficient to produce Nash equilibrium behavior. As we saw in Chapter 2, a rational agent also needs to develop predictive models of other rational agents to play the Nash equilibrium (CKR). The results in Figs. 4.1 and 4.2 then seem to indicate that through learning the behavior of other agents, some sort of common knowledge of rationality is being developed in a group of all rational agents and as a consequence, the agents play the Nash equilibrium.

Finally in Fig. 4.2 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. 4.5.2, where the issue is explored in more detail.

4.4.2 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

$$\langle G | K, \gamma_p \rangle = \int_{\mathcal{D}} G P(G | K, \gamma_p) dG. \tag{4.2}$$

The gain of a particular agent is the same as in Eq. 2.2. The conditional expected reward is as shown in Fig. 4.3. Much in line with our intuition, the conditional gain is maximized by higher values of K, γ_p, i.e. agents with a higher computational budget and lesser recency bias earn the most reward when playing against a group of randomly chosen agents.

Learning-Planning tradeoff

It is interesting to note that the data in Fig. 4.3 suggest that there is a trade-off between learning (γ_p) and planning (K). This shows up as a negative slope of

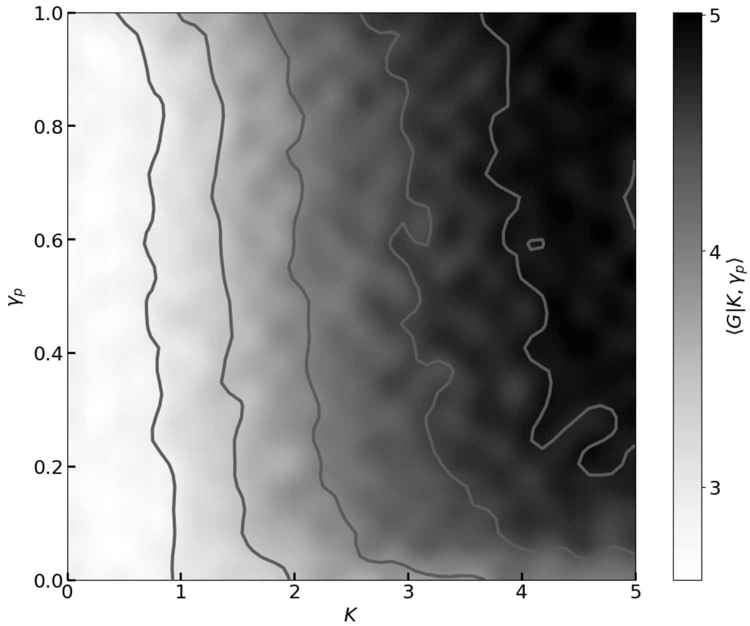


FIGURE 4.3: Conditional expected gains $\langle G|K, \gamma_p \rangle$ (colorbar) and contours (solid grey curves) at $\langle G|K, \gamma_p \rangle = 3, 3.5, 4, 4.5, 4.8$.

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 recently (Moerland et al., 2020), 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 found 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. 4.3 this 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.

4.5 Cooperation amongst learning and planning agents

Equipped with some intuitions of the dependence of agent behavior on the agent parameters, we now focus on certain computational experiments which

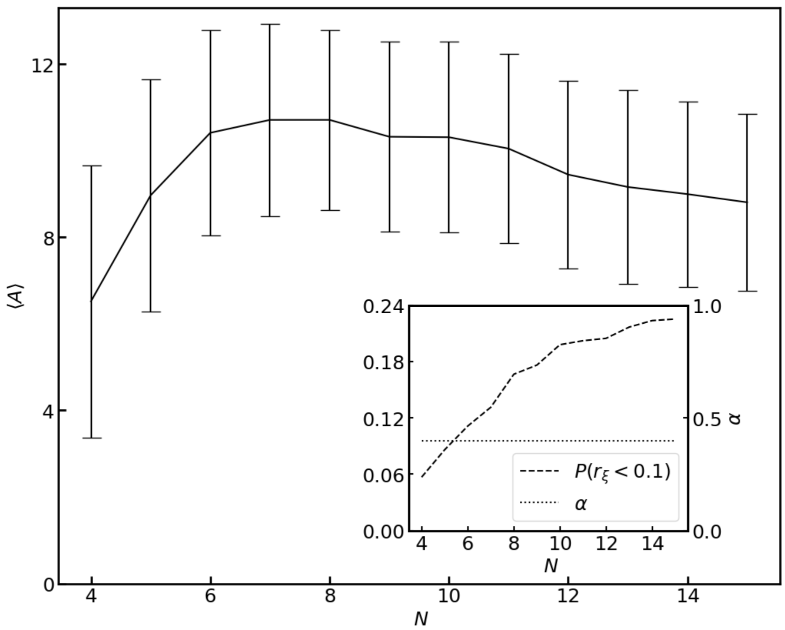


FIGURE 4.4: Average contributions as a function of group size N and the variance (errorbars) for constant α. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

are relevant to experimentally observed behavior in human players playing PGG. In sec. 4.5.1 we observe the impact of group size on cooperation and in sec. 4.5.2 we study how noise in game trajectories might impact the average contributions.

4.5.1 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 (Pereda María and Angel, 2019), some claim that cooperation is harder in larger groups whereas others claim a non-monotonic impact of group size on cooperation (Isaac and Walker, 1988; Yang et al., 2013).

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. 4.4.2), playing PGG for $T = 100$ periods. Figs. 4.4 and 4.5 show the average contribution $\langle A \rangle$ as a function of group size. In fig. 4.4 we keep $\alpha = 0.4$ as a constant and we see that the cooperation is impacted non-monotonically by group size. Cooperation peaks for intermediately sized groups.

In order to explore reasons why cooperation may behave non-monotonically, 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. 4.4). Here

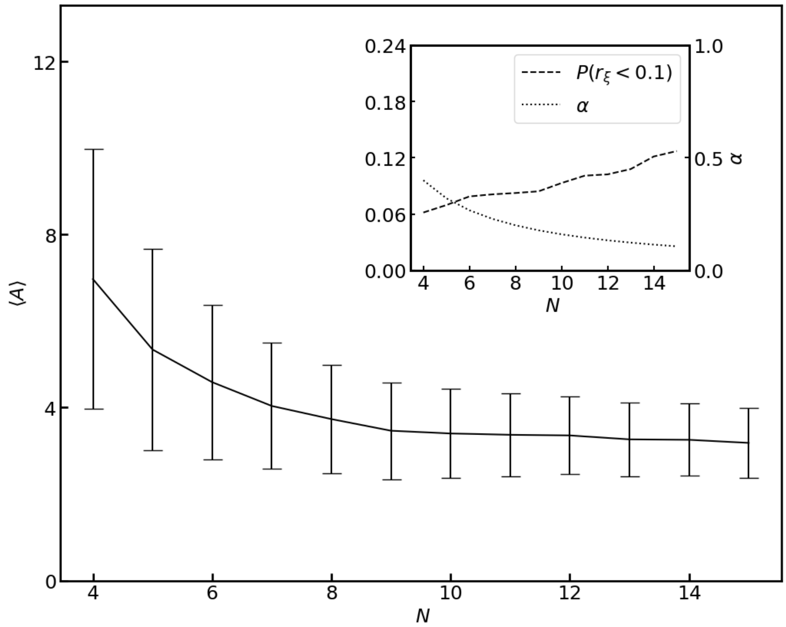


FIGURE 4.5: Average contributions as a function of group size N and the variance (errorbars) for $\alpha \sim 1/N$. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

$r_\xi = \sqrt{\xi_+^2 + \xi_-^2}$. We observe that $\langle r_\xi \rangle$ monotonically decreases with group size.

In the limit $r_\xi \to 0$ we can see that the transition function becomes independent of player contribution $f_{i,t}$ (see from Eq. 3.15 $\lim_{r_\xi \to 0} p = \mu_{i,t-1}$). This essentially (from the perspective of the ith agent) decouples the agent from other players. We can see this effect more precisely in Eq. 3.13. For simplicity we ignore the bounded rationality term and consider that $T = t + 1$. Upon substituting for the value function and expanding we have,

$$V_t^* = \max_{P(f_t^{t+1})} \sum_{\bar{f}_t} G_{k,t} \Big[[PQ]_t + \gamma [PQ]_t \sum_{\bar{f}_{t+1}} G_{k,t+1} [PQ]_{t+1} \Big], \tag{4.3}$$

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

$$\max_{P(f_{k,t+1})} \sum_{\bar{f}_{t+1}} \alpha(N-1)\mu_{t+1}[PQ]_{t+1} - \sum_{\bar{f}_{t+1}} (1-\alpha) f_{k,t+1} [PQ]_{t+1}, \tag{4.4}$$

which further simplifies to,

$$\max_{P(f_{k,t+1})} \sum_{\mu_{k,t+1}} \alpha(N-1)\mu_{t+1} Q_{t+1} - \sum_{f_{k,t+1}} (1-\alpha) f_{k,t+1} P_{t+1}. \tag{4.5}$$

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. 4.3 we get

$$V_t^* = \max_{P(f_{k,t})} \sum_{\bar{f}_t} G_{k,t}[PQ]_t + \gamma[PQ]_t\langle\mu_{k,t+1}|\mu_{k,t}\rangle\alpha(N-1). \tag{4.6}$$

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. Therefore, lower values of r_ξ indicate that the agents are decoupled from the rest of the group. This seems to be natural for larger groups, as an individual agent's action tends to have a lesser impact on the group behavior as the group size increases.

If the decrease in r_ξ were the only process at play here, 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. 2.2). This increases linearly with N for constant α (see inset Fig. 4.4). 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. 4.5). 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 to

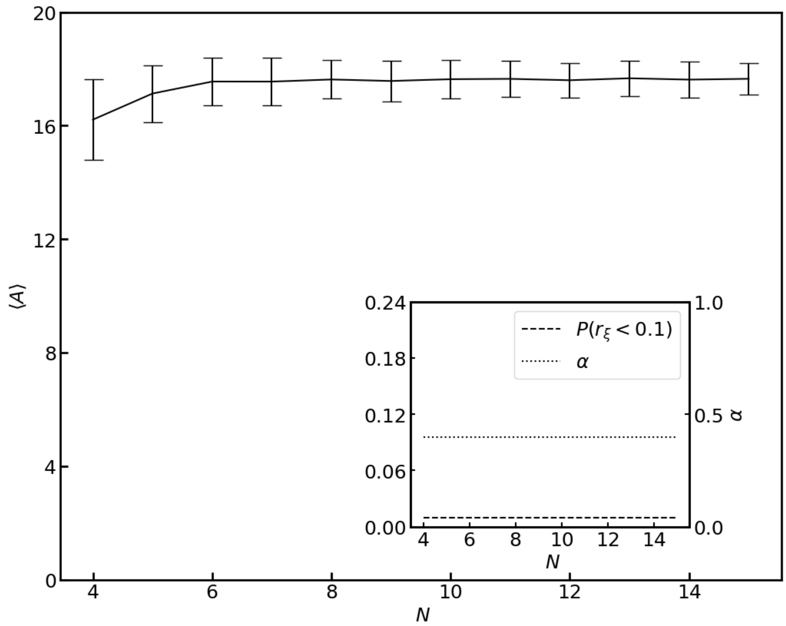


FIGURE 4.6: Average contributions as a function of group size N and the variance (error bars) for constant α and $\gamma_p = 0$. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

$\langle A \rangle$ to become monotonically increasing with group size. This can be seen in Fig. 4.6 where we show average contributions as a function of the 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.

4.5.2 Noise induced cooperation

Anomalous behavior of $K = 3$ agents

In Fig. 4.1 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 ≈ 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 $\xi_\pm > 0$. This can be seen in the following.

When r_ξ is not close to 0, the second summand in Eq. 4.6 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. 3.15 one can see that this is the case when $\xi_\pm > 0$. Therefore when $\xi_\pm > 0$ the agents contribute the most.

One can then say that $\langle A \rangle$ correlates with the occupation probability of the said quadrant of the $\xi_\pm$ space (see Fig 4.7). This can be defined as $f = \frac{\sum_t \langle I_t \rangle}{T}$ where I_t is the indicator function given by

$$I_t = \begin{cases} 1, & \xi_\pm(t) \geq 0 \\ 0, & \text{else.} \end{cases} \tag{4.7}$$

For $\gamma_p = 0$ we start and stay in the aforementioned quadrant as the agent's model is not updated (Eq. 4.1), 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. 4.7). Here the mean step length l is defined as

$$l = \frac{1}{T-1} \sum_{t=2}^{T} \sqrt{(\boldsymbol{\xi}(t) - \boldsymbol{\xi}(t-1))^2}, \tag{4.8}$$

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. 4.1. This then further reduces the occupation probability and also the contribution $\langle A \rangle$ as a consequence.

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 adding one noisy agent with $K = 0$, a fixed mean of contribution $m = 10$ and varying

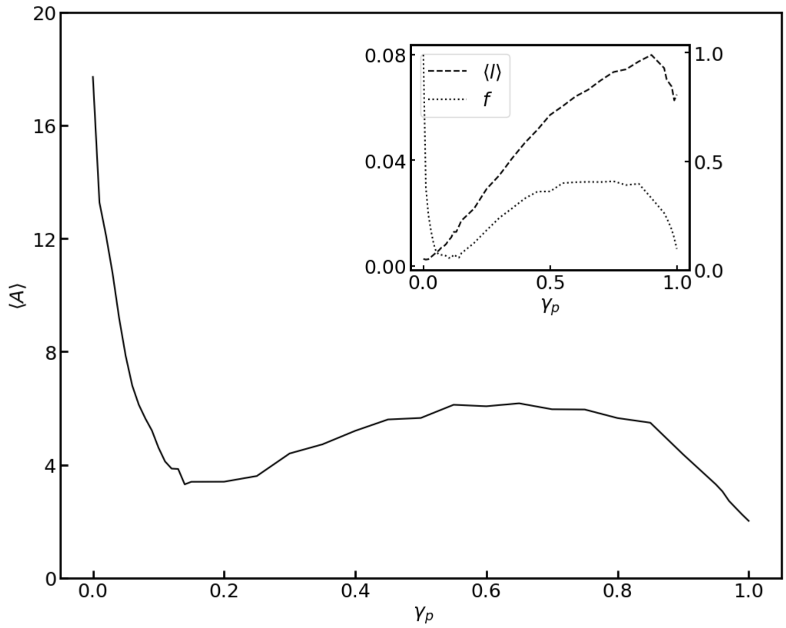


FIGURE 4.7: Average contribution of groups of identical agents with $K = 3$. Inset shows the corresponding values of $\langle l \rangle$ and f as a function of γ_p.

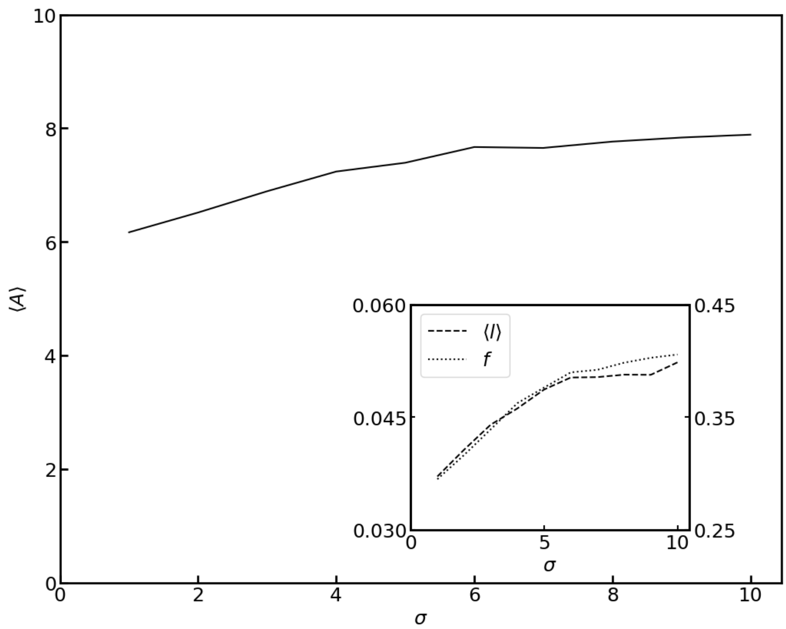


FIGURE 4.8: Average contribution of groups of three randomly chosen agents and one noisy agent with the variance of contributions given by σ. Inset shows the corresponding values of $\langle l \rangle$ and f as a function of σ.

variances σ of the prior distribution ($P_0(f_{k,t})$), to a group of three other randomly chosen agents, as done in Sec. 4.4.2. We then observe how the group average contributions $\langle A \rangle$ are impacted as we increase σ.

In Fig. 4.8 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.

4.6 Conclusions

We have demonstrated the exclusive impact of learning on the behavior of bounded rational agents in PGG through two major computational experiments. First, 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. It is also easy to see that this behavior cannot be possible in the agents of Chapter 3 as the increase in contributions is caused by the increase in the occupation probability of the $\xi_{\pm} > 0$ quadrant. In the case of bounded rational agents without learning this occupation probability is fixed. This prediction remains to be tested via experiments.

Second, we provide a theoretical explanation of 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 (i.e., α) 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 with a quantifiable way of predicting cooperation in PGGs with varying numbers of players. Yet again, this variability in behavior is an exclusive consequence of learning as evidenced by Figs. 4.4, 4.5 and 4.7.

With the addition of the learning mechanism, we have shown that our model is able to replicate (to our knowledge) all known experimental data on the vanilla version of PGG[7]. This demonstrates the enormous explaining power our model holds as it starts from very few and abstract assumptions about human behavior and yet is able to replicate some stylized facts as

[7] There are other variants of PGG beyond the one considered in this thesis. For instance, including a punishment mechanism, or PGG where the play is not anonymous so some reputation effects come into play, etc..

shown in this chapter, without needing to explicitly hard-code them into the agent model.

Our model provides an alternative to the ad-hoc cellular automata (CA) type models that are commonly found in sociophysics literature (Szolnoki, Perc, and Szabó, 2009; Szabó and Hauert, 2002; Wu, Fu, and Wang, 2018; Su, Wang, and Stanley, 2018), which employ these idealized CA agents and focus on collective effects, while mostly ignoring available human behavioral data on PGG. In the next chapter therefore we will extend our data validated model to the domain of spatial PGG or SPGG and compare it to the state of the art in sociophysics.

Albeit, one criticism of our 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 our model be treated as a more fine-grained model of player behavior in games as it has been used to replicate microscopic player behavior (at the level of individual player trajectory). One should then systematically find more coarse-grained CA type models which are effective descriptors of some coarse grained observables, possibly following the ideas of Wolpert et al., 2014. We will discuss this in greater detail in Chapter 6 of the thesis, for now, we move on to studying the collective effects of bounded planning and learning agents in the domain of SPGG.

Chapter 5

Collective effects

In the previous chapters, we motivated and developed an agent based model that replicates human behavior in PGG. In this chapter, we extend our model to the more general setting of spatial PGG or SPGG. SPGG may be taken as a model of a vast number of real situations of human interaction that happen concurrently. In a shared household, participants fulfill chores as an investment (f_i) into a pleasant atmosphere as the common good (G), members of a sports club invest donations (f_i) for enjoying common goods such as a well-maintained playing court (G), members of an association invest personal commitment (f_i) in the common good of a thriving association (G), and so on. A substantial fraction of societal interaction can, in this paradigm, be viewed as an enormously complex network of PGG, which interact through agents participating in several PGG simultaneously. In this setting, beyond just the intra-group dynamics (which we focused on in chapters prior to this), the network of connections between groups and players also becomes of relevance through the fact that players play in several groups. Therefore in this chapter, we wish to explore what novel and exclusive effects emerge from the system as a consequence of the topology of the network.

5.1 Networks of PGG

The interaction structure for players playing multiple PGGs simultaneously with different groups of players is best captured by *hypergraphs*. In what follows, we give a brief account of the concept hypergraphs, and how it connects to our system of interest. We will then describe how to adapt the behavioral dynamics, as described for single groups in the previous chapters, to the spatially extended setting of an SPGG.

5.1.1 Hypergraphs

A hypergraph (Ouvrard, 2020) is a graph whose edges are allowed to connect more than two nodes. An edge in a hypergraph is called a *hyperedge*. Hence a hypergraph $\mathcal{H}$ can be defined as a tuple $\mathcal{H} = (V, E)$, where V is a finite set of nodes v_i indexed by $i \in \{1, \ldots, |V|\}$ and E is a finite set of hyperedges e_j indexed by $j \in \{1, \ldots, |E|\}$, where each hyperedge is a non-empty subset of V. One can completely specify a hypergraph by an *incidence matrix* H_{ij} where $i \in \{1, \ldots, |V|\}$ and $j \in \{1, \ldots, |E|\}$ and

$$H_{ij} = \begin{cases} 1 & v_i \in e_j, \\ 0 & \text{else.} \end{cases} \tag{5.1}$$

We can identify the nodes in $\mathcal{H}$ as the agents and the edges can be identified as the various groups the agents play in. To be consistent with the work presented in the previous chapters, in this work, we consider agents playing on 4-*regular* hypergraphs, where a k-regular hypergraph H is a hypergraph such that $|e_i| = k$ for all $i \in \{1, \ldots, |E|\}$. Each agent, although, is allowed to play in arbitrarily many groups. Other than this we consider no restrictions on the topology of the hypergraphs.

5.1.2 The SPGG

Recall that eq. 2.2 describes the gain for a given player playing in one period, in one group. In a network of interacting PGG groups, the corresponding gain becomes the sum of gains over all groups the agent plays in. Notice that the gains in each of these groups are independent. Therefore, in order to find a favourable policy in a particular group, each agent needs to keep in mind the actions of other players in that group alone. The dynamical interaction between groups comes about through the learning of the agents, i.e., their gradually updating their internal model of other player's behaviour. We assume that each agent bears one such model for players in general, i.e., for the players of all groups it is playing in.

5.1.3 Agent dynamics on hypergraphs

With the interaction structure as afforded by hypergraphs, we focus on how to adapt the agent behavioral dynamics from Chapters 3 and 4 to the spatially extended setting. For the following, let us consider an SPGG being played on a hypergraph $\mathcal{H}$ with $|V| = N_p$ players and $|E| = N_g$ groups. In order to carry out the discussion on the agent model we describe it from the perspective of an arbitrary agent given by an index $i \leq N_p$, which plays in $k \leq N_g$ different groups given by the indices $i_1, \ldots, i_k$.

As already mentioned, agent i plans in each of the k groups independently. This means that agent i in period t, makes use of Eq. 3.13 to find the optimal policy for itself, and then, for each group it plays in, it independently samples an investment from the conditional distribution $P(f_{i,t}|\mu_{i,t-1})$ by conditioning on the observed $\mu_{i,t-1}$ of that group.

Learning, on the other hand, is not done independently in each group. Essentially, the exponentially weighted log likelihood in Eq. 4.1 is summed not only over all the observed past of one group but over all the groups. By suitably modifying the notation to also include the group identity, one can write the analogous equation as

$$\begin{aligned} {\xi^i_\pm}^*(t) = \arg\max_{\xi^i_\pm} \sum_{m=1}^{k} \Bigg[\sum_{w=t-1}^{2} \Big(\gamma_p^{t-w} \log Q(\mu_{i,i_m,w}|\mu_{i,i_m,w-1}, f_{i,i_m,w-1})\Big) \\ - (1-\gamma_p)(\xi^i_\pm(t) - \xi^i_\pm(t-1))^2 \Bigg], \end{aligned} \tag{5.2}$$

where the subscript i_m denotes that the quantity is being evaluated in group i_m with $m \leq k$.

The learning and planning mechanisms can be then combined into a single agent which is defined by a quadruple of parameters, (m, K, γ, γ_p), where m parameterizes the prior policy of the agent[1]. In each round $t \in \{2, 3, ...T\}$, the agent

1. **plans**: by considering the game state at $t-1$ in all the k groups it plays in and $\xi^i_\pm(t-1)$, the agent making use of Eq. 3.13 finds the best policy for itself.

2. **acts**: by sampling an investment, $f_{i,i_m,t}$, from the evaluated policy, after conditioning on the observed $\mu_{i,i_m,t-1}$ of each group $m \in \{1, \cdots, k\}$.

3. **learns**: after observing the state of the game in the current period t in all the k groups the agent learns and updates its model of the players by evaluating $\xi^i_\pm(t)$ from Eq. 5.2.

In period $t = 1$ the investments of the agent are sampled from its prior distribution $P_0(f_{i,1})$ for all the groups it plays in. In period $t = 2$ there is certainly planning based on observed behavior, but there is no learning, as the agent has not yet observed a transition.

5.2 Topology and dynamics

As mentioned above, in this chapter we wish to investigate the specific impact of the interaction network structure on the agent dynamics, and thereby

[1]We assume that the prior policy of the agent is given by the same distribution $P_0(f_{i,t})$ in each period t and the distribution is a truncated Gaussian centered around m and a fixed variance $\sigma_m = 5$.

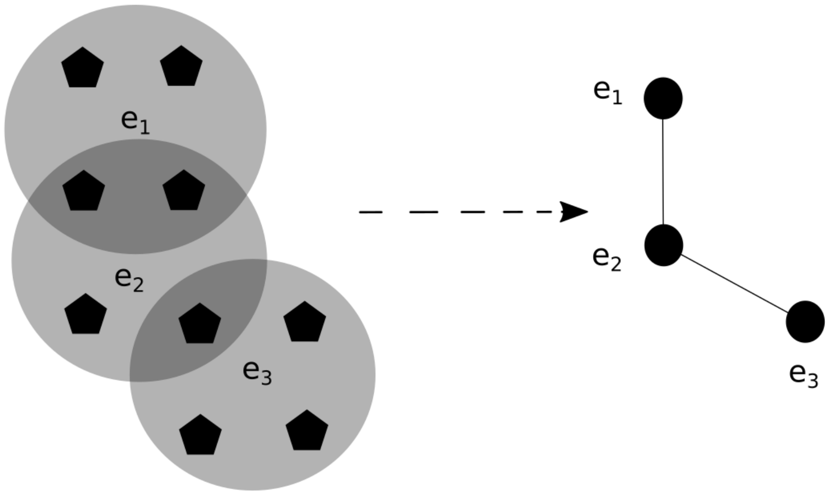


FIGURE 5.1: Projecting a hypergraph to a line graph. The hyperedges (e_1, e_2, e_3) of the hypergraph correspond to the vertices of the line graph. Black pentagons correspond to agents, black circles correspond to groups.

on possible collective phenomena in the system. This will be the main question we pursue in the remained of the chapter. Before we proceed, however, we need to clarify our concepts of "topology" and "dynamics", and introduce the key descriptors of topology of the hypergraph as well as the relevant observables we will use to describe the agent behavioral dynamics.

5.2.1 Topological features

Hypergraphs (as opposed to graphs) provide the opportunity to consider two distinct types of interactions: inter-group and intra-group. While our previous studies have focused primarily on intra-group interactions, we focus here on inter-group interactions and their effect on contributions. The smallest unit of interest in this paradigm is a group. This allows us to project the uniform hypergraph to a corresponding *line graph* $\mathcal{L}$[2]. Fig. 5.1 shows a schematic of the projection of a hypergraph to the corresponding line graph. Notice that the hyperedges e_1, e_2, e_3 become vertices in the line graph and the information about the location of vertices of the hypergraph (black pentagons) gets lost when performing this projection. Therefore in a line graph all of the intra-group structure is lost and it remembers only non-zero overlaps between groups.

Our task now remains to choose the appropriate descriptor of the graph topology. Studying SPGG, we are interested in evaluating the degree of influence (measured through the dynamical observables) a particular group (i.e., a node in the line graph) has on other groups in the graph depending, e.g.,

[2]A line graph of a hypergraph $\mathcal{H} = (V, E)$ is a graph $\mathcal{L}(\mathcal{H}) = (V_{\mathcal{L}}, E_{\mathcal{L}})$, where $V_{\mathcal{L}} = E$ and two vertices e_i, e_j in are connected in $\mathcal{L}$ iff. $e_i \cap e_j \neq$

on the distance between them. This is to determine how 'far' the influence of a group travels within the graph. Hence we need a suitable descriptor that represents the notion of a distance in a graph topology and is suitable to quantify such propagation of influence.

A well-established class of such descriptors is the *Bonacich-Katz* class of centrality measures (Bonacich, 1987). The Bonacich-Katz centrality of a node i in the line graph $\mathcal{L}$ is parameterized by two parameters ω, η and is given by

$$c_i(\omega, \eta) = \omega(I - \eta J)^{-1} J 1, \tag{5.3}$$

where I is the identity matrix, J is the adjacency matrix of $\mathcal{L}$ and 1 is a column vectors of all ones. Here ω is a constant multiplying factor and therefore doesn't impact the centrality ranking of the nodes and can therefore be ignored via an appropriate normalization. η parameterizes the expected radius of influence a particular group has on other groups, which is proportional to $(1 - \eta)^{-1}$. It will be the main parameter of concern for us in this chapter. While for $\eta \to 0+$, c_i is equivalent to degree centrality and it corresponds to eigenvector centrality if $\eta \to \frac{1}{\lambda_{\max}} -$, where $\lambda_{\max}$ is the largest eigenvalue of the adjacency matrix of $\mathcal{L}$ (Benzi and Klymko, 2015).

5.2.2 Dynamical features

In order to quantify cooperation in PGG, a natural observable of interest (as before) is the average contribution of the group or player. Let us consider an agent i where $i \leq N_p$. Let the set of groups the agent plays in be given by $p_i = \{k | H_{ik} \neq 0\}$. Then we define the average agent contribution as

$$A_{i,t}^{\text{agent}} = \frac{\sum_{j \in p_i} f_{i,j,t}}{|g_i|}, \tag{5.4}$$

and the corresponding ensemble averaged quantity given by $\langle A_{i,t}^{\text{agent}} \rangle$. The corresponding average group contribution for a group j where $j \leq N_g$ is defined as

$$A_{j,t}^{\text{group}} = \frac{\sum_{i \in g_j} f_{i,j,t}}{4}, \tag{5.5}$$

where $g_j = \{i | H_{ij} \neq 0\}$ and the ensemble average quantity is given by $\langle A_{j,t}^{\text{group}} \rangle$. For both these quantities the corresponding time average quantity is given by $A_i' = \frac{\sum_{t=0}^{T} A_{i,t}'}{T}$.

The above quantities are averaged over time and therefore hold no information regarding the interactions between groups unfolding over time. The latter can be investigated by considering temporal correlations between group trajectories. To this end, we consider the correlation matrix C_{ij}, in

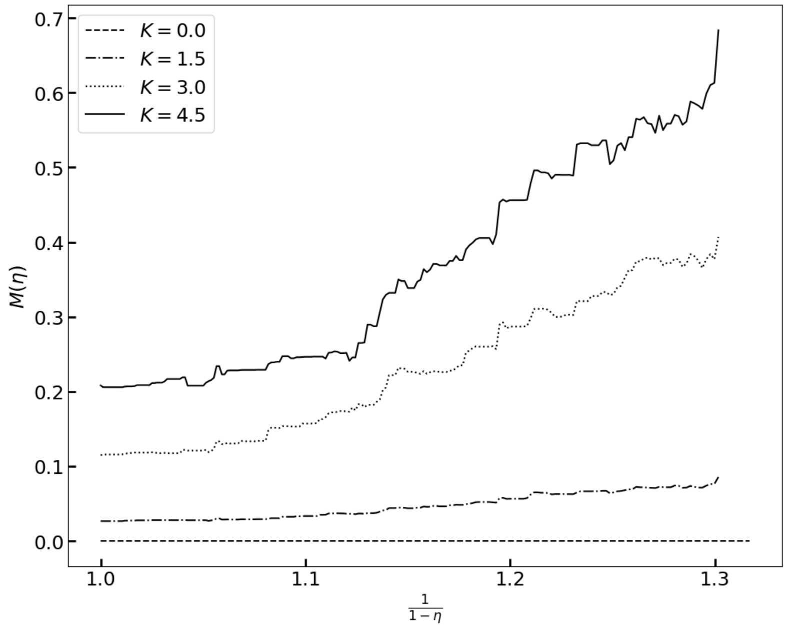


FIGURE 5.2: Dependence of conditional expected contribution variance, $M(\eta)$, upon radial distance of influence, as expressed by $(1-\eta)^{-1}$. The curves are averaged over ten hypergraphs of size $(N_p, N_g) = (64, 25)$ for agents with $K = 0$ through $K = 4.5$.

which the i, j entry is the correlation between trajectories of groups i and j. It is given by

$$C_{i,j} = \frac{1}{T}\sum_{t=0}^{T} \frac{\sigma_{A^{\text{group}}_{i,t} A^{\text{group}}_{j,t}}}{\sigma_{A^{\text{group}}_{i,t}} \sigma_{A^{\text{group}}_{j,t}}}, \tag{5.6}$$

where σ_{XY} is the covariance of random variables X, Y given by $\langle XY\rangle - \langle X\rangle\langle Y\rangle$ and σ_X is the variance of X. Because the correlation between two group averaged trajectories is symmetric over the two groups, it is a natural measure for us as the coupling between groups which is mediated by the learning mechanism of the common player(s) also has no way to break the symmetry between the groups.

5.3 Results

As we intend to evaluate the specific impact of network topology on the dynamics, we will keep the agent characteristics homogeneous and consider heterogeneity only through graph topology. Therefore we consider identical agents that are embedded in a random hypergraph (for details on how these random hypergraphs are generated see App. C).

For these agents we fix $(m, \gamma) = (10, 0.9)$ for similar reasons as in Chapter 4 and additionally, we consider $\gamma_p = 0.9$. This choice reflects the fact that in our model the interactions between neighboring groups are mediated by

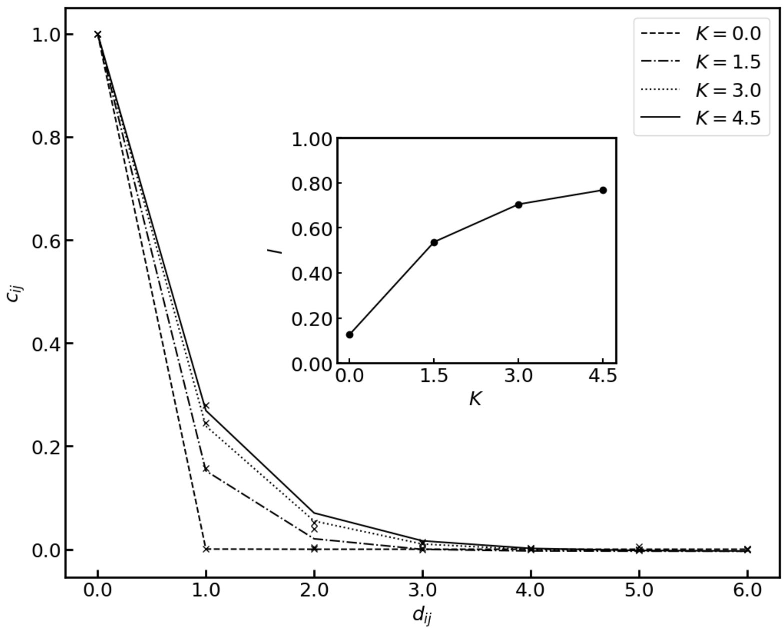


FIGURE 5.3: Group trajectory correlations as a function of the shortest distance between the groups for groups of agents with $K = 0$ through $K = 4.5$. Crosses are data points, polygons are exponential fits.

learning. Lower values of γ_p would weaken the group interactions, thereby rendering the network structure pointless. On the other hand for $\gamma_p \approx 1$ it is observed that something similar occurs for longer times (recall Sec. 4.5.2), as every new observation has only a diminishing impact on the agents' preferences. As we will see in the following, even though we create a setup (choice of parameters) that qualitatively maximizes the group interactions, the resulting dynamics still seem mostly independent of the various topological features.

We consider ten randomly generated hypergraphs with $(N_p, N_g) = (64, 25)$ with all the agents described by $(m, \gamma, \gamma_p) = (10, 0.9, 0.9)$ and $K \in \{0, 1.5, 3, 4.5\}$. For each random hypergraph, we perform the ensemble average of 5000 simulation runs. Following the question asked in the beginning of Sec. 5.2 we investigate, in a random hypergraph with all identical agents, what topological descriptor is the most appropriate to predict average group contribution. As already mentioned we consider the class of descriptors given by Bonacich-Katz centrality and we consider the most "appropriate" centrality measure (or the most appropriate η) as the one that minimizes the function

$$M(\eta) = \int \theta_{\langle A_j^{\text{group}} \rangle | (c_j(\eta) = c)} dc. \tag{5.7}$$

Here the variance is defined as

$$\theta_{\langle A_j^{\text{group}} \rangle | (c_j(\eta) = c)} = \text{VAR}\{\langle A_j^{\text{group}} \rangle | c \leq c_j \leq c + dc\}, \tag{5.8}$$

for some choice of discretization. Essentially, every value of η is an assignment of a centrality to each node. We wish to find that assignment η such that given a centrality $c(\eta) = c_0$, the variance of the average contributions corresponding to the nodes with the centrality close to c_0 is minimal when integrated over all c_0. In other words, we wish to find η for which the scatter of the scatter plot between $\langle A_j^{\text{group}} \rangle$ and $c_j(\eta)$ is minimal.

In Fig. 5.2 we plot $M(\eta)$ curve averaged over 10 different random hypergraphs as a function of η and one can see that the global minimum of $M(\eta)$ is given by $\eta \approx 0$ for various values of agent parameters given by $K \in \{0, 1.5, 3, 4.5\}$ thereby indicating that the centrality measure with very small values of η best predicts the group average contribution independent of K. Recalling the definition of Katz centrality, smaller values of η correspond to smaller radii of influence. This then seems to indicate that it is the local topological features (in this case, node centrality i.e. number of neighbors) that are the best predictors of average group contribution.

In the above analysis, there is no "dynamics" as such, as we have only considered the distribution of $\langle A^{\text{group}} \rangle$ across nodes of varying centrality. Therefore any claims of groups have a smaller "radius of influence" could be misguided if the above result is considered in isolation. In order to create a more robust picture of the radius of influence of a group, we consider how quickly do inter-group correlations decay as a function of the shortest distance between the groups.

Therefore, we proceed by comparing the correlation matrix C with the distance matrix D[3] to evaluate how the correlations between group average trajectories scale with the shortest distance between the groups. In Fig. 5.3 one can see that the correlation between group average trajectories falls exponentially with the (shortest) distance between the groups with correlation lengths $l < 1$ (as shown in the inset). Thereby meaning that substantial correlations of a group's behavior are only with its immediate neighbors. This further corroborates that group contributions in SPGG are mostly governed by local interactions.

The locality of interactions in SPGG is a piece of good news for two major reasons. First and the more obvious reason is that locality simplifies the analysis of the system, thereby allowing for the possibility of developing simpler effective dynamics that replicate these observations. The second reason is that if group contributions are mostly governed by local interactions (in this case group centrality $c(\eta = 0)$), then even if we have graphs with different global structures but similar local structures we should observe similar behavior. The latter would seem to indicate a universality (i.e. global topology

[3]A distance matrix D for a graph $\mathcal{L} = (V, E)$ is a matrix of size $|V| \times |V|$ where the i, j entry $D_{i,j}$ is the length of the shortest path between node i and node j in $\mathcal{L}$.

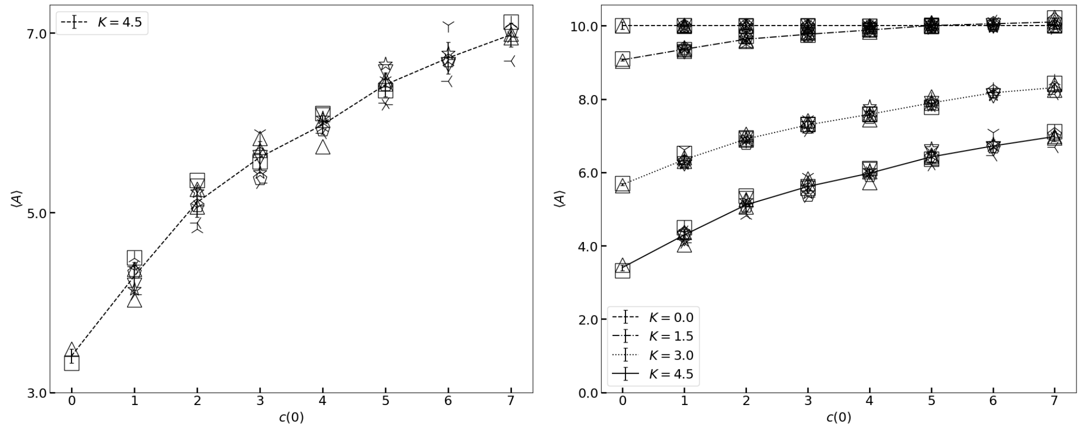


FIGURE 5.4: Impact of group centrality on the group average contribution for ten randomly generated hypergraphs. Left panel: results for $K = 4.5$, with different symbols corresponding to different hypergraphs. Right panel: same as left panel for different values of K. Polygons connect averages over all hypergraphs, respectively.

independent) in cooperation behavior across a large set of networks.

Fig. 5.4 presents the impact of group centrality ($c(\eta = 0)$) on the group average contribution for different values of K, for ten randomly generated hypergraphs. Different symbols correspond to different hypergraphs. For each value of K, the results are found not to differ significantly for different hypergraphs. This demonstrates that the topological features of the hypergraphs are irrelevant to this end. Cooperation levels of groups are predominantly determined by the number of neighbors of the group. Quite surprisingly, one does not even need to consider how many players are shared between two neighbor groups (recall that $c(0)$ is measured from the line graphs and not the hypergraphs).

5.4 Discussion

Another observation from Fig. 5.2 is that $M(\eta)$ monotonically increases with K, attaining a minimal value at $\eta = 0$ for all K investigated. Having a higher scatter for higher K would seem to indicate that agents with higher K accommodate their contributions to more details of topology rather than just the number of groups they play in[4]. This is a view that also gets supported by looking at the variation of correlation lengths with K. In the inset of Fig. 5.3, it can be seen that correlation lengths increase with K, thereby indicating that as K increases, more distant neighbors become relevant as compared to lower values of K. Hence agents with higher values of K are more sensitive to the surrounding topology of the interaction network. However, the data suggest

[4]The other end of the extreme is the case of $K = 0$, where the agents disregard any topological or dynamical features and play according to independent samples from a stationary prior distribution.

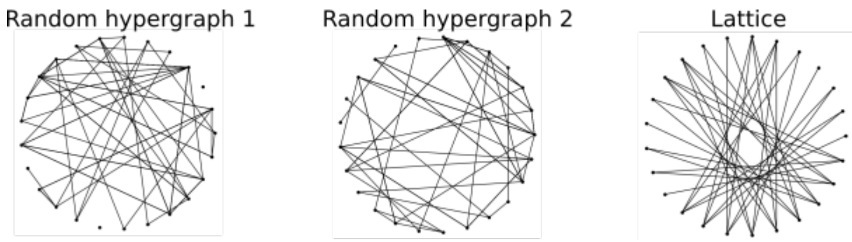


FIGURE 5.5: Line graphs corresponding to three hypergraphs. The left two correspond to random hypergraphs and the rightmost one corresponds to square lattice.

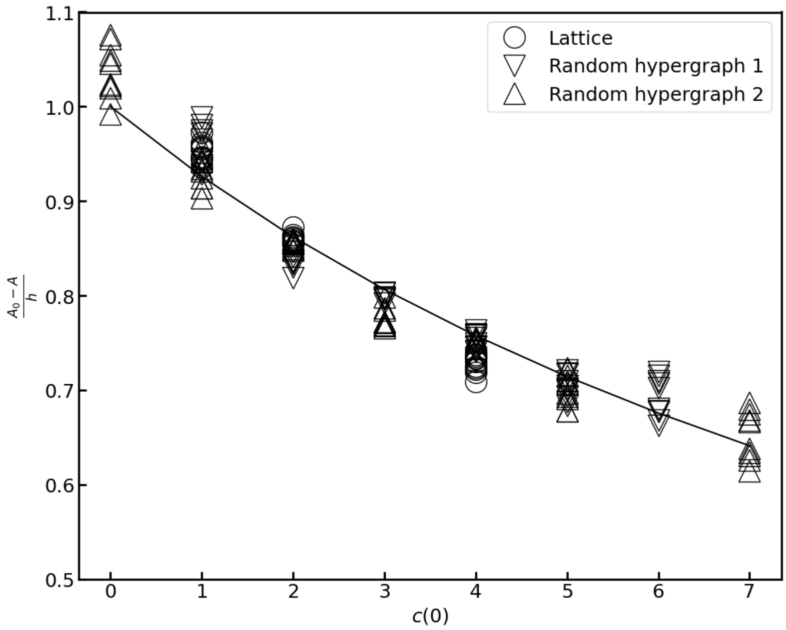


FIGURE 5.6: $(A_0 - A)/h$ as a function of group centrality for the three hypergraphs. The corresponding data from the hypergraphs are represented by circles, triangle-up and triangle down. The solid line corresponds to the function $(1 + c(0)/\beta)^{-1}$ for $\beta = 12.5$.

that the range of this sensitivity saturates at higher K, with the decay length staying below unity.

This sensitivity can also be seen in Fig. 5.4. Note that group centrality has a positive impact on group contributions irrespective of agent parameters, although agents with higher values of K experience an appreciably bigger increase in their contributions as compared to their lower K counterparts. In the following, we quantify this sensitivity (to centrality) as a function of K.

We consider 3 different hypergraphs (their corresponding line graphs can be seen in Fig. 5.5), two of them generated randomly and one uniform square lattice, all with $(N_p, N_g) = (64, 25)$. We consider identical agents with K varying from 0.5 to 5 and average over 5000 simulation runs for each configuration (i.e. a pair of a value of K and one of the three hypergraphs). We then perform three-parameter fits on the average contribution curves with

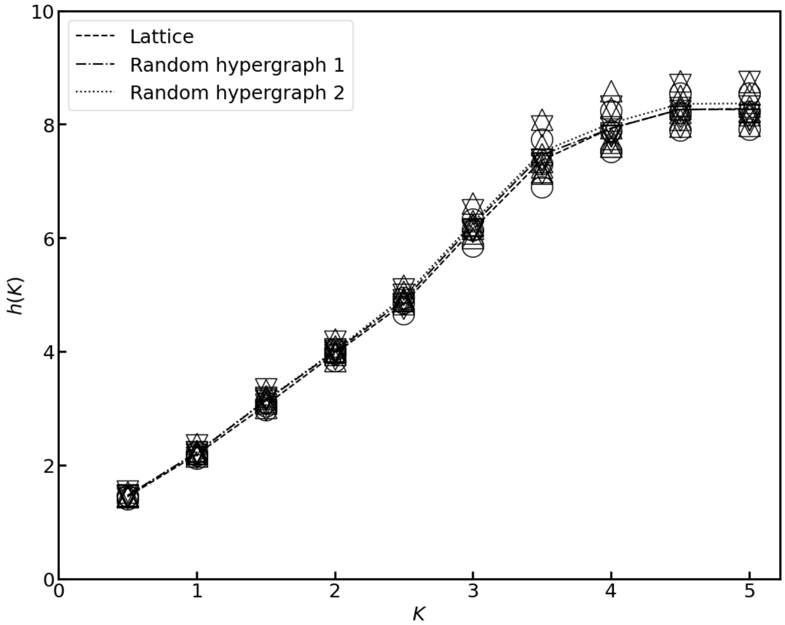


FIGURE 5.7: $h(K)$ for three hypergraphs. Polygons (dashed, dash dotted and dotted) correspond to the best fit. The corresponding data from the hypergraphs are represented by circles, triangle-up and triangle down, respectively.

the ansatz,

$$A(c(0), K) = A_0(K) - \frac{h(K)}{1 + c(0)/\beta}. \tag{5.9}$$

As it turns out, A_0 does not vary appreciably with K[5]. Therefore we remove its dependence on K and treat it as a constant. Therefore, parameters A_0 and β are obtained by performing a fit on all configurations and h is fitted to each configuration separately. It turns out that setting $A_0 = 12.25$ and $\beta = 12.5$, the (fitted) values of $h(K)$ collapse onto a characteristic curve independent of the network topology. In Fig. 5.7 we plot three curves $h(K)$ for the corresponding hypergraphs as obtained from the fits. The empty symbols correspond to $(A_0 - A)/(1 + c(0)\beta)$ for each individual group, where A_0 and β take the aforementioned values. A and $c(0)$ for each group are obtained from the simulations. Within scatter, the data are well represented by eq. (5.9). In a similar fashion we also show the characteristic value of β to be well descriptive of the impact of group centrality (see Fig. 5.6). To conclude, the deviation from the 'trivial' case $K = 0$ seems to factorize as expressed by eq. (5.9), into a part depending on $c(0)$ and a function $h(K)$ which depends only on K. The latter starts off roughly linearly but saturates at higher values of K.

[5]Here we ignore the case of $K = 0$, as it represents a qualitatively trivial case, and the corresponding fitting procedure has non-unique global minima (both $h(0) \to 0$ and $\beta \to \infty$ lead to a flat curve).

5.5 Comparisons to other models

Contemporary literature on SPGG (Szolnoki, Perc, and Szabó, 2009; Szabó and Hauert, 2002; Wu, Fu, and Wang, 2018; Su, Wang, and Stanley, 2018) mostly considers simplistic agents that either always cooperate or defect and update their strategies through "imitation learning" (where an agent imitates the strategy of the most successful individual in its vicinity). In these models, the authors commonly report phase transitions depending on the initial fraction of cooperators in the population. In contrast, our results suggest that phase transition in fixed group size games is not a possibility in SPGG as can be seen from the correlation lengths asymptoting to values less than 1.

Despite the differences in our approaches, there are still some similarities in the observed outcomes. For instance, our work suggests that cooperation in SPGG can be driven by making the players more diverse i.e. increasing the number of groups they play in and consequently increase inter-group connections. This observation is in line with results previously reported in Santos, Santos, and Pacheco, 2008, despite the fact that the authors follow a completely different modelling route.

5.6 Conclusions

Based on previously developed model agents that boundedly learn and plan, we have explored collective behavior in SPGG on a variety of hypergraphs of different topologies. What we find that collective investment behavior is determined essentially by local descriptors, with correlations decaying exponentially in space. Furthermore, the impact of local connectivity, $c(0)$, and rationality of the agent, K, on the expected average investment factorize in a universal way, independent of network topology. Its behavior can be quantified with a characteristic function as shown in Fig. 5.7 and lends itself to experimental test.

Chapter 6

Conclusions and Outlook

6.1 Conclusions

In this thesis we have presented an alternate modelling route to cellular automaton (CA) based models, to model human behavior in games. We demonstrate the explaining power of our model by comparing it to experimental data while only incrementally increasing its complexity (plan $\rightarrow$ plan + learn $\rightarrow$ plan + learn + topology). We demonstrated in Chapter 3 that bounded rational agency is sufficient to replicate human behavior in short games, by fitting our model to individual player trajectories, something that no other known model does. By doing this we challenged the dominant explanations for the cooperative, non-Nash, and erratic behavior of human players. Subsequently we demonstrated that in order to account for human behavior in varying group sizes and longer games, learning is a necessary ingredient, in Chapter 4. We also provided a testable prediction (which comes about only rarely in sociophysics) that noise in contributions could increase cooperation levels. It must be noted that through these two works, we show that cooperation amongst selfish agents can be attributed to bounded rationality and learning and that we don't need to explicitly modify the utility function of the agents (i.e. hard coding altruism or inequity aversion) to mimic cooperative behavior as has been done in prior works. Only after having justified our model on human behavior, we explore the collective effects of agents in SPGG in Chapter 5. There we demonstrated that our model agents show a topology independent behavior, where cooperation is directly linked to player diversity. Finally, through our work, we hope to have communicated to sociophysicists the need to move to more humanized agents in order to develop robust theories of collective social behavior.

6.2 Outlook

There are three major future directions that our work can head towards - going higher in the abstraction hierarchy by developing effective agents, going parallel by using our model agents to replicate human behavior in other games, and going deeper by augmenting our model with more human properties. In the following we will elaborate a little on each of these directions.

6.2.1 Going higher - effective dynamics

Recall from Chapter 1 that through the work presented in the thesis, all we have done is present an agent function $f = f_p \circ f_l$ that we have compared with experimental data and it shows to replicate human behavior in PGG. In order for us to scale up the simulations to any realistic setting (larger group sizes or larger hypergraphs in SPGG), we will have to think about the complexity of the algorithms that we employ. More specifically, the planning function involves selecting paths that maximize some path integral, which is done by solving the Bellman equation that has a complexity of $\mathcal{O}(Tn^3)$ where n is the number of states (of the game) and T represents the time horizon (see Papadimitriou and Tsitsiklis, 1987). This is the complexity of one game. For an SPGG with multiple groups overlapping in non-trivial ways, using the algorithms adopted in this thesis becomes unfeasible rather quickly (this is why in Chapter 5 we restricted ourselves to $(N_p, N_g) = (64, 25)$). Therefore, for such scaling up to be possible, one would need to focus on finding effective dynamical rules that are sufficient to reproduce some coarse-grained observable of interest.

One can take inspiration from the state space compression (SSC) method introduced in Wolpert et al., 2014. The basic idea is rather simple. Let's consider a microscopic state space X (we could identify with space of all $\bar{f}$s in our case) and a dynamical rule D_X representing the evolution law on this space (our agent + environment paradigm). Correspondingly we have macroscopic state space Y and a macroscopic dynamical rule D_Y which are to be determined. Let us say that for our purposes all we care about in PGG and SPGG is the temporal evolution of the average of contributions A_t. In such a case, our objective is to find a "suitable" Y and D_y such that A_t^X (the trajectory generated by D_X on X) is as close to A_t^Y as possible. The suitability of the macroscopic state space and the dynamics depends on how much easy they are to compute (complexity wise). Therefore, with a suitable formalization of the computation costs and accuracy costs (as presented in Wolpert et al., 2014) one can find effective behavioral descriptions of our model.

6.2.2 Going parallel - other games

Another direction to proceed further could be by testing the bounded planning and learning model on other games. Recall that our model parameters (m, K, γ, γ_p) represent properties of the agents that are unchanging during the game, equivalently they could be considered the identity of the player. Therefore, a true test of the model would be if it could not only generate the observed behavior in these games but also if the inferred parameters of the players are constant across these games.

However, for the model to be applied to other games, we will have to get rid of the context specific assumptions (anonymity of play, etc..) we've made throughout the thesis. The most important one has to do with the functional form of the transition function (being parameterized by $\xi_\pm$). This would also consequently impact the learning mechanism as it relies on performing a maximum likelihood estimation (MLE) over $\xi_\pm$. Regardless, the framework of learning and planning with bounds upon them should remain unchanged.

Moreover, what could also be interesting is if we could develop experimental methods to measure player properties before the game is played. For instance, if we allow close to zero reaction time in a single shot PGG, then multiple samples of the player could represent their prior policy. Then by allowing for more reaction time, one could find the posterior policy and consequently get an idea of K. In so far as measuring the discounting factor γ is concerned, there are already existing experimental paradigms (for instance see Andersen et al., 2008). The true test of the model would be if it could predict how players with the obtained preferences play iterated PGG.

6.2.3 Going deeper - cognitive science

The final direction to proceed could be to develop even more accurate models of human agency. There are certain aspects of human agency that we have ignored in our pursuit of modelling PGG behavior, for instance, have a look at Fig. 6.1. The figure shows the density of contributions of all the players in the city of Samara. The distribution has 5 very well pronounced peaks and they coincide with contributions that are exact multiples of 5. As you can imagine, this preference for multiples of 5 is not specific to Samara, rather is demonstrated by players of all the cities. In fact, tendencies such as this are not only limited to PGGs alone. For instance, recent experiments involving human subjects performing serial reaction time tasks (Wu et al., 2022) seem to show a behavior that the authors call "chunking". Here the participants remember sequences in chunks, the sizes of which depend on the time allowed to remember the sequences (which could be viewed as some analogue of K).

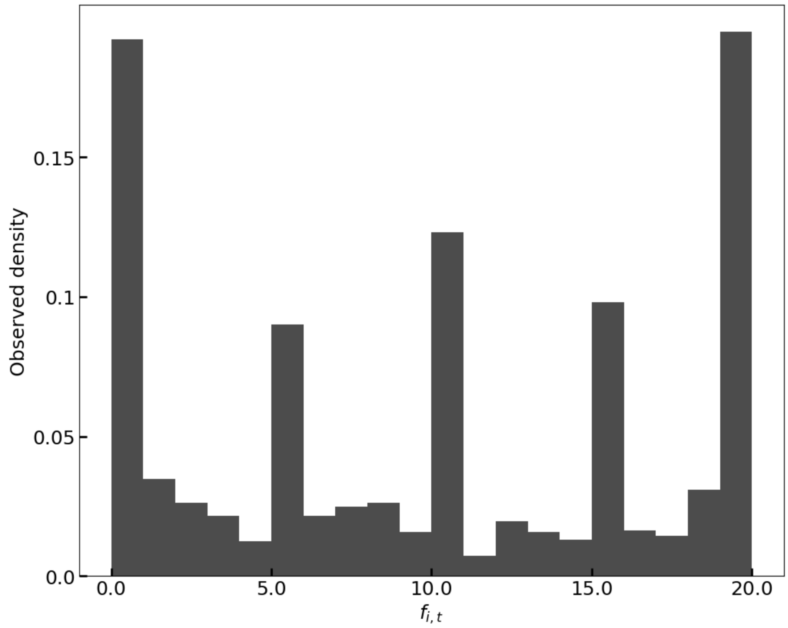


FIGURE 6.1: Density of contributions of all players and all periods in the city of Samara. Data was obtained from the authors of Herrmann, Thöni, and Gächter, 2008.

It is not a coincidence that humans show very similar tendencies across very different domains. One may model this behavior as a coarse-graining of the action space in order to reduce the complexity of the task at hand, in order for it to be doable within the computational bounds of the human. An agent needs to figure out "optimal" coarse-graining of state and action spaces, as too much coarse-graining may lead to having imprecise models of the environment and too fine-grained models may become incomputable in the given bounds. This may be formalized using the information bottleneck method of Tishby, Pereira, and Bialek, 2000. Also, because the coarse-graining must be impacted by the agent's understanding (learning) of the environment it is in (knowing what details to ignore), the Lagrange parameter in the information-bottleneck method must be related to the computational bounds. Regardless of the manner of modelling this, the link between learning and coarse-graining must be formalized.

Another aspect that got ignored in the bigger scheme of things was the trade-off between learning and planning resources as demonstrated in Sec. 4.4.2. These trade-offs should be explicitly described in a future work, potentially in a natural manner. If the trade-off is to be decided dynamically (online) as opposed to passively (offline) from a look-up table, then there might be a need to invoke meta-cognitive mechanisms which learn to trade-off (for instance see Lin et al., 2015) learning resources with planning resources optimally.

Appendix A

Publication: Bounded rational agents playing a public goods game

Bounded rational agents playing a public goods game

Prakhar 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

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

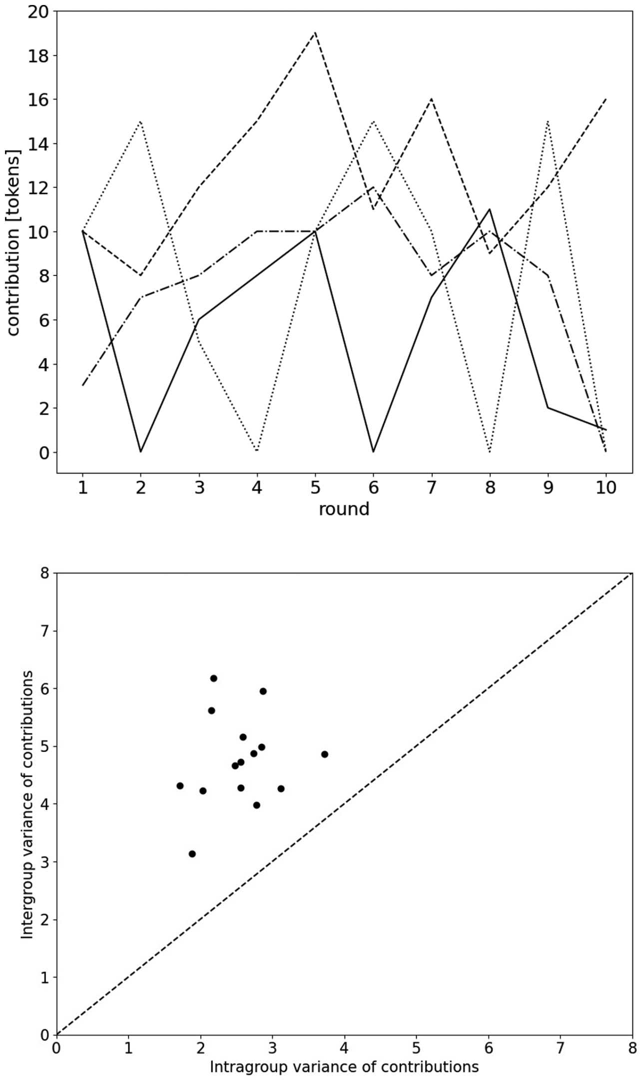


FIG. 1. Top: Trajectory of a single group game played among citizens of Athens [17]. Contributions of all four players are shown, in different styles. Bottom: Intragroup vs intergroup variance in the average contributions of players. Each data point corresponds to one city. The significant offset of the data above the fist 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 kth agent we shall often refer to the other agents as the *system*, and we call the kth 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 kth agent at turn $t \in \{1, \dots, T\}$.

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

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

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

(5) $\mathcal{G}_k[\theta_t^T] = \sum_{t'=t}^{T} G_k(\bar{f}_{t'})$ 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 kth 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) $\bar{f}_{-k,t} :=$ The contribution of all agents except the kth agent at turn t (we may identify $\bar{f}_t = (f_{k,t}, \bar{f}_{-k,t})$).

(2) $\pi_t^T = (\bar{f}_{-k,t}, \bar{f}_{-k,t+1}, \dots, \bar{f}_{-k,T})$, 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

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

Note that there are two kinds of processes at play here. First, the stochastic process describing the system (from the perspective of the kth agent) which generates $\bar{f}_{-k,t}$. 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

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

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

$$V_t[P(f_t^T)] \longrightarrow \max, \tag{3}$$

where

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

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

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

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

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

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* $0 \leqslant \gamma \leqslant 1$, which is the factor by which the future gains are discounted in comparison to immediate rewards. One can then write

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

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

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

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

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

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

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

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

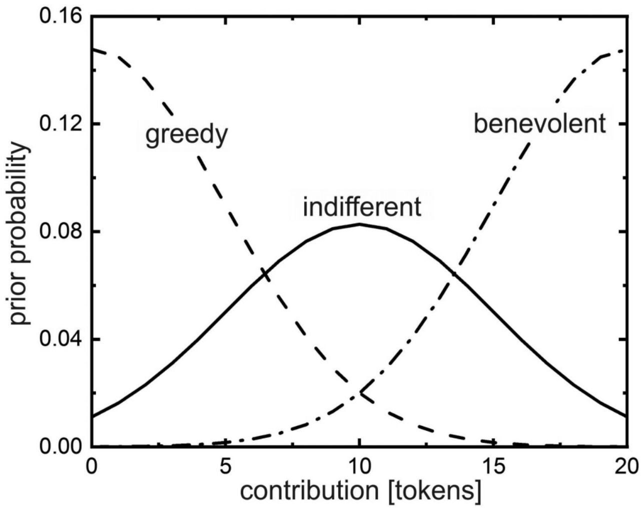


FIG. 2. Basal tendencies of the agents $P_0(f_{k,t})$ as given by truncated Gaussian prior distributions with variable m.

likely value of the Gaussian given by

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

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 $\xi_\pm$ 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 $\xi_\pm$.

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 POLICY M, t_max
2:     t = t_max
3:     E = 0
4:     V = 0
5:     while t > 0
6:         for 0 ⩽ f_{1,t−1} ⩽ τ do
7:             for 0 ⩽ μ_{t−1} ⩽ do
8:                 for 0 ⩽ a ⩽ τ do
9:                     E[a] = GAIN(Q, a, μ_{t−1}, f_{1,t−1}, V)
10:                end for
11:                j = arg max E
12:                policy[t, f_{1,t−1}, μ_{t−1}] = (δ_{kj})_{k∈{0,...,(N−1)τ}}
13:                V = policy[t, f_{1,t−1}, μ_{t−1}] · E
14:            end for
15:        end for
16:        t = t − 1
17:    end while
18:    RETURN policy
19: end function
```

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 $\xi_\pm$. 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

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

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 $D_{KL}(P_0||\delta) \leqslant K$ 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 $P^*(f_{k,t}) = cP_0(f_{k,t})\exp(\beta E[f_{k,t}]), \quad f_{k,t} \in \{0, \dots, \tau\}$ 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 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 $(\xi_\pm, m, K, \gamma)$.

1. Choice of $\xi_\pm$

In order to understand the significance of $\xi_\pm$, 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 ($r \leqslant 0.3$), where the dependence of $\langle A\rangle$ is steep, and levels off at minimal contributions for $r \leqslant 0.1$. $r = 0$ 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 $\xi_\pm$ 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

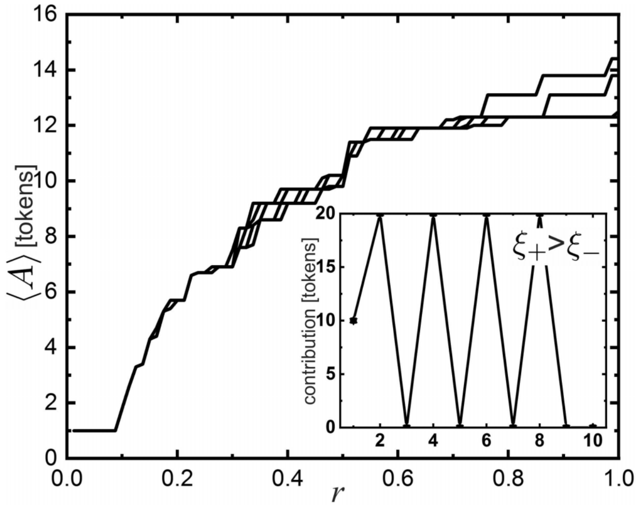


FIG. 3. Group average contribution, $\langle A\rangle$, as a function of $r = \sqrt{\xi_-^2 + \xi_+^2}$ for four different values of the polar angle, $\theta \in \{0, \frac{\pi}{6}, \frac{\pi}{3}, \frac{\pi}{2}\}$. The data collapse shows that the polar angle (hence the ratio ξ_+/ξ_-) is not relevant 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 $\xi_\pm$. 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 $\bar{f}_0 = (10, 10, 10, 10)$, 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 $\xi_\pm$, $r = 0.26$ 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) with $(m, K, \gamma) = (5, \infty, 1)$ and one agent (solid circle) with $(m, K, \gamma) = (5, \infty, 0.7)$.

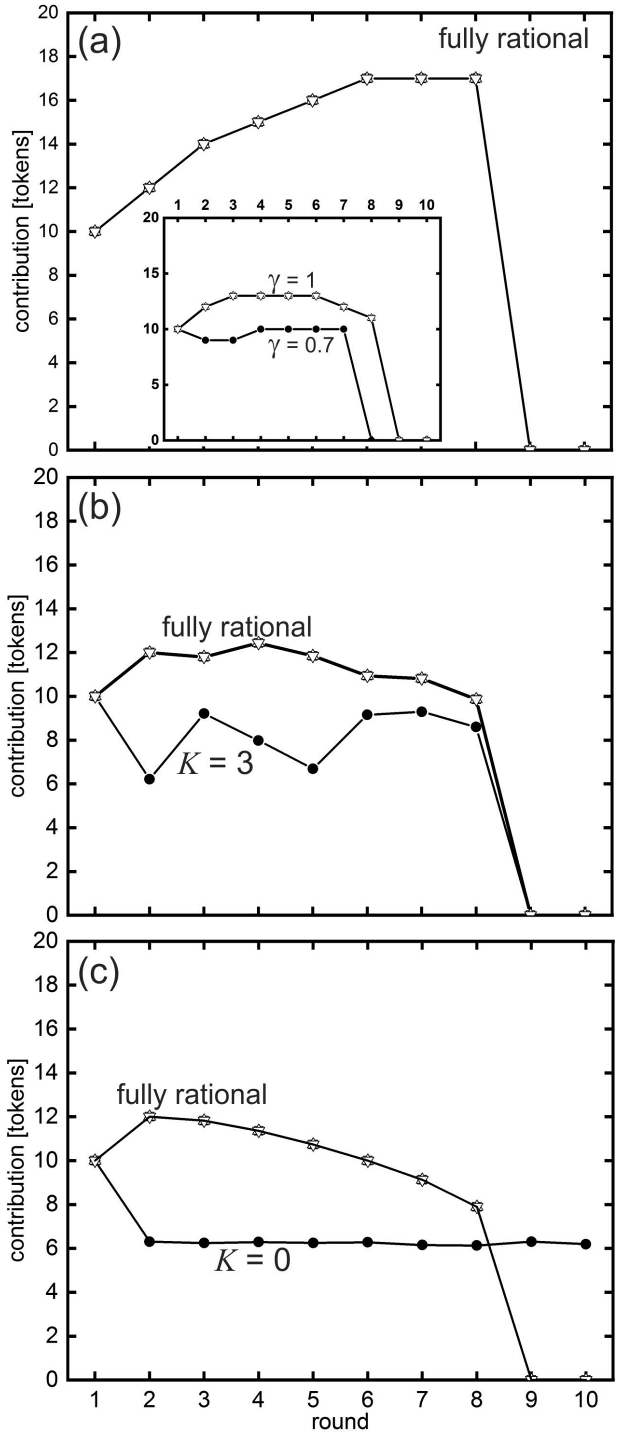


FIG. 4. Full game trajectories of a group of four fully rational agents. In (a) all agents are $(m, K, \gamma) = (5, \infty, 1)$. For the inset, one agent has different $\gamma = 0.7$ (solid circles). Panels (b) and (c) show ensemble average trajectories over 10 000 simulation runs with three agents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

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 $\approx$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

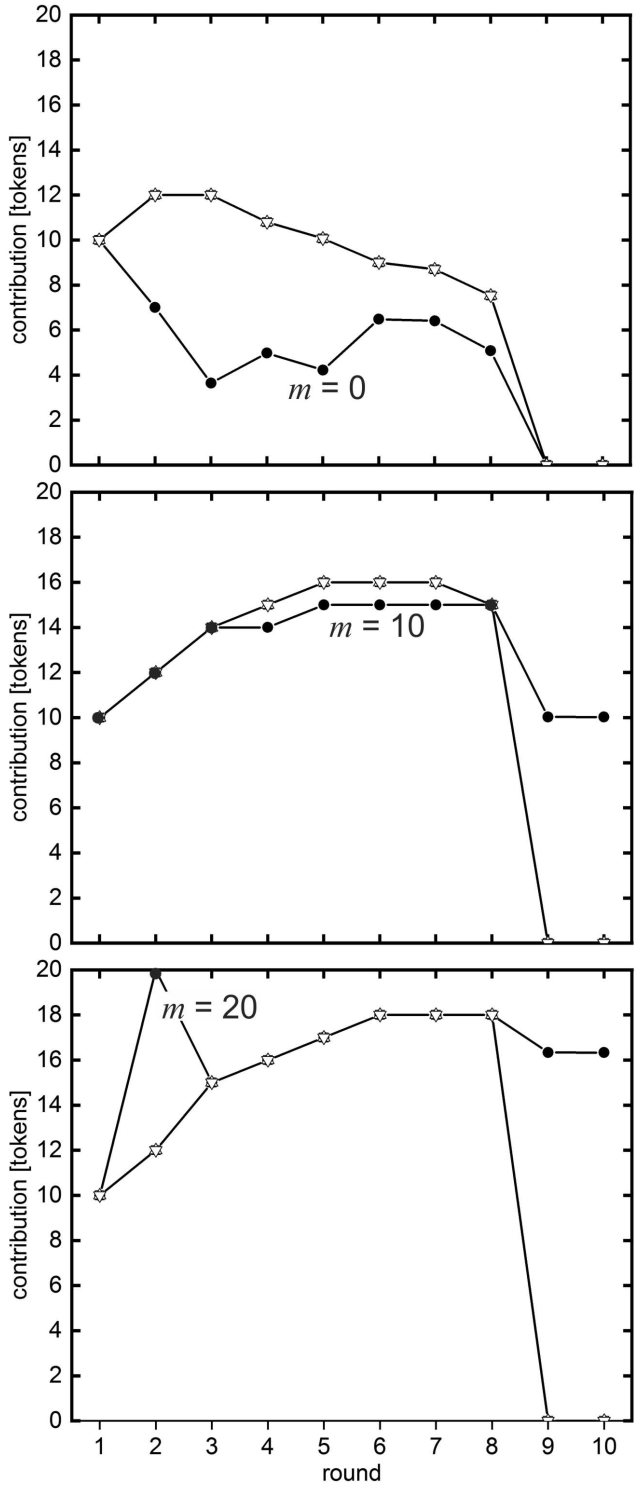


FIG. 5. Full game ensemble average trajectories of group of four agents. There are three identical agents (open circles) with $(m, K, \gamma) = (5, \infty, 1)$ and one different agent (solid circle) with (a) $(m, K, \gamma) = (0, 3, 1)$, (b) $(m, K, \gamma) = (10, 3, 1)$, and (c) $(m, K, \gamma) = (20,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

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 $\langle A \rangle$.

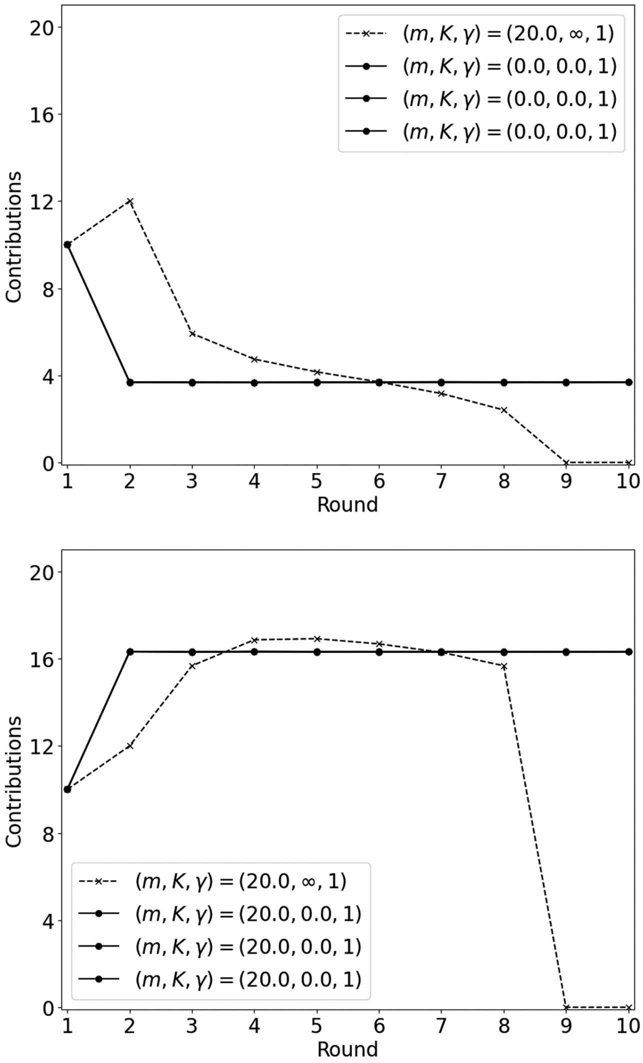


FIG. 6. Demonstration of coupling among agents within one group. A benevolent rational agent (dashed curve) with $(m, K, \gamma) = (20, \infty, 1)$ is made to play in two different systems, with either $m = 0$ (greedy, top, solid curves) or $m = 20$ (benevolent, bottom, solid curves). The system agents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

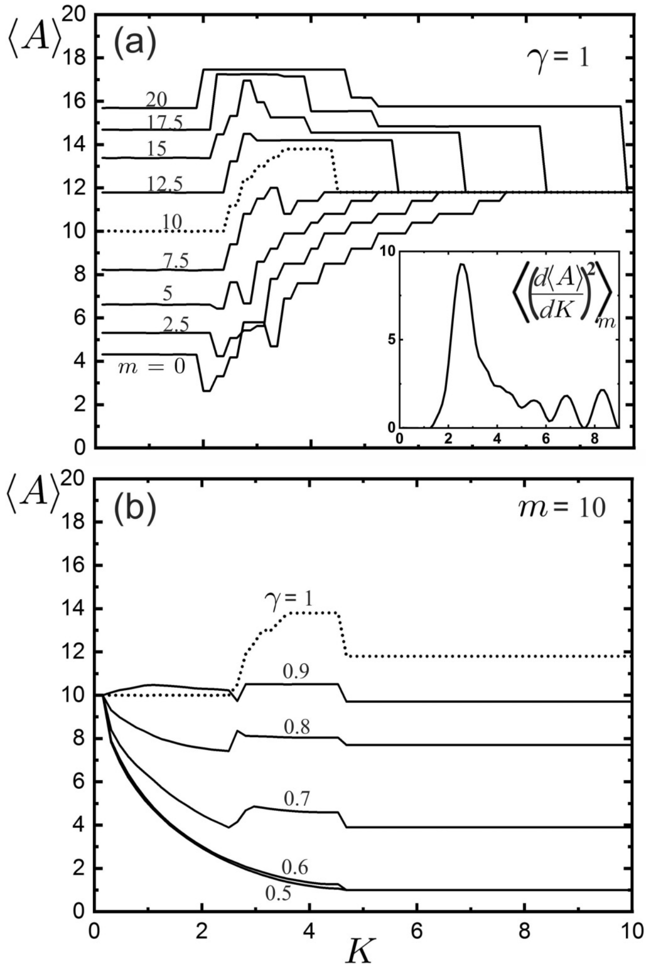


FIG. 7. The dependence of the average contribution $\langle A \rangle$ on agent parameters in a group of identical agents. The ensemble average was determined over 10 000 runs. In each panel, the dotted curve is for $(m, \gamma) = (10, 1)$.

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.

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 $(Q(\bar{f}_{-k,t}|\bar{f}_{t-1}), K_k, m_k, \gamma_k)$, 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

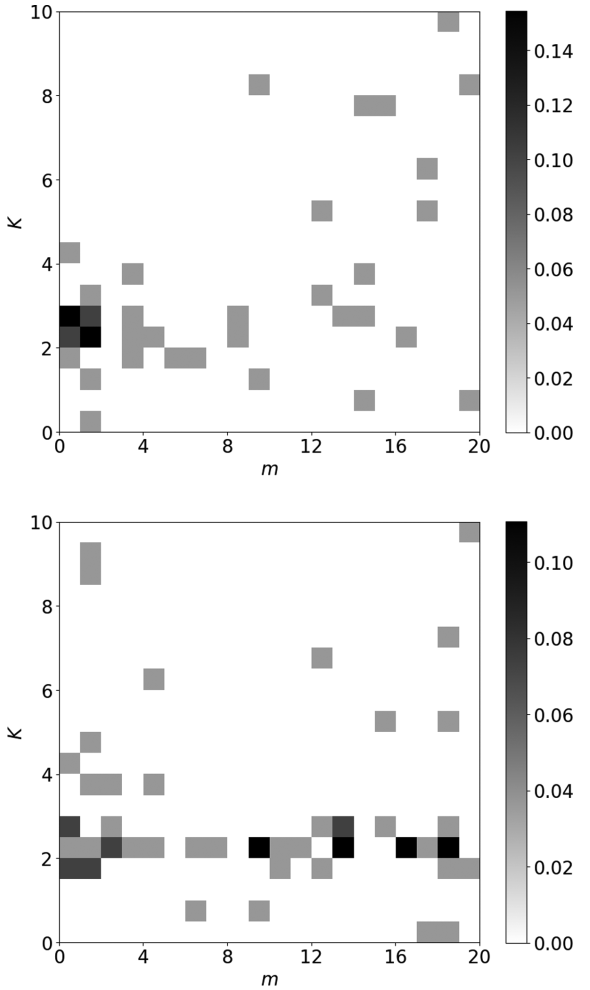


FIG. 8. Joint distributions of K and m for Melbourne (top) and Boston (bottom). There seems to be a clear preference for $K \approx 2.5$.

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 $\vec{f}_0$ 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

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

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

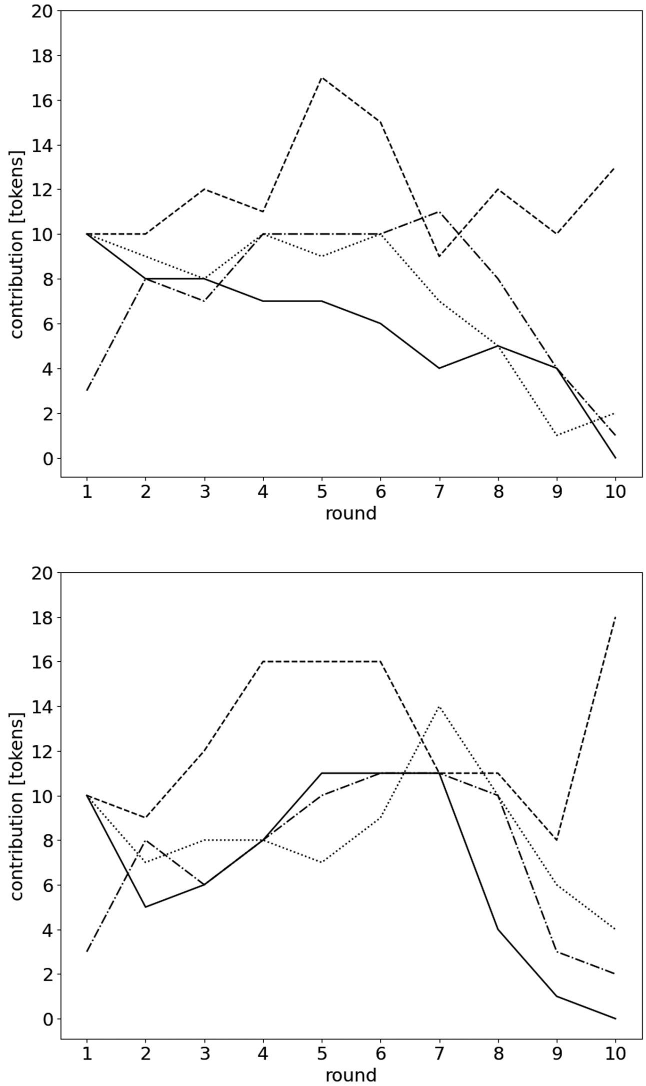


FIG. 9. Two trajectories simulated under identical conditions, with parameters obtained from fitting the group trajectory displayed in Fig. 1(a) (data from a Athens group). The same group of agents yields a different trajectory each time the simulation is run, due to the inherent randomness of the model. 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, $\langle \cdot \rangle_{(\gamma, m_1, \ldots, m_4)}$ denotes the average over multiple simulation runs of the group defined by the parameters $(\gamma, m_1, \ldots, m_4)$.

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

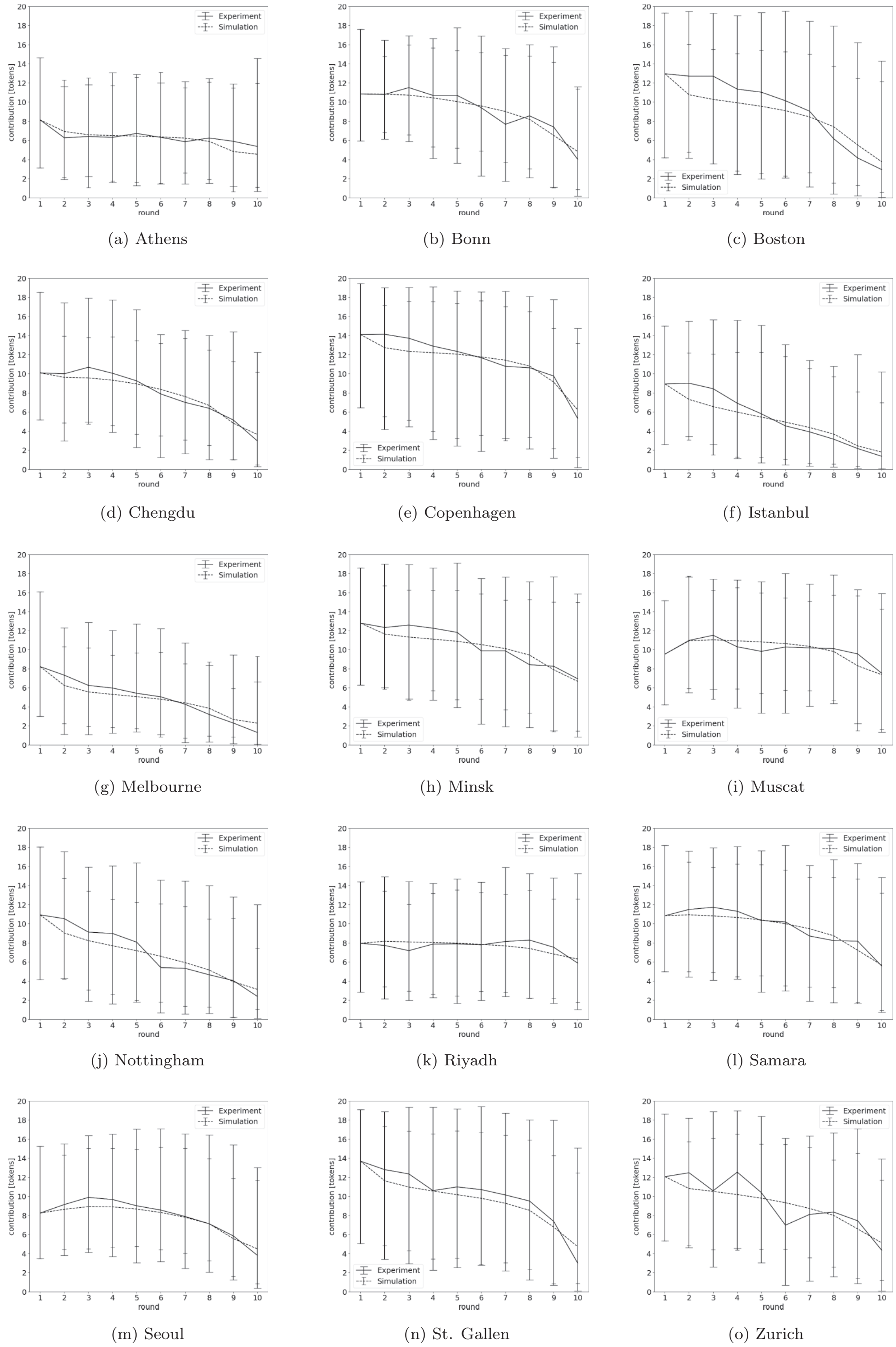


FIG. 10. Actual and simulated city-averaged contributions. The error bars indicate the variance of contributions.

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

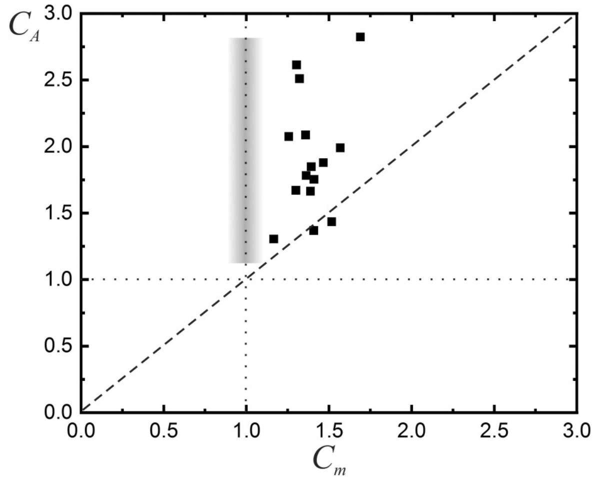


FIG. 11. Coupling in A vs coupling in m as obtained from the fitted simulations. The gray bar represents the data from Fig. 1(b).

$K = 2.5$ 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

$$C_m = \frac{\mathrm{var}\{m|\mathrm{intergroup}\}}{\langle \mathrm{var}\{m_k|\mathrm{intragroup}\}\rangle_l} \tag{12}$$

and

$$C_A = \frac{\mathrm{var}\{\langle A\rangle|\mathrm{intergroup}\}}{\langle \mathrm{var}\{\langle A_k\rangle|\mathrm{intragroup}\}\rangle_l}, \tag{13}$$

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-

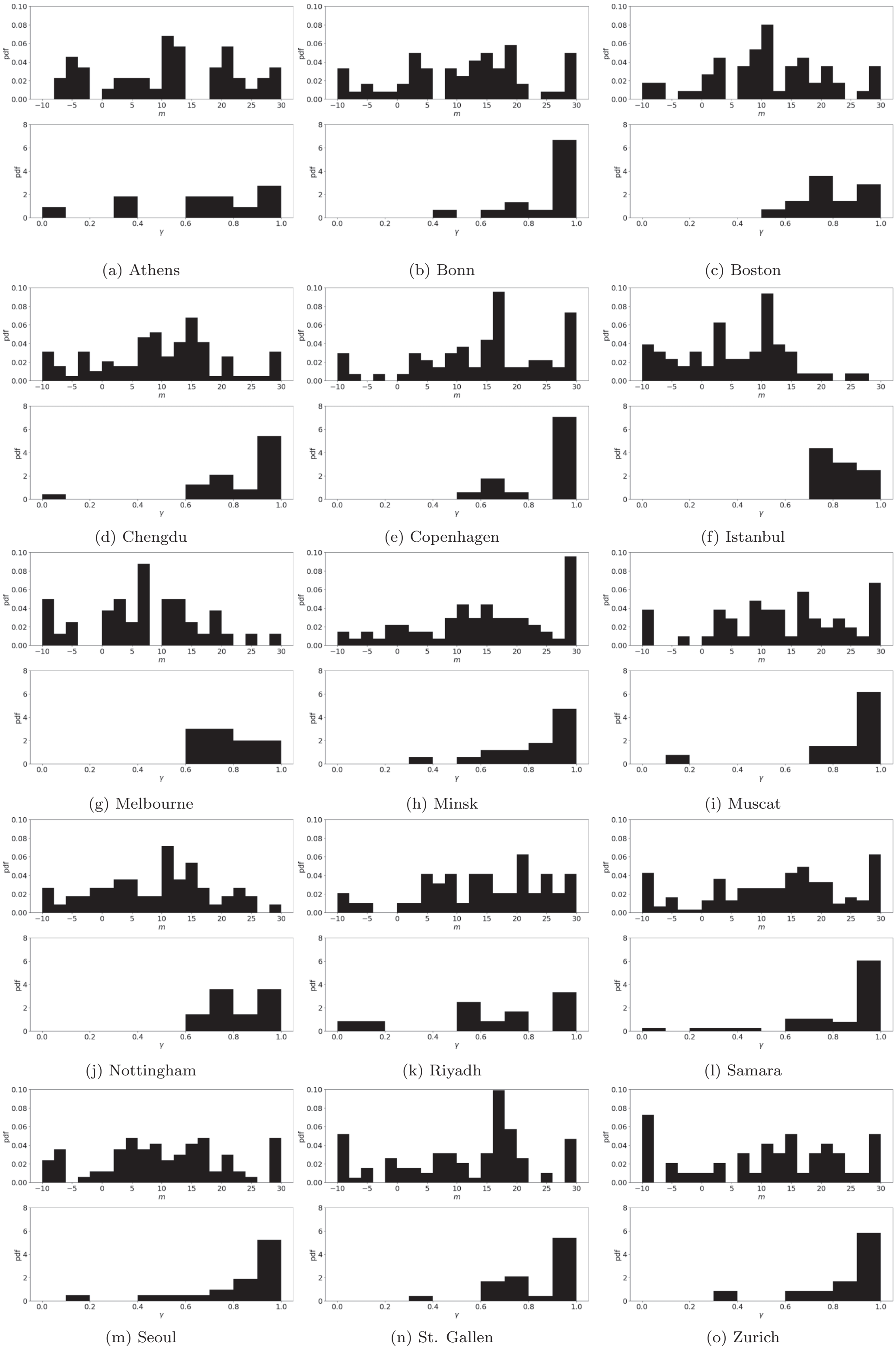


FIG. 12. Marginal distributions of γ and m as obtained by performing the fits for various cities.

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 $\xi_{\pm}$ 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 Multidemensional 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. Gunnthorsdottir, 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).

Appendix B

Manuscript: Bounded learning and planning in public goods games

PHYSICAL REVIEW E **107**, 054140 (2023)

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.

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

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

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

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

(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:

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

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

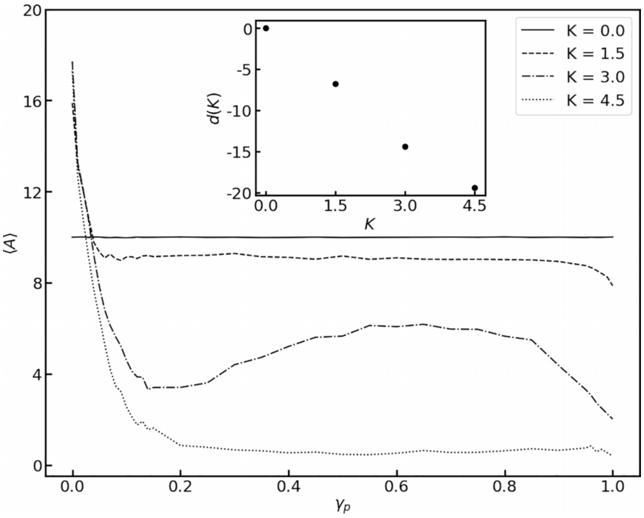


FIG. 1. Average contributions $\langle A \rangle$ as a function of learning strength γ_p for various values of K. Inset depicts the dependence of the decay rate of $\langle A \rangle$ with respect to K.

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.

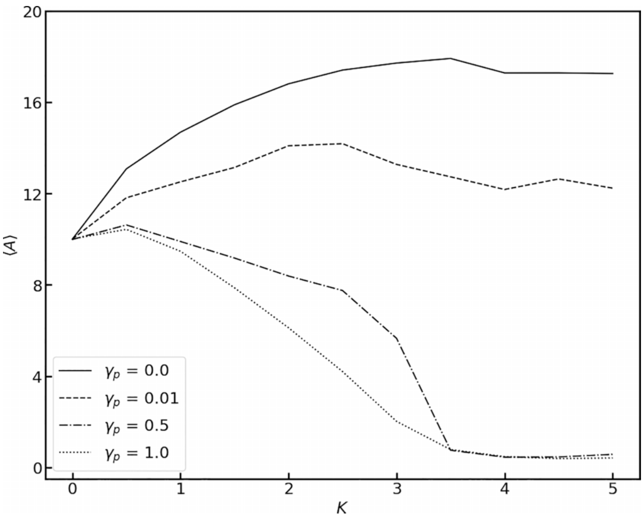


FIG. 2. Average contributions $\langle A \rangle$ as a function of computational budget K for various values of γ_p.

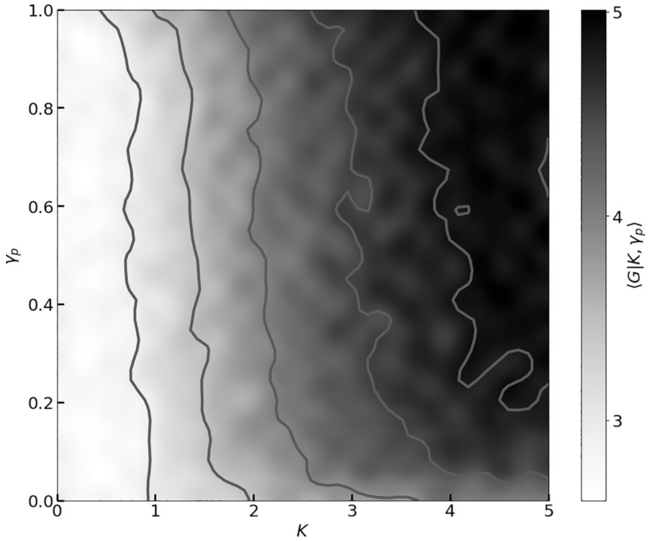


FIG. 3. Conditional expected gains $\langle G|K, \gamma_p\rangle$ (colorbar) and contours (solid grey curves) at $\langle G|K, \gamma_p\rangle = 3, 3.5, 4, 4.5, 4.8$.

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

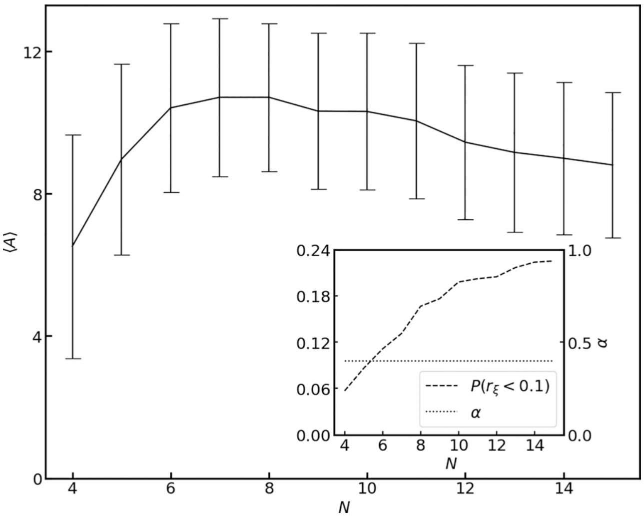


FIG. 4. Average contributions as a function of group size N and the variance (errorbars) for constant α. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

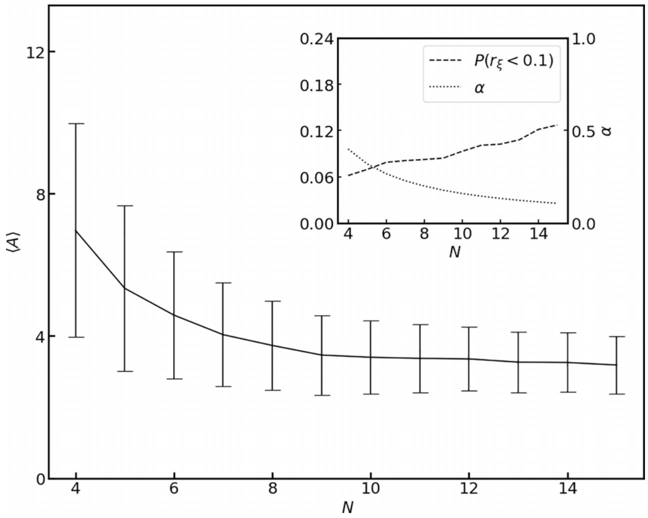


FIG. 5. Average contributions as a function of group size N and the variance (errorbars) for $\alpha \sim 1/N$. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

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).

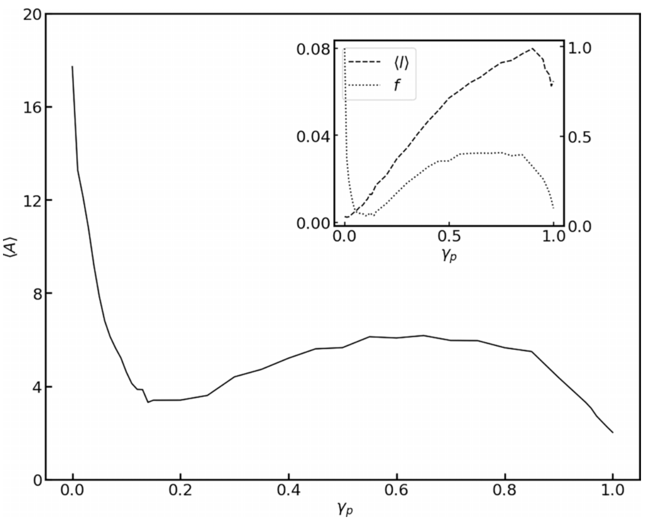


FIG. 6. Average contribution of groups of identical agents with $K = 3$. Inset shows the corresponding values of $\langle l \rangle$ and f as a function of γ_p.

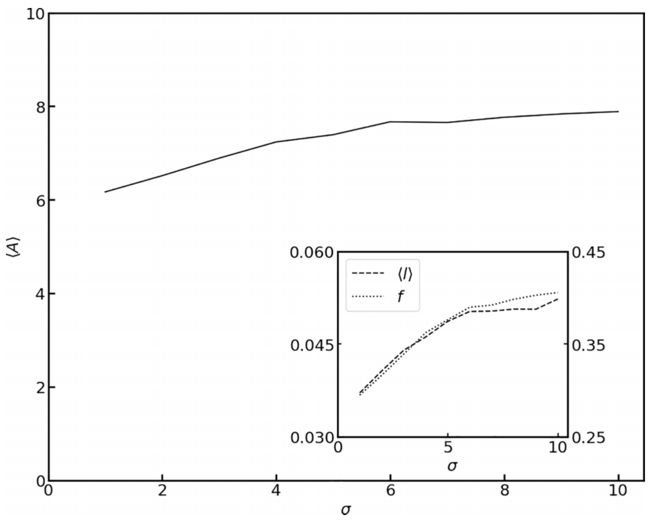


FIG. 7. Average contribution of groups of three randomly chosen agents and one noisy agent with variance of contributions given by σ. Inset shows the corresponding values of $\langle l \rangle$ and f as a function of σ.

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

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

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

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

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 $r_\xi \to 0$

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

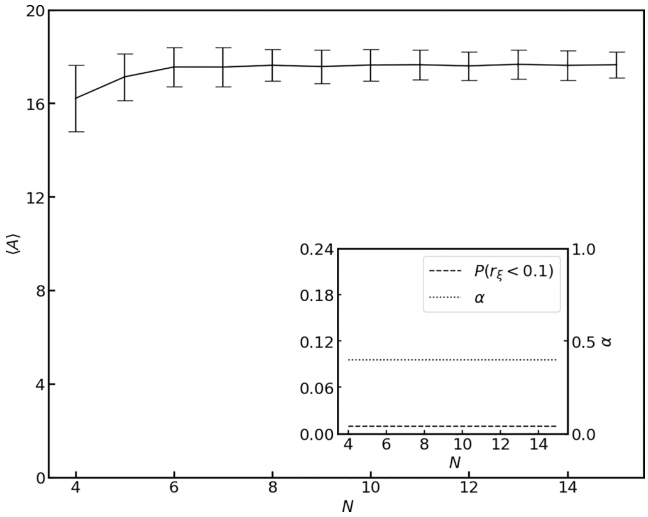


FIG. 8. Average contributions as a function of group size N and the variance (errorbars) for constant α and $\gamma_p = 0$. Inset shows $P(r_\xi < 0.1)$ and MPCR α as a function of group size N.

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).

Appendix C

Manuscript: Public goods games played on hypergraphs, by agents with bounded learning and planning

Public goods games played on hypergraphs, by agents with bounded learning and planning

Prakhar Godara* and Stephan Herminghaus†

Max Planck Institute for Dynamics and Self-Organization (MPIDS), Am Faßberg 17, D-37077 Göttingen, Germany

(Dated: March 6, 2023)

Public goods games between model agents with bounded rationality and a simple learning rule, which have been previously shown to represent experimentally observed human playing behaviour, are studied by direct simulation on various lattices with different network topology. Despite strong coupling between playing groups, we find that average investments do not significantly depend upon network topology, but are determined solely by immediate local network environment. Furthermore, the dependence of investments on characteristic agent parameters factorizes into a function of individual cognitive budget, K, and a simple function $1/(1 + c(0)/\beta)$, where $c(0)$ is the group centrality and $\beta = 12.5$ for all networks investigated. Given the good agreement of agent behaviour with available experiments, this seems to indicate that even complex societal networks of investment in public goods may be accessible to predictive simulation with limited effort.

I. INTRODUCTION

The sustainable management of the ecological niche of humankind on planet Earth is increasingly being recognized as a sizeable problem, attracting growing interest of scientific research across disciplines [1–3]. Aside from limited planet resources and rapidly changing climate conditions, a topic of major concern is the possible response of human societies to such stimuli [3–5]. In order to anticipate these responses, and to potentially advise policy makers to come up with appropriate legislatory precautions, a thorough understanding of the collective behaviour of humans in dense societies needs to be achieved.

Given the complexity of the behaviour even of a single human, this may seem entirely out of reach at first glance. However, it is well known that in systems consisting of sufficiently many similar entities in mutual interaction, quite precise predictions may be possible on collective phenomena, even if little is known about the individual entities. We can precisely predict, e.g., the critical exponents of all singular quantities of a condensed matter system close to a phase transition solely from the number of degrees of freedom of the order parameter and its dimensionality, without even knowing which molecules the system is composed of [6]. Similar statements hold even if none of the many constituents (i.e., molecules) are identical, as may be the case in polymers [7]. In fact, many similarities between phase transitions and collective phenomena in societies have been demonstrated [8–13]. Consequently, methods of statistical physics are meanwhile widely applied to social systems, with ever growing success [11–14]. The main challenge, and subject of lively debate [3–5, 11, 15–20], is to come up with model agents which are sufficiently simple to allow for large scale simulations, and at the same time model traits of human interaction behaviour which are relevant for the respective study.

* prakhar.godara@ds.mpg.de (corresponding author)

† stephan.herminghaus@ds.mpg.de

II. PUBLIC GOODS GAMES

A well established, and much studied, paradigm of interaction of human agents in societies is the well-known public goods game (PGG). In its standard game-theoretic setting, the PGG is played among N players over a total of T periods, where T is known to all players. In each period $t \leq T$, each player i is given a fixed integer number, τ, of tokens, a fraction $f_{i,t} \in [0, \tau]$ of which they can invest anonymously into a public pot[1]. The immediate reward of the ith player in that period is given by

$$G_{i,t} = \alpha \sum_{j=1}^{N} f_{j,t} - f_{i,t} \tag{1}$$

where α is multiplying factor which is known to all players[2]. This can be written as

$$G_{i,t} = \alpha(N-1)\mu_{i,t} - (1-\alpha) f_{i,t}, \tag{2}$$

where $\mu_{i,t} = \frac{\sum_{k \neq i} f_{k,t}}{N-1}$ is the average contribution of other players. The total gain for the ith player can then be defined as $G_i = \sum_t^T G_{i,t}$.

[1] τ thereby just provides a scale of coarse-graining of the investments. Following a widely followed convention in the field [20, 21], and without loss of generality, we use $\tau = 20$ throughout this article.

[2] It can be seen from Eq. 2 that the game has very trivial optimal strategies for $\alpha \geq 1$ (contribute everything) or $\alpha \leq \frac{1}{N}$ (contribute nothing) and therefore the only "interesting" case where the personal and collective gains conflict and give rise to a social dilemma is the case when $\frac{1}{N} < \alpha < 1$. Outside this range, α does not have much qualitative impact on game dynamics. Therefore we set $\alpha = 0.4$ throughout this article, which is a value commonly found in literature.

This game may be taken as a model of a vast number of real situations of human interaction. In a shared household, participants fulfill chores as an investment (f_i) into a pleasant atmosphere as the common good (G), members of a sports club invest donations (f_i) for enjoying common goods such as a well-maintained playing court (G), members of an association invest personal commitment (f_i) in the common good of a thriving association (G), and so on. A substantial fraction of societal interaction can, in this paradigm, be viewed as an enormously complex network of PGG, which interact through agents participating in several PGG simultaneously. Judging from the mentioned analogies to condensed matter systems, one might anticipate that the topology of this network has a major impact upon the collective investment behaviour of the agents.

Eq. 2 describes the gain for a given player playing in one period, in one group. In a network of interacting PGG groups, the corresponding gain becomes the sum of gains over all groups the agent plays in. Notice that the gains in each of these groups are independent. Therefore, in order to find a favourable policy in a particular group, each agent needs to keep in mind the actions of other players in that group alone. The dynamical interaction between groups comes about through the learning of the agents, i.e., their gradually updating their internal model of other player's behaviour. We assume that each agent bears one such model for players in general, i.e., for the players of all groups it is playing in.

The PGG networks modeling a society are very complex as compared to the networks of nearest-neighbour interaction in a condensed matter system. In the latter, interactions are all of one kind (or a few) and extend over a set of spatially nearest (perhaps including next nearest) neighbors. Coordination numbers (i.e., network degrees) are small and homogeneous, and the topology of the network is strictly determined by the dimensionality. In contrast, the number of games agents participate in may vary strongly, and the topology of the network of interactions between groups, the size of which may also vary vastly, does not need to bear any relation to a network of proximity in a space of fixed (integer) dimensions. Hence one must be prepared to find qualitative differences to collective behaviour in condensed matter systems when simulating complex networks of PGG, and universality classes as in phase transitions may not hold. As we will show, the investment behaviour of agents can most probably be anticipated solely on the basis of rather local network descriptors, and network topology is found to be surprisingly irrelevant.

A. PGG agents

We employ model agents we have developed before [20, 22], guided by progress in intelligence research [23, 24]. While it has been claimed that reinforcement learning is *the* essential ingredient of agents reproducing experimentally observed PGG playing behaviour [18, 25], we could show that agents with bounded rational foresight with no learning capabilities at all may provide an even better account of experimental data [20, 21]. Not only do they quantitatively reproduce the commonly observed downward trend of average investments during a game, they are also capable of accounting for the substantial period-to-period variance of these investments [21]. When bestowing a simple learning mechanism to these agents, as appears plausible for long-time playing settings, we furthermore obtained complex dependencies of average investment upon group size [22], as had been observed in a number of real-world settings [26, 27]. The resulting model agent is still sufficiently simple to be used in large scale simulations such as those needed to investigate collective behaviour in complex PGG networks.

Consequently, in the present paper we consider bounded learning and planning agents as described before [20, 22], playing a spatially extended PGG (SPGG). The agent model is composed of two parts - *learning* and *planning*. Learning refers to agents observing the past game behavior of other agents in order to make a predictive model of them, while planning refers to making use of this model to chart out an optimal course of action in a particular game. In the following we describe the learning and planning mechanisms of an agent in a single group.

B. Planning mechanism

To describe the planning mechanism, we assume that agent i has a predictive model of other agents incorporated as a transition function,

$$Q(\mu_{i,t}|\mu_{i,t-1}, f_{i,t-1}) = g(\mu_{i,t}; p, \sigma). \tag{3}$$

It describes the likelihood of other agents' next action (expressed by $\mu_{i,t}$), given the previous state of the game $(\mu_{i,t-1}, f_{i,t-1})$. $g(\mu; p, \sigma)$ is the truncated Gaussian distribution on the interval $[0, \tau]$, where σ is the variance of the distribution, which we fix to $\sigma = 3$ throughout the article. $p(\mu_{i,t-1}, f_{i,t-1})$ is the peak position of the distribution as given by

$$p = \begin{cases} \mu_{i,t-1} + \xi^i_+ |\mu_{i,t-1} - f_{i,t-1}|, & \mu_{i,t-1} - f_{i,t-1} < 0 \\ \mu_{i,t-1} - \xi^i_- |\mu_{i,t-1} - f_{i,t-1}|, & \mu_{i,t-1} - f_{i,t-1} > 0. \end{cases} \tag{4}$$

where ξ^i_+ and ξ^i_- parameterize Q. As the ith agent can influence the actions of others solely through its contributions (because players play anonymously and do not interact otherwise), the parameters ξ^i_+ and ξ^i_- describe to which degree agent i believes to be able to encourage or discourage other agents to contribute. In that sense, $\xi^i_\pm$ is the model the agent has of the other agents. As far as planning is concerned, we do not bother about how the agent comes up with a particular set of values $\xi^i_\pm$, but rather what decisions does the agent come up with, given Q.

Let us assume that the agent has a policy of its own actions from period t until the end of the game (period T) given by $P(f_t^T) = \prod_{t'=t}^{T} P(f_{i,t'}|\mu_{i,t'-1})$. By making use of Q and $P(f_t^T)$ one can write the path probabilities of various game trajectories. Combining it with the gain of the agent associated with a trajectory, one can write down the expected utility for the policy as a path integral. This expected utility can then be maximized to find the optimal policy. Finally, the path integral maximization can be written in a recursive form, described by a Bellman equation [28] given by

$$V_t^* = \max_{P(f_t^T)} \sum_{f_{i,t},\mu_{i,t}} P(f_{i,t}|\mu_{i,t-1}) \Big[Q \cdot G_{i,t}(f_{i,t}, \mu_{i,t}) + \gamma Q \cdot V_{t+1} \Big], \tag{5}$$

where * is to indicate a maximized quantity. $\gamma \in [0, 1]$ represents a foresight which exponentially 'decays' into the future, discounting for future rewards. If we also include bounds on the computational capabilities of the agent [20], an additional term appears in Eq. 5. which then becomes

$$V_t^* = \max_{P(f_t^T)} \sum_{f_{i,t},\mu_{i,t}} P(f_{i,t}|\mu_{i,t-1}) \Big[Q \cdot G_{i,t}(f_{i,t}, \mu_{i,t}) - \frac{1}{\beta} \log \frac{P(f_{i,t}|\mu_{i,t-1})}{P_0(f_{i,t}|\mu_{i,t-1})} + \gamma Q \cdot V_{t+1} \Big]. \tag{6}$$

The Lagrange parameter β comes about from the additional constraint $\frac{1}{\beta}(D_{KL}(P^*(f_{i,t}|\mu_{i,t-1})||P_0(f_{i,t}|\mu_{i,t-1})) - K) = 0$, where K is the computational budget of the agent. Essentially, the additional term penalizes high Kullback-Leibler (KL) divergence of the posterior (P) and the prior (P_0) policies, for some given P_0 for the agent. The solution of the optimization problem in Eq. 6 then provides us with the bounded optimal policy $P(f_t^T) = \prod_{t'=t}^{T} P(f_{i,t'}|\mu_{i,t'-1})$.

In conclusion, in period t, given a model $\xi_\pm^i$, a prior policy P_0, a computational budget K and exponentially decaying foresight γ that describe agent i's properties, the agent makes use of Eq. 6 to evaluate the best policy. Now we concern ourselves with how the agent comes up with its model of other agents i.e. $\xi_\pm^i$.

C. Learning mechanism

Following [22] we consider the learning mechanism as a maximum likelihood estimation over the transitions observed by the agent. The task of learning for agent i in period t then is the task of finding the tuple $\xi_\pm^i$ that is most representative of the state transitions observed by the agent. This can be written as

$$\xi_\pm^{i\,*}(t) = \arg\max_{\xi_\pm^i(t)} \Bigg[\sum_{w=t-1}^{2} \gamma_p^{t-w} \log Q(\mu_{i,w}|\mu_{i,w-1}, f_{i,w-1}) - (1-\gamma_p)(\xi_\pm^i(t) - \xi_\pm^i(t-1))^2 \Bigg]. \tag{7}$$

The summation is over all observed past and the summand is the weighted log likelihood of having observed a particular transition from period $w-1$ to period w. The exponential weighting factor γ_p represents the recency bias or the bounded memory of the agent [29] as it weights recent transitions more than earlier transitions. Finally the last summand represents the tendency of the agent to not update it's model. Therefore when $\gamma_p = 0$ the agent performs no learning and maintains it's previously believed model, and when $\gamma_p = 1$ the agent looks arbitrarily back in the past and performs the maximum likelihood estimation and updates it's model.

As we are interested in collective effects here, we allow all players to play in multiple groups simultaneously. Their rewards in single playing groups are evaluated independently and then summed over all groups they play in. In this setting, beyond just the intra-group dynamics, the network of connections between groups and players also becomes of relevance through the fact that players play in several groups.

III. PGG NETWORKS

The interaction structure for players playing multiple PGGs simultaneously with different groups of players is best captured by *hypergraphs*. In what follows, we give a brief account of the concept hypergraphs, and how connects to our system of interest. We will then describe how to adapt the behavioral dynamics, as described for single groups in the previous section, to the spatially extended setting of an SPGG.

A. Hypergraphs

A hypergraph [30] is a graph whose edges are allowed to connect more than two nodes. An edge in a hypergraph is called a *hyperedge*. Hence a hypergraph $\mathcal{H}$ can be defined as a tuple $\mathcal{H} = (V, E)$, where V is a finite set of nodes v_i indexed by $i \in \{1, \ldots, |V|\}$ and E is a finite set of hyperedges e_j indexed by $j \in \{1, \ldots, |E|\}$, where each hyperedge is a non-empty subset of V. One can completely specify a hypergraph by an *incidence matrix* H_{ij} where $i \in \{1, \ldots, |V|\}$ and $j \in \{1, \ldots, |E|\}$ and

$$H_{ij} = \begin{cases} 1 & v_i \in e_j, \\ 0 & \text{else.} \end{cases} \tag{8}$$

In this work, we identify the nodes in $\mathcal{H}$ as the agents and the edges are identified as the various groups the agents play in. As most literature on PGG considers four player games, in this work we consider agents playing on 4-*regular* hypergraphs[3], but each agent is allowed to play in arbitrarily many groups. Other than this we consider no restrictions on the topology of the hypergraphs.

B. Agent dynamics on hypergraphs

Now that we have introduced the interaction structure, our task remains to adapt the agent behavioral dynamics from Sec. II to the spatially extended setting. Let us consider an SPGG being played on a hypergraph $\mathcal{H}$ with $|V| = N_p$ players and $|E| = N_g$ groups. In order to carry out the discussion on the agent model we describe it from the perspective of an arbitrary agent given by an index $i \leq N_p$, which plays in $k \leq N_g$ different groups given by the indices $i_1, \ldots, i_k$.

As already mentioned, agent i plans in each of the k groups independently. This means that agent i in period t, makes use of Eq. 6 to find the optimal policy for itself and then, for each group it plays in, it independently samples an investment from the conditional distribution $P(f_{i,t}|\mu_{i,t-1})$ by conditioning on the observed $\mu_{i,t-1}$ of that group.

Learning, on the other hand, is not done independently in each group. Essentially, the exponentially weighted log likelihood in Eq. 7 is summed not only over all observed past of one group, but over all groups. By suitably modifying the notation to also include the group identity, one can write the analogous equation as

$$\xi^{i}_{\pm}{}^{*}(t) = \arg\max_{\xi^i_\pm} \sum_{m=1}^{k} \Bigg[\sum_{w=t-1}^{2} \Big(\gamma_p^{t-w} \log Q(\mu_{i,i_m,w}|\mu_{i,i_m,w-1}, f_{i,i_m,w-1}) \Big) - (1-\gamma_p)(\xi^i_\pm(t) - \xi^i_\pm(t-1))^2 \Bigg], \tag{9}$$

where the subscript i_m denotes that the quantity is being evaluated in group i_m with $m \leq k$.

The learning and planning mechanisms can be then combined into a single agent which is defined by a quadruple of parameters, (m, K, γ, γ_p), where m parameterizes the prior policy of the agent[4]. In each round $t \in \{2, 3, ...T\}$, the agent

1. **plans**: by considering the game state at $t-1$ in all the k groups it plays in and $\xi^i_\pm(t-1)$, the agent making use of Eq. 6 finds the best policy for itself.

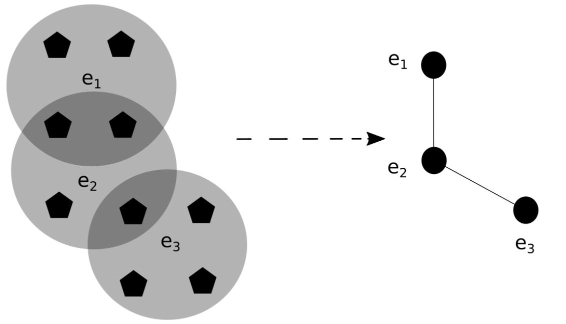


Figure 1: Projecting a hypergraph to a line graph. The hyperedges (e_1, e_2, e_3) of the hypergraph correspond to the vertices of the line graph. Black pentagons correspond to agents, black circles correspond to groups.

2. **acts**: by sampling an investment, $f_{i,i_m,t}$, from the evaluated policy, after conditioning on the observed $\mu_{i,i_m,t-1}$ of each group $m \in \{1, \cdots, k\}$.
3. **learns**: after observing the state of the game in the current period t in all the k groups the agent learns and updates it's model of the players by evaluating $\xi^i_\pm(t)$ from Eq. 9.

In period $t = 1$ the investments of the agent are sampled from it's prior distribution $P_0(f_{i,1})$ for all the groups it plays in. In period $t = 2$ there is certainly planning based on observed behavior, but there is no learning, as the agent has not yet observed a transition.

IV. TOPOLOGY AND DYNAMICS

In our previous studies [20, 22] we focused on how agent characteristics (m, K, γ, γ_p) impact the behavioral dynamics in PGG. Here we have, as an additional feature, namely, the interaction structure imposed by the hypergraph. It is then natural to investigate the specific impact of the interaction network structure on the agent dynamics, and thereby on possible collective phenomena in the system. This will be the main question we pursue in the remainder of this article. Before we proceed, however, we need to clarify our concepts of "topology" and "dynamics", and to introduce the key descriptors of topology of the hypergraph as well as the relevant observables we will use to describe the agent behavioral dynamics.

A. Topological features

Hypergraphs (as opposed to graphs) provide the opportunity to consider two distinct types of interactions: inter-group and intra-group. While our previous studies have focused primarily on intra-group interactions, we focus here on inter-group interactions and their effect

[3] A k-regular hypergraph H is a hypergraph such that $|e_i| = k$ for all $i \in \{1, \ldots, |E|\}$.

[4] We assume that the prior policy of the agent is given by the same distribution $P_0(f_{i,t})$ in each period t and the distribution is a truncated Gaussian centered around m and a fixed variance $\sigma_m = 5$.

on contributions. The smallest unit of interest in this paradigm is a group. This allows us to project the uniform hypergraph to a corresponding *line graph* $\mathcal{L}$[5]. Fig. 1 shows a schematic of the projection of a hypergraph to the corresponding line graph. Notice that the hyperedges e_1, e_2, e_3 become vertices in the line graph and the information about the location of vertices of the hypergraph (black pentagons) gets lost when performing this projection. Therefore in a line graph all of the intra-group structure is lost and it remembers only non-zero overlaps between groups.

Our task now remains to choose the appropriate descriptor of the graph topology. Studying SPGG, we are interested in evaluating the degree of influence (measured through the dynamical observables) a particular group (i.e., a node in the line graph) has on other groups in the graph depending, e.g., on the distance between them. This is to determine how 'far' the influence of a group travels within the graph. Hence we need a suitable descriptor which represents the notion of a distance in a graph topology, and is suitable to quantify such propagation of influence.

A well-established class of such descriptors is the *Bonacich-Katz* class of centrality measures [31]. The Bonacich-Katz centrality of a node i in the line graph $\mathcal{L}$ is parameterized by two parameters ω, η and is given by

$$c_i(\omega, \eta) = \omega(I - \eta J)^{-1} J 1, \tag{10}$$

where I is the identity matrix, J is the adjacency matrix of $\mathcal{L}$ and 1 is a column vectors of all ones. Here ω is a constant multiplying factor and therefore doesn't impact the centrality ranking of the nodes and can therefore be ignored via an appropriate normalization. η parametrises the expected radius of influence a particular group has on other groups, which is proportional to $(1-\eta)^{-1}$. It will be the main parameter of concern for us in this article. While for $\eta \to 0+$, c_i is equivalent to degree centrality and it corresponds to eigenvector centrality if $\eta \to \frac{1}{\lambda_{\max}}-$, where $\lambda_{\max}$ is the largest eigenvalue of the adjacency matrix of $\mathcal{L}$ [32].

B. Dynamical features

In order to quantify cooperation in PGG, a natural observable of interest is the average contribution of the group or player. Let us consider an agent i where $i \leq N_p$. Let the set of groups the agent plays in be given by $p_i = \{k | H_{ik} \neq 0\}$. Then we define the average agent contribution as

$$A_{i,t}^{\text{agent}} = \frac{\sum_{j \in p_i} f_{i,j,t}}{|g_i|}, \tag{11}$$

and the corresponding ensemble averaged quantity given by $\langle A_{i,t}^{\text{agent}} \rangle$. The corresponding average group contribution for a group j where $j \leq N_g$ is defined as

$$A_{j,t}^{\text{group}} = \frac{\sum_{i \in g_j} f_{i,j,t}}{4}, \tag{12}$$

where $g_j = \{i | H_{ij} \neq 0\}$ and the ensemble average quantity is given by $\langle A_{j,t}^{\text{group}} \rangle$. For both these quantities the corresponding time average quantity is given by $A_i^{'} = \frac{\sum_{t=0}^{T} A_{i,t}^{'}}{T}$.

The above quantities are averaged over time, and therefore hold no information regarding the interactions between groups unfolding over time. The latter can be investigated by considering temporal correlations between group trajectories. To this end, we consider the correlation matrix C_{ij}, in which the i, j entry is the correlation between trajectories of groups i and j. It is given by

$$C_{i,j} = \frac{1}{T} \sum_{t=0}^{T} \frac{\sigma_{A_{i,t}^{\text{group}} A_{j,t}^{\text{group}}}}{\sigma_{A_{i,t}^{\text{group}}} \sigma_{A_{j,t}^{\text{group}}}}, \tag{13}$$

where σ_{XY} is the covariance of random variables X, Y given by $\langle XY \rangle - \langle X \rangle \langle Y \rangle$ and σ_X is the variance of X. Because the correlation between two group averaged trajectories is symmetric over the two groups, it is a natural measure for us as the coupling between groups which is mediated by the learning mechanism of the common player(s) also has no way to break the symmetry between the groups.

V. RESULTS

As we intend to evaluate the specific impact of network topology on the dynamics, we will keep the agent characteristics homogeneous and consider heterogeneity only through graph topology. Therefore we consider identical agents that are embedded on a random hypergraph (for details on how these random hypergraphs are generated see App. A).

For these agents we fix $(m, \gamma) = (10, 0.9)$ for similar reasons as in [22] and additionally, we consider $\gamma_p = 0.9$. This choice reflects the fact that in our model the interactions between neighboring groups are mediated by learning. Lower values of γ_p would weaken the group interactions, thereby rendering the network structure pointless. On the other hand for $\gamma_p \approx 1$ it is observed that something similar occurs for longer times [22], as every new observation has only a diminishing impact on the agents' preferences. As we will see in the following, even though we create a setup (choice of parameters) that qualitatively maximizes the group interactions, the resulting dynamics still seems mostly independent of the various topological features.

We consider ten randomly generated hypergraphs with $(N_p, N_g) = (64, 25)$ with all the agents described by $(m, \gamma, \gamma_p) = (10, 0.9, 0.9)$ and $K \in \{0, 1.5, 3, 4.5\}$. For

[5] A line graph of a hypergraph $\mathcal{H} = (V, E)$ is a graph $\mathcal{L}(\mathcal{H}) = (V_\mathcal{L}, E_\mathcal{L})$, where $V_\mathcal{L} = E$ and two vertices e_i, e_j in are connected in $\mathcal{L}$ iff. $e_i \cap e_j \neq \emptyset$

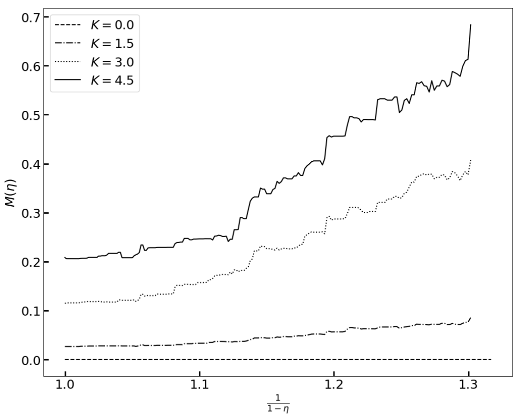


Figure 2: Dependence of conditional expected contribution variance, $M(\eta)$, upon radial distance of influence, as expressed by $(1-\eta)^{-1}$. The curves are averaged over ten hypergraphs of size $(N_p, N_g) = (64, 25)$ for agents with $K = 0$ through $K = 4.5$.

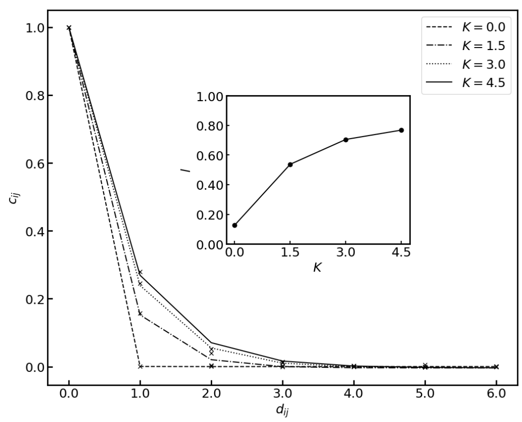


Figure 3: Group trajectory correlations as a function of the shortest distance between the groups for groups of agents with $K = 0$ through $K = 4.5$. Crosses are data points, polygons are exponential fits.

each random hypergraph we perform the ensemble average of 5000 simulation runs. Following the question asked in the beginning of Sec. IV we investigate, in a random hypergraph with all identical agents, what topological descriptor is the most appropriate to predict average group contribution. As already mentioned we consider the class of descriptors given by Bonacich-Katz centrality and we consider the most "appropriate" centrality measure (or the most appropriate η) as the one that minimizes the function

$$M(\eta) = \int \theta_{\langle A_j^{\text{group}} \rangle | (c_j(\eta)=c)} dc. \tag{14}$$

Here the variance is defined as

$$\theta_{\langle A_j^{\text{group}} \rangle | (c_j(\eta)=c)} = \text{VAR}\{\langle A_j^{\text{group}} \rangle | c \leq c_j \leq c + dc\}, \tag{15}$$

for some choice of discretization. Essentially, every value of η is an assignment of a centrality to each node. We wish to find that assignment η such that given a centrality $c(\eta) = c_0$, the variance of the average contributions corresponding to the nodes with the centrality close to c_0 is minimal when integrated over all c_0. In other words we wish to find η for which the scatter of the scatter plot between $\langle A_j^{\text{group}} \rangle$ and $c_j(\eta)$ is minimal.

In Fig. 2 we plot $M(\eta)$ curve averaged over 10 different random hypergraphs as a function of η and one can see that the global minimum of $M(\eta)$ is given by $\eta \approx 0$ for various values of agent parameters given by $K \in \{0, 1.5, 3, 4.5\}$ thereby indicating that the centrality measure with very small values of η best predicts the group average contribution independent of K. Recalling the definition of Katz centrality, smaller values of η correspond to smaller radii of influence. This then seems to indicate that it is the local topological features (in this case, node centrality i.e. number of neighbors) that are the best predictors of average group contribution.

In the above analysis there is no "dynamics" as such, as we have only considered the distribution of $\langle A^{\text{group}} \rangle$ across nodes of varying centrality. Therefore any claims of groups have a smaller "radius of influence" could be misguided if the above result is considered in isolation. In order to create a more robust picture of the radius of influence of a group, we consider how quickly do intergroup correlations decay as a function of the shortest distance between the groups.

Therefore, we proceed by comparing the correlation matrix C with the distance matrix D[6] to evaluate how the correlations between group average trajectories scale with the shortest distance between the groups. In Fig. 3 one can see that the correlation between group average trajectories falls down exponentially with the (shortest) distance between the groups with correlation lengths $l < 1$ (as shown in the inset). Thereby meaning that substantial correlations of a group's behavior is only with it's immediate neighbors. This further corroborates that group contributions in SPGG are mostly governed by local interactions.

The locality of interactions in SPGG is a good news for two major reasons. First and the more obvious reason is that locality simplifies the analysis of the system, thereby allowing for the possibility of developing simpler effective dynamics that replicate these observations. The second reason is that, if group contributions are mostly governed by local interactions (in this case group centrality $c(\eta = 0)$), then even if we have graphs with different

[6] A distance matrix D for a graph $\mathcal{L} = (V, E)$ is a matrix of size $|V| \times |V|$ where the i, j entry $D_{i,j}$ is the length of the shortest path between node i and node j in $\mathcal{L}$.

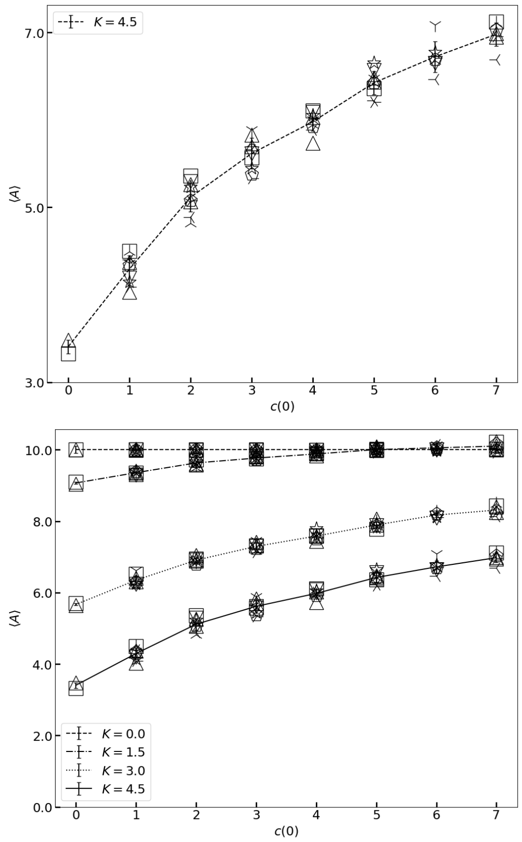


Figure 4: Impact of group centrality on the group average contribution for ten randomly generated hypergraphs. Top panel: results for $K = 4.5$, with different symbols corresponding to different hypergraphs. Bottom panel: same as top panel for different values of K. Polygons connect averages over all hypergraphs, respectively.

global structures but similar local structures we should observe similar behavior. The latter would seem to indicate a universality (i.e. global topology independent) in cooperation behavior across a large set of networks.

Fig. 4 presents the impact of group centrality ($c(\eta = 0)$) on the group average contribution for different values of K, for ten randomly generated hypergraphs. Different symbols correspond to different hypergraphs. For each value of K, the results are found not to differ significantly for different hypergraphs. This demonstrates that topological features of the hypergraphs are irrelevant to this end. Cooperation levels of groups are predominantly determined by the number of neighbors of the group. Quite surprisingly, one does not even need to consider how many players are shared between two neighbor groups (recall that $c(0)$ is measured from the line graphs and not the hypergraphs).

VI. DISCUSSION

Another observation from Fig. 2 is that $M(\eta)$ monotonically increases with K, attaining a minimal value at $\eta = 0$ for all K investigated. Having a higher scatter for higher K would seem to indicate that agents with higher K accommodate their contributions to more details of topology rather than just the number of groups they play in[7]. This is a view that also gets supported by looking at the variation of correlation lengths with K. In the inset of Fig. 3, it can be seen that correlation lengths increase with K, thereby indicating that as K increases, more distant neighbors become relevant as compared to lower values of K. Hence agents with higher values of K are more sensitive to the surrounding topology of the interaction network. However, the data suggest that the range of this sensitivity saturates at higher K, with the decay length staying below unity.

This sensitivity can also be seen in Fig. 4. Note that group centrality has a positive impact on group contributions irrespective of agent parameters, although agents with higher values of K experience an appreciably bigger increase in their contributions as compared to their lower K counterparts. In the following we quantify this sensitivity (to centrality) as a function of K.

We consider 3 different hypergraphs (their corresponding line graphs can be seen in Fig. 5), two of them generated randomly and one uniform square lattice, all with $(N_p, N_g) = (64, 25)$. We consider identical agents with K varying from 0.5 to 5 and average over 5000 simulation runs for each configuration (i.e. a pair of a value of K and one of the three hypergraphs). We then perform three-parameter fits on the average contribution curves with the ansatz,

$$A(c(0), K) = A_0(K) - \frac{h(K)}{1 + c(0)/\beta}. \tag{16}$$

As it turns out, A_0 does not vary appreciably with K[8]. Therefore we remove it's dependence on K and treat it as a constant. Therefore, parameters A_0 and β are obtained by performing a fit on all configurations and h is fitted to each configuration separately. It turns out that setting $A_0 = 12.25$ and $\beta = 12.5$, the (fitted) values of $h(K)$ collapse onto a characteristic curve independent of the network topology. In Fig. 7 we plot three curves $h(K)$ for the corresponding hypergraphs as obtained from the fits. The empty symbols correspond to $(A_0 - A)/(1 + c(0)\beta)$

[7] The other end of the extreme is the case of $K = 0$, where the agents disregard any topological or dynamical features and play according to independent samples from a stationary prior distribution.

[8] Here we ignore the case of $K = 0$, as it represents a qualitatively trivial case, and the corresponding fitting procedure has non-unique global minima (both $h(0) \to 0$ and $\beta \to \infty$ lead to a flat curve).

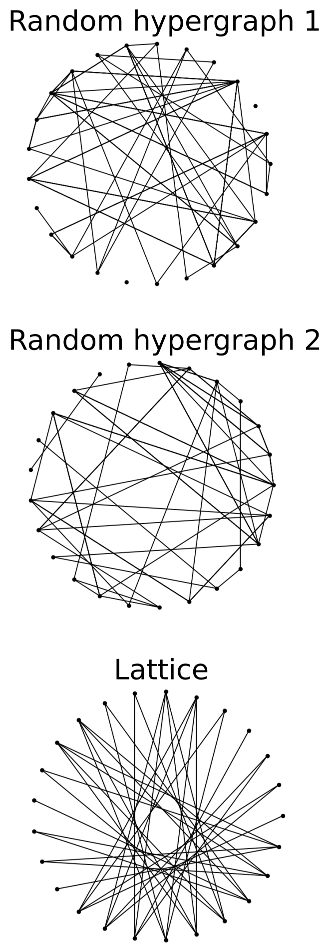


Figure 5: Line graphs corresponding to three hypergraphs. Top two correspond to random hypergraphs and the bottom corresponds to square lattice.

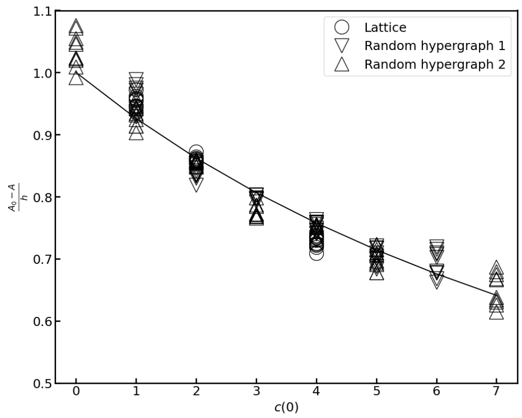


Figure 6: $(A_0 - A)/h$ as a function of group centrality for the three hypergraphs. The corresponding data from the hypergraphs are represented by circles, triangle-up and triangle down. The solid line corresponds to the function $(1 + c(0)/\beta)^{-1}$ for $\beta = 12.5$.

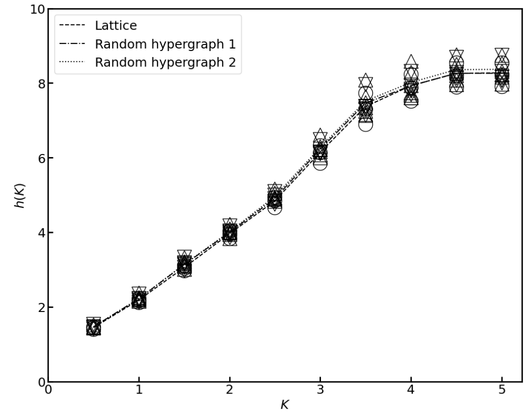


Figure 7: $h(K)$ for three hypergraphs. Polygons (dashed, dash dotted and dotted) correspond to the best fit. The corresponding data from the hypergraphs are represented by circles, triangle-up and triangle down, respectively.

for each individual group, where A_0 and β take the aforementioned values. A and $c(0)$ for each group are obtained from the simulations. Within scatter, the data are well represented by eq. (16). In a similar fashion we also show the characteristic value of β to be well descriptive of the impact of group centrality (see Fig. 6). To conclude, the deviation from the 'trivial case $K = 0$ seems to factorize as expressed by eq. (16), into a part depending on $c(0)$ and a function $h(K)$ which depends only on K. The latter starts off roughly linearly but saturates at higher values of K.

VII. CONCLUSIONS

Based on previously developed model agents that boundedly learn and plan, we have explored collective behavior in SPGG on a variety of hypergraphs of different topology. What we find that collective investment behavior is determined essentially by local descriptors, with correlations decaying exponentially in space. Furthermore, the impact of local connectivity, $c(0)$, and rationality of the agent, K, on the expected average investment factorize in a universal way, independent of network topology. Its behavior can be quantified with a characteristic function as shown in Fig. 7 and lends itself to experimental test.

Finally, our work suggests that cooperation in SPGG can be driven by making the players more diverse i.e. increasing the number of groups they play in and consequently increase inter-group connections. This observation is in line with results previously reported in [33], despite the fact that the authors follow a completely different modelling route.

[1] Daniel W. O'Neill, Andrew L. Fanning, William F. Lamb, and Julia K. Steinberger. A good life for all within planetary boundaries. *Nature Sustainability*, 1:88–95, 2018.

[2] Louis J. Kotzé and Rakhyun E. Kim. Earth system law: the juridical dimensions of earth system. *Earth System Governance*, 1:100003, 2019.

[3] Unai Alvarez-Rodriguez, Federico Battiston, Guilherme Ferraz de Arruda, Yamir Moreno, Matjaž Perc, and Vito Latora. Evolutionary dynamics of higher-order interactions in social networks. *Nature Human Behaviour*, pages 1–10, 2021.

[4] Christoph Hauert, Arne Traulsen, Hannelore Brandt, Martin A. Nowak, and Karl Sigmund. Via freedom to coercion: The emergence of costly punishment. *Science*, pages 1905–1907, 2007.

[5] Ricarda Winkelmann, Jonathan F. Donges, Keith E. Smith, Manjana Milkoreit, Christina Eder, Jobst Heitzig, Alexia Katsanidou, Marc Wiedermann, Nico Wunderling, and Timothy M. Lenton. Social tipping processes towards climate action: a conceptual framework. *Ecological Economics*, 192:107242, 2022.

[6] Kenneth G. Wilson and J. Kogut. The renormalization group and the ϵ expansion. *Physics Reports*, 12(2):75–199, 1974.

[7] P.G. de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, 1979. URL `https://books.google.de/books?id=Gh1TcAAACAAJ`.

[8] A. Schadschneider. Statistical physics of traffic flow models. *Physica A*, 285:101–120, 2000.

[9] Dirk Helbing and Sergi Lozano. Phase transitions to cooperation in the prisoner's dilemma. *Physical Review E*, 81:057102, 2010.

[10] D. Garcia, A. Abisheva, S. Schweighofer, U. Serdült, and F. Schweitzer. Ideological and temporal components of network polarization in online political participatory media. *Policy and Internet*, 7:46, 2015.

[11] Matjas Perc, Jillian J. Jordan, David G. Rand, Zhen Wang, Stefano Boccaletti, and Attila Szolnoki. Statistical physics of human cooperation. *Physics Reports*, 687: 1–51, 2020.

[12] Fabian Baumann, Philipp Lorenz-Speen, Igor M. Sokolov, and Michele Starnini. Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters*, 124:048301, 2020.

[13] Fabian Baumann, Philipp Lorenz-Speen, Igor M. Sokolov, and Michele Starnini. Emergence of polarized ideological opinions in multidimensional topic space. *Physical Review X*, 11:011012, 2021.

[14] José Moran, Antoine Fosset, Davide Luzzati, Jean-Philippe Bouchaud, and Michael Benzaquen. By force of habit: self-trapping in a dynamical utility landscape. *Chaos*, 30:053123, 2020.

[15] Quian Yu, Debin Fang, Xiaoling Zhang, Chen Jin, and Qiyu Ren. Stochastic evolution dynamics of the rock-scissors-paper game based on a quasi birth and death process. *Scientific Reports*, 6:28585, 2016.

[16] Te Wu, Feng Fu, and Long Wang. Coevolutionary dynamics of aspiration and strategy in spatial repeated public goods games. *New Journal of Physics*, 20:063007, 2018.

[17] Marco Tomassini and Alberto Antonioni. Computational behavioral models for public goods games on social networks. *Games*, 10:35, 2019.

[18] Maxwell N. Burton-Chellew and Stuart A. West. Payoff-based learning best explains the rate of decline in cooperation across 237 public-goods games.

[19] Wolfram Barfuss. Dynamical systems as a level of cognitive analysis of multi-agent learning. *Neural Computing and Applications*, 34:1653–1671, 2022.

[20] Prakhar Godara, Tilman Diego Aléman, and Stephan Herminghaus. Bounded rational agents playing a public goods game. *Phys. Rev. E*, 105:024114, Feb 2022. doi:10.1103/PhysRevE.105.024114. URL `https://link.aps.org/doi/10.1103/PhysRevE.105.024114`.

[21] Benedikt Herrmann, Christian Thöni, and Simon Gächter. Antisocial punishment across societies. *Science*, 319(5868):1362–1367, 2008. ISSN 0036-8075. doi:10.1126/science.1153808. URL `https://science.sciencemag.org/content/319/5868/1362`.

[22] Prakhar Godara and Stephan Herminghaus. Bounded learning and planning in public goods games (*In review*).

[23] Marcus Hutter. *Universal Algorithmic Intelligence: A Mathematical TopDown Approach*, pages 227–290. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-68677-4. doi:10.1007/978-3-540-68677-4˙8. URL `https://doi.org/10.1007/978-3-540-68677-4_8`.

[24] Daniel Alexander Ortega and Pedro Alejandro Braun. Information, utility and bounded rationality. In Jürgen Schmidhuber, Kristinn R. Thórisson, and Moshe Looks, editors, *Artificial General Intelligence*, pages 269–274, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

[25] Maxwell N. Burton-Chellew, Heinrich H. Nax, and Stuart A. West. Payoff-based learning explains the decline in cooperation in public goods games. *Proceedings of the Royal Society B*, 282, 2015. URL `http://dx.doi.org/10.1098/rspb.2014.2678`.

[26] R. M. Isaac and J. M. Walker. Group size effects in public goods provision: The voluntary contributions mechanism. *Quarterly Journal of Economics*, 103(1):179–199, 1988.

[27] M. Pereda, V. Capraro, and A. Sanchez. Group size effects and critical mass in public goods games. *Scientific Reports*, 9:5503, (2019). doi:10.1038/s41598-019-41988-3.

[28] Richard Bellman. On the theory of dynamic programming. *Proceedings of the National Academy of Sciences*, 38(8):716–719, 1952. ISSN 0027-8424. doi:10.1073/pnas.38.8.716. URL `https://www.pnas.org/content/38/8/716`.

[29] Drew Fudenberg and David K. Levine. Recency, consistent learning, and nash equilibrium. *Proceedings of the National Academy of Sciences*, 111(supplement_3): 10826–10829, 2014. doi:10.1073/pnas.1400987111. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1400987111`.

[30] Xavier Ouvrard. Hypergraphs: an introduction and review. *arXiv preprint arXiv:2002.05014*, 2020.

[31] Phillip Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987. ISSN 00029602, 15375390. URL `http://www.jstor.org/stable/2780000`.

[32] Michele Benzi and Christine Klymko. On the limiting be-

havior of parameter-dependent network centrality measures. *SIAM Journal on Matrix Analysis and Applications*, 36(2):686–706, 2015.

[33] Francisco C Santos, Marta D Santos, and Jorge M Pacheco. Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454(7201):213–216, 2008.

Appendix A: Generating a random hypergraph

In Algorithm 1 we show our method of generating a random hypergraph. The function takes as input (N_p, N_g) and returns the incidence matrix for the hypergraph. It has to be kept in mind that $N_p \geq 4$ otherwise we will not be able to generate a 4-regular hypergraph and also $N_p \leq 4N_g$, because if $N_p > 4N_g$ there will be at least one player which doesn't play in any group, thereby effectively decreasing N_p until $N_p = 4N_g$.

Algorithm 1 Random hypergraph generation

```
function HYPERGRAPH(N_p, N_g)
    H = zeros(N_p, N_g)                          ▷ Incidence matrix
    for 1 ≤ i ≤ N_p do                           ▷ Each player gets a group
        filled = 0
        while filled = 0 do
            j = random (1, N_g)                  ▷ choose a random j ∈ [1, N_g]
            if ∑_k H_kj < 4 then
                H_ij ← 1
                filled ← 1
            end if
        end while
    end for
    for 1 ≤ j ≤ N_g do                           ▷ Each group gets 4 players
        while ∑_k H_kj < 4 do
            i = random(1, N_p)
            H_ij ← 1
        end while
    end for
    RETURN H
end function
```

Bibliography

Amado, André et al. (2015). "Learning process in public goods games". In: *Physica A: Statistical Mechanics and its Applications* 430, pp. 21–31. ISSN: 0378-4371. DOI: https://doi.org/10.1016/j.physa.2015.02.077. URL: https://www.sciencedirect.com/science/article/pii/S0378437115002009.

Andersen, Steffen et al. (2008). "Eliciting risk and time preferences". In: *Econometrica* 76.3, pp. 583–618.

Arrow, Kenneth J. (1972). "Gifts and exchanges". In: *Philosophy & public affairs* 1, pp. 343–362.

Baumann, Fabian et al. (2020). "Modeling Echo Chambers and Polarization Dynamics in Social Networks". In: *Phys. Rev. Lett.* 124 (4), p. 048301. DOI: 10.1103/PhysRevLett.124.048301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.124.048301.

Bellman, Richard (1952). "On the Theory of Dynamic Programming". In: *Proceedings of the National Academy of Sciences* 38.8, pp. 716–719. ISSN: 0027-8424. DOI: 10.1073/pnas.38.8.716. eprint: https://www.pnas.org/content/38/8/716.full.pdf. URL: https://www.pnas.org/content/38/8/716.

— (1954). "The theory of dynamic programming". In: *Bulletin of the American Mathematical Society* 60.6, pp. 503 –515. DOI: bams/1183519147. URL: https://doi.org/.

Benzi, Michele and Christine Klymko (2015). "On the limiting behavior of parameter-dependent network centrality measures". In: *SIAM Journal on Matrix Analysis and Applications* 36.2, pp. 686–706.

Bonacich, Phillip (1987). "Power and Centrality: A Family of Measures". In: *American Journal of Sociology* 92.5, pp. 1170–1182. ISSN: 00029602, 15375390. URL: http://www.jstor.org/stable/2780000 (visited on 12/13/2022).

Boyd, Robert et al. (2003). "The evolution of altruistic punishment". In: *Proceedings of the National Academy of Sciences* 100, pp. 3531–3535.

Burton-Chellew, Maxwell and Stuart A. West (2021). "Payoff-based learning best explains the rate of decline in cooperation accross 237 public-goods games". In: *Nature Human Behaviour*. DOI: 10.1038/s41562-021-01107-7.

Burton-Chellew, Maxwell N., Heinrich H. Nax, and Stuart A. West (2015a). "Payoff-based learning explains the decline in cooperation in public goods games". In: *Proceedings of the Royal Society B* 282, p. 20142678. DOI: 10.1098/rspb.2014.2678.

— (2015b). "Payoff-based learning explains the decline in cooperation in public goods games". In: *Proceedings of the Royal Society B: Biological Sciences* 282.1801, p. 20142678. DOI: 10.1098/rspb.2014.2678. eprint: https://royalsocietypublishing.

org/doi/pdf/10.1098/rspb.2014.2678. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2014.2678.

Camerer, Colin F (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton university press.

Castellano, Claudio, Santo Fortunato, and Vittorio Loreto (2009). "Statistical physics of social dynamics". In: *Rev. Mod. Phys.* 81 (2), pp. 591–646. DOI: 10.1103/RevModPhys.81.591. URL: https://link.aps.org/doi/10.1103/RevModPhys.81.591.

Chen, Chia-Ching et al. (2013). "Too smart to be selfish? measures of cognitive ability, social preferences, and consistency". In: *Journal of Economic Behavior and Organization* 90, pp. 112–122.

Church, Alonzo (1936). "An Unsolvable Problem of Elementary Number Theory". In: *American Journal of Mathematics* 58.2, pp. 345–363. ISSN: 00029327, 10806377. URL: http://www.jstor.org/stable/2371045 (visited on 09/09/2022).

Cournot, A.A., N.T. Bacon, and I. Fisher (1897). *Researches Into the Mathematical Principles of the Theory of Wealth*. Economic classics. Macmillan. URL: https://books.google.de/books?id=eGgPAAAAYAAJ.

Croson, Rachel TA (2007). "Theories of commitment, altruism and reciprocity: Evidence from linear public goods games". In: *Economic Inquiry* 45.2, pp. 199–216.

Dohmen, Thomas et al. (2008). "Representative trust and reciprocity: prevalence and determinants". In: *Economic Inquiry* 46, pp. 84–90.

Dufwenberg, Martin and Georg Kirchsteiger (2004). "A theory of sequential reciprocity". In: *Games and economic behavior* 47.2, pp. 268–298.

Falk, Armin and Urs Fischbacher (2006). "A theory of reciprocity". In: *Games and economic behavior* 54.2, pp. 293–315.

Falk, Armin et al. (2018). "Global evidence on economic preferences". In: *Quarterly Journal of Economics* 133.4, pp. 1645–1692. DOI: 10.1093/qje/qjy013.

Fehr, Ernst and Klaus M Schmidt (1999). "A theory of fairness, competition, and cooperation". In: *The quarterly journal of economics* 114.3, pp. 817–868.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr (2001). "Are people conditionally cooperative? Evidence from a public goods experiment". In: *Economics letters* 71.3, pp. 397–404.

Fudenberg, Drew and David K. Levine (2014). "Recency, consistent learning, and Nash equilibrium". In: *Proceedings of the National Academy of Sciences* 111.supplement_3, pp. 10826–10829. DOI: 10.1073/pnas.1400987111. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1400987111. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1400987111.

Harris, Charles R. et al. (2020). "Array programming with NumPy". In: *Nature* 585, 357–362. DOI: 10.1038/s41586-020-2649-2.

Herrmann, Benedikt, Christian Thöni, and Simon Gächter (2008). "Antisocial Punishment Across Societies". In: *Science* 319.5868, pp. 1362–1367. ISSN: 0036-8075. DOI: 10.1126/science.1153808. URL: https://science.sciencemag.org/content/319/5868/1362.

Hofbauer, Josef and Karl Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press. DOI: 10.1017/CBO9781139173179.

Hutter, Marcus (2007). "Universal algorithmic intelligence: A mathematical top→down approach". In: *Artificial general intelligence*, pp. 227–290.

Isaac, R. Mark and James M. Walker (1988). "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism". In: *The Quarterly Journal of Economics* 103.1, pp. 179–199. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/1882648 (visited on 07/07/2022).

Janssen, Marco and TK Ahn (2003). "Adaptation vs. anticipation in public-good games". In: *annual meeting of the American Political Science Association, Philadelphia, PA*.

Jusup, Marko et al. (2022). "Social physics". In: *Physics Reports* 948. Social physics, pp. 1–148. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2021.10.005. URL: https://www.sciencedirect.com/science/article/pii/S037015732100404X.

Kuhn, Harold W. (June 1982). "Nonlinear Programming: A Historical View". In: *SIGMAP Bull.* 31, 6–18. ISSN: 0163-5786. DOI: 10.1145/1111278.1111279. URL: https://doi.org/10.1145/1111278.1111279.

Lam, Siu Kwan, Antoine Pitrou, and Stanley Seibert (2015). "Numba: A LLVM-Based Python JIT Compiler". In: *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*. LLVM '15. Austin, Texas: Association for Computing Machinery. ISBN: 9781450340052. DOI: 10.1145/2833157.2833162. URL: https://doi.org/10.1145/2833157.2833162.

Legg, Shane and Marcus Hutter (2007). "Universal intelligence: A definition of machine intelligence". In: *Minds and machines* 17, pp. 391–444.

Leike, Jan and Marcus Hutter (2015). "Bad universal priors and notions of optimality". In: *Conference on Learning Theory*. PMLR, pp. 1244–1259.

Lin, Christopher H et al. (2015). "Metareasoning for planning under uncertainty". In: *arXiv preprint arXiv:1505.00399*.

Moerland, Thomas M et al. (2020). "Think too fast nor too slow: The computational trade-off between planning and reinforcement learning". In: *arXiv preprint arXiv:2005.07404*.

Neumann, J v. (1928). "Zur theorie der gesellschaftsspiele". In: *Mathematische annalen* 100.1, pp. 295–320.

Ortega, Daniel Alexander and Pedro Alejandro Braun (2011). "Information, Utility and Bounded Rationality". In: *Artificial General Intelligence*. Ed. by Jürgen Schmidhuber, Kristinn R. Thórisson, and Moshe Looks. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 269–274.

Ouvrard, Xavier (2020). "Hypergraphs: an introduction and review". In: *arXiv preprint arXiv:2002.05014*.

Papadimitriou, Christos H and John N Tsitsiklis (1987). "The complexity of Markov decision processes". In: *Mathematics of operations research* 12.3, pp. 441–450.

Pereda María, Capraro Valerio and Sánchez Angel (2019). "Group size effects and critical mass in public goods games". In: *Scientific Reports* 9.5503. ISSN: 00335533, 15314650. URL: https://doi.org/10.1038/s41598-019-41988-3 (visited on 07/07/2022).

Russell, Stuart J and Devika Subramanian (1994). "Provably bounded-optimal agents". In: *Journal of Artificial Intelligence Research* 2, pp. 575–609.

Santos, Francisco C, Marta D Santos, and Jorge M Pacheco (2008). "Social diversity promotes the emergence of cooperation in public goods games". In: *Nature* 454.7201, pp. 213–216.

Sigmund, Karl et al. (2010). "Social learning promotes institutions for governing the commons". In: *Nature* 466.7308, pp. 861–863.

Simon, Herbert A (1955). "A behavioral model of rational choice". In: *The quarterly journal of economics*, pp. 99–118.

Solomonoff, R.J. (1964). "A formal theory of inductive inference. Part I". In: *Information and Control* 7.1, pp. 1–22. ISSN: 0019-9958. DOI: https://doi.org/10.1016/S0019-9958(64)90223-2. URL: https://www.sciencedirect.com/science/article/pii/S0019995864902232.

Su, Qi, Long Wang, and H Eugene Stanley (2018). "Understanding spatial public goods games on three-layer networks". In: *New Journal of Physics* 20.10, p. 103030.

Szabó, György and Christoph Hauert (2002). "Phase Transitions and Volunteering in Spatial Public Goods Games". In: *Phys. Rev. Lett.* 89 (11), p. 118101. DOI: 10.1103/PhysRevLett.89.118101. URL: https://link.aps.org/doi/10.1103/PhysRevLett.89.118101.

Szilard, Leo (1929). "Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen". In: *Zeitschrift für Physik* 53.11-12, pp. 840–856.

Szolnoki, Attila, Matjaž Perc, and György Szabó (2009). "Topology-independent impact of noise on cooperation in spatial public goods games". In: *Physical Review E* 80.5, p. 056109.

Tishby, Naftali, Fernando C Pereira, and William Bialek (2000). "The information bottleneck method". In: *arXiv preprint physics/0004057*.

Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Von Neumann, J. and O. Morgenstern (2004). *Theory of Games and Economic Behavior*. Princeton classic editions. Princeton University Press. ISBN: 9780691119939. URL: https://books.google.de/books?id=8nTWtgAACAAJ.

Wolpert, David H et al. (2014). "Optimal high-level descriptions of dynamical systems". In: *arXiv preprint arXiv:1409.7403*.

Wu, Shuchen et al. (2022). "E pluribus unum but how? Chunking as a rational solution to the speed-accuracy trade-off". In:

Wu, Te, Feng Fu, and Long Wang (2018). "Coevolutionary dynamics of aspiration and strategy in spatial repeated public goods games". In: *New Journal of Physics* 20.6, p. 063007.

Yang, Wu et al. (2013). "Nonlinear effects of group size on collective action and resource outcomes". In: *Proceedings of the National Academy of Sciences* 110.27, pp. 10916–10921. DOI: 10.1073/pnas.1301733110. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1301733110. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1301733110.