STRATEGIC AND SELFLESS INTERACTIONS

a study of human behaviour

felipe maciel cardoso

December 22, 2020

Não sou nada.
Nunca serei nada.
Não posso querer ser nada.
À parte isso, tenho em mim todos os sonhos do mundo.

I am nothing.
I will never be anything.
I couldn't want to be something.
Apart from that, I have in me all the dreams of the world.
— Tabacaria, Álvaro de Campos (Fernando Pessoa)

À Zu, minha mãe, e à Amanda

ABSTRACT

Humans are unique animals, cooperating in scale unrivalled by any other species. We built societies composed of non-kins, and empirical results have shown that people have social-preferences and might be willing to perform costly actions in the benefit of others. On the other hand, humans also compete among themselves leading at times to negative outcomes, such as the overuse of Earth's natural resources. Yet, competition between economic agents underlies the well functioning of markets, and its destabilisation – such as in an unbalanced distribution of market power – can harm trade efficiency. Accordingly, analysing how people cooperate and compete is of prime importance in the understanding of human behaviour, especially considering the impending challenges threatening the future welfare of our societies.

In this thesis, we present works exploring people's behaviour in social dilemmas – situations in which self-interested decisions are at variance with the social optimum – and in other strategic scenarios. Using the theoretical framework of game theory, their interactions take place in games abstracting these situations. Specifically, we performed behavioural experiments in which people played adaptations of common-pool resources, public goods, and other tailor-made games. Moreover, in an attempt to understand the existence of cooperation in humans, we propose a theoretical approach to model its evolution via a dynamics of heuristics selection.

We begin by introducing the theoretical and empirical foundations in which this thesis is based upon, namely, game theory, experimental economics, network science, and the evolution of cooperation. Subsequently, we illustrate the practical aspects of performing experiments using software implementations.

To understand people's behaviour in collective action problems – such as climate change mitigation, which requires a global level of coordination and cooperation – we performed public goods and common-pool resources games among Chinese and Spanish participants. The obtained results provide some insights onto the

variances and universalities of people's responses in these scenarios.

In this line, in recent years, individuals and institutions are increasingly concerned with social and environmental issues. Contributions in these scenarios, nonetheless, requires a substantial level of altruism by agents who have to make costly decisions. We performed two experiments to understand the drivers behind such decisions in two contemporarily relevant situations, namely, charity donations and socially responsible investments. Their results indicate that framing and other socio-demographic characteristics are significantly associated with pro-social and altruistic decisions.

Furthermore, we also explore people's behaviour in a competitive and complex scenario wherein subjects played as intermediaries in price formation experiments. We do so by performing an experiment implementing a generalization of the bargaining game in complex networks. Our findings indicate significant effects of network topology both in experimental results as also in theoretical models based on the observed behaviour.

Lastly, we expose a theoretical work attempting to understand the emergence of cooperation through a novel approach to study the evolution of strategies in structured populations. This is accomplished by modelling agents' decisions as results of heuristics, which are selected by a process inspired by evolutionary algorithms. Our analyses show that, when these agents have memory from previous interactions, cooperative strategies will thrive. Yet, those strategies will function according to different heuristics depending on which information they take into consideration.

RESUMEN

Los seres humanos son animales únicos, cooperando en una escala sin par en cualquier otra especie. Construimos sociedades compuestas de individuos no emparentados, y resultados empíricos nos han demostrado que las personas tienen preferencias sociales y pueden estar dispuestas a tomar acciones costosas que beneficien a otros. Por otro lado, los seres humanos también compiten entre ellos mismos, lo que en ocasiones conlleva consecuencias negativas como la sobreutilización de recursos naturales. Sin embargo, la competición entre agentes económicos subyace el funcionamiento adecuado de los mercados, y su destabilización – tal como en una distribución desbalanceada de poder de mercado – puede ser dañina a la eficiencia comercial. Por consiguiente, analizar cómo las personas cooperan y compiten es de importancia primordial para el entendimiento del comportamiento humano, especialmente al considerar los desafíos inminentes que amenazan el bienestar futuro de nuestras sociedades.

En esta tesis, se presentan trabajos analizando el comportamiento de las personas en dilemas sociales – situaciones en las cuales decisiones egoístas discrepan del optimo social – y en otros escenarios estratégicos. Utilizando el framework de la teoría de juegos, sus interacciones tienen lugar en juegos abstrayendo estas situaciones. Específicamente, realizamos experimentos conductuales en los cuales las personas participaron en juegos adaptados de recursos comunes, de bienes públicos y otros juegos hechos a medida. Además, con la intención de comprender la existencia de la cooperación en humanos, proponemos un enfoque teórico para modelar su evolución a través de una dinámica de selección de heurísticas.

Empezamos presentando los fundamentos teóricos y empíricos en los que se basa esta tesis, a saber, la teoría de juegos, la economía experimental, la ciencia de redes y la evolución de la cooperación. Posteriormente, ilustramos los aspectos prácticos de la realización de experimentos mediante implementaciones de software.

Para comprender el comportamiento de las personas en problemas de acción colectiva – como la mitigación del cambio climático,

que requiere un nivel global de coordinación y cooperación – realizamos juegos de bienes públicos y recursos comunes entre participantes chinos y españoles. Los resultados obtenidos proporcionan algunas ideas sobre las variaciones y universalidades de las respuestas de las personas en estos escenarios.

En esta línea, durante los últimos años, las personas e instituciones están cada vez más preocupadas por los temas sociales y ambientales. Sin embargo, las contribuciones en estos escenarios requieren un nivel sustancial de altruismo por parte de los agentes que tienen que tomar decisiones costosas. Realizamos dos experimentos para comprender los factores que impulsan dichas decisiones en dos situaciones de relevancia contemporánea: las donaciones benéficas y las inversiones socialmente responsables. Sus resultados indican que el encuadre y otras características sociodemográficas están asociadas significativamente con decisiones prosociales y altruistas.

Además, también hemos analizado el comportamiento de las personas en un escenario competitivo y complejo en el cual los sujetos participaron como intermediarios en experimentos de formación de precios. Lo hacemos a través de un experimento que implementa en redes complejas una generalización del juego de negociación. Nuestros hallazgos indican efectos significativos de la topología de la red tanto en resultados experimentales como también en modelos teóricos basados en el comportamiento observado.

Por último, exponemos un trabajo teórico que intenta comprender el surgimiento de la cooperación a través de un enfoque novedoso para estudiar la evolución de estrategias en poblaciones estructuradas. Esto se logra modelando las decisiones de los agentes como resultados de heurísticas, siendo estas heurísticas seleccionadas mediante un proceso inspirado en los algoritmos evolutivos. Nuestros análisis muestran que, cuando estos agentes tienen memoria de sus interacciones anteriores, las estrategias cooperativas prosperarán. Sin embargo, esas estrategias funcionarán de acuerdo con diferentes heurísticas según la información que tomen en consideración.

ACKNOWLEDGMENTS

This thesis would not be possible without the support of my two advisors, Yamir Moreno and Carlos Gracia Lázaro. Without Yamir's help, I would not be able to pursue my PhD in Spain and enjoy the research opportunities I had in the past years. It would neither be possible for me to have realized my thesis without the guidance and aid by Carlos, which have taught me much of what I have learnt in my PhD.

I am also very grateful to all my friends from the COSNET group that helped me with my research through insightful discussions and provided me with many joyful moments.

I am also deeply in gratitude to my Brazilian and Spanish families. Without them, it would be tough to endure the challenges of the PhD program and the troubles of living in a different country. Most of all, I am thankful to Amanda, whose love never allowed me to forget what is important in life and has always kept me going in the most strenuous moments.

Finally, I would like to acknowledge the support by IBSEN/Bridging the gap: from Individual Behaviour to the Socio-tEchnical MaN, Funder: European Union, Grant Agreement Nº 662725; and by TAPLEX/ABORDANDO LA COMPLEJIDAD DE SISTEMAS SOCIOTÉCNICOS, BIOLÓGICOS Y NATURALES, MINECO, FIS2017-87519-P.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Part I

PRELIMINARIES

1 INTRODUCTION

We are all apprentices in a craft where no one ever becomes a master.

Ernst Hemingway, New York Journal-American

Detail from ***Stone Henge, Wiltshire****, engraved by* `Robert Wallis`*, after* `J. M. W. Turner`

This thesis is concerned with the behaviour of interacting agents. They live, have a noticeable cognitive capacity, and make decisions. Specifically, this thesis focus on the most intelligent and successful living beings that science knows so far: humans. Humans are ecologically dominant on the planet, having spread over diverse environments like no other species [1]. Driven by the struggle for survival in the harshness of the real world, humans conquered the earth by building unique societies of unrelated individuals [2], cooperating in a scale not rivalled by any other animal [3, 4].

"To a large extent the future of the only place where life is known to exist is being determined by the actions of humans. Yet, the power that humans wield is unlike any other force of nature, because it is reflexive and therefore can be used, withdrawn or modified." `Lewis and Maslin` [5]

Humans don't have to worry about predators anymore, although we are often worried about ourselves and our institutions. In this regard, contemporary societies endure challenges that are unique in history. Not because they harder or have worse consequences, such a statement cannot escape the realm of subjectivity. What is clear is that contemporary challenges are a result of, or are intensively affected by, the ever-changing Anthropocene [5]. In less than 0.01% of the time of earth's existence, we have been able to alter the dynamics of ecosystems and even the whole earth climate [6]. Even if we end up extinct, our presence and influence on this planet will be available in the geological record for aeons [7]. At the time of writing, society is facing a global crisis as a pandemic propagates by the aether of connections we built in the last centuries. Even the resilience of our society depends on how we interact with our constructs and artefacts, as it is seen in the uncertainty associated with the effects of social media in the forms of government [8]. Whatever we have to face in the future, we need to understand how humans behave and how they respond to the forthcoming challenges.

"We have created a Star Wars civilization, with Stone Age emotions, medieval institutions, and godlike technology." Edward O. Wilson [9]

In this regard, it is crucial to have in mind that humans, although unique in the animal kingdom [1, 2], live in groups as other social animals and manifest behaviour which evolved both biologically and culturally while they interacted with themselves [10]. This resulted in individuals that sometimes have to be *strategic* but can also behave *selflessly*. Humans can selfishly compete for resources, but also can cooperate to reach a common goal [11] and even act altruistically in favour of others [12]. People have *social preferences* and might choose sub-optimal decisions for them for the benefit of another, yet they also have bias and prejudices and show favouritism towards their own group [13, 14]. Those behaviours, moreover, can be influenced by environmental and cultural factors, by what people see, their culture, and the structures underlying their interactions [15–18]. Arguably, thus, identifying people's decisions while interacting with others is chief for examining contemporary society's current and imminent affairs.

In this thesis, we present our work exploring the behaviour of humans in scenarios of competition, cooperation, and altruism. These interactions take place in *games*, a simple abstraction to study strategy and conflict among interacting agents [19, 20]. We observe and analyse their behaviour through behavioural experiments and also explore models based on observed behaviour as well on theoretical hypotheses. Our goal is twofold, first to investigate human behaviour and increase our collection of people's responses in strategic games and, second, to explore both emergence and implications of human behaviour in theoretical models. We begin in Chapter 2 by providing a brief overview of the foundations of this thesis. As behavioural experiments are essential to this thesis, in Chapter 3, we summarise the methodology behind performing them. Subsequently, we present our work arranged in chapters according to their context, namely:

COLLECTIVE ACTION PROBLEMS In Chapter 4, we discuss two collective action problems, namely public goods and common-pool resources, among participants from two countries. Section 4.1 presents a social dilemma experiment in which participants gain profit by harvesting a virtual forest vulnerable to over-exploitation.

In Section 4.2, we explore how information and the distribution of targets influence a collective risk climate-change dilemma game.

"You say you got a real solution
Well, you know
We'd all love to see the plan
You ask me for a contribution
Well, you know
We're doing what we can"
Revolution 1,
`The Beatles`

FRAMING & ALTRUISM In Chapter 5, we present experiments exploring the influence of framing and information on donations and responsible investments. Section 5.1 present experiments using multiple public goods with an associated donation. In Section 5.2, we present one experiment to understand the willingness of choosing impact investment options among experts and non-experts in this field.

TRADING IN NETWORKS In Chapter 6, we present an experiment exploring a generalization of the bargaining game in complex networks. Further, we propose a novel theoretical model based on the observed behaviour and discuss the results.

A HEURISTIC MODEL OF COOPERATION In Chapter 7, we expose a theoretical work attempting to answer the problems underlying cooperation in humans by using an original model of heuristics selection.

Finally, we conclude in Chapter 8 by summarising the findings of this thesis and exposing some prospective remarks. Necessary results that we deemed not sufficiently relevant to the main text are discussed in the Appendices [1].

1 Details of each experimental session were omitted for compactness, such as socio-demographic characteristics, instructions and ethical statements. They can be found in the supplementary materials of the corresponding publications.

2

FOUNDATIONS

Let Pascal say that man is a thinking reed. He is wrong; man is a thinking erratum. Each period in life is a new edition that corrects the preceding one and that in turn will be corrected by the next, until publication of the definitive edition, which the publisher donates to the worms.

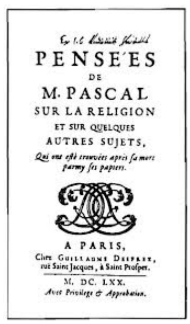
Machado de Assis,
The Posthumous Memoirs of Brás Cubas

Cedalion standing on the shoulders of Orion from ***Blind Orion Searching for the Rising Sun****,* `Nicolas Poussin`

In this chapter, we present the theories and results forming the basis of this thesis. The reader might notice that the descriptions are brief and centre around the topics which are most related to our work, as these fields have a myriad of branches, and a thorough description would be out of scope. Nevertheless, they are sufficient for a proper comprehension of the next chapters. We begin by presenting the approaches by mathematicians in the fields of decision theory and game theory in Section 2.1. Subsequently, we detail results obtained in the field of experimental economics concerning human behaviour in games (Section 2.2), and Section 2.4 then describes some of the theories explaining the observed behaviour. As some of these findings and even our own of experimental work (Chapters 6 and 7) requires a basic understanding of networks, we briefly introduce some of its theory in Section 2.3. Finally, we conclude in Section 2.5.

2.1 THEORIES ON DECISIONS AND GAMES

In the 17th-century, Blaise Pascal published the *Pensées*, a collection of texts including the famous *Pascal's Wager* wherein he provides an argument in favour of theism. In it, Pascal considers that humans bet their lives on the existence of God. If God exists, a true believer should receive an infinite benefit: salvation for eternity. Conversely, a non-believer would be doomed for eternity, an infinite punish-



PENSEES
DE
M. PASCAL
SUR LA RELIGION
ET SUR QUELQUES
AUTRES SUJETS,
Qui ont esté trouvées aprés sa mort
parmy ses papiers.

A PARIS,
Chez GUILLAUME DESPREZ,
ruë Saint Jacques, à Saint Prosper.
M. DC. LXX.
Avec Privilege & Approbation.

Pensées *("Thoughts")* `Blaise Pascal.`

-	IT RAINS	IT DOES NOT RAIN
Take the umbrella	-1	-1
Leave the umbrella at home	-10	0

Table 2.1: **Decision theory example.** Illustration of utility values when deciding to take or not take the umbrella when going out.

ment. If God does not exist, on the other hand, as humans live a finite life, humans can only receive a finite benefit, living without the constraints imposed by religion, or a finite loss, wasting their lives with the believer chores. Thus, humans are gamblers without the knowledge of God's existence and obliged to bet in one option during their lifetime. His argument follows a probabilistic reasoning, by weighing *expected value* of each consequence the only rational choice would be to bet that God exists and live accordingly[1].

Pascal's essay probably constitutes the first text on *decision theory* [20, 21], which studies how agents make decisions. It posits that agents have beliefs and desires and act rationally according to them. More specifically, their preferences determine a corresponding *utility* that will be maximized by rational decision-makers, i.e., they will consistently choose the option with the maximum expected utility value. When deciding involves uncertainty, it is often assumed that agents have a *subjective probability distribution* concerning the unknowns. As an illustration, imagine one person has to decide if she takes her umbrella with her when going out, carrying it implies a cost, although smaller than the cost of not having an umbrella if it rains. The utilities associated with each outcome in this example are described in Table 2.1. Agent's decision will depend on their beliefs about the raining probability – e.g., if there is 0.5 chance of raining, a rational decision-maker should take the umbrella with her.

Decision theory reaches its boundaries when agents are interacting with others, and their decisions depend on the actions of

1 This, of course, does not consider the problem of choosing the right deity to believe in, and of knowing what It would want from its believers.

other rational individuals. In these cases, the subjective probability distribution of one agent will depend on the decision-making process of the others it is interacting with, and vice-versa. In other words, the subjective probability distribution of one rational agent is an input of the distributions of all the others, leading thus to a system of equations. To solve such systems emerged the field of *game theory*, according to Roger Myerson, the *essential logical fulfilment* of decision theory [20].

In game theory, the word *games* refers to any situation of strategic interaction between self-interested parts. They are elementary components of social groups, as put by Colin Camerer: "a rough equivalent for social science of the periodic table of elements in chemistry" [15]. Games are abstract situations wherein involved agents, or *players*, obtain a benefit according to the combination of their actions; despite their simplicity, they provide a powerful framework to study real-world interactions. Game theory considers rational and *intelligent* players, the latter characteristic meaning that they can find the optimum decision if such exists [20]. Thus, if there is a mathematical solution to the decision-making problem, it will correspond to the agent decision. Commonly, studies focus on what is known as *non-cooperative game theory*, which deals specifically with the game wherein agents are not able to make deals or arrangements [22]. In these situations, the question is thus, what strategies agents will agents employ.

Researchers' imaginativeness has created a plethora of games emulating the most diverse situations, with each research field focusing on the ones most appropriate to them. Inside and outside of academia, the game that probably received the most attention is the Prisoners' Dilemma [23]; the *mother of all games* according to Karl Sigmund [24]. It belongs to a class of *2x2* games, pairwise interactions of individuals with two possible decisions, or in the game theory jargon, *strategies*. Its is attributed to Albert W. Tucker and had the following initial formulation (from Luce and Raiffa [25, p.95]):

> *Two suspects are taken into custody and separated. The district attorney is certain that they are guilty of a specific crime, but he does not have adequate evidence to convict them at a trial. He points out to each prisoner that each has two*

	NOT CONFESS	CONFESS
NOT CONFESS	-1,-1	-5, -0.25
CONFESS	-0.25, -5	-2, -2

Table 2.2: **Prisoners' Dilemma payoffs.** An example of payoffs in the Prisoners' Dilemma game. Each cell contains a tuple, wherein the left value corresponds to the row player's payoff, while the right value to the column player's payoff.

Si es delincuente que muera presto
`Francisco Goya.`

alternatives: to confess to the crime the police are sure they have done, or not to confess. If they both do not confess, then the district attorney states he will book them on some very minor trumped-up charge such as petty larceny and illegal possession of a weapon, and they will both receive minor punishment; if they both confess they will be prosecuted, but he will recommend less than the most severe sentence; but if one confesses and the other does not, then the confessor will receive lenient treatment for turning state's evidence whereas the latter will get "the book" slapped at him.

The suspects, thus, are playing a game wherein they can *cooperate* (not confessing according to our previous formulation) or *defect* (confessing)[2]. The utilities, or *payoffs*, associated with each outcome can be represented in a matrix form, such as shown in Table 2.2. Each cell shows two values: the payoff of the row player on the left, and the column player on the right. Unequivocally, both players will be better off if they chose to cooperate, instead of both defecting. Looking from the players' perspective, however, it seems more advantageous to defect as it will always yield a bigger payoff. This is easily seen for the row player by checking his payoffs column-wise: for each decision of the column player, confessing is more profitable. Therefore, rational decision-makers would end up defecting, an outcome far from the social optimum.

2 Recently it also became common to use the donation game representation [24, 26]. Nonetheless, the one presented here is a more general version.

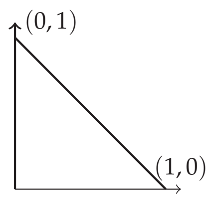
2.1.1 *Equilibrium Strategies*

To understand rigorously the result of the Prisoners' Dilemma we need to introduce the basic mathematical formulation of games and strategies. The definitions of the next paragraphs are nonetheless generic and can be used for the analysis of a broad range of games [20].

NORMAL FORM GAMES The values from Table 2.2 serve just as an example. Generically, if both players cooperate, they will receive a *reward* R, if just one defects, it will obtain the higher payoff T (*temptation*) while the other receives S (the *suckers' payoff*), finally, if both defect they will both receive the *punishment* P. This results in the matrix 2.1, such that any real valued combination satisfying that $T > R > P > S$ is considered to be a Prisoners' Dilemma game.

$$\begin{array}{c c} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R,R & S,T \\ T,S & P,P \end{pmatrix} \end{array} \tag{2.1}$$

Games in this form are labelled *normal form games*. Specifically, a normal form game is a tuple (N, A, u), wherein N is a finite set of n players, $A = A_1 \times \cdots \times A_n$ is a n-tuple of finite sets of decisions to all players $(1, \ldots, n)$ and $u = (u_1, \ldots, u_n)$ is a n-tuple payoff function for each player i, wherein $u_i : A \mapsto \mathbb{R}$. This correspond the most fundamental game representation [27], and it is sufficient for the analysis of players' strategies as studied in this thesis.



$\mathcal{S}$ *in* $\mathbb{R}_2$

PURE AND MIXED STRATEGIES If players always decide for one strategy, such as only cooperating or only defecting, they are said to follow a *pure strategy profile*. Nonetheless, anticipation is crucial while playing games and players might want to randomize their decisions, i.e., to follow a *mixed strategy profile*. In this case, player 1 might randomize the $(A_{1,1}, \ldots, A_{1,m})$ decisions in A_1, such that the probability of playing each option is given by a vector $\mathbf{x} = (x_1, \ldots, x_m)$, in which $\sum x_i = 1$ and $x_i \geq 0$. The set of all such mixed strategies is denoted by $\mathcal{S}$: a simplex in $\mathbb{R}_m$, spanned by the unit vectors of the standard base corresponding to each of the m pure strategies.

In a two-person game, such as the Prisoners' Dilemma, if player 1 follows the mixed strategy $\mathbf{x}$ and player 2 follows the strategy $\mathbf{y}$, player 1's expected payoff is given by the payoff function π_1:

$$\pi_1(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{U}\mathbf{y} = \sum_i x_i(\mathbf{U}\mathbf{y}) = \sum_{i,j} \mathbf{U}_{i,j} x_i y_j \tag{2.2}$$

Wherein $\mathbf{U}$ is the payoff matrix for player 1, such that $\mathbf{U}_{i,j} = u_1(A_{1,i}, A_{2,j})$. Conversely, if we assume for the sake of simplicity that the game is symmetric, $\mathbf{U}$ would also be the payoff matrix for player 2. Therefore, the payoff for player 2 would be given by:

$$\pi_2(\mathbf{y}, \mathbf{x}) = \mathbf{y} \cdot \mathbf{U}\mathbf{x} = \sum_i y_i(\mathbf{U}\mathbf{x}) = \sum_{i,j} \mathbf{U}_{i,j} y_i x_j \tag{2.3}$$

BEST-RESPONSE STRATEGIES If, by any chance, player 1 happened to know that player 2 acted according to the strategy $\mathbf{y}$, player 1 could adapt its strategy in order to maximize its benefits. This process can be done by finding the *best response* ($\mathbf{B}(\mathbf{y})$) to player 2 decisions, which is given by equation 2.4. Remarkably, unless $\mathbf{B}(\mathbf{y})$ only contains one unique pure strategy, the number of best responses will be infinite. Specifically, $\mathbf{B}(\mathbf{y})$ will contain all $\mathbf{x}^*$ such that $\pi_1(\mathbf{x}^*, \mathbf{y}) \geq \pi_1(\mathbf{x}, \mathbf{y})$ for all possible strategy $\mathbf{x} \in \mathcal{S}$:

$$\mathbf{B}(\mathbf{y}) = \arg\max_{\mathbf{x}} \pi_1(\mathbf{x}, \mathbf{y}) = \arg\max_{\mathbf{x}} \mathbf{x} \cdot \mathbf{U}\mathbf{y} \tag{2.4}$$

Naturally, if player 1 is playing with more players in a *n-person* game, its best response to a strategy profile $\mathbf{x}_{-1} = (\mathbf{x}_2, \ldots, \mathbf{x}_n)$ containing the strategy of every player other than 1 is given by:

$$\mathbf{B}(\mathbf{x}_{-1}) = \arg\max_{\mathbf{x}_1} \pi_1(\mathbf{x}_1, \mathbf{x}_{-1}) \tag{2.5}$$

NASH EQUILIBRIUM In a two-person game, if player 1 can adapt its strategies to the profile $\mathbf{y}$ of player 2, as by definition both have the same capacities, the latter should also do the same. This implies that, if player 2 does not change its strategy, it is due to $\mathbf{y}$ already being the best response to player 1's strategy profile $\mathbf{x}$. In this case, when $\mathbf{x} \in \mathbf{B}(\mathbf{y})$ and $\mathbf{y} \in \mathbf{B}(\mathbf{x})$, $\mathbf{x}$ and $\mathbf{y}$ form a *Nash equilibrium* [28], in which neither player has an incentive to deviate from its current decisions. More generally, a strategy profile $\mathbf{x}^*$ is said to form a *weak* Nash equilibrium if:

$$\pi_i(\mathbf{x}_i^*, \mathbf{x}_{-i}^*) \geq \pi_i(\mathbf{x}_i, \mathbf{x}_{-i}^*), \forall i \in N, \mathbf{x}_i \in \mathcal{S} \tag{2.6}$$

Given that $\mathbf{x}_i^*$ and $\mathbf{x}_{-i}^*$ correspond the strategies in profile $\mathbf{x}^*$ of player i and players other than i, respectively. Particularly, if there is a unique best response for each player, it is considered to be a *strict* Nash equilibrium:

$$\pi_i(\mathbf{x}_i^*, \mathbf{x}_{-i}^*) > \pi_i(\mathbf{x}_i, \mathbf{x}_{-i}^*), \forall i \in N, \mathbf{x}_i \in \mathcal{S}, \mathbf{x}_i \neq \mathbf{x}_i^* \tag{2.7}$$

"It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own self-interest."
`Adam Smith` [29]

THE PRISONERS EQUILIBRIUM As aforementioned, rational players are expected to defect in the Prisoners' Dilemma game, independently of the interacting player's strategy. In this case, defection (given by the strategy profile D) is said to be a *strong dominant strategy*: its expected payoff is greater than that of any other strategy ($D \cdot \mathbf{Uy} > \mathbf{x} \cdot \mathbf{Uy}, \forall \mathbf{x}$) [3]. This implies that players will be driven to the sub-optimal outcome, while they could be better off by both cooperating. Notably, this result stands against the idea that individuals pursuing their self-interest would lead to the welfare of all parts. Nonetheless, as it is presented in the next sessions, not always a theory based on axiomatic behaviour is going to succeed in describing reality. Individuals can cooperate, even when playing the Prisoners Dilemma, we only have to look for why and when.

INTERACTING THROUGH TIME Until now, we have just dealt with one-shot games, i.e., games in which players interact only once. Nevertheless, individuals, especially humans, will stack up a large number of interactions during their lifetime, and the best decision, in this case, does not necessarily agree with the one from the one-shot game. For instance, would the strategy of two rational

3 If it was the case that its expected payoff was greater or equal than any other strategy, it was going to be considered a *weak dominant strategy*.

individuals playing Prisoners' Dilemma for 100 rounds be different than from the one-shot version? According to *backward induction*, unfortunately no; players will keep defecting. In the 100th round, as players are rational and intelligent, they know they should defect, as it won't have any impact in the next rounds. By the same logic, it should not cooperate in the 99th round, in the 98th and so on. Ultimately, agents will defect in all rounds. Nonetheless, as we detail in section 2.4, the *shadow of retaliation* might enable cooperation if games are indefinitely repeated.

2.1.2 *Deciding in groups: The Public Goods Game*

One of the pivotal characteristics of Prisoners' Dilemma is epitomizing *social dilemmas*, circumstances in which there is a conflict between collective and individual interests. Logically, social dilemmas are suitably studied in situations wherein individuals behave in groups. The Public Goods Game is especially important in this regard, as it constitutes the generalization of the Prisoners' Dilemma between more than two players. It is one of the games most replicated in experimental situations, both in experimental psychology, sociology and political science [30]. It considers the case where n players have access to a common fund which will multiply their investments by a factor $r > 1$. In experiments, usually, players are endowed with some money, let's say 10 €, and they can decide to contribute some fraction c_i from this amount to the common pool. After all players have contributed, the total amount is multiplied and shared equally among all participants. Therefore, the resulting payoff of player i will be given by:

$$\pi_i = 10 - c_i + r\frac{\sum_j^n c_j}{n} \tag{2.8}$$

As the total contribution will be multiplied by r, the best for the whole group is for all players to contribute all their endowment into the common pool. Nevertheless, the best action for each player is to *free ride*: contribute nothing and get her share from the total contributed by the remaining players. For instance, being r =1.5 and n =6, if player i contributes nothing and the other five contribute their whole endowment, she would end up with 10 + 12.5 =

22.5 €. Meanwhile, if all players contributed everything they had, she would end up with 15 €. Clearly, independently of how much the other players contribute, player *i* has no incentive to contribute any amount greater than zero. In summary, the Nash equilibrium of the Public Goods Game corresponds to individuals contributing nothing from their endowment.

If we assume that this is the expected behaviour from humans, this result will entail direct implications for public policies when considering real-world public goods, such as public infrastructure, public parks, public scientific research, clean air, and so on. As an illustration, by commuting through public transportation, every individual ensures the quality of the air in a city, however, if one individual commutes by car, she will be able to enjoy the air quality without paying the costs (more commuting time, crowded metros, etc.). Notably, public goods are defined by two important elementary characteristics: they are *non-rivalled*, as the consumption of its benefits by an individual does not impede any other from consuming it; and *non-excludable*: it is not possible to deny a participant the consumption of its benefits. Those characteristics would lead to the demise of public goods, resulting in the *Tragedy of the Commons*, as it was coined by Garret Hardin [31]. Nonetheless, people use public transportation, pay taxes, they often bring their garbage back with them when going to parks, at least some of them. How likely then is for people to contribute to a public good? How much they deviate from rational economic behaviour? In the next section, we briefly introduce some results from the economic literature, and in Chapter 4 we present our works addressing those questions.

2.2 HUMAN BEHAVIOUR IN EXPERIMENTS

"there is a property common to almost all the moral sciences, and by which they are distinguished from many of the physical; that is, that it is seldom in our power to make experiments in them"
John Stuart Mill [34]

As we deal with problems closely related to economic theory, we mostly focus on experiments performed in economics, which nonetheless had influences of other social fields, such as psychology [32, 33]. Despite being absent from the genesis of economic sciences, nowadays, experimental economics has become one of the most important branches of mainstream economics as noted by the 2002 Nobel prize in economics awarded to Daniel Kahneman and Vernon Smith, two noteworthy pioneers of behavioural economics:

> *"Economics has also been widely considered a non-experimental science, relying on observation of real-world economies rather than controlled laboratory experiments. Nowadays, however, a growing body of research is devoted to modifying and testing basic economic assumptions; moreover, economic research relies increasingly on data collected in the lab rather than in the field."*

Economic experiments have allowed the testing on the basic assumption of economic theories, such as that individuals behave according to the axiomatic view of game theory. As Richard Thaler puts it, behavioural economics replaced *homo economicus* for *homo sapiens* in the economic theory [33]. As the layman could expect, human behaviour did not correspond to fully rational and selfish behaviour.

One of the unequivocal results on how humans deviate from classical economic theory prediction is seen in experiments with the *Dictator Game*. In it, a participant, the *Dictator*, receives an endowment to split in the proportion she wants with another participant. According to traditional economic theory, the participant should keep the full endowment, giving none to the other participant. In its initial formulation, as devised by Kahneman, Knetsch, and Thaler [35], participants had to choose between two options: evenly split 20$ (10$ for each one) or keep 18$ and share 2$ with the other participant. In their experiment, three-quarters of the participants chose the even split; subsequent replications allowing participants to choose how to split have shown that only 40% of them kept the full endowment and the amount shared averaged 20% of the total [36, 37]. These results demonstrate that humans deviate from rational selfish behaviour and have *social preferences*, i.e., they care to some extent about others' performance [38].

"Show me the axiom and I'll design the experiment that refutes it", supposidly by Amos Tversky *[39]*

People, however, are not only fair-minded. The setup devised by Kahneman and colleagues intended to test another influential experiment from the economic literature, the *Ultimatum Game*. In it, participants can have two roles: of a *Proposer*, who receives the endowment and decides how much to share, or of a *Responder*, that can accept or refuse the proposal. A refusal will result in a trade failure: no one receives anything. Again, economic theory predicts that the Proposers should offer the minimum amount possible (e.g., 1 cent), while the Responders should accept any positive offer. Experiments, nonetheless, show that Proposers are willing to share significantly more than the minimum, with amounts usually being greater than Dictators' offers [15]. This shows that, as Proposers are concerned with Responders rejecting their offer, they tend to propose higher values than Dictators. Therefore, these results provide a clear indication of the two essential features of people: being both *strategic* and *altruistic*.

Importantly, people's behaviour is affected by the specifics of the situations in which they are interacting. Besides, not everybody behaves equal [40] and changing the game structure and culture of the participants deeply affect game outcomes [15]. Experimental economics provides, therefore, one powerful paradigm to further uncover how people behave and why they behave as they do. As noted by Colin Camerer, *"the goal is not to disprove game theory but to improve it, by establishing regularity, which inspires new theory" [15].* Knowledge obtained empirically can be used then to create new theories, which can be further tested by new experiments. This has been the case in our work presented in this thesis, such as described in Chapters 4 and Chapter 6.

In the following subsection we will briefly introduce the methodology behind experimenting in economics and illustrative experimental results of the Public Goods Game, as it relates the most to the work presented in this thesis [4].

2.2.1 *The methodology of experiments*

Alvin Roth [43] categorize economic experiments according to their goals in three types: i) *Speaking to theorists:* experiments with the goal of testing hypothesis originated from theory; ii) *Searching for facts:* experiments looking for new phenomena in settings with some modified situation; iii) *Whispering in the ears of princes:* experiments planned for policymaking. Initially experimental economics was mostly concerned with testing economic theory predictions, nowadays, however, it mostly concerns with testing how variations in the experimental setup can influence outcomes. This implies that experiments can lack an underlying rigorous theory and can be

4 See [15, 41, 42] for a broad overview of experimental economics results.

used to test informal hypothesis [30]; in other words, they would be searching for facts (*ii*). Notably, our experiments reported in this thesis fall on this category. Specifically, our experiments can be divided into two types: looking for differences in behaviour between different populations (Chapter 4); or looking for differences between two experimental treatments in the same population (Chapters 5 and 6).

2.2.1.1 *The limitations of experiments*

Experimenters are often looking for phenomena that can be characterized qualitatively, as precise quantitative predictions are virtually impossible to be obtained [30]. Experimental conditions can have profound effects on the final result, being generally unfeasible to reproduce precisely data from previous experiments. This forces experimenters when testing a hypothesis to reproduce previous results to some extent; otherwise, it would not be feasible to distinguish the causes behind the observations. Specifically, results will depend on the specific context of each execution: its initial conditions (e.g., socio-demographic characteristics of the population), and the auxiliary assumptions about the experiments (e.g, the participants have understood the instructions). Therefore, an experimental conclusion obtained by evidence e, which supports or refutes a hypothesis H, is also determined by the initial conditions I and auxiliary assumptions K:

$$(K \land I \land H) \implies e \tag{2.9}$$

This implies that it is not only impossible to prove the general truth[5] of a hypothesis when e is observed, but also to refute it in a deductive approach when e is not observed. This is known as the Duhem-Quine problem, and it has direct implications when driving conclusions from experimental results. The controlled experiment enables the testing of hypothesis, but by being controlled it is far from the messiness of real-world interactions; hypotheses are not confirmed in isolation. Therefore, inferences have to be made with

5 According to the deductivist view [30, 44], we never can confirm that H is true, only that it is false. Consequently, science would evolve by disproving false theories.

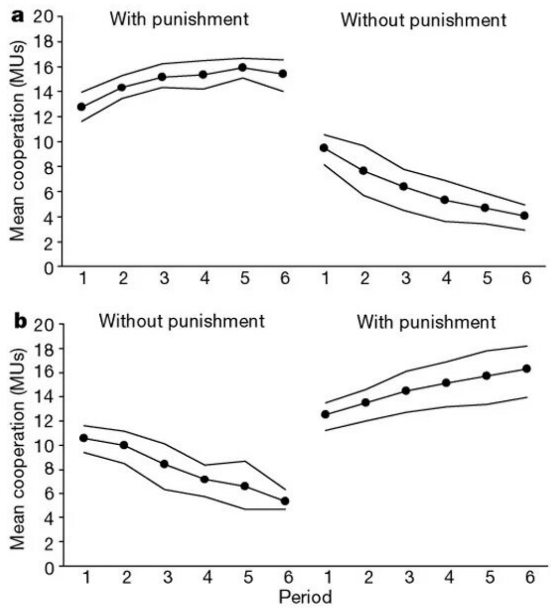


Figure 2.1: **Experimental results of a repeated Public Goods Game, with and without punishment.** Each panel shows mean co-operation at each time period. Top panel (**a**) shows sessions wherein participants started in the punishment treatment and the bottom panel (**b**) in a treatment without punishment. Figure from [47].

care when extrapolating experimental results. They are a valuable source of insights, but only can take us so far, as Guala puts it: *"data – no matter how useful – cannot ultimately replace the evidence collected in the field"* [45].

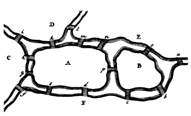
"Public goods and dilemma experiments are like using ping-pong balls; sensitive enough to be really informative but only with adequate control."
`John Ledyard` [46]

2.2.1.2 *The case of Public Goods*

The Public Goods Game provides a good case study of how knowledge about human behaviour is obtained in the lab. Aside from being one of the most paradigmatic experiments in economics, it is connected to the majority of works presented in this thesis. Results from early experiments have shown that individuals would not *free ride*; they would contribute something to the public good. Generally, participants would be willing to contribute something between 40% and 60% of their endowment [46]. Nonetheless, if experimenters allowed the game to run repeatedly for some periods, contributions would eventually decay towards the Nash equilibrium [48]. This phenomena has sometimes been referred to as the *overcontribution and decay* and has been intensively replicated [30]. On one hand, the positive contribution shows again that humans tend to act pro-socially, on the other hand, the decay also demonstrates that this is not an unconditional tendency.

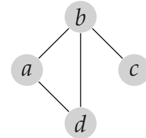
One explanation for the decay phenomena is the impossibility of direct retaliation by participants, as withdrawing contributions will also impact non-free-riders. In this line, an experiment by Fehr and Gachter allowed participants to punish others after the contribution phase [47]. To assess the effect of punishment, they had to perform two treatments: one *with*, and one *without* punishment; the requirement of a controlled experiment. They observed significant differences between the two treatments: contributions were higher, and the decay was not observed when participants could punish, as shown in Fig. 2.1. This phenomena was connected with a new hypothesis concerning human pro-sociality [47, 49, 50], as we comment in Section 2.4. Nonetheless, these results have to remain constrained to a narrow laboratory scope until evidence is obtained in the field [45].

2.3 STRUCTURE: THE SINE QUA NON OF INTERACTIONS



The Seven Bridges of Königsberg

If we are to study interactions, we are obliged to understand the structure underlying them. One useful approach is to represent the interacting entities as *vertices (nodes)* of a *graph* (or *network*) and represent their connections as its *edges (links)*, being this representation crucial for some of our work (Chapters 6 and 7). Graph theory is a branch of mathematics considered to have begun with the work on the seven bridges of the Prussian city of Königsberg by the mathematician Leonhard Euler. In it, Euler mapped river islands and the bridges between them as the vertices and the edges of a graph, respectively. This representation allowed him to show that there was no path, more specifically no *Eulerian path*, that could visit each node (island) without visiting an edged (crossing a bridge) twice in the Königsberg bridges' graph. Recently, this approach for modelling real systems as graphs have converged to the field of *network science* [51, 52], commonly referring to them as networks.



A graph is usually represented by having its nodes as circles and edges as lines.

In the social sciences, the first known use of networks to represent interactions between individuals is often attributed to the school mapping by Jacob Moreno [53], wherein he mapped interactions between boys and girls students in a *sociogram*. With this representation, focus on individuals was supplanted by a new holistic perspective: nodes as interdependent units connected by edges representing channels for the flow of resources, opportunities/-constraints of interactions, or lasting relation between individuals [54]. This change in perspective gained more momentum in the 20th century as data from social and economic interactions [55] became available. Concurrently, different fields began to study the structure interactions of diverse type of systems, such as the world wide web [56], power grids [57], and protein networks [58].

Remarkably, properties from different types of systems could be explained by abstract and generic models [59, 60], indicating that the underlying structure might have deep effects in the resulting phenomena, which is also the case for the systems studied here. Thus, we will introduce basic definitions from graph theory and the network models required for the understanding of our results [6].

2.3.1 *Graphs & Networks: Definitions and Properties*

A graph $\mathcal{G} = (V, E)$ is a structure composed of $|V| > 0$ vertices (nodes) connected according to the set $E = \{e_1, e_2, \dots\}$ of edges (links). Commonly, graphs are represented by an $|V|x|V|$ *adjacency matrix* $\mathcal{A}$ wherein each cell $\mathcal{A}_{i,j} = 1$ if there is an edge between vertex i and j, and $\mathcal{A}_{i,j} = 0$ otherwise. Moreover, if the edges have directionality – i.e., they are ordered pairs – $\mathcal{G}$ is denominated a *directed graph*, and an *undirected graph* otherwise. Each vertex has an associated degree k, which corresponds to the number of edges it belongs to:

6 For a general introduction on networks, see [61–63]. Moreover, [64] provides a summary of basic models properties, [54] gives an introduction of social networks, and [55] provides an introduction to the study of economic networks.

$$k_v = \sum_{j \in V} \mathcal{A}_{v,j} \tag{2.10}$$

If $\mathcal{G}$ is directed, the degree can be decomposed into the in-degree (k_v^{in}), edges arriving at v; and the out-degree (k_v^{out}), edges outgoing from v:

$$k_v^{in} = \sum_{j \in V} \mathcal{A}_{v,j} \tag{2.11}$$

$$k_v^{out} = \sum_{j \in V} \mathcal{A}_{j,v} \tag{2.12}$$

The average degree $\langle k \rangle$ of a network is often used to characterize it. Moreover, different graph generating processes are expected to generate different *degree distributions*, i.e., a distribution such that $P(k)$ corresponds to the probability of a randomly chosen vertex having degree k. Consequently, the empirical degree distribution often provides substantial information about the underlying mechanisms of the connections in a real network.

2.3.1.1 *Paths*

Networks often represent structures for traversal, as occur with transportation [65–68] and computer networks [69–71]. In other types of systems, distances between nodes might be also important as they can indicate the strength between the indirectly connected entities. In this regard, networks' *paths* are one of their most useful attributes. A path corresponds to a sequence of distinct vertices such that each consecutive vertex pair corresponds to an edge in the graph. A similar type of sequence, without the vertex distinctiveness restriction, is called a *walk* and can consequently be infinite. The length of a path is given by the number of edges it traverses, namely, its number of vertices -1. Most importantly, the distance (d) between two vertices is given by the minimum path length between them, which correspond to the length of the *shortest path* or the *geodesic* between them. Concerning the whole network, one informative metric is its average path length $\langle d \rangle$, which corresponds to the average geodesic between all vertices pairs:

$$\langle d \rangle = \frac{\sum_{u \neq v}^{V} d(u,v)}{|V|(|V|-1)} \tag{2.13}$$

2.3.1.2 *Centralities*

One of the recurrent questions while studying networks concerns the individual importance of its vertices. Certain nodes can be critical for some process [72], and usually this can be seen by their position in the graph, i.e., by how central it is. Different metrics have emerged to measure vertices centralities according to the specific context, as the analysis described in Chapter 6. The most straightforward centrality metric corresponds to the vertices' degree, despite simple it can nevertheless be very powerful, as the number of connections of a node is likely to be very informative of its role in the system. Some of the commonly used centralities metrics relevant to our work are [7]:

BETWEENNESS Betweenness centrality b measures how important a vertex v is, with respect to the possible paths between pair of nodes. Namely, it measures, the relative number of shortest paths a vertex belongs to:

$$b(v) = \sum_{v \neq s \neq d}^{|V|} \frac{\sigma_{s,d}(v)}{\sigma_{s,d}} \tag{2.14}$$

This summation takes place for all pair of nodes without v, wherein $\sigma_{s,d}(v)$ stands for the number of shortest paths between vertices s and d containing v, and $\sigma_{s,d}$ to the total number of shortest paths between s and d.

CLOSENESS Closeness centrality C measures how close a vertex is to all the other vertices. Specifically, it corresponds to the average distance to all the vertices of the graph, thus, a low value should indicate a more central vertex. It is given by:

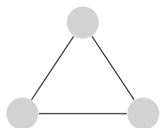
$$C(v) = \frac{\sum_{u \in V} d(u,v)}{n} \tag{2.15}$$

2.3.1.3 *Clustering*

If vertice a is connected to vertice b, and b is connected to vertice c, how likely is for a and c to be connected? In probabilistic terms,

7 For a broader list of metrics, see:[63, Chapter 7]

the answer would be always 1 if all the connections were transitive, which would only occur in a complete graph. This is extremely unlikely to be the case in real networks, although they often exhibit some partial transitiveness. Transitivity can be measured by taking into account the fraction of times the relation $(a, b)\&(b, c) \implies (a, c)$, i.e., the number of times the existence of a path of length 2 implies a loop of length 3. Commonly, this will be done by calculating the graph's *clustering coefficient* (CC), which corresponds to the number of closed triangles (3-vertices clique) for every path of length two, or *triple*:



A loop of length 3 or a 3-vertices clique.

$$CC(G) = \frac{3|\triangle(G)|}{|\tau(G)|} \tag{2.16}$$

Wherein $\triangle(G)$ corresponds to the set of all closed triangles in G and $\tau(G)$ to the set of all triples in G.

2.3.2 *Graph models*

In this subsection, we introduce some mathematical models used to generate graphs. Here the reader might observe that one essential characteristic distinguishing graph models is their expected *degree distribution* or *degree sequence*.

2.3.2.1 *Erdős–Rényi Graph*

The Erdős–Rényi random graph, usually referred to as *ER* graph, was introduced by mathematicians Paul Erdős and Alfréd Rényi. In the ER model, a graph $G(n, p)$ has n vertices connected randomly to each other with a probability p. As each edge occurs independently from every other, G is expected to have $\binom{n}{2}p$ edges and its degree distribution is given by a binomial distribution of the form:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \tag{2.17}$$

Wherein $P(k)$ corresponds to the probability of a node having degree k. For large n, the distribution can be approximated by a Poisson distribution:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \tag{2.18}$$

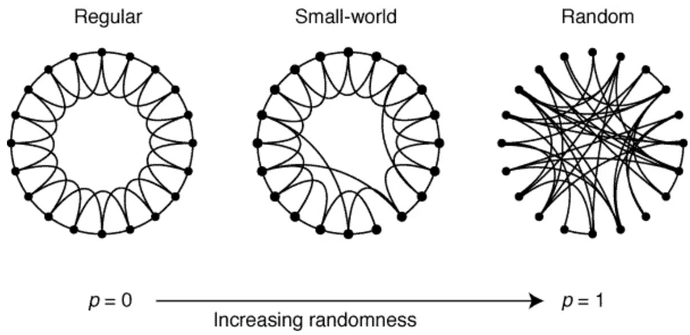


Figure 2.2: **Transition from a ring lattice to a random graph.** For small values of p, a small-world regime exists, in which graphs will have a small average path length and a large clustering coefficient. Figure from [59].

As the probability of two vertices being connected is the same for every pair of vertices, a closed triple will be connected with probability p. In other words, the probability of being connected is given by:

$$CC(G) = \frac{\langle k \rangle}{n-1} \tag{2.19}$$

This implies that for large networks, the clustering coefficient tends to vanish. Real networks, nonetheless, tend to have a relatively large clustering coefficient, thereby demonstrating that connections do not occur randomly. On the other hand, the average path length in an ER graph tends to relatively small, such as is encountered in real networks:

$$\langle d \rangle \sim \frac{\ln n}{\ln \langle k \rangle} \tag{2.20}$$

2.3.2.2 *Small-World Networks*

In 1967, Stanley Milgram published the result of a series of experiments in the social sciences, wherein participants were given an unknown recipient that should receive a letter sent by them [73]. As the target was a complete stranger, they were instructed to send the letter to one person they knew in the first name basis who should forward the letter following the same method. Remarkably,

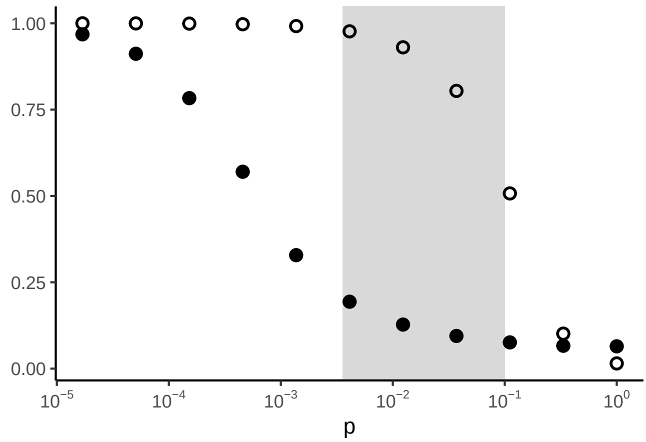


Figure 2.3: **Average path length and clustering coefficient in Watts-Strogatz networks.** Normalized average path length (solid black circles) and clustering coefficient (open circles). Values correspond to a mean of 10^2 of 10^3 nodes networks.

letters that reached the recipient had only been forwarded 6 times on average. From this result emerged the notion of the "6 degrees of separation", in that we are only are 6 steps away from every other person in the world; thus, we live in a *small world*.

One compelling explanation of this phenomenon was published in 1998 by Duncan Watts and Steven Strogatz in one of the most influential papers in network science [59]. Their model starts from a regular ring lattice, in which every node is connected with their $\langle k \rangle$ nearest neighbours, such as illustrated in the left graph of Fig. 2.2. Its edges are rewired according to a probability p, incrementing the disorder of the system as p grows. Thus, when $p = 1$ a random graph is obtained, such as the right graph of Fig. 2.2. Interestingly, for small $p > 0$, the resulting graph has a small average path length, comparable with a random graph, but with a large clustering coefficient, as shown by the grey area of Fig. 2.3. This is a result of rewiring adding shortcuts to the graph, while the graph continues to be highly clustered as p is small, as illustrated by the middle graph of Fig. 2.2.

2.3.2.3 *Barabasi-Albert*

One important characteristic of real networks, not fully developed by the previous models, is their high degree heterogeneity. In real networks, some nodes can have significantly more connections than others, an unlikely result of a random process. Specifically, it has been observed that the degree distribution of some networks follows a power-law distribution [56, 60, 74, 75] of the form $P(k) \sim k^{-\gamma}$, such that typically $2 \leq \gamma \leq 3$ [63]. Interestingly, this type of distribution is scale-invariant, leading to such network being refereed as *scale-free networks*.

The *preferential attachment* process is the most well-known explanation for such distribution [60]. It corresponds to a feedback-loop between connections, wherein nodes with high connectivity are more likely to obtain new connections. This proposal was firstly posited by Price in 1976 [76] by the name of *cumulative advantage*, explaining how highly cited research papers are more likely to be cited. Nonetheless, the work by Lazlo Barabasi and Reka Albert in 1999 [60] became the most famous using this approach. In their model, they consider a growing network, in which a node arrives with m new edges at each time step. It will connect those m edges to existing nodes in proportion to their degrees, a preferential attachment process, as coined by them. This will generate a network with a degree distribution following exactly $P(k) \sim k^{-3}$, as lustrated by Fig. 2.4.

"It is common in bibliometric matters and in many diverse social phenomena, that success seems to breed success."
`Derek de Solla Price` [76]

2.3.2.4 *Configuration Model*

The configuration model was proposed by Béla Bollobás [77] to study graphs that have a degree sequence fixed beforehand. This allows specifying that every node has degree greater than zero, which is not possible with the ER graph. Moreover, it enables the specification of any degree sequence, being also useful for generating graphs with a power-law degree distribution without correlated degrees [78]. Given a fixed degree sequence $(k_1, k_2, \cdots, k_n)$, it generates a graph according to the following algorithm:

1. Generate $\sum_i^n k_i$ half-edges, wherein for each vertice i there is k_i half edges connecting to it;

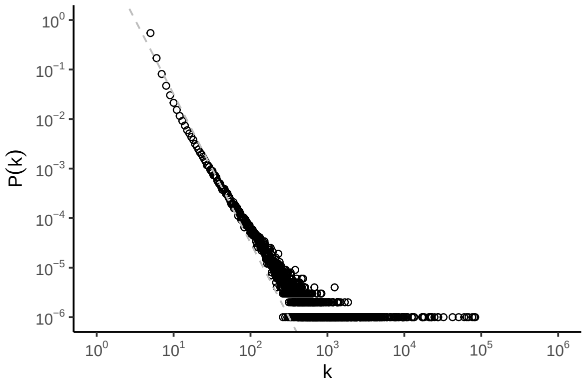


Figure 2.4: **Degree distribution in a scale-free network.** Degree distribution of a network with 10^6 nodes generated by the Barabasi-Albert model, for $m = 5$ [60]. The dashed line corresponds to a line with slope 3 to guide the eye.

2. Randomly connects the pairs of half-edges to form an edge of the graph.

Clearly, $\sum_i^n k_i$ has to be even as $\sum_i^n k_i = 2|E|$. This procedure does not guarantee the generation of a simple graph, as self-loops and multiple edges are possible. Nevertheless, both tend to relatively wane when $n \to \infty$ [64]. For the generation of networks following a power-law degree distribution it has been shown that, if the maximum degree is smaller than the square root of its size ($k_{max} \leq \sqrt{n}$), nodes' degree will not be correlated [78].

RANDOM REGULAR NETWORK Also known as a uniform random regular graph, it is a subset of the *k-regular* graphs, i.e, graphs wherein all nodes have the same degree k. It corresponds to a random graph generated by the configuration model by setting a fixed degree of k for every node.

2.4 EXPLAINING COOPERATION

Previous sections have demonstrated that the initial predictions of game theory poorly predicted the behaviour of humans. People cooperated in experiments in which they should defect, they shared and donated money expecting no benefit from it. Indeed, humans cooperate in a unique scale, living in societies wherein trust and support among unrelated individuals are imperative [2]. Moreover, cooperation is also observed in the whole nature, from bacterias and cells to primates and other mammals [79]. This posed a challenge for evolutionary biology, as at first glance costly behaviour in favour of other individuals should not be selected [80]. As it turns out, cooperative behaviour can provide fitness advantages in certain conditions. In this section, we provide a brief description of the main explanations for the biological evolution of cooperation and in which conditions humans tend to cooperate.

2.4.1 *Kin Selection*

The term kin selection was coined by John Maynard Smith in 1964 [81], nonetheless, the work most associated with this concept was published in the same year by William D. Hamilton [82]. Analysing the evolution of social behaviour in groups, Hamilton came to the conclusion that behaviour not producing direct fitness benefits[8] could evolve by natural selection when it sufficiently increases the fitness of related individuals; in other words, increases their *"inclusive fitness"*. Specifically, he considered that if one individual performing an altruistic act of cost c generates a benefit b to the recipient's fitness, this behaviour would be selected if their genetic relatedness r is large enough:

$$r > \frac{b}{c} \tag{2.21}$$

This expression is now known as the *Hamilton's rule* [26, 83, 84] and it was one of the revolutionary ideas in the biology of the 20th century, explaining how evolution can select for altruistic traits. It formalized the famous phrase of J.B.S. Haldane, *"I would lay down my life for two brothers or eight cousins."*, proclaiming cooperation with kin as one of the elementary constituents of evolution. Still, for it to take place, individuals have to be able to discriminate their kin [85] or to be more likely to interact with them, such as in limited dispersal [86].

8 When the action is beneficial to both the recipient and the actor, it is usually referred as mutualism [80].

2.4.2 *Reciprocity*

"Models that attempt to explain altruistic behavior in terms of natural selection are models designed to take the altruism out of altruism."
`Robert Trivers` *[87]*

According to Robert Trivers, if altruism is a product of natural selection, it should not be labelled altruism [87]: if individuals are helping others to maximize their gene pool in future generations they are not being properly altruistic; they are behaving according to their selfish genes [88]. Trivers proposed a model for altruism based on reciprocity, in which individuals interacting for long enough time would maintain cooperation by helping others that have helped them previously. He defined this as *reciprocal altruism*, despite this definition being not precisely accurate, as this relation provides a mutual befit for both parts. Accordingly, it is more sensible to refer to those type of relationships as reciprocal; if they were altruistic, the actor should not obtain any direct fitness benefit from it [79].

Reciprocity, nonetheless, is an undeniable enforcer of cooperation if people[9] interact repeatedly. The tournaments performed by Robert Axelrod [90] present and insightful demonstration of how reciprocity can be fixed in a population. Axelrod received submissions of strategies coded by game theorists from different fields: psychology, political science, economics, sociology, and mathematics. Each strategy would play against one another in a round-robin tournament with an unknown number of rounds so that they could not profit from backward induction - i.e., knowing the end of the game and when they could defect.

The code sent by each participant could be arbitrarily complex, surprisingly, however, the winner of this tournament happened to be the simplest strategy: *Tit-for-Tat (TFT)*, submitted by Anatol Rapoport. Rapoport's strategy followed two simple rules: i) start by being nice, cooperating in the first round; ii) reciprocate in subsequent rounds, copying the other's decision from the previous round. Despite being so simple, it epitomizes the mechanisms of self-defence; furthermore, an evolutionary analysis shows that they will be stable in a population if agents don't make mistakes [24]. Nonetheless, if agents make mistakes, defecting when they shouldn't, TFT players will enter in a loop of retaliation, punishing

9 Reciprocity seems to be a very human characteristic, being rare among non-human animals [89].

one another repeatedly. In this scenario, it is far from an optimal strategy, and a population of TFT players can be invaded by defectors [24, 26].

"Anyone who injures their neighbor is to be injured in the same manner: fracture for fracture, eye for eye, tooth for tooth." ***Leviticus*** *24:19–20*

Cooperation can still be maintained if TFT pivots the way to a more generous strategy, the *Generous Tit-for-Tat (GTFT)*, which cooperates with defectors with a small probability [91]. Nevertheless, it will still be prone to invasion by full-cooperators, which in turn will be invaded by full-defectors and so on, creating a cycle of defection and cooperation. *Win-stay loose-shift (WSLS)*, although also very simplistic, was the one strategy that demonstrated capable of resisting invasions [26, 92]. WSLS is a *pavlovian* strategy, repeating its decisions if they yielded large payoffs, and switching otherwise. It can persist in a population even in the presence of errors, although WSLS still requires TFT to pave its way in a population with a majority of defectors [26]. Unfortunately, however, it has not been observed as a strategy implemented by participants in economic experiments, being participants more likely to play variants of TFT [93, 94].

"For direct reciprocity you need a face. For indirect reciprocity you need a name." David Haig [24]

2.4.3 *Indirect Reciprocity*

Arguably, a significant number of interactions, especially in humans, are feeble and does not repeat [26, 95, 96]. In those cases, there is no direct benefit in the long run justifying reciprocity, and individuals cannot punish who defected with them. Nonetheless, if agents can discriminate others by their *reputation* in previous interactions, punishing defectors can sustain cooperation even if two players never interact twice, as shown by theoretical models [26, 97, 98]. This mechanism is termed *indirect reciprocity*, as individuals don't discriminate others by what they have done to them directly. Experiments with humans have shown that if people have a score of participants past actions, cooperation will be sustained by conditional cooperation – i.e., people cooperating with others in function of their reputation [96, 99]. Indirect reciprocity is notoriously more sophisticated than its direct counterpart, as individuals have to able to communicate and store other players' history, for this reason, some believe that evolution of indirect reciprocity is connected with the evolution of intelligence and language [96].

2.4.4 *Strong Reciprocity*

Reciprocity as we have explored in Section 2.4.2 is sometimes classified as *weak reciprocity* [45, 100] since individuals reciprocate because it is in their best interest to do so; the threat of retaliation by the other player will make cooperation the optimum choice in the long run. In some situations, however, direct reciprocity is not possible, as in one-shot interactions with multiple other players [49]. For instance, in a Public Goods Game, if players start to retaliate by not contributing, every player would free ride in the long run [48]. One solution for this problem is enabling participants to directly punish free-riders, however, if this is costly, punishment will be altruistic and prone to a *second free-rider problem*. Thus, by requiring a higher commitment to fairness, punishment at a cost would only be performed by *strong reciprocators* [50].

In a highly influential paper, Fehr and Gächter enabled participants in a Public Goods Game to punish other players after the contribution phase [47], as we have seen in Section 2.2.1.2. The punishment was costly to participants and groups changed at every round, such that participants never played in the same group twice. Thus, despite there was no direct benefit in punishing, most of the players punished free-riders and contributions were significantly larger when participants could punish others. People in this experiment declared felling anger toward free riders, indicating a natural predisposition to react negatively to free riding. This result was then used by researchers as a basis justifying how large scale cooperation can exist in societies [47, 49, 50]. The natural human tendency for cooperating, and for punishing the ones who don't, would enforce the cooperation of the whole population.

NORMS AND INSTITUTIONS The argument in favour of strong reciprocity to explain cooperation in societies was not free from criticism. Critics point out that, although strong reciprocity has shown effects inside the lab, evidence from real societies were poor [45]. Essentially, anthropological evidence from costly punishments come from small societies wherein individuals tend to interact repeatedly, what would be characterized as weak reciprocity. Although anti-social behaviour would be negatively judged and frowned upon by shared norms [2, 101], this is not necessarily a costly sanction. Arguably, the punishment would usually be cheap, taking the form of symbolic sanctions or by sharing its costs in coalitions [45]. In this view, the real mechanism for enforcing cooperation would come from institutionalized third parties capable of promoting weak reciprocity (e.g., guarantee that stakeholders interactions last indefinitely), and of enforcing sanctions [102–104]. Nonetheless, strong reciprocity cannot be ruled out, results in the lab indisputably indicate that humans incur costs when others free-ride, what might justify our predisposition to establish norms and institutions. Moreover, the lack of evidence in the field might be a result of real societies already living in a highly cooperative state [105]. Thus, as most of the mechanisms explaining cooperation [84, 106–110], still is open to debate whether strong reciprocity is behind the high level of human pro-sociality.

2.4.5 *Structured Populations*

Until now we have only focused on the strategic point of view of individuals in *well-mixed* populations, i.e, we assumed that individuals interacted randomly and/or repeatedly with each other. Nonetheless, individuals live in structured spaces constraining their interactions. This can alter the evolution of cooperation significantly, as is demonstrated by the seminal work of evolutionary strategies in lattices by Nowak and May in 1992 [111]. They have shown that if pure strategies arranged in a lattice are playing the Prisoners' Dilemma, cooperators can form clusters and defend themselves against defectors. Needless to say, evolution highly depends on the specifics for cooperation to be fixed [112, 113]. Nevertheless, this shows that even without retaliatory strategies, cooperation can emerge if the structure of the population is taken into account.

There is nowadays a vast literature of works exploring the interaction between evolutionary game dynamics and population structure extensively both in theoretical models [114–121] as in experiments [16, 17, 122]. Some models have shown that heterogeneous networks can sustain cooperation [115, 123, 124] in more costly situations than in random and regular graphs. Curiously, experimental work with humans playing Prisoners' Dilemma in large networks found no significant difference between heterogeneous networks and lattices [17]. People seem to respond conditionally to the level of cooperation of their neighbours, deeming the level of cooperation independent of the network structure [125].

On another hand, it also might be the case that the heterogeneity of the network is not as important as previously thought, and cooperation would be influenced by the so-called *network reciprocity* [26, 116]. It posits that cooperation will prevail if the benefit b provided to the recipient, divided by the cost c, exceeds the average number of neighbours, k. Leading to an expression similar to Hamilton's rule:

$$k < \frac{b}{c} \tag{2.22}$$

Subsequent experimental work has shown that cooperation is greater when this equality is true [16], indicating, jointly with the evidence from the previous works, that the structure of the population can promote cooperation in the proper conditions. Those precise conditions, however, are still open to debate, given that humans not necessarily respond to differences in payoff [126], deeming rules such as network reciprocity inadequate to explain their behaviour in spatial games [125].

2.5 IN SUMMARY

The last sections should have provided enough background for the understanding of the next chapters. It should be clear for the reader how much different factors are entangled, and the intrinsic difficulties in extracting knowledge on human behaviour. It is very difficult, if not impossible, to obtain a universal rule of human behaviour. For instance, two enforcers of cooperation, punishment and population structure, have shown to interact negatively with each other in experiments [122]. Such difficulties are nonetheless what justifies the need for more research in this area. Only with more diverse data and new theoretical insights can we hope to arrive at a general understanding of how people behave with each other.

3 PERFORMING EXPERIMENTS

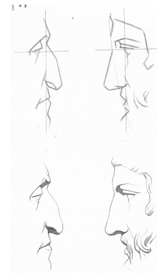


Plate 1,3 from `Charles Bargue's` ***Cours de dessin***

> *"There are two possible outcomes: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery."*
>
> Enrico Fermi

Behavioural experiments do not require a large apparatus to be performed, good research can be done with pencil and paper [127, 128] or even with a bowl of beans [129]. Nonetheless, unless the specifics of the experiment require otherwise, it is preferred to run the experiment using a computer platform. In this way, session effects [130] due to experimenter influence and human mistakes are reduced. Moreover, non-computerized experiments are significantly more time consuming, especially if participants are playing in groups. For instance, in a repeated public goods game at the end of every round, the groups' total contribution and participants' payoff have to be calculated, which would be overwhelming if participants are not playing over a computer network.

For the aforementioned reasons, all the experiments reported in the next chapters have been performed using a software platform. Specifically, excepting experiments reported in Sections 5.1 [1] and 6.1[2], all the experiments reported have been developed and performed by the author.

3.1 PLANNING

The first stage of an experiment usually is elaborating the ideas behind it, consolidating them in a practical setup. What scenario is intended to be reproduced or emulated in the lab? What hypotheses are going to be tested? What behaviour do we want to observe?

1 It has been developed by the LINEX institute in Valencia.
2 It has been developed by the software developers at the BIFI Institute.

After that has been decided, it is necessary to outline what are the requirements in terms of software and data. The former will determine how elaborated the system will have to be, and the latter how many participants and groups will be needed to have enough statistical power. Subsequently, it is necessary to prepare the instructions for the subjects explaining the experiment and start the software development process. Ideally, the project starts with a requirement analysis followed by the design of the software guaranteeing a clear structure and the possible reusability of the program [131].

3.2 DEVELOPING

The software for performing an experiment has several requirements: it needs a clear interface with understandable instructions; it has to manage groups of players playing simultaneously, synchronizing rounds through waiting pages; participants inputs have to be stored in a central database. This might seem rather daunting, fortuitously, however, reusability is spread in software development. The availability of code from people that confronted similar problems removes the burden of reinventing the wheel in every enterprise. In our case, almost all of our experiments were based on the oTree platform [132], which itself depends on several other free software, such as Django[3], Bootstrap [4] and Redis [5]. In oTree, an experiment is structured according to the following nested components[6]:

- **`Session`:** structure corresponding to the experimental session, it is divided into a list of subsessions.
  - **`Participant`:** entity corresponding to the participant playing the experiment, thus it references all the actions performed by her in the inner classes.
  - **`Subsession`:** the elementary constituent of each session, usually it corresponds to one round in the game.

3 `https://www.djangoproject.com/`
4 `https://getbootstrap.com/`
5 `https://redis.io/`
6 More details on how to develop an experiment with oTree are available at `https://otree.readthedocs.io/en/latest/index.html`

Nonetheless, one round also might be fractioned into more subsessions.

`Group:` structure containing a set of players, for instance, in public goods games each group of players with access to a common fund would be contained in a separate `group`.

`Player:` stores the instance of play by the participant in one particular subsession. Thus, it allows for a participant to have different player roles in each subsession. For instance, in subsession *A*, the player can be the Proposer in an Ultimatum game, while in subsession *B* she will be the Responder.

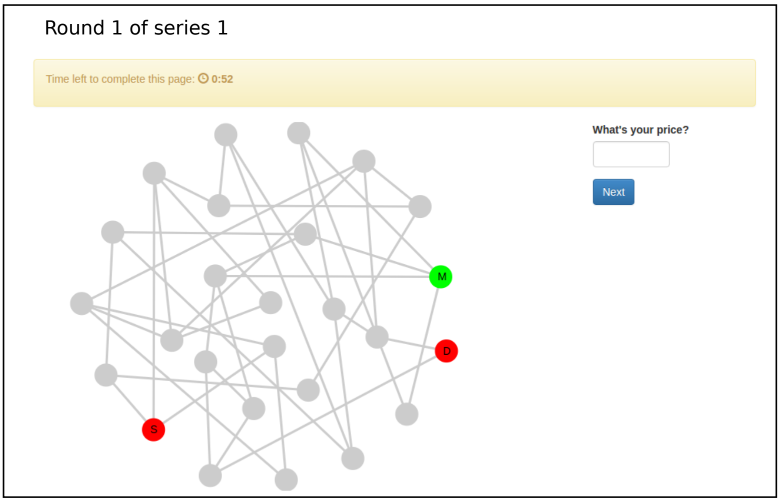
`Page:` the elementary division of a subsession, each page is visited by a player and constitutes one step of the experiment.

Therefore, given that the core of the software requirements is already taken care of by oTree, most of the development consists in developing the specifics of the experiment: data modelling, configuration and layout of the pages, any calculus performed at each time step, control of participants sequence of play, translations (in the case of multilingual experiments), etc. Hence, most of the effort is generally spent with the experimental interface, which nonetheless can be very time-consuming. For instance, the forest representation displayed in the experiment presented in Section 4.1 required the development of a library to control the map drawing and the resource states. The instructions also have to been developed carefully, it should state in clear and simple sentences what the experiment is about and provide a tutorial on how to play the game. As an illustration, we display below the instructions of the experiment presented in Chapter 6.

Instructions

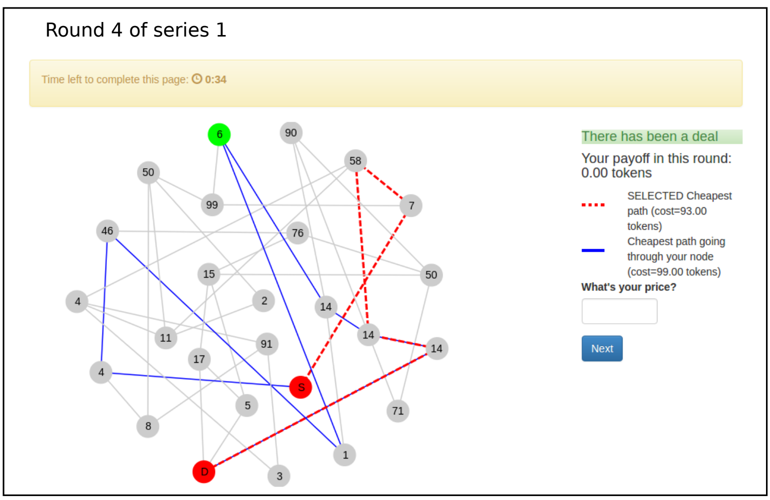
Thank you for participating in this experiment, that is part of a research project in which we try to understand how individuals make decisions. You re not expected to behave in any particular way.
At this moment the experiment begins. Please keep quiet until the end, turn your cell phone off, and remember that the use of any material foreign to the experiment is not allowed (including pen, pencil or paper).
Your earnings will depend on your own decisions and those of the other participants. Additionally, you will receive 5 € for participating in the experiment until the end.
Please keep quiet during the experiment. If you need help, raise your hand and wait to be assisted. Please do not ask any question aloud.
You participate along with other people with whom you interact according to the rules explained below. The session lasts about an hour and a half. The following instructions are the same for every participant of this experiment.
Once completed the session, you will receive 5 € for participating, along with your earnings corresponding to the rounds, once converted into euros. For convenience, the total earnings are rounded up to the nearest 50 cents.
You will access the experiments after reading these instructions. When all participants have accessed, the rounds will begin.
You are going to participate in 4 experiments. Each experiment consists of 15 rounds. Before starting each experiment, all the players, you included, will be randomly located in the nodes of the network shown below.
Your position in the network will be denoted with the letter '**M**' (for me). In the same way, two different nodes will be chosen as Source (**S**) and Destination (**D**) respectively. Their position in the network will be denoted with the letters '**S**' and '**D**'.
All the players will remain in the same position

during each experiment of 15 rounds. In the same way, the source and destination will remain in the same position throughout each experiment of 15 rounds.
The players will play the role of intermediaries.
A good must be transported from **S** to **D** generating a benefit of 100 tokens for all players involved (**S**, **D**, and all nodes in the path between them).
Intermediaries (that is, the players) simultaneously have to post the fraction of these 100 tokens they would like to charge if selected, which must be between 0 and 100 tokens. **You will have 60 seconds to post your price.** If you do not post a price, the computer will decide for you: please do not run out your time and make your own decision.
This is the screen you will see in the first round (this screenshot is only an example):



The sum of prices along any given path between **S** and **D** determine a total cost. Once all the intermediaries have posted a price, the cheapest path (with lowest total cost) from **S** to **D** will be selected. If the total cost of the cost cheapest path is less than or equal to 100 tokens, the good will be taken from **S** to **D**. Otherwise, that is, if all the paths from **S** to **D** cost more than 100 tokens, there will be no deal and no value will be generated. Ties are broken randomly, that is, if there are more than one cheapest path, one of them will be selected at random.

Your payoff in this round will be:
a) If you are located on the selected cheapest path, you will receive your price as payoff. b) Otherwise, that is, if you are not on the selected path, you will not receive any payoff in that round. S and D will receive, equally distributed, the rest of the 100 tokens.
From the second round on, you will be informed about whether there was a deal in the previous round, and if so what was the selected cheapest path, and the costs of this path. You will also be informed about the cheapest path through your node regardless of whether this was the selected cheapest path. The selected path will be highlighted by a dashed red line, while the cheapest path through your node will be highlighted by a blue solid line. Note that the cheapest path through your node may contain loops, i.e., it may pass more than one time through some nodes.
With this information on the screen, you must set a price for the current round. At the end of each experiment, the positions of all players, the source, and the destination will be randomly reassigned, and a new experiment of 15 rounds will begin.
This is the screen you will see in the subsequent rounds (this screenshot is only an example):



Please, click the below NEXT button to start:
[NEXT]

3.3 testing & fixing

"All code is guilty, until proven innocent."
`Anonymous author`

When a first version of the experimental software is ready, it has to be tested several times for finding possible programming errors or misspelled text. Testing can generate new ideas or the recognition of new demands, which will be inserted into a new version. This loop repeats until software is recognized to be ok. Hopefully, the final version will be bug-free.

3.4 running

Lastly, a batch of experiments will be planned and participants will be recruited for the sessions. Ideally, a pilot is initially programmed with a small subset of participants to verify if some change in the experimental setup would be needed or beneficial for the experiment. In our work, most of the times, we recruited participants through the volunteer pool of the IBSEN project (http://www.ibsen.eu), which, at the time of writing, contains 28234 registered participants. Participants have to sign an informed consent to participate, besides their anonymity is always preserved in the experiment. Moreover, the call for participants will occur only after it has been checked and approved by a research ethics committee, ensuring the procedure is performed following the relevant guidelines and regulations.

Furthermore, monetary incentives are used to motivate participants, and they are directly tied to their performance in the game. Given that we are not able to observe their other intrinsic preferences, it is reasonable, *ceteris paribus*, to assume they prefer a larger payoff over a smaller one. This procedure follows the methodology of paying participants from experimental economics, which is at variance with some other fields, such as experimental psychology [7]. This practical detail will nonetheless constrain the number of participants, as an experimental budget is limited. Usually, payment per subject is calculated such that the average payoff is around the country average hourly wage. Accordingly, at the end of the experiment participants will be paid in cash in the case of a lab

7 See [30, Chapter 11] for an overview and discussion in this subject.

experiment, while in online experiments payment is often done by using an external service such as PayPal[8].

In a lab experiment, people can enrol through open calls in the IBSEN recruitment platform, and they are instructed to attend at a scheduled time and place. As some people might not be able to attend, participants are usually over recruited. Unfortunately, depending on the experimental setup, extra participants might not be able to play, in this case, they will be paid the game expected payoff and be offered apologies. It is not a perfect solution, but still better than not performing the session. Participants are often paid according to a show-up fee (a fixed amount guaranteeing that they receive at least something) plus some quantity in the function of their earnings in the game.

Experiments can also be performed online, with each participant accessing the platform remotely. This approach reduces controllability of the experiment but allows a higher flexibility in the number of participants. As participants' efforts are reduced (they can play from their homes for just some minutes), the payment can follow a lottery, such that only some fraction of the players will be paid some substantive amount. Hence, by this approach, there is virtually no limitation in the maximum number of participants. Nonetheless, participants probably do not have the same incentives as when paid directly by their earnings in the game. Thus, to motivate them, the lottery selects winners in proportion to their performance in the experiment. The main issue with online experiments is that, unfortunately, it is not trivial to synchronize play among participants, which makes repeated games hard to be performed, although some solutions are possible [133].

When all the experimental sessions have been performed, backup copies of the experimental data are made. The next steps are preprocessing and analysing the data, which will generate results such as the ones presented in the following chapters.

8 https://www.paypal.com/

Part II

EXPERIMENTS AND THEORETICAL MODELS

4 COLLECTIVE ACTION PROBLEMS

At the approach of danger there are always two voices that speak with equal force in the heart of man: one very reasonably tells the man to consider the nature of the danger and the means of avoiding it; the other even more reasonable says that it is too painful and harassing to think of the danger, since it is not a man's power to provide for everything and escape from the general march of events; and that it is therefore better to turn aside from the painful subject till it has come, and to think of what is pleasant. In solitude a man generally yields to the first voice; in society to the second.

Leo Tolstoy, War and Peace

The Wood Sawyers, Jean-Francois Millet

In the Leviathan, Thomas Hobbes, who was deeply impacted by the English Civil War, posited that humans in their *state of nature* would live in constant conflict between themselves, in his own words: *"If any two men desire the same thing, which nevertheless they cannot both enjoy, they become enemies."* [134]. According to Hobbes' perspective, we would be only able to maintain peace by being under the control of an external authority through a *social contract*, which would constrain ourselves. This external authority should have a higher power, be a *Leviathan*, or a state, otherwise, there would be no guarantee that people would not violate the imposed norms: *"Covenants, without the sword, are but words, and of no strength to secure a man at all."* [134].

Cover from Thomas Hobbes' ***Leviathan***

Particularly relevant in this regard are the situations wherein there is a lack of central control, such as it is the case of public goods. As we have seen in Section 2.1.2, public goods are nonrivalled and nonexcludable, which makes them especially vulnerable to invasion by free riders. In this line, Garret Hardin posited that this constitutes a situation wherein no technological solution[1] is possible [31]. Hardin exemplified this by an open pasture wherein

1 As an example where a technological benefit makes the system go worse, see: [135]

herdsman could raise cattle without control, in this case, herdsman would try to maximize their gains by increasing their herd until the system collapsed. According to his view, total freedom of the commons would imply ruin for all, unless we could constrain ourselves. According to Hardin, this should be done by employing mutually agreed coercion enforced by some external authority, an approach close to Hobbes's perspective.

Gardin's view, however, was not free from criticisms. Some pointed out that his historical examples were inaccurate [136, 137], and that there were more solutions than the one proposed by him [138–141]. These points got more relevance as research upon the so-called *common pool resources (CPR)* grown, which constitute a type of public goods that can be overused by its stake-holders, such as fish and lumber. Thus, they differ from classical public goods as their resources are *rivalrous*: the consumption by one participant affects its availability to others. Accordingly, if users, e.g., fishers or loggers, are free from surveillance, they face a dilemma to either exploit the resource sustainably as a form of cooperation or overuse the resource for immediate profit as a form of defection.

Although this social dilemma is not without successful resolutions, there are no panaceas either [142]. Instances of overused CPRs abound in human history. Among the more famous are the crash of the Peruvian anchoveta fishery in the early 1970s [143] and the overfishing-induced ecosystem regime shift off the coast of Newfoundland in the early 1990s [144]. One in four of the world's fisheries collapsed between 1950 and 2000 according to the Food and Agriculture Organization of the United Nations [145]. Protected Bornean rainforests in Kalimantan lost over 56% of their geographic span between 1985 and 2001, much of it due to unsanctioned logging [146]. In fact, over 10% of worldwide timber trade is illegal, amounting to a staggering $15 bn annually based on estimates from the early 2000s [147].

"Don't compete! — competition is always injurious to the species, and you have plenty of resources to avoid it!", ***Mutual Aid: A Factor of Evolution*** `Pyotr Kropotkin`

Nevertheless, as pointed by Elinor Ostrom, *"the tragedies the common are real, but not inevitable"* [138]. Even in the absence of an external authority, common-pool resources have proven manageable if communication channels between participants are made available [103, 148, 149] or rent dissipation from harmful competition can be reduced with a proper rights-based management protocol [129]. In this line, it is also important to recognize that selfishness is not necessarily a chief human impulse [150] as evidenced by various experimental treatments that entice cooperation [151–153].

To ensure sustainable exploitation common-pool resources require a dose of self-restraint, but this has often proven elusive in practice. According to Ostrom, i) restricting access, and ii) creating incentives in favour of resource investment over overexploitation are fundamental to solve CPR problems. Nonetheless, there is no type of regulating regime that works efficiently with respect to all CPR [138]. Therefore, the behaviour of humans in those types of dilemmas deserves a paramount consideration. As mentioned in section 2.2, individuals have complex psychology and have a puzzling decision-making process, being strongly influenced by their morals, culture, and social impulses [2, 12, 15, 154, 155].

Accordingly, given that behaviour depends on factors such as education and culture, it is unclear how individuals from different countries will behave in these systems. Global efforts have shown to be hard to coordinate, and targets are often missed [156]. It is painfully evident the ongoing struggle to save perhaps the most valuable of the commons, the Earth's climate system [157]. In this regard, it is noteworthy that appropriation history might affect the willingness to accept the burden of emissions reduction among participants from different countries, impairing climate change mitigation [158, 159].

In this chapter, we focus on our own work to understand how humans make decisions with respect to common-pool resources sustainability and climate change mitigation, taking into account confounding factors such as age, education, and culture. We present two experiments we have performed among Spanish and Chinese participants: i) in the first, players had to manage a realistic CPR, and we provide an analysis characterizing their behaviours towards settings of common goods consumption and ecological systems in general(Section 4.1); in the second, participants played an adaptation of the public goods game simulating a fund to mitigate climate change [160] with simultaneous contributions from the two different countries (Section 4.2).

4.1 A COMMON POOL OF DYNAMIC RESOURCES

Behavioural patterns behind the demise of the commons across different cultures [161].

M. Jusup, **F. Maciel-Cardoso**, C. Gracia-Lázaro, C. Liu, Z. Wang, & Y. Moreno

One of the fundamental characteristics of common pool resources is being, by definition, difficult to control [162, 163], implying a need for user self-restraint to ensure sustainability. This is somewhat critical when considering that full information on the underlying dynamics of a common-pool resource is often unavailable, being human decision makers required to identify the optimal level of exploitation.

To address this question, we incorporated resource dynamics into an experimental platform which simulated the resource dynamics realistically through a virtual forest. Participants had the roles of loggers, a situation that closely mimics epistemic and socio-economic realities of resource exploitation [164, 165]. Specifically, when dealing with biological resources, e.g., a fish stock or a forest, a broad outline of the resource's dynamics is typically knowable, yet many key details, such as the point of optimal population growth or the population's tipping points, remain unknown [166–168]. From the socio-economic perspective, exploitation is most often done for-profit, with comparisons in terms of various profitability indicators being of utmost importance to business owners. The experiment described in this session not only demonstrates that the demise of the commons is a serious threat in these conditions, but also pinpoints a unifying cause behind robust behavioural patterns displayed by two geo-socially distant populations.

4.1.1 *Ecological Model*

To emulate a virtual forest responsive to human decisions, we ran a background ecological model that evaluated tree regrowth against the posted logging efforts. Inputs from participants in the experiment were passed into a mathematical model, such that the resources state reflected tree logging by participants and the forest natural regrowth. At each time step, the number of trees in the forest is updated by removing the total trees logged C_i by each player i. Simultaneously, trees regrow naturally at rate a g, which is fastest when the current number of trees in the environment, R, is small relative to the carrying capacity, M.

To increase realism of the ecological model, we also incorporated an Allee effect, a phenomenon whereby population size correlates with the mean individual fitness of the population [169]. In particular, if the number of trees is very low, then reproductive success becomes highly unlikely because, e.g., trees are on average too far from one another for a pollinator to carry pollen. We implemented this by having the growth rate abruptly drop to zero when the number of trees is below the no-recovery threshold, i.e., when $R < R_c$. Given that $H(\cdot)$ corresponds the Heaviside step function – i.e., zero if the argument is negative and one if it is positive –, and N to the total number of players, the forest's dynamics is given by:

$$\frac{dR}{dt} = g\,(M - R)\,H\,(R - R_c) - \sum_{i=1}^{N} C_i, \tag{4.1}$$

Participants provided as input the effort T they would desire to expend logging trees. It determined the number C of collected trees according to the time needed to find and process a suitable tree, τ. As finding and processing becomes more difficult as the number of trees gets small compared to the carrying capacity, C_i is given by:

$$C_i = \frac{R}{\tau M} T_i, \tag{4.2}$$

Furthermore, we incorporated basic economic aspects in the form of revenue from selling logged trees at price p, as well as the cost of logging per unit effort, c. Profit π_i is positive if revenue exceeds the cost, while excessive logging in a heavily exploited forest could have generated losses.

$$\pi_i = pC_i - cT_i, \tag{4.3}$$

If we assume that participants' effort is constant, in the long run the resource state will tend to an equilibrium value R'. Therefore, by letting $\frac{dR}{dt} = 0$ we can obtain from Eq. (4.1) that:

$$R' = \frac{M}{1 + \frac{1}{gM}\sum_{i=1}^{N}\frac{T_i}{\tau}},$$

Thus, the number of trees left for cutting in an equilibrium is some fraction of carrying capacity M, where this fraction decreases (resp., increases) with more effort (resp., faster regrowth). If we further assume that all players exert the same effort, T^*, and earn the same profit, π^*, we can examine the conditions that lead to maximum equitable profit, or the *maximum sustainable yield (MSY)*, such that $\left.\frac{d\pi^*}{dT}\right|_{T=T^*} = 0$. From the last condition, it follows that

$$T^* = \frac{\tau g M}{N}\left(\sqrt{\frac{p}{\tau c}} - 1\right). \tag{4.4}$$

And the corresponding equilibrium resource state is

$$R^* = M\sqrt{\frac{\tau c}{p}}.$$

4.1.2 *Experimental Setup*

Participants played for 50 rounds, each round representing a year. The exact number of rounds, as participants were made aware, was undisclosed to avoid the final-round effects, i.e., a change in behavior due to the impending end of the game. Each round consisted of inputting a decimal number between 0 and 7 corresponding to their weekly logging effort. The carrying capacity was set to $M = 400$ trees and the forest's regrowth rate was such ($g = 0.0504$) that keeping the number of trees at around 151 would have produced the maximum sustainable yield (MSY). If the resources went below $R_c = 100$ trees, regrowth was blocked emulating the Allee effect. Fig. 4.1 shows the expected outcomes in terms of resources (panel **A**) and profit (panel **B**) for an constant average effort. According to these parameter values, MSY is obtained with an average effort of $T^* = 2.765$ logging days per week.

Details concerning the ecological model was hidden from participants, including including the existence of MSY and the exact value of the no-recovery threshold. Nonetheless, the forest's state could be monitored at all times via a detailed interface (Fig. 4.2).

Participants could also compare their own performance in terms of effort, yield, and profit with others. Using this setup, we performed experimental sessions in two pool of participants: i) 96 undergraduate students in Xi'an, China; ii) 90 individuals from the general population in Zaragoza, Spain. Participants were instructed that they would receive a monetary payoff tied to their end profit in the game. The resulting payoffs amounted to an average of ¥62.6 in China and €15.1 in Spain, further details are shown in a.

4.1.3 *Results*

We used the MSY value as natural performance classifier for 16 Chinese and 15 Spanish player groups. Namely, we defined the optimal exploitation as any number of trees left for cutting after 50 rounds that was within ±10% of the MSY number (136–166 trees). Only one group from each of the countries was able to optimally exploit the resource. None of the groups underexploited the resource by ending the experiment above the optimal range, whereas a total of 14 Spanish and 15 Chinese groups overexploited the resource by ending below this range. The virtual forest's time evolution suggests that the former groups performed much worse (Fig. 4.3A). Seven groups from Spain drove the number of trees in the forest below the no-recovery threshold (=100 trees), whereas only one group from China did the same, and it did so only in the last round of the game.

Nevertheless, a time-series regression analysis reveals that multiple Chinese groups kept depleting the resource, and given more time, would have likely crossed the no-recovery threshold too (Fig. 4.3B). Denoting with R_t the virtual forest's state at time step t, where $t_0 = 20 \leq t \leq 50$, we fitted the following model to the time-series data pertaining to the groups who overexploited, but did not deplete the resource.

$$R_t = c_0 + c_1 (t - t_0) + (1 + c_2) R_{t-1} + c_3 (R_{t-1} - R_{t-2}) \qquad (4.5)$$

Parameters c_0, c_1, and c_3 correspond to the constant, the trend, and the auto-regressive term, respectively. Parameter c_2 reflects time series stationarity. The results for the *Sustained* groups are shown in Table 4.1. Six additional Chinese groups, , but none of the Spanish groups, kept depleting the resource ($c_1 < 0$) until the end of the experiment. Given more time, these groups would have likely crossed the no-recovery threshold.

A total of seven groups on each side sustained the overexploited resource, i.e., their exerted effort was sustainable, but they kept earning suboptimal profits (Fig. 4.3C). Interestingly, none of the groups managed to fully reverse the decline and finish with a recovering resource ($c_1 > 0$). A conclusion is that the outcome in both countries, especially when considering together the groups who kept depleting or already depleted the resource (red rectangle in Fig. 4.3C), was remarkably similar and rather dismal.

4.1.3.1 *Behavioural Model*

To understand how participants were behaving in the experiment, we constructed a statistical regression model of participants behaviour. The model's dependent (i.e., response) variable was effort, which we tried to explain using following independent (i.e., explanatory) variables:

- **The virtual forest's state**: we expected participants to exhibit different behaviours when the resource is abundant as opposed to when the resource is depleted.

- **Lagged own efforts**: included to account for potential autocorrelations in the play of individual participants; positive autocorrelations, in particular, would be an indication of decision-making "inertia" whereby high (resp., low) past efforts increase the likelihood of high (resp., low) present effort.

- **Lagged average efforts of others**: included to account for potential cross-correlations as a reflection of mutual influences between participants.

Model parameters, i.e., regression coefficients, accompanying these three types of explanatory variables were kept constant among participants from a given country, thus characterizing a collective behavioural focus. Individual differences entered the model by allowing constant terms and residual variances to be participant-specific via *fixed effects*, and via participant-specific residual variances, respectively. We interpreted the former as individualistic propensities to exert effort irrespective of the state of the explanatory variables. Accordingly, players with larger fixed effects were more likely to cut trees even if the number of trees left for logging was small, or even if other players refrained from logging. Residual variances, by contrast, quantified individualistic propensities to randomly vary effort. We introduced participant-specific residual variances because we expected that human participants would exhibit a wide spectrum of behaviours. With these ideas in mind, a general model formulation was

$$\begin{aligned} T_i(t) = \beta_R R(t) + \sum_{s=1}^{S_1} \beta_T^{-s} T_i(t-s) + \\ + \sum_{s=1}^{S_2} \beta_{\langle T \rangle}^{-s} \langle T(t-s) \rangle + \beta_i + \epsilon_i(t), \end{aligned} \tag{4.6}$$

where dependent variable $T_i(t)$ corresponds to the ith player's effort in round t. Among the three types of explanatory variables, $R(t)$ is the virtual forest's state in round t, $T_i(t-s)$ is the ith player's lagged effort s rounds prior to t, and $\langle T(t-s) \rangle$ is the lagged average effort of others, also s rounds prior to t. The numbers of lagged terms in the model, S_1 and S_2, were unknown prior to parameter estimation. Quantity β_i is the model's constant term, i.e., a fixed effect specific to the ith player. Finally, $\epsilon_i(t)$ are the model's normally distributed residuals with zero mean and residual variance σ_i^2, again specific to the ith player. Assuming the normal distribution here implied a lack of autocorrelative structure in residuals. This was reasonable given that the lagged own efforts in Eq. (4.6) should account for potential autocorrelations in player decisions.

The model is able to explain the posted efforts. Fig. 4.4 shows that predictions fit observations well as it is seen in observation-vs-prediction scatter plots: points gather around the "diagonal", i.e., the line with intercept 0 and slope 1. The coefficients of determination further indicate that the model accounts for nearly 60% (resp., 70%) of the total variance in the Chinese (resp., Spanish) data.

The behavioural regression model offers plausible explanations on why outcomes in China and Spain were similarly dismal. We found among the Spanish participants that, while the virtual forest's state and the average effort of others inform decisions on the current effort, a key determinant in this context is one's own lagged efforts (Fig. 4.5A). We thus witnessed a form of decision-making "inertia" by which past choices heavily weigh on the present choice. The effect is significant up to five lags in the past. Interestingly, the Chinese participants exhibit qualitatively the same behavioural patterns; again the forest's state and the average effort of others inform decisions, but these are much less influential than one's own lagged efforts (Fig. 4.5B). Even quantitatively the results are remarkably similar because only the effect of own effort at lag 2 is slightly weaker among the Chinese participants, while the effect at other lags is statistically indistinguishable between the two countries (Fig. 4.5B). The same is true for the average effort of others. Based on these results, one conclusion force itself upon us. Given that a decent number of participants from vastly different countries performed in a remarkably similar fashion, behavioural patterns behind the demise of the commons are, if not universal, then at least robust to a myriad of confounding factors.

The one substantial difference between the two countries is that the virtual forest's state correlates negatively with the effort of the Chinese, but positively with the effort of the Spanish participants (Fig. 4.5). The former start exploiting the resource more cautiously, but then compensate for a steady resource degradation with more effort. The latter, by contrast, start more aggressively, but then curtail their zeal in response to a disappearing resource. The described difference between the two countries helps to explain the faster resource depletion in Spain than in China (Fig. 4.3), and is fully consistent with the clustering results (Fig. 4.7). Analysing the participant-specific model terms further complements this explanation (Appendix Section a.1.1). Meticulous regression diagnostics show that we avoided the common pitfalls of this type of analysis, and thus that the model's results are credible (Appendix Section a.1.2).

4.1.3.2 *Clustering*

To gain a deeper insight about the differences among participants from both countries, we resorted to the k-means clustering algorithm. We used four quantitative characteristics as a basis for clustering with the idea that these characteristics would reflect behaviours exhibited in each of the two halves of the game experiment. They are cumulative efforts and total profits from both the first and the second half of the game taken separately. We surmised that behavioural changes between the two game halves would be of particular interest given that the resource state deteriorates as time passes, causing profits to decline as well.

In such an analysis, the optimal number of clusters into which the dataset should be partitioned is not a priori known. As many as 11 different optimality measures for addressing this problem are commonly found in literature [170]. Among these, we selected the silhouette method for its conceptual clarity [171]. The silhouette method contrasts cluster cohesion (i.e., how similar data points are to their respective clusters) to cluster separation (i.e., how dissimilar data points are to other clusters). The larger the average silhouette value of the dataset depending on the number of clusters, the better is the given partitioning into clusters. Using the silhouette method on Chinese and Spanish data separately, we first found that the Chinese participants are best partitioned into three clusters (Fig. 4.6A). The Spanish case is somewhat ambiguous because partitioning into two clusters yields only a marginally larger average silhouette value than partitioning into four clusters (Fig. 4.6B). A closer inspection of both options reveals that the results are more informative in the context of our game experiment when the Spanish participants are partitioned into four clusters.

The Chinese participants exhibit three prominent behaviours broadly describable as aggressive, moderate, and timid (Fig. 4.7A). Effort and profit gradually decrease from aggressive to moderate to timid players. Remarkably, performing independent clustering on the Spanish data reveals considerably similar behaviours patterns, with the addition of a fourth one, dubbed flipping (Fig. 4.7A). This last behaviour is aggressive or moderate in the first half of the game, but turns timid in the second half. We furthermore found that aggressive and timid behaviours are almost equally abundant in both countries, encompassing ≈25% and ≈10% of players, respectively (Fig. 4.7A). The Chinese case is enough to demonstrate that with such a distribution of players overexploitation is the most likely outcome. Adding the rather aggressive first-half behaviour of flipping players to this only contributes to the faster resource decline in Spain than in China, thus helping to explain why multiple Spanish groups managed to even cross the no-recovery threshold.

Prominent player behaviours show what separates optimal harvesting from sustained overexploitation from resource depletion. Groups who harvest optimally have almost the same composition in both countries (Fig. 4.7B), characterized by a relative scarcity of aggressive (≈17%) and a disproportional abundance of timid (≈33%) players. Groups responsible for sustained overexploitation also have almost the same composition in both countries (Fig. 4.7B), only here aggressive players are abundant (≈30%) and timid players are scarce (≈7%). The Chinese group who depleted the resource has the highest proportion of aggressive players (≈33%) and no timid ones whatsoever (Fig. 4.7B), while the corresponding Spanish groups have only a few stray timid players (≈2.5%). The latter groups also harbour almost all flipping players (≈45%), who act rather aggressively in the first half of the game and contribute to resource decline alongside aggressive players (≈21%). The four identified prominent behaviours thus go a long way in explaining the subtle differences in the virtual forest's time evolution between China and Spain, as well as the overall bias towards overexploitation. Particularly intriguing is a number of remarkable similarities between the two countries hinting at the existence of robust behavioural patterns behind the demise of the commons.

4.1.4 *Discussion*

Having asked participants from China and Spain to exploit a virtual forest while facing the same epistemic and socio-economic obstacles as real-world operators, we found that seemingly different outcomes are, in fact, remarkably similar and bode ill for the fate of common-pool resources. An exploratory data analysis in the form of clustering reveals that the results are largely attributable to three behavioral types (also called phenotypes in the literature), dubbed aggressive, moderate, and timid. Although the nature of the game in our experiment is different from those in previous experiments that report behavioral phenotypes [172–174], we see clear parallels between aggressive, moderate, and timid players herein and defectors, cooperators, and supercooperators in Ref. [174], respectively. The consistency of previously identified behavioral phenotypes [173, 174] further suggests that the types we found are also a consistent feature of human behavior rather than a peculiarity of the specific experimental setup. In fact, having worked with two geo-socially distant populations, and in a novel and relatively complex context, our results go a long way in fortifying the conclusions of the cited studies that human behaviors in social dilemmas are divisible into a small number of stable phenotypes.

A previous study [175] using a similar setup, albeit with explicit resource "dynamics" such that every 10 standing trees yielded one new tree per round, reported the outcome of the game experiment compared to other situations. Here, by contrast, we implemented a more realistic dynamic—whose qualitative characteristics, but not quantitative details, are known by the participants—and identified collective behavioral mechanisms that underpin decisions on exploitation, thus pointing to one main culprit for similarly dismal outcomes in both countries. Instead of prioritizing the resource state when deciding the current effort, participants operate under decision-making "inertia" by which they are much more concerned with their own past efforts. A surprising aspect here is that this mechanism materializes in two populations not only separated geographically, but also influenced by a myriad of confounding factors such as age, education, and culture. The Chinese participants shared comparatively young age, exposure to higher education, and upbringing in the midst of a quintessential East Asian cultural heritage. The Spanish participants mirrored the general population in terms of age and educational background, while socio-culturally belonging to a typical western democracy. Given that the same mechanism materialized despite these large differences, we concluded that behavioral patterns behind the demise of the commons are highly robust to confounding factors. It remains open for future research to explore whether changes in the experimental design would yield significant differences between populations. For example, changing the profit per tree by adjusting the price of trees or the unit cost of effort would affect the strength of the underlying dilemma, and thus provoke more or less logging. Whether participants from different countries would be equally sensitive to variations in dilemma strength remains unclear at the moment.

Global environmental risks can no longer be contained without cooperation at an unprecedented scale in human history [176–178], but does humankind have what it takes to achieve such cooperativeness? The existence of collective behavioral patterns that are robust given a specific contextual situation is a reason for cautious optimism. In the case of common-pool resource exploitation, for example, encouraging a shift in focus from one's own past decisions to the resource state should reduce overexploitation in China and Spain alike. The aim here is to raise awareness of problematic behaviors, unlike experimental treatments that try to evoke a cooperative state of mind by indirect suggestion, e.g., by exploiting a known cognitive bias [151]. More generally, robustness promises that precautionary policies or educational programs, when crafted with great care, may curb risky behaviors across continents and cultures. Pursuing this promise, therefore, has the potential to become an attractive research agenda for a wide variety of multidisciplinary studies on the origin of human cooperation.

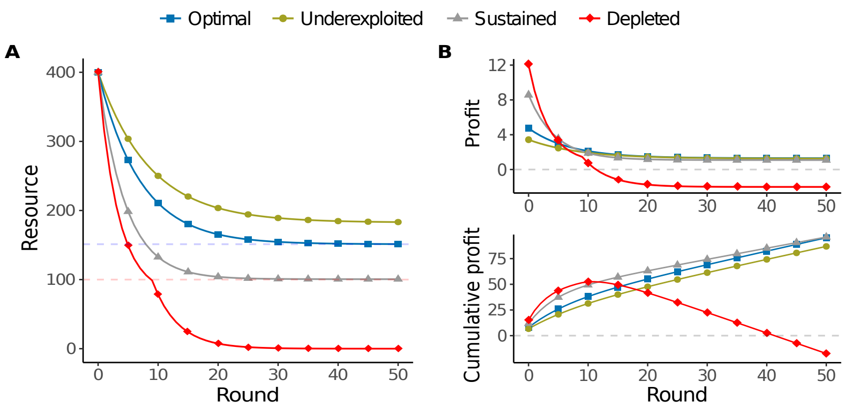


Figure 4.1: **Resource state and profits are driven by efforts. A,** Resource dynamics is a combination of regrowth and exploitation. Constant effort leads in the long run to an equilibrium state wherein the number of, in our case, trees left for logging decreases with more effort. If effort is extremely high, regrowth ceases due to the Allee effect and the resource gets depleted. **B,** Profit in a single round is higher when more of the resource is exploited. As the resource approaches an equilibrium, it is optimal in the long run to find the equilibrium that maximizes exploitation, and thus profit. The differences in profit per round among the different equilibria may not be large (upper panel), but they accumulate over time (lower panel). Extremely high effort, by contrast, generates high short-term profits that eventually turn into losses once the resource gets depleted. Parameter values correspond to the ones used in the experiment, namely: $g = 0.0504\,\mathrm{d}^{-1}$, $M = 400$, $\tau = \frac{1}{14}\,\mathrm{d}$, $R_c = 100$, $p = 1$, and $c = 2$. Consequently, $T^* = 2.765$ days per week and $R^* \approx 151$ trees.

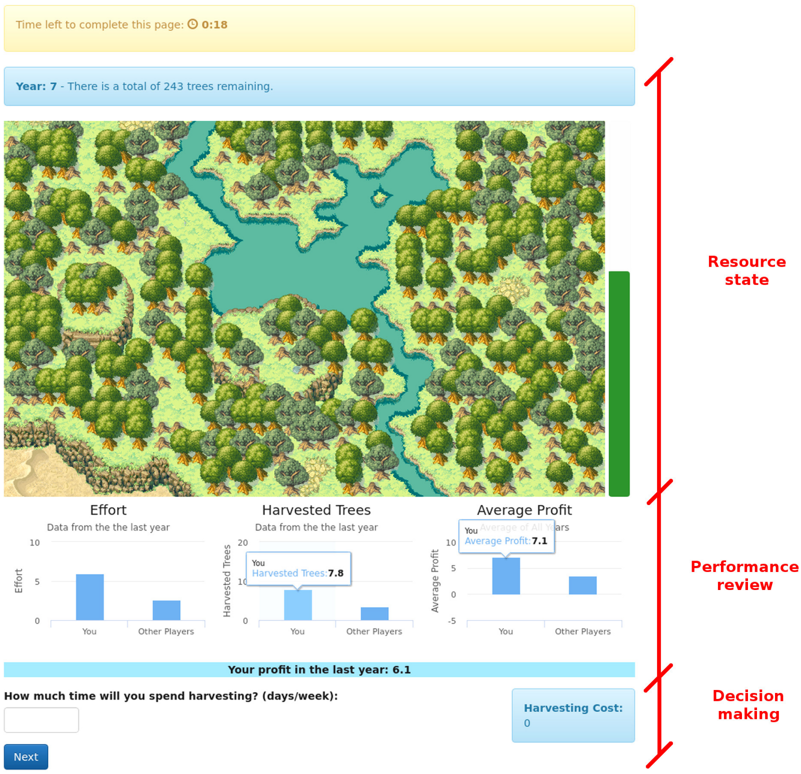
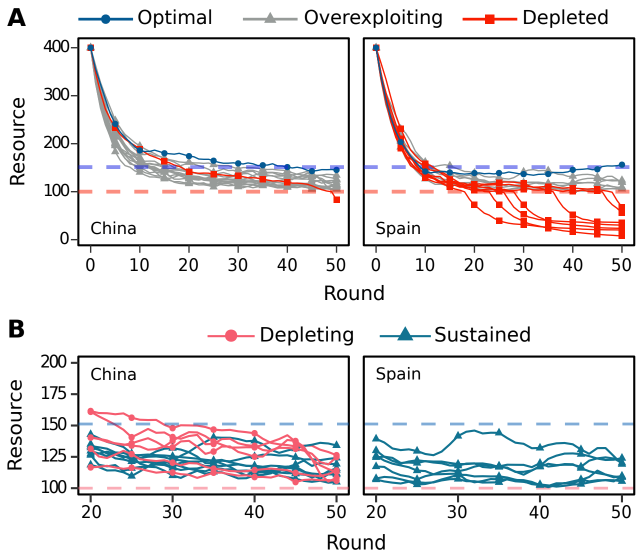


Figure 4.2: **Screenshot of the gameplay page**. We divided the gameplay page into three distinct parts: *resource state*, *performance review*, and *decision making*. The resource state part consisted of a game-like visualization of the virtual forest plus a status (resp., progress) bar showing the number (resp., fraction) of remaining trees. The performance review part focused on effort, harvest, and profit bar charts with a hover effect such that moving the cursor over any of the bars triggered a tool tip displaying the corresponding numerical value. Lastly, the decision-making part comprised a simple input form asking for the desired effort and a message box that automatically converted effort into the harvesting cost. Final decisions had to be confirmed by clicking the *Next* button. For the actual sessions of the experiment, we used Chinese or Spanish translations.



C

	China	Spain
Underexploited	0	0
Optimal	1	1
Overexploited	15	14
Recovering	0	0
Sustained	7	7
Depleting	6	0
Depleted	1	7

Figure 4.3: **Overexploitation is a trend. A,** Out of 16 Chinese and 15 Spanish groups who exploited the common-pool resource, only one group from each country was able to keep the resource at an optimum. We defined the optimum as ±10% from the number of trees maximizing the sustainable yield (≈151 trees). Worryingly, all other groups overused the resource, and what is more, one Chinese and seven Spanish groups depleted it below the no-recovery threshold (=100 trees). **B,** Chinese groups seemingly do better than their Spanish counterparts, but is this truly so? To examine the likely fate of overexploited, but non-depleted virtual forests beyond round 50, we tested whether after a transitory period of about 20 rounds the number of trees was recovering, sustained, or depleting (Table 4.1). We found that six Chinese groups kept depleting the resource until the very end, while seven groups from each country sustained a relatively constant number of trees. No groups from either country managed to overturn the negative trend and allow the resource to recover. **C,** Notably, none of the groups from the two countries underexploited the resource, while a total of seven groups from each country depleted or would have likely ended up depleting the resource (red rectangle).

	Parameters				Goodness of fit	
Country	c_0	c_1	c_2	c_3	R^2	R^2_{adj}
China	114.047***	-0.394***	-0.816	0.770***	0.89	0.87
China	41.970**	-0.036	-0.365***	0.319	0.70	0.66
China	21.193	-0.145	-0.163***	0.376*	0.98	0.98
China	30.339**	-0.101*	-0.215***	0.871***	0.94	0.94
China	21.944	-0.070	-0.183***	0.085	0.93	0.93
China	18.049	-0.075	-0.147***	0.150	0.98	0.98
China	83.821***	-0.274**	-0.717	0.305	0.92	0.91
China	67.756*	-0.476**	-0.418***	0.280	0.98	0.98
China	50.327***	-0.047	-0.430***	0.568**	0.72	0.69
China	27.221*	-0.002	-0.229***	-0.216	0.84	0.83
China	17.294**	0.047	-0.137***	-0.706***	0.95	0.94
China	64.589***	-0.300**	-0.508***	0.516**	0.94	0.94
China	80.330***	-0.561***	-0.520***	0.788***	0.94	0.94
China	41.448**	-0.049	-0.324***	0.521**	0.77	0.74
Spain	30.941*	-0.078	-0.279***	0.398*	0.90	0.89
Spain	22.099*	-0.021	-0.190***	0.235	0.82	0.80
Spain	26.041**	-0.052	-0.223***	0.515***	0.89	0.87
Spain	30.323*	0.036	-0.290***	0.237	0.63	0.59
Spain	20.294	0.017	-0.193***	0.327	0.72	0.68
Spain	25.381*	-0.063	-0.184***	0.671	0.85	0.83
Spain	27.124*	-0.001	-0.227***	0.259	0.74	0.71

Table 4.1: **Time-series analysis of the virtual forest's state to determine the presence of significant trends.** Star symbols *, **, and *** signify 5%, 1%, and 0.1% statistical significance, respectively. We tested if $c_i \neq 0$, $i = 0, 1, 3$, and $c_2 > -1$.

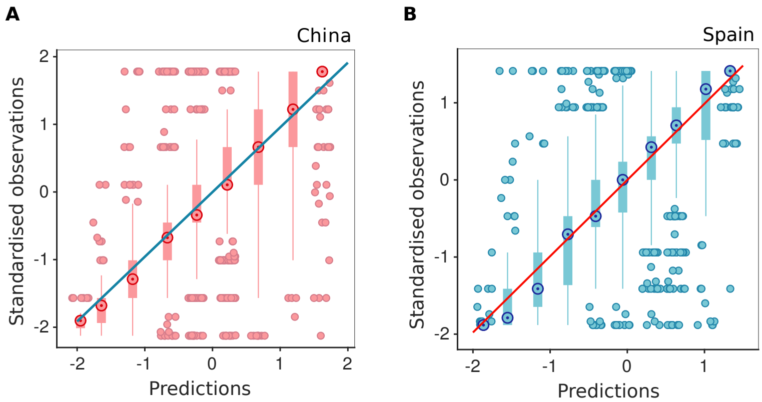


Figure 4.4: **Behavioural regression performance.** Observation-vs-prediction scatter plots and the accompanying statistics intuitively display and quantify the performance of statistical regression models. In such plots, the scattered points should group around the "diagonal", meaning that the line fitted to these points should be statistically indistinguishable from the line with intercept 0 and slope 1. **A,** For the Chinese data, the intercept is indeed indistinguishable from 0 (estimate -0.0018; 95% CI $[-0.0221, 0.0186]$), but the slope is slightly lower than 1 (estimate 0.9544; 95% CI $[0.9279, 0.9809]$), thus suggesting that the model somewhat overpredicts (resp., underpredicts) low (resp., high) efforts. The coefficients of determination is $R^2 = 0.592$. **B,** For the Spanish data, these minor performance issues disappear because not only the intercept is indistinguishable from 0 (estimate -0.0006; 95% CI $[-0.0193, 0.0180]$), but also the slope is indistinguishable from 1 (estimate 0.9888; 95% CI $[0.9659, 1.0116]$). The coefficients of determination is $R^2 = 0.689$. Due to a large number of data points (>4000 per plot), we grouped them into bins as evenly as possible, and then displayed the medians (circles), the interquartile ranges (boxes), the limits that would encompass 99.3% of normally distributed data (whiskers), and "outliers" (individual points).

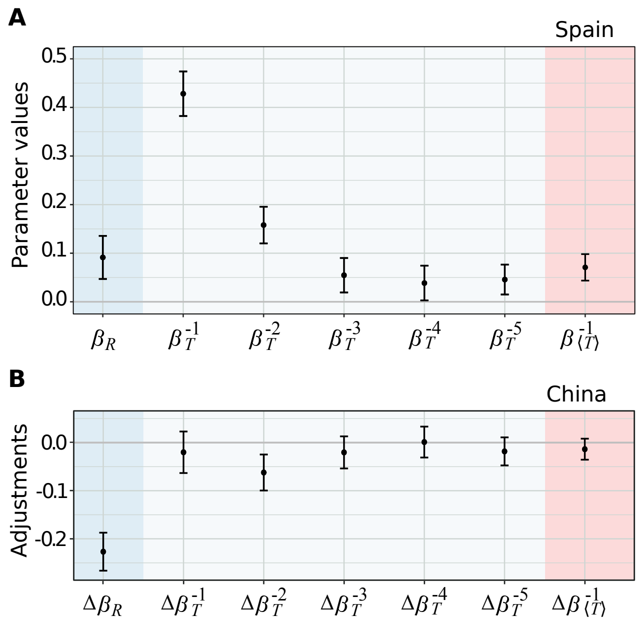


Figure 4.5: **Behavioural patterns behind the demise of the commons are robust across nations. A,** Estimated parameter values show that while the virtual forest's state (parameter β_R) and the effort of others (parameter $\beta_{\langle T\rangle}^{-1}$) inform participant decisions, the Spanish participants exhibit a form of decision-making "inertia" by which the current effort strongly reflects previous own efforts (parameters β_T^{-1} to β_T^{-5}). The effect is significant up to five lags in the past. Here, shown are the parameter estimates (points) and the corresponding 95% confidence intervals (error bars). **B,** Adjustments of the Spanish parameter values to fit the data from China indicate that the Chinese participants exhibit the same decision-making "inertia" as their counterparts in Spain. The effect is only slightly weaker at lag 2 (parameter $\Delta\beta_T^{-2}$), but otherwise statistically indistinguishable between the two countries. The effort of others also has a statistically indistinguishable effect. The only qualitative difference is reflected in the $\Delta\beta_R$ parameter, revealing that the Chinese (resp., Spanish) participants exert more effort when the resource is scarce (resp., abundant). This is consistent with a gentler (resp., steeper) initial decline of the resource in China (resp., Spain). The negative relationship between resource abundance and effort in China backs up our conclusion from the time-series analysis (Table 4.1) that six additional Chinese groups would have eventually depleted the resource.

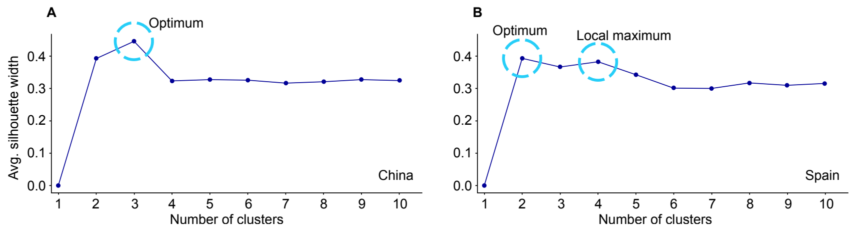


Figure 4.6: **Determining the optimal number of clusters with the average silhouette width.** The silhouette value is a measure of how well a data point fits to its own cluster as opposed to other clusters (cohesion vs. separation), ranging from -1 for a poor fit to 1 for a good fit. Averaging silhouette values over an entire dataset produces an aggregate measure, called the average silhouette width, of how well the data have been clustered. This measure is a function of the number of clusters. The best clustering is achieved with the number of clusters for which the average silhouette width is maximal. **A,** For the Chinese data, the optimal number of clusters is three. **B,** For the Spanish data, partitioning into two or four clusters yields nearly an equal average silhouette width. We opted for the latter number because four clusters proved to be very informative in the context of our game experiment.

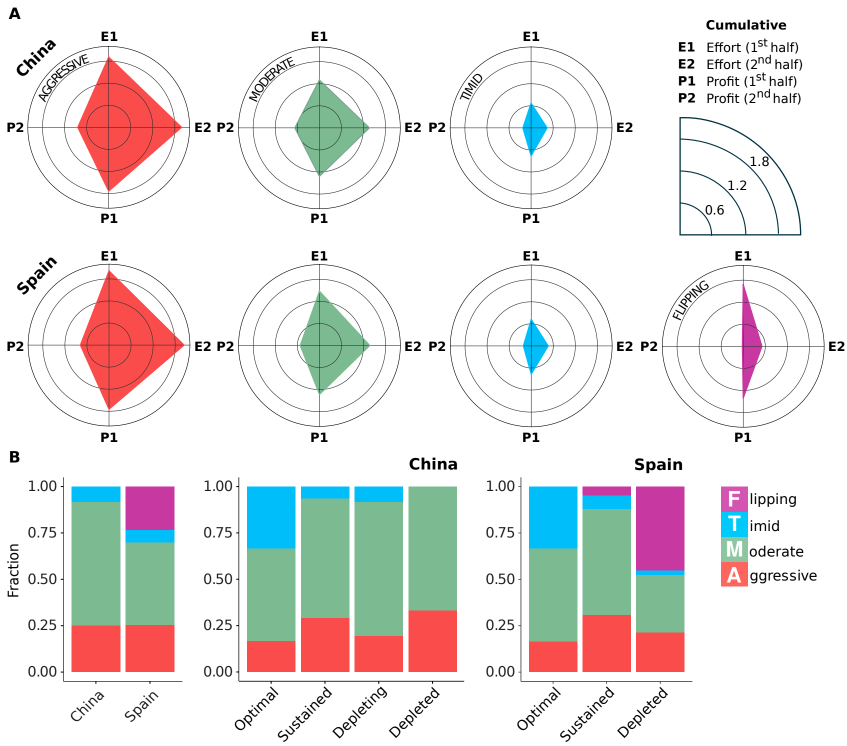


Figure 4.7: **Interplay of prominent behaviours explains overexploitation. A,** Running a clustering algorithm on data from China and Spain separately, we identified three (resp., four) distinct prominent behaviors among the Chinese (resp., Spanish) participants. Apart from the behavior unique to Spain, the three remaining behaviors are nearly identical irrespective of the country. In terms of effort, these can be described as aggressive, moderate, and timid. With the scale set relative to the MSY effort and the corresponding profit, we see that aggressive players exceed the MSY effort by over 80%, earning large profits in the first half of the game. Moderates stay closer to the MSY effort, nonetheless exceeding it by about 20%. Timid players start cautiously at 60% of the MSY effort and, unlike aggressive or moderate players, reduce effort in response to resource deterioration. The fourth Spanish behavior flips from an aggressive initial stance to a timid subsequent one, earning almost no profit late in the game. **B,** Overall abundance of aggressive and timid players is remarkably similar across countries (left panel), as is the abundance of these players in groups that played optimally and groups that sustained the resource in an overexploited state (middle and right panels). Optimal play clearly requires a much more favorable aggressive-to-timid ratio than is present in the overall abundance, thus explaining overexploitation. The flipping behavior is nearly exclusive to groups that depleted the resource in Spain, indicating that many players become responsive to the resource state only when it is too late.

4.2 TARGETS AND BIASES

Targets and biases: a collective-risk social dilemma between two countries, *in preparation*

F. Maciel Cardoso, M. Jusup, C. Gracia-Lázaro, Z. Wang, A. Sánchez, & Y. Moreno

Anthropomorphic causes are increasing our emissions of greenhouse gasses, and there is no compelling indication that a global agreement will reduce them sufficiently [179, 180]. This is not surprising as the reduction of emissions directly implies economic constraints, and current manufacture is overwhelmingly dependant on fossils fuels as food production is on livestock. In this regard, reducing the pace of climate change epitomize a global social dilemma played by states and corporations [31]. A peculiarity of this dilemma is the associated collective risk, leading to all parties being severely punished if climate change is not successfully mitigated. These characteristics were embodied in an experiment by Milinski et al. [160], in which participants contributed to a common pool with a common target. If the target was missed, participants were very likely to lose all their money, simulating the dangerous risk of climate change.

In this *"collective-risk social dilemma"*, at variance with Public Goods Games, players contribute to avoid a loss and not to increase gains. Specifically, according to Milinski et al. [160], the collective risk social dilemma has defining characteristics that distinguish it from other types of social dilemmas: (i) people have to make decisions repeatedly before the outcome is evident, (ii) contributions to the common fund are lost, (iii) it is unclear what is the effective value of the public good, and (iv) people's remaining money will be lost with some probability if the sum of the contributions does not reach the target.

Experiments have shown that subjects are very responsive to high risk, but a significant number of groups is not able to reach their goal [159, 160, 181]. In this line, it has been shown that communication increased the probability of reaching the target, as it allows for participants to indicate their future actions [181], such as observed in common-pool resources management [103, 148, 149].

Nonetheless, these solutions are only possible when participants can trust each other. Moreover, communicating is not always possible on a global scale. Economy globalization implies stakeholders of different cultural and socio-economic backgrounds, and it is unclear how peoples' responses differ when confronted with foreigners. Cultural differences, prejudices, and miss-comprehension can be all at play when people from different countries interact with each other.

Here, we devise an experiment to check how people of two different countries respond in the climate change dilemma game. Participants had to contribute to the collective risk social dilemma jointly with subjects from a different country. To see if the nationality of the other country influenced contributions, we provided this information in some sessions. Other works have explored this game with participants from different countries [159], however, to the best of our knowledge, our setup is the first that allows observing if bias towards a different country conditioned participants' contributions.

Moreover, the equity of contributions is intrinsic to climate change mitigation discussions [182]. In this line, countries with different past resource appropriation and rate of emissions should have different loads [183, 184]. Nonetheless, citizens not necessarily respond well to different responsibilities. In the collective-risk social dilemma, this would correspond to unequal group targets, which might affect participants willingness to contribute. To check if this was the case we performed two types of sessions: i) *Homogeneous target*, participants from each country had the same target; ii) *Heterogeneous target*, one country's target was twice the other country's target. Furthermore, each participant played two treatments: in the Homogeneous sessions, they played in a treatment with a global threshold of GT=€60 for the whole group, followed by another treatment in which, in addition to the global threshold, each country had a local threshold of LT=€30; in the Heterogeneous sessions, each country played in a treatment with a local target of €20, while the other had a local target of €40, and subsequently both played with inverted targets. Thus, every participant played two treatments, as described in Table 4.2.

Session Type	Treatment 1	Treatment 2
Homogeneous	$GT = 60$	$LT_{China} = 30$
		$LT_{Spain} = 30$
Heterogeneous	$LT_{China} = 20$	$LT_{China} = 40$
	$LT_{Spain} = 40$	$LT_{Spain} = 20$

Table 4.2: **Targets in each session type.** We performed two types of sessions, one wherein Chinese and Spanish participants had the same target (*Homogeneous sessions*), and another wherein they had unequal targets (*Heterogeneous sessions*). Moreover, sessions could be *Informed* (participants new the other country nationality) or *Uninformed* (participants did not know their nationality).

We also performed sessions differing in the information provided to participants. Players were assigned to an *Informed* or to an *Uniformed* session, such that in the former they would have information about the other country nationality, while in the latter they only knew they were from a different country. This experimental setup allows us to verify if providing the nationality information of other players influenced participants' contributions.

We ran experimental sessions with 96 participants from the general population of Zaragoza, Spain, and 96 undergraduate students from Xi'an, China. They played in groups of 12, which further contained a subgroup of 6 Chinese and one of 6 Spanish participants. In each treatment, participants started each treatment with an endowment of €10 and in each round had to decide to contribute €0, €0.25, €0.5, €0.75, or €1 to the common fund. Accordingly, in the Homogeneous sessions, €0.5 was the fair contribution. In the Heterogeneous sessions, an average of €0.33 (resp. €0.66) per round was necessary for reaching the local target of €20 (resp. €40). During the game, subjects had full information on other players contribution, how much was necessary to reach the target as also an indication of how many rounds were left. If at the end of the treatment the global target was not achieved, subjects had a 0.9 % probability of losing all their money. Furthermore, when there was a local target (all treatments except the first one in the Homogeneous sessions), when it was not achieved participants in

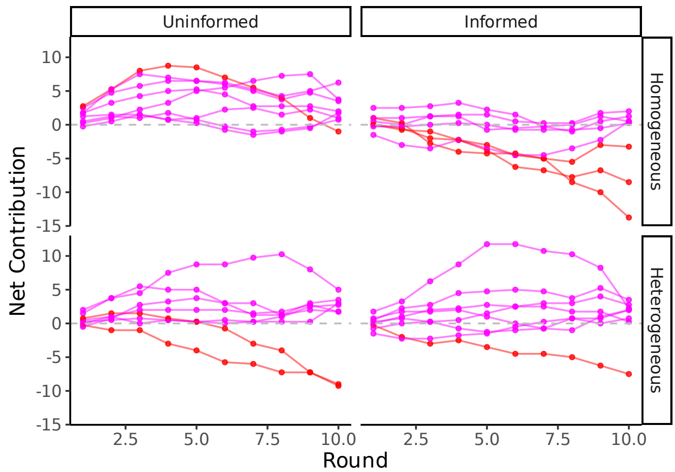


Figure 4.8: **Net contributions per round.** Vertical panels show values for each treatment and horizontal panels for each session type. Blue (resp. red) curves correspond to the cases wherein the target was (resp. was not) achieved.

the subgroup received a penalty of 2€, ensuring that participants also had an incentive to reach the local target.

4.2.1 *Results*

Fig. 4.8 shows the evolution of net contributions according to the distribution of targets and the information provided to participants. Most groups were able to reach the global target (25 out of 32). Moreover, regarding outcomes, it seems that there was no major difference between the heterogeneous and homogenous sessions (3 versus 4 failed groups), both in the Informed (Fischer's exact test: P =0.56) and Uninformed (Fischer's exact test: P =1) conditions. The total contributions also seem close to each other, as suggested by unequal variances t-tests of differences in total contributions, both in the Informed (P=0.18, $t_{11.7} = -1.4$) and Uninformed (P=0.31, $t_{9.1} = 1.1$) conditions, although with small statistical power.

Although there is no significant difference between groups' outcomes, the pattern of contributions per round suggests that investments to the common fund are smaller in the Homogeneous-

Informed sessions with respect to the Uninformed ones. None of their groups seems to have reached a high positive net contribution, such as it is observed in the other session types. Furthermore, it might be the case that participants from the two countries behave differently in each session despite the apparent agreement between outcomes. For a proper examination of these questions, we have to take into account the unobserved heterogeneity of individuals, as they might affect their contribution patterns. This is salient in our setup, as individuals play repeatedly and their decisions are not independent. To address those issues, we perform a random effects linear regression model [185, 186] taking into account the individual heterogeneity. For the Homogeneous sessions it is specified by the following equation:

$$\begin{aligned} C_{it} = \beta_0 + \beta_1 Informed_i + \beta_2 Spanish_i + \\ \beta_3 Informed_i * Spanish_i + \\ \beta_4 LT + \beta_5 LT * Spanish_i + \\ \beta_6 Round_t + \\ \alpha_i + \epsilon_{it} \end{aligned} \tag{4.7}$$

The dependent variable C_{it} corresponds to the contribution given by participant i at round t. The dummy variable $Informed_i$ indicate if participant i was playing in an Informed treatment, thus, the reference group are the participants in the Uninformed treatment. We also include a dummy variable for the participant country ($Spanish_i$), being the Chinese participants the reference group. Moreover, as they possibly respond differently to information, we also included an interaction term between these two variables. Similarly, *LT* is a dummy controlling the treatment participants were in (ref. *GT*) – i.e, *LT* is equal to 1 when participants are playing the *LT* treatment and 0 otherwise. We also controlled time effects through the variable $Round_t$. Finally, α_i corresponds to the participants' individual effect, and ϵ_{it} to other unobserved factors.

For the Heterogeneous sessions, we normalized participants contributions according to their local targets. The regression model is similar to the Homogeneous, and it is specified according to the following equation:

	Dependent variable:
	Player Contribution
Spanish	0.036 (0.030)
Informed	-0.081^{**} (0.033)
LT	0.010 (0.021)
Round	-0.009^{***} (0.003)
Informed x Spanish	0.081^{*} (0.043)
LT x Spanish	-0.053^{**} (0.027)
Constant	0.558^{***} (0.027)
Observations	1,920
R^2	0.018
Adjusted R^2	0.015
F Statistic	35.875^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4.3: **Homogeneous sessions regression**. Estimates from a random effects model for the Homogenous sessions' participants. Robust standard errors are clustered at the individual level.

$$\begin{aligned}\tilde{C}_{it} = \beta_0 + \beta_1 Informed_i + \beta_2 Spanish_i + \\ \beta_3 Informed_i * Spanish_i + \\ \beta_4 LT_{40} + \beta_5 LT_{40} * Spanish_i + \\ \beta_6 Round_t + \\ \alpha_i + \epsilon_{it}\end{aligned} \tag{4.8}$$

In this case, $\tilde{C}_{it}$ corresponds to the normalized contribution and LT_{40} is a dummy variable controlling for the target of the participant's group ($LT = 20$ is the reference group) – i.e, it equals to 1 in the LT_{40} treatment and 0 otherwise.

The results of the regression for the Homogeneous and Heterogeneous sessions are shown in Tables 4.3 and 4.4, respectively. The

	Dependent variable:
	Normalized Player Contribution
Spanish	0.102 (0.150)
Informed	0.021 (0.072)
$(LT = 40)$	-0.162^{***} (0.035)
Round	-0.035^{***} (0.008)
Informed x Spanish	-0.084 (0.136)
$(LT = 40)$ x Spanish	-0.152^{*} (0.084)
Constant	1.314^{***} (0.083)
Observations	1,920
R^2	0.056
Adjusted R^2	0.053
F Statistic	112.892^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4.4: **Heterogeneous sessions regression**. Estimates from random-effects regression for the Heterogeneous sessions' participants. Robust standard errors are clustered at the individual level.

decay in contributions with time is common to both session types, as generally it is observed in repeated Public Goods Games [46, 48]. In the homogeneous sessions, Chinese participants contributed less when information of Spaniard's nationality was disclosed (-0.08 in contributions per round). Interestingly, this effect is not observed for Spanish participants (0=0.08-0.08), and although their contributions are smaller when their group had a local target, they were still around the fair (0.51=0.56-0.05). These effects are relatively small but they might be enough to lead groups to collapse, as their baseline contributions are only slightly over the fair (0.56). Importantly, the effect of information might be enough for groups in the Informed treatment to be closer to the tragedy of the commons. This is evidenced by comparing the total contributions in each condition, which indicates that they were smaller when participants were informed (one-sided unequal variances t-test: $t_{9.1} = -2.3, P = 0.023$), though with modest statistical power. Likewise, comparing the means of the two countries groups when playing together suggest that Chinese groups' total contributions were significantly smaller than from Spanish (one-sided unequal variances paired t-test: $t_7 = 2.5, P = 0.04$). These differences are due to some participants having paltry mean contributions in the Informed sessions, as shown by Fig. 4.9. Without information, most participants will likely contribute the fair; with information, however, some of them will free ride.

This pattern suggests that Chinese participants might be less willing to contribute when they know they are playing 'against' Spanish participants. This hypothesis, however, is unsupported by the results of the Heterogeneous sessions: they indicate no difference between Spanish and Chinese participants with respect to the disclosed information. Naturally, it shows that the relative contributions are smaller when participants have larger targets, a likely result of having a higher toll on their endowment. In this case, the Spanish participants seem to have contributed less when they were obliged to make higher contributions than the other group (-0.15), however, we are not able to distinguish this from an experimental artefact. In the Heterogeneous sessions, Spanish participants always started with a larger target, which might have induced them to contribute less, which is not the case of Chinese

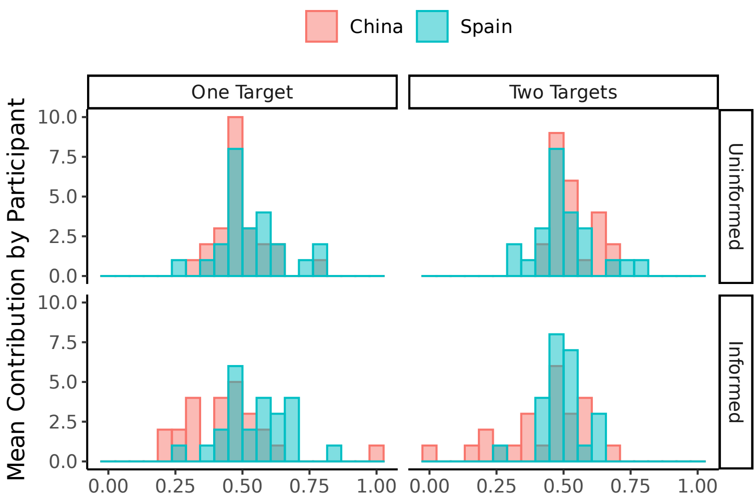


Figure 4.9: **Distributions of players' average contribution in the Homogeneous sessions.** Histogram of players' mean contributions according to the information provided and treatment in the Homogeneous sessions. The presence of participants with averages smaller than the fair results in a reduction of the group's average contribution.

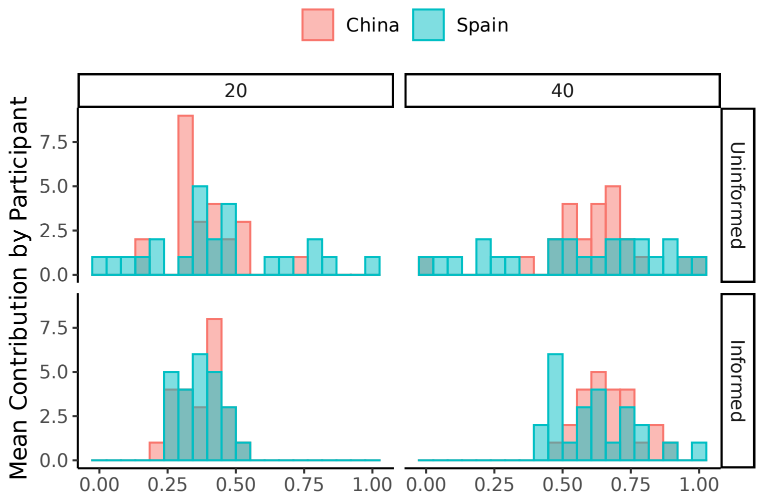


Figure 4.10: **Distributions of players' average contribution in the Heterogeneous sessions.** Histogram of players' mean contributions according to the information provided and the group target in the Heterogeneous sessions. Left (resp. right) panels correspond to groups with a target of 20 (resp. 40).

participants. Thus, it remains an open question whether this effect is valid. Nonetheless, it is remarkable that players from both countries complied adequately with the inequity, even if we consider the negative effect on the Spanish participants' contributions (1=1.31-0.16-0.15). Curiously, providing information seem to have clustered the contributions around the target in the Heterogeneous sessions, as shown by Fig. 4.10. When players did not know the other group nationality, their average contributions spread almost uniformly over the whole possible range, especially for Spaniards. Suggesting that unknowing the other country nationality might influence the appearance of both free-riding and altruistic behaviour.

4.2.2 *Discussion*

The outcomes of Spanish and Chinese participants while playing a collective-risk social dilemma does not seem to be influenced by whether targets are homogeneously or heterogeneously distributed among them. Our analysis indicates that, in general, most participants will contribute around the fair, except by some participants contributing less in the Informed sessions. This leads to a reduction in the total contributions of Chinese participants, raising the risk of their groups not meeting the target in the Homogeneous sessions. Notably, this effect is not observed in the Heterogeneous sessions, indicating that, although outcomes are similar, participants might react to information differently in both experimental setups. In general, nonetheless, all participants seem to contribute around or over the fair, even when they have the burden of a larger target. Spaniards' contribution is smaller than the Chinese's in this latter case – i.e., LT_{40} treatment in the Heterogeneous sessions – but not enough to nullify this trend.

Moreover, in our setup groups could also have a local threshold, implying that participants are concerned with two groups, one with 12 participants and a subgroup with 6 participants, which in turn could increase the likelihood of reaching the target according to some hypotheses [187]. Nevertheless, we observe just a small effect in the other direction for the Spanish participants, i.e., they contribute less when a local target was introduced in the second treatment of the Homogeneous sessions. Seemingly, thus, contributions can be smaller when a local target is present. However, further replications are necessary to confirm whether this result is affected by order effects, as in our setup participants always played with a local target after playing the *GT* treatment. Either way, the Chinese participants are not affected by the order or local target in the Homogeneous sessions, contributing around the same in both treatments.

It seems participants respond well to the targets imposed on them, even when they are larger than the other participating group. This suggests that the application of differentiated responsibilities might be well accepted by rich countries, which should be confirmed by more experiments and work outside the lab. Nonetheless, our results indicate that participants behaviour can differ significantly between heterogeneous and homogeneous sessions, which demonstrates the necessity to investigate what factors might underlie different responses to information. Future work might also be able to unveil whether players from other countries respond differently to information and inequity in targets.

One of the difficulties of performing synchronized experiments between two countries, such as the ones presented here, is obtaining large samples for more statistical power. Despite finding statistically significant evidence, our tests at group level rely on small samples, implying caution while interpreting these results. It remains open whether future replications can reproduce our results with more groups.

5 FRAMING & ALTRUISM

How selfish so ever man may be supposed, there are evidently some principles in his nature, which interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it.

Adam Smith, Theory of Moral Sentiments

Seated beggar and his dog, Rembrandt

Policymakers, legislators, and public institutions in general must know how people respond to incentives and constraints. For quite some time, at least since Machiavelli, it was believed that humans would always behave selfishly, hence public policy should aim at providing ways to turn human selfishness into social welfare[1]. This justified the widespread uses of material incentives to motivate the supposed *homo economicus* to act in some specific way. Nevertheless, as we have illustrated in Section 2.2, people have social preferences and disregarding this fact can lead to suboptimal policies [188], or even result in public policy to backfiring [189].

"it is necessary for anyone who organizes a republic and establishes laws in it to take for granted that all men are evil and that they will always act according to the wickedness of their nature whenever they have the opportunity" N. Machiavelli [190]

Incentives seldom are orthogonal and additive to people intrinsic responses, indeed, they interact with their moral and psychological motives, possibly increasing their pro-social response synergistically or, on the contrary, undermining it [12]. Ideally, therefore, it is necessary to be aware of the resulting effect of the incentive, or, if not possible, at least observe if it works as intended. Moreover, it also should be taken into account that policies efficacy will also depend on people's imprecise decision-making process [191]. People behave according to elusive heuristics [191–193] which doubtfully will correspond to the behaviour of a self-interested and rational agent. In fact, humans demonstrate pro-social responses from an early age, which can be undermined by material incentives [4]. Moreover, there is widespread evidence that people contribu-

1 Samuel Bowles provides a pithy account of the origin of this view in [12]. Interestingly, he shows that even the proponents of the legislating for the selfish man believed that humans had social preferences.

tions and investments are significantly affected by framing effects [191, 194]. Thus, to devise effective public policies and sustainable business practices it is imperative to recognize humans' natural propensity to act altruistically [47] and further identify how people respond in each specific scenario.

"it is time to fully embrace what I would call evidence-based economics" Richard Thaler *[33]*

Naturally, applying behavioural experiments are a suitable approach to enhance our collection of people responses in different scenarios. It allows us to grasp how people are expected to behave without the burden of the unintended consequence of an ill-devised policy. In this regard, socially responsible investments and charitable donations are particularly relevant nowadays as global interest in them increases. To uncover people's responses in these two contemporary investment situations, we have performed two experiments looking in how people's pro-sociality can be affected by the specifics of *i)* public goods with donations and *ii)* impact investing funds – investments whose goal is to generate social and environmental benefits alongside economic returns. In Section 5.1 and in Section 5.2 we present experiments to uncover peoples' choices in the first and second cases, respectively.

5.1 FRAMING EFFECTS IN CONTRIBUTIONS AND DONATIONS

Framing in multiple goods games and donations to charities, *under review.*

F. Maciel Cardoso, S. Meloni, C. Gracia-Lázaro, A. Antonioni, J. A. Cuesta, Á. Sánchez, & Y. Moreno

The number and economic relevance of charities and non governmental organizations has rapidly grown in the last few decades. For instance, more than 1.5 million nonprofits were registered with the US IRS in 2015, contributing around 5.4% to the US GDP [195]; in 2016/17, there were 166,854 voluntary organizations in the UK, employing about 878 000 people [196]. This growth has been fueled by the subsidies of many governments around the world, either by transferring funds directly to organisations or through tax deduction policies to donors[197–204]. At the same time, more than 1 billion people give money to charities [205]. In view of this volume of activity, philanthropy and voluntary contributions to charities have aroused the interest of a growing number of researchers in the last decades [206, 207], leading to theoretical models [208], qualitative research [209], and experimental studies on the economics of charity [210], fundraising events [211], different forms of fundraising [212], and the effect of status [213], lead donors [214], rebates [215], subsides [216], and message framing [217] on charitable giving.

Secondly, when studying altruistic behaviour in humans, gender differences deserve special attention. Empirical evidence suggests that women give more to charities than men [218]. Socio-cultural and evolutionary theories predict sex-differentiated behaviour [219], although they often disagree on how men and women will behave in specific circumstances. Socio-cultural theory stresses the role of cultural stereotypes [220] whereas evolutionary theory explains sex behavioural differences as adaptations [221]. Particularly, both theories agree on the existence of behavioural differences with respect to cooperation or altruism. Many experiments have been conducted to assess these differences, and in general, women show higher levels of cooperation and altruism than men [222–225], although other studies show that gender does not affect these traits [226, 227].

One of the most frequently used frameworks to experimentally address donations to charities is that of public good games (PGG)[46]. The representation of donations to charities through a PGG is far from perfect, but approximate enough to have been considered often in the literature [228–230]. In this context, the research question we address in this work focuses on the effects of framing on both contributions to PGGs and donations to charities [2]. In order to compare the effectiveness of different fundraising schemes, we have carried out an experiment involving contributing to multiple PGG simultaneously. In this type of experiment, subjects can choose between two or more common pots to allocate their endowments, and the choices made by them are used to assess the effects of different framings [231–234]. In our case, we have compared two distinct methods for raising funds: direct versus indirect donations. To this end, we have devised a special-purpose PGG with two different treatments: a first setup involving an explicit social fee, or tax (*Direct-Donation*, henceforth DD), and another one involving an implicit social fee (*Indirect-Donation*, henceforth ID). As we will see below, our setup allowed us to simultaneously measure two variables: the contributions to public goods and the amounts donated to charity. Regarding those donations, the very existence of a direct self-benefit precludes measuring altruism, and, therefore, we have given the players the chance to contribute to several PGGs, which differed in the fraction of the benefit that goes to charity. Furthermore, the existence of funds with different social taxes enables us to study the pattern of contributions and their corresponding framing effect.

Our experiment provides several relevant conclusions concerning how people respond to framings intending to increase contributions with a social impact channeled through charities. We have observed that framing affects the choice of contributions depending on the donation structure: Indirect donations led to greater total contributions than social taxes. Conversely, there was no influence of framing on donations to charity: the fraction of the contributions devoted to charity is not affected by how those donations

2 See [194] for a recent review on framing in PGG.

are presented, i.e., as indirect or as direct donations. Regarding gender influence, we have found that women contribute to public goods and donate to charity more than men. All these findings may have implications of interest for the design of socially responsible investing strategies.

5.1.1 *Experimental design*

Figure 5.1 shows a schematic representation of the experimental setup. Experiments were conducted on groups of 10 participants. Each group played an iterated PGG with 5 funds, which differed in the fraction of profit donated to charity (0%, 5%, 10%, 15%, 20%, respectively). In a standard PGG, participants contribute to a common pot, and the total of the pot is multiplied by the so-called multiplication factor, being subsequently equally distributed among all participants irrespective of their contribution. In every round, subjects were given 100 experimental currency units (hereafter, ECU) which they could distribute among the five funds at will.

Arguably, there are two natural approaches to implement donations in a PGG scenario: donations coming from taxes on the contributions or coming from decreases in the profitability. To study the effects among these two framings, we split participants into two treatments. In the *direct-donation* (DD) treatment, once the contribution of the round was made, a fraction to be donated to charity was removed from each fund and the remaining amount was multiplied by 1.5 and equally distributed among all participants. The fraction destined to the charity was 0% (no donation whatsoever), 5%, 10%, 15%, or 20%, according to the chosen fund, whereas the multiplication factor was the same for all the funds. Conversely, in the *indirect-donation* (ID) treatment subjects were informed that the experimenters (the 'bank') would make the donation. In order to ensure that each fund would yield exactly the same payoff in any of the two setups, different multiplying factors were used, specifically 1.5, 1.425, 1.35, 1.275, 1.2, respectively.

Moreover, to capture the effects on donations and contributions, every group played two phases in a row: one in which subjects had to contribute all 100 ECUs (*Forced-Contribution*, henceforth FC),

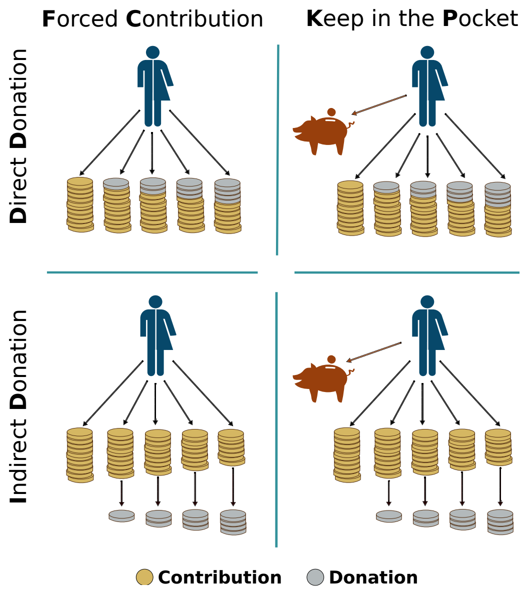


Figure 5.1: **Experimental setup.** Participants played a PGG adaptation wherein they could contribute to 5 different funds. In the *Direct Donation* (DD) treatment, each fund involved a different charitable donation rate that was deducted from the contributions, while in the *Indirect Donation* (ID) setup, each fund had different profitability according to the donation rate. Funds are designed such that associated benefits and donations are the same in both treatments, and participants were randomly assigned to one of them. In each treatment, participants played two consecutive phases: in *Forced Contribution* (FC), participants were required to contribute all their endowment to the available funds; in *Keep in the Pocket* (KP), participants chose how much to contribute to the funds, keeping the remaining for them. Accordingly, there were two cohorts: half of the participants played first the FC (FFC order), while the other half played first the KP (FKP).

and a second one in which they were allowed to keep as much of those 100 ECUs as they wished and contribute the rest (*Keep-in-the-Pocket*, henceforth KP). Each one of these two phases consisted of 20 rounds, and subjects played these two phases consecutively. Although all the players played both FC and KP, the order of both phases was not the same for all the groups: half of the groups played first the FC phase (First Forced Contribution order, henceforth FFC) and the rest of the groups played first the KP (First Keep in the Pocket order, henceforth FKP). Note that the order may play a role in framing: FFC participants are only concerned initially with their fund options, whereas FKP ones first have to decide between saving or contributing since the very beginning, ending up later with the distribution decision only.

Accumulated payoffs could not be reinvested: the maximum amount that subjects could contribute every round was the 100 ECUs that they received afresh at the start of the round. Players could see, at each round, the total amount contributed to each fund among all the players of their group. In the DD treatment, players could also see the fraction destined to charity by each fund (respectively, the corresponding multiplying factors in the ID treatment.) Before the experiment, the researchers informed the players about the destination of the charity donations: Médecins Sans Frontières (Doctors Without Borders). At the end of the experiment, each player received the payoffs accumulated along all the rounds played converted to euros, plus a fixed show-up fee.

5.1.2 *Results*

We measure the effectiveness of the two different fundraising methods, DD and ID, through the differences in contributions and donations. Contributions to public goods are measured in the KP treatment as the fraction of the 100 ECUs that a player contributes in all five funds, while donations can be measured in FC as the fraction of the 100 ECUs that goes to charity. Just for clarity, we remind the reader that in FC the entire endowment must go to funds and the only choice subjects can make is how it is distributed, i.e, which fraction to donate.

	Indirect Donation	Direct Donation
First Forced Contribution	30	30
First Keep in the Pocket	30	30

Table 5.1: **Number of participants in each cohort**. Each participant was designated to one of two treatments: *Indirect Donation (ID)* or *Direct Donation (DD)*. Furthermore, all participants played two phases, namely, *Forced Contribution (FC)* and *Keep in the Pocket (KP)*.

5.1.2.1 *Contributions*

The results of the contribution to public goods are shown in Figure 5.2. Boxplots of panel A show the average total contribution by subject averaged over the 20 rounds of the KP treatment, while panel B displays the evolution over time of the averaged group contribution. As shown in both panels, the FKP order exhibits an influence of framing on contributions that is not present in the FFC. To evaluate the significance of this dependence, we have performed a random-effects model [185, 186]. Equation (M1) describes the model for subjects' contribution (C_{it}) at time t, given that the participant i was playing the ID treatment (ID_i), being contributions in DD the reference.

$$C_{it} = \beta_0 + \beta_1 ID_i + u_{it} \quad , \tag{M1}$$

where the error term u_{it} is composed by an unobserved individual effect (α_i), a time effect (λ_t), and an idiosyncratic error (ϵ_{it}) which naturally is not correlated with the regressor:

$$u_{it} = \alpha_i + \lambda_t + \epsilon_{it} \tag{M1.1}$$

The results of this model are shown in column (1) of Table 5.2. Moreover, given that participants played in two different orders, we should add an order term to the model. Equation (M2) adds the FKP_i term to indicate if participant i started playing KP, as

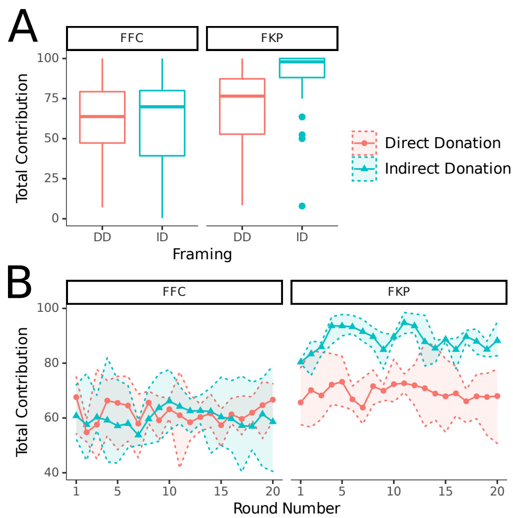


Figure 5.2: **Contributions to public goods are sensitive to framing only in the FKP order. A)** Boxplot of the average total contribution to PGGs by subject during the course of the experiment. Each colour corresponds to a treatment: one involving an explicit social fee (*Direct Donation*, DD), and the other one involving an implicit social fee (*Indirect Donation*, ID). In the left panel (FFC), individuals played first the FC treatment in which they had to allocate all their endowments into the different public goods and, subsequently, the KP treatment. In the right panel, subjects (FKP) played the KP treatment first, in which they chose how much to contribute to the funds, saving the remaining. In both panels contributions were measured in the KP treatment. The lower and upper hinges correspond to the first and third quartiles. The upper (resp. lower) whisker extends from the hinge to the largest (resp. smallest) value no further than 1.5 * IQR from the hinge. **B)** Groups' average contribution to PGGs at each round for each order (FFC and FKP). The shaded area corresponds to 0.95 bootstrapped confidence interval.

well as an interaction term between the order and treatment effect ($ID_i * FKP_i$):

$$C_{it} = \beta_0 + \beta_1 ID_i + \beta_2 FKP_i + \beta_3 ID_i * FKP_i + u_{it} \quad \text{(M2)}$$

The results are shown in column (2) of Table 5.2. Furthermore, as participants indicated their gender, we could use this information to analyze how it affects their contribution. The resulting model is given by equation (M3), wherein W_i indicates if the participant was a woman, being men subjects the reference group.

$$C_{it} = \beta_0 + \beta_1 ID_i + \beta_2 FKP_i + \beta_3 ID_i * FKP_i + \beta_4 W_i + u_{it} \quad \text{(M3)}$$

The results of this last model are shown in column (3) of Table 5.2. Accordingly, given that this data come from a controlled randomized experiment, the independent variables of these models do not correlate with the error term.

The analyses show that there is a framing effect on contributions, as participants from different treatments do not contribute the same amount. Participants from ID contribute significantly more, although this effect is only observed when they begin the experiment playing KP. Furthermore, it is shown that women contribute significantly more than men. To conclude, participants tend to contribute more when the donation is done by an external agent instead of directly extracting the amount out of their earnings. Nevertheless, this effect is only observed when they have not previously participated in a phase forcing them to contribute.

5.1.2.2 *Donations*

Regarding donations to charity, although present in both FC and KP treatments, measuring them in the FC treatment allows removing the effect of the contributions to the public goods. The results of the contributions to charity in FC are shown in Figure 5.3. Boxplots of panel A display the average total donation by subject averaged over the 20 rounds of each phase and panel B shows the averaged group contribution as a function of the round number. Here, the donation is measured as the fraction of the 100 ECUs that goes to charity. Both panels suggest that there is no difference in donations

	Dependent variable:		
	Contribution		
	(1)	(2)	(3)
Indirect Donation	8.886*	−1.655	−1.657
	(5.051)	(7.369)	(7.133)
FKP		7.636	8.451
		(6.632)	(6.512)
Woman			12.186**
			(5.091)
Indirect Donation x FKP		21.089**	21.088**
		(9.332)	(9.078)
Constant	65.635***	61.817***	54.100***
	(3.357)	(5.051)	(6.510)
Observations	2,363	2,363	2,363
R^2	0.017	0.115	0.146
Adjusted R^2	0.017	0.113	0.145
F Statistic	41.235***	305.333***	403.005***
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5.2: **Regression results for the contributions to public goods.** Random-effects (Wallace and Hussain estimator) with cluster robust standard errors at the individual level . Column (1) refers to the model for subjects' contributions being DD the reference (Equation M1). Column (2) refers to the model (1) after adding the FKP term to take into account the order plus an additional term for the interaction between the order and the treatment, being DD×FFC the reference (Equation M2). Column (3) refers to the model (2) after adding a *W* term for the gender, being the reference a male subject playing DD×FFC (Equation M3).

between FC and KP treatments regardless of the order they were played (either FFC or FKP).

To confirm the lack of framing effect in donations to charity, we have performed a random-effects model as in the case of contributions to public goods. Similarly as in the analysis of contributions, we performed a regression using a random effects model for the total donation (D_{it}) in the FC phase, i.e., D_{it} corresponds to the donation by subject i accumulated in all funds at time t. Accordingly, equations (M4), (M5), and (M6) are analogous versions of (M1), (M2), and (M3), respectively. The results are shown in Table 5.3, where columns (4), (5), and (6) correspond respectively to the models described by equations (M4), (M5), and (M6).

$$D_{it} = \beta_0 + \beta_1 ID_i + u_{it} \tag{M4}$$

$$D_{it} = \beta_0 + \beta_1 ID_i + \beta_2 FKP_i + \beta_3 ID_i * FKP_i + u_{it} \tag{M5}$$

$$D_{it} = \beta_0 + \beta_1 ID_i + \beta_2 FKP_i + \beta_3 ID_i * FKP_i + \beta_4 W_i + u_{it} \tag{M6}$$

The analysis confirms that, regarding donations to charity, there is neither difference between treatments nor order effects. Nonetheless, women donate significantly more (*i.e.*, contribute to funds with higher donation rate) than men.

5.1.2.3 *Distribution of contributions*

The experiment was designed with five different funds with different social taxes to allow us to study the pattern of contributions and the effects of framing on it. Besides, in the case of a framing effect, being able to extract a pattern in the contributions can help us to investigate the possible drivers behind the differences between the two framings.

To study the distribution of contributions in the five different funds, we have carried out regression analyses of the contributions in both FC and KP phases. We performed individual regressions to check whether the variables' effects on contributions varied across funds in the two treatments. Equation M7 describes the regression for fund f, where the dependent variable C_{itf} corresponds to the

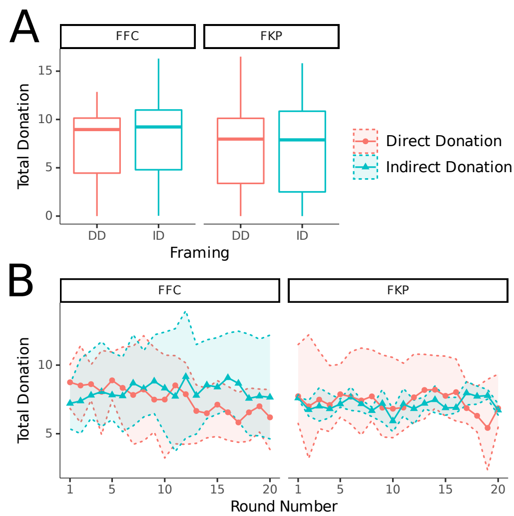


Figure 5.3: **Total donations to charity in the FC phase. A)** Boxplot of the average total donation by subject during the course of the experiment. The lower and upper hinges correspond to the first and third quartiles. The upper (resp. lower) whisker extends from the hinge to the largest (resp. smallest) value no further than 1.5 * IQR from the hinge. **B)** Group averages at each round. The shaded area corresponds to 0.95 bootstrapped confidence interval.

	Dependent variable:		
	Total Donation		
	(4)	(5)	(6)
Indirect Donation	0.198	0.548	0.549
	(0.803)	(1.002)	(0.963)
FKP		−0.273	−0.010
		(1.129)	(1.013)
Women			3.943***
			(0.772)
Indirect Donation x FKP		−0.700	−0.701
		(1.601)	(1.438)
Constant	7.446***	7.583***	5.086***
	(0.565)	(0.660)	(0.778)
Observations	2,347	2,347	2,347
R^2	0.0004	0.005	0.139
Adjusted R^2	−0.0001	0.003	0.138
F Statistic	0.546	10.542**	379.479***
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 5.3: **Regression results for the donations to charity.** Random-effects (Wallace and Hussain estimator) with cluster robust standard errors at the individual level. Column (4) refers to the model for subjects' donations with DD as the reference (Equation M4). In (5), two terms have been added to (4): the FKP term taking into account the order, plus an additional term for the interaction between the order and the treatment, being DD×FFC the reference (Equation M5). Column (6) refers to the model (5) plus a *W* term for the gender, being a male subject playing DD×FFC the reference (Equation M6).

amount contributed by subject i, at time t to the fund f. Note that a different regression has been performed for each fund in a given phase. Results of the regressions for the FC (*resp.*, KP) phases are shown in table 5.4 (*resp.*, 5.5).

$$C_{itf} = \beta_0 + \beta_1 ID_i + \beta_2 FKP_i + \beta_3 ID_i * FKP_i + \beta_4 W_i + u_{it} \quad \text{(M7)}$$

Table 5.4 shows the results for the FC phase, wherein participants could only decide how to distribute their contributions. As shown, participants from different treatments seem to contribute in a similar fashion when funds are considered in isolation. The only significant effect is observed for gender, women being more likely to contribute a larger share to funds with a positive social tax while contributing less to the fund with no social tax. Thus, they end up donating to charity more than men.

Table 5.5 displays the result of the same regressions for the KP phase, wherein participants can decide the amount they contribute to public goods. Clearly, women also donate to charity more than men in this phase, as they are more likely to choose funds with a higher social tax. There is a higher contribution associated with participants playing the ID treatment in the FFC, nonetheless, this was expected as participants contributed more in total as shown in previous sections.

Summarizing, differences in individual funds contributions are a consequence of our previous findings. Specifically, being a woman or playing the ID treatment in order FFC is associated with higher contributions.

5.1.3 *Discussion*

In order to explain the observed differences in contributions between the two frames, as well as the observed gender differences, we have studied their possible causes. On the one hand, we have performed regression analyses to evaluate possible differences between treatments with respect to the responses of subjects to the behaviour of the rest of the players in their group, as well as differences in the conditional contribution between genders. The details of these analyses can be found in Section c.1 of the Appendix.

	Dependent variable: C_{itf}				
	0	5%	10%	15%	20%
ID	-3.774	3.521	-3.229	0.047	3.439
	(7.048)	(3.343)	(2.418)	(2.339)	(3.052)
FKP	2.813	-1.925	-3.435	1.632	0.910
	(7.122)	(1.928)	(2.438)	(3.641)	(3.585)
Women	-25.784^{***}	2.738	5.441^{***}	5.199^{**}	12.411^{***}
	(5.744)	(2.545)	(1.774)	(2.428)	(2.358)
ID x FKP	1.132	1.236	4.611	-3.437	-3.529
	(10.292)	(4.389)	(3.175)	(4.400)	(4.677)
Constant	58.036^{***}	11.673^{***}	10.405^{***}	10.228^{***}	9.642^{***}
	(6.121)	(2.154)	(2.584)	(2.020)	(2.541)
Observations	2,347	2,347	2,347	2,347	2,347
R^2	0.133	0.021	0.047	0.029	0.095
Adjusted R^2	0.131	0.020	0.045	0.027	0.093
F Statistic	358.453^{***}	51.261^{***}	115.645^{***}	70.157^{***}	245.595^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01		

Table 5.4: **Random Effects regression with cluster robust standard errors at the individual level for the Forced Contribution phase.** Each column corresponds to a fund of a determined social tax, namely: 0, 5%, 10%, 15%, 20%, from left to right. The reference is a male subject playing DD×FFC.

	Dependent variable: C_{itf}				
	0	5%	10%	15%	20%
ID	−3.948	0.812	−2.066	−2.292	5.839*
	(5.711)	(1.646)	(1.776)	(1.882)	(3.079)
FKP	0.017	2.295	0.857	0.846	4.434
	(5.480)	(1.571)	(1.716)	(1.764)	(3.196)
Women	−11.060**	4.103**	4.643***	5.156***	9.336***
	(4.718)	(1.778)	(1.256)	(1.377)	(2.275)
ID x FKP	11.976	5.853*	4.935**	5.068*	−6.757
	(8.265)	(3.366)	(2.474)	(2.658)	(4.570)
Constant	31.870***	4.616***	5.227***	5.906***	6.479***
	(5.665)	(1.414)	(1.350)	(1.639)	(2.050)
Observations	2,363	2,363	2,363	2,363	2,363
R^2	0.057	0.077	0.065	0.063	0.070
Adjusted R^2	0.056	0.076	0.063	0.061	0.069
F Statistic	143.432***	197.424***	162.838***	157.617***	178.595***
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 5.5: **Random Effects regression with cluster robust standard errors at the individual level for the Keep in the Pocket phase.** Each column corresponds to a fund of a determined social tax, namely: 0, 5%, 10%, 15%, 20%, from left to right.

Regression results indicate that participants do not condition their contribution to other players' behaviour. There is neither evidence that men or women would react differently to this general trend, nor significant differences between different framings in this respect.

A plausible explanation of the observed influence of framing on contributions should lie in how information is presented to the players. In this regard, taxes are only shown to participants playing the DD treatment, being the presence of taxes the main difference between the two treatments. According to this explanation, ID players react negatively to the tax while contributing. Conversely, DD players are not be affected by the reduction in the profitability to the same degree. Furthermore, the fact that the framing effect is observed only in the FKP order suggests that subjects that play first the FC phase are conditioned by this learning effect, being their contributions in the subsequent KP phase independent of the framing.

5.1.3.1 *Conclusions*

Summarising our results, by using a setup based on a PGG modified to include a social responsibility factor, we have found that framing will affect fund contributions depending on how the donation procedure is implemented. On the one hand, contributions are higher when the associated social donations are presented as indirect donations than as social taxes. On the other hand, the fraction of the contributions devoted to charity is not affected by the framing effect. This result is not unrelated to the work of Krieg and Samek [234], where they observe that a return of a 20% of the contribution back to the donor increases significantly the contribution level, whereas recognition or sanctions have no effect. We have also found that, on average, women contribute to the public goods and donate to charity more than men, which is observed in some philanthropy contexts [235]. The implications of these findings are crucial for policy-makers in the design of socially responsible investing strategies and fair policies, *e.g.,* when the government or a charity intends to promote socially responsible conducts, or compete successfully for the limited amount of funds available to the different charities. People are not only self-interested, nonetheless, but their likelihood of acting prosocially can also be influenced by the type of incentive and economic context [12]. In this regard, the results of Corazzini *et al.* [233] point to the relevance of avoiding miscoordination among donors by making particular options salient. The mechanism we have identified here could then be one option to provide such saliency.

5.2 UNDERSTANDING DRIVERS WHEN INVESTING FOR IMPACT

Understanding drivers when investing for impact: an experimental study [236].

L. De Amicis, S. Binenti, **F. Maciel Cardoso**, C. Gracia-Lázaro, A. Sánchez, & Y. Moreno

In recent years, impact investing has risen to prominence in a global business environment that is increasingly concerned, and at times even pressured, to take into account social and environmental issues. Impact investing is thus marking a new trend among traditional practitioners, institutions and policymakers worldwide, and the range of impact investment options and opportunities at global level has naturally grown in parallel to the expanding interest in social investment. The Global Impact Investing Network [237] estimates that the sector has grown from \$4.3 billion in 2011 to \$502 billion in 2018[3] and, at the upper end of the market, impact investing is estimated to reach as much as \$1 trillion in value by 2020 [238].

In light of this new trend, a growing body of research emerged to define the theory and practice of social finance. The GIIN [239] defines impact investments as a form of investment that is "made into companies, organisations, and funds with the intention to generate social and environmental impact alongside a financial return". According to this interpretation of the term and phenomenon - arguably the most accredited and quoted one - the two defining elements of impact investments are the *expectation* of financial returns on capital, or at minimum a return of capital, and *intentionality*, namely the intention of having a positive impact as a direct consequence of a deliberate action. Despite the centrality of expectations and intentionality, current research has mainly focused on impact assessment and measurement frameworks aimed at capturing the environmental and social returns generated by investments ([240], [241]; [242], [243–249]). While such a focus is critically important to matters of effectiveness, accountability and transparency, it represents a debated and contested field that dominates and largely

3 The estimate is based on the responses provided by 266 leading impact investing organizations from around the world, managing collectively \$239 billion.

monopolise research on social investments. Compared to other instances of socially responsible business practices that have been widely investigated through the lens of reputation/brand building and consumption theories ([250], [251], [252]), little research has been conducted on the socio-demographic characteristics and the behavioural drivers pushing investors to choose impact funds over traditional investments.

Yet, if we are to make social finance a "standard practice", it is crucial to look at *what* might render impact-oriented funds a more appealing investment options and to *whom* – in this light, this study aims to contribute to this research gap through an experiment-based investigation. The value of exploring investors' behaviour and their decision-making process is two-fold. First, in the context of behavioural economics and game theory, such a focus can add significant value to existing research by shedding light on the nudging factors and determinants influencing the choices of economic actors (*i.e.*, intrinsic value of the research focus). Second, behavioural insights can have implications for normative initiatives or incentive actions aimed at pushing the impact investing trend into the mainstream, such as awareness-raising campaigns, marketing strategies and policy-making (*i.e.*, instrumental value of the research).

Within this wider scope of investigation and focus, the present experiment-based research aims to address the following questions:

RQ1 What is the effect of previous knowledge about impact investing? Do economic actors invest differently if they are already familiar with the concept of *impact investing* as opposed to those who have never heard of it?

RQ2 How do investors' preferences change depending on the way different investment instruments are proposed to them?

RQ3 How much financial return are investors willing to sacrifice for social impact, considering different risk factors?

RQ4 Do external factors affect the behaviour of economic actors (*i.e.*, could incentives from the government change investors' behaviour)?

The experiment consists of a multiple-choice game envisaging different investment scenarios. According to their performance in an effort task at the beginning of the game, participants are given a budget to simulate investment decisions under different incentive circumstances while controlling for different variables, such as prior knowledge about impact investing. The experiment is directed at two different sample groups: non-experts, who are likely to have no prior knowledge on the concept of impact investing, and "experts", namely professionals working in the impact investing sector. As an incentive to elicit truthful behaviour, and at variance with traditional surveys, participants are economically rewarded according to the earnings they make through their investment decisions. Following the experiment, the data is analysed through logistic regressions. This approach was preferred over percentages as the regression analysis allowed to isolate the effect of each variable.

The research design allows to draw a number of conclusions that will be of interest for stakeholders and policy-makers aiming to promote impact investing. The study concludes that people operating in the sector (experts) and female participants tend to favour the impact investing option. Furthermore, the older people are, the more attracted to impact investing they appear to be. External factors such as fiscal incentives influence positively, although only marginally, the respondents' behaviour in choosing Impact Investing Funds (IIF) over Traditional Investing Funds (TIF). No clear correlation has been found between the participants' educational level and their disposition to invest for impact. Providing additional details or, more effectively, images on the social purpose and impact of the IIF has proved to be critical in substantially increasing the probability of opting for an IIF over a TIF, both for male and female participants. Furthermore, participants were less likely to choose the IIF option when this was associated with higher risk (for both male and female participants). Finally, when considering participants' prior knowledge on the topic, the difference between control groups was relatively small - yet, it appears that providing participants with key information on social finance (by showing them a video) had a positive impact with normative

	Non-experts		Experts	
	Female	Male	Female	Male
3-year Bachelor	33	14	1	1
4-year Bachelor	79	45	2	4
5-year Bachelor	59	24	2	3
Lower secondary education	15	10	0	0
Master (1 year)	35	20	4	7
Master (2 years)	31	15	11	11
Other (non-listed)	20	7	2	0
PhD	7	4	3	8
Post-secondary, non-tertiary ed.	18	20	0	0
Short-cycle, tertiary education	21	18	0	1
Upper secondary education	23	23	1	0
Total	341	200	26	35

Table 5.6: **Participants' level of education**. Level of formal education of Non-Experts (two leftmost columns) and Experts (two rightmost columns).

implications for current incentive structures, awareness campaigns and educational programmes about impact investing.

5.2.1 *Research Design and Methodology*

5.2.1.1 *Sample description and experiment preliminaries*

We designed an experimental set-up wherein subjects could interact individually with the experiment through a web landing page, supported by an application based on the oTree platform [132]. Participants were given the opportunity to make individual decisions in their own time and environment, limiting the potential impact of endogenous biases. Two pools of participants were selected for this experiment. The first subject group consisted of a non-probability sampling of 541 individuals who are likely to have

	Non-experts		Experts	
	Female	Male	Female	Male
Argentina	7	1	0	0
Italy	0	0	7	11
Australia	1	0	1	0
Luxembourg	0	0	1	0
Austria	0	0	1	1
Mexico	8	4	0	0
Belgium	1	1	0	0
Netherlands	0	0	2	1
Bolivia	0	1	1	1
Paraguay	1	0	0	0
Brazil	1	0	0	0
Peru	1	1	0	0
Chile	10	1	0	0
Poland	0	0	1	0
Colombia	7	7	0	0
Portugal	2	2	0	0
C. Rica	0	1	0	0
Russia	0	1	0	0
Croatia	1	0	0	1
Serbia	0	0	0	1
Ecuador	0	2	0	0
Spain	282	166	0	2
El Salvador	1	0	0	0
Switzerland	0	0	1	3
France	0	2	2	1
Tunisia	0	0	1	0
France+	1	0	0	0
U.K.	1	2	7	11
Georgia	1	0	0	0
U.S.A.	2	0	0	0
Greece	0	1	0	1
Uruguay	3	1	0	0
Hungary	0	0	1	0
Venezuela	9	5	0	0
Ireland	1	0	0	1
Zambia	0	1	0	0

Table 5.7: **Country of Residence of participants.** France+ stands for Overseas France.

	Non-experts	
	Female	Male
Prior knowledge	66	45
No prior knowledge - Did see the video	129	74
No prior knowledge - Did not see the Video	146	81

Table 5.8: **Prior Knowledge**. Non-experts prior knowledge about impact investing.

no previous knowledge on the concept of impact investing. This non-discriminatory group presented great practical advantages without constituting an inferential risk on the research outcomes, as shown by and adopted in several other experiment-based researches [253]. The second subject group consisted of 61 experts and practitioners in the field of impact investing - they were recruited through the researchers' wide professional network in the world of social finance and thanks to a referral sampling system.

At the beginning of the experiment, participants were asked to fill in a standard demographic questionnaire. Out of a sample of 602 participants, 367 were female (61%, 341 non-experts and 26 experts) and 235 were male (39%, 200 non-experts and 35 experts). Tables 5.6 and 5.7 provide information on the participants' level of formal education and country of residence, respectively; Figure 5.4 shows the age distribution and gender across the two sample groups. Following the demographic questionnaire, subjects were directly asked whether they held any prior knowledge on "impact investment". In case of a positive answer, the participant could proceed to the game; in case of a negative answer, with a 50% chance, participants were shown a two-minute video tutorial briefly introducing them to the concept and practice of impact investment (for video see [254]) - if shown, participants could proceed to the next step of the experiment only after having watched the video until the end. In this way, we secured a diverse sample in which prior knowledge on impact investing could be factored in and controlled for. This is summarised in Table 5.8, showing the answers of non-experts.

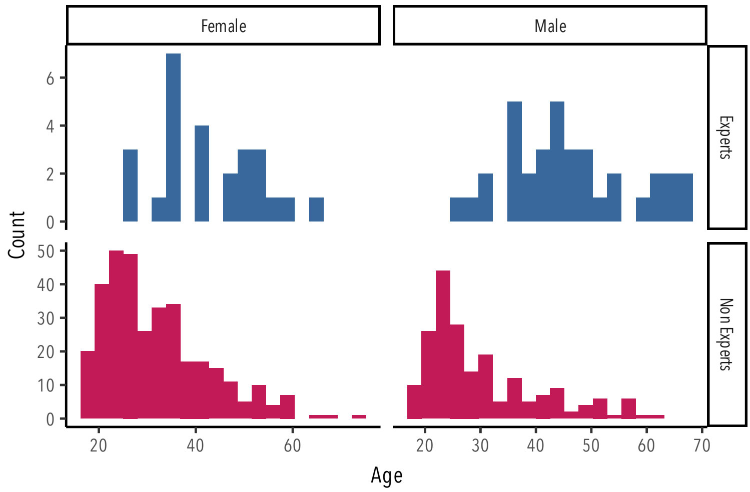


Figure 5.4: **Age distributions of the participants in the experiment.** Top left: female experts. Top right: male experts. Bottom left: female non-experts. Bottom right: male non-experts.

5.2.1.2 *Effort task*

The core part of the experiment consisted of an investment game in which individuals were asked to make scenario-sensitive investment decisions generating different financial returns and impact outcomes, including no impact at all. In order to simulate the way investors have a real stake in their investment decisions by relying on their own financial capital, the experiment began with a simple *effort task*, in which the subject was asked to count the number of "ones" displayed in a sequence of 1s and 0s (see Figure 5.5). Participants were then assigned a budget as a function of how well they performed in the effort task - this budget represents the "capital" they could invest in each of the scenarios proposed during the game. Most participants obtained the maximum score in the effort task, and hence the largest endowment/capital for the investment game. More specifically, 90.3% of experts and 94.4% of non-experts indicated the correct solution in the effort task.

As already mentioned, contrary to fixed or randomised initial endowments, the psychological implications of performing an effort task help strengthening the external validity of the game as participants are expected to feel more 'attached' to the money

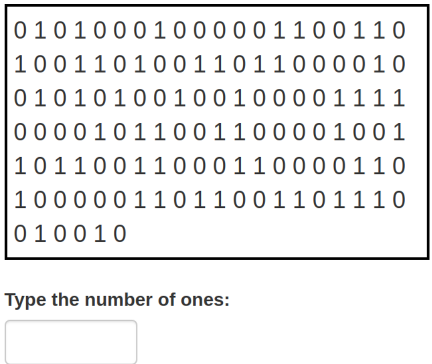


Figure 5.5: **Effort task.** Screenshot of the effort task proposed to participants.

they actually earn and invest - this more accurately resembles real-life conditions [255]. Moreover, by the rules of the game, subjects were made aware that, thanks to a lucky draw selecting about 10 % of participants, their earnings for each investment decision could be turned into real money. At the end of every question, participants were provided the following reminder to ensure they paid equal attention to all questions: “After you have answered all the questions, one question will be randomly selected and that will be the question used to calculate your earnings.” A similar setup was used by [256] to study the effect of overheads on the donations to charities. In this way, the lottery not only helped recruit more participants (this applies specifically to non-experts, who do not necessarily hold a particular interest in the field of social finance and hence in outcomes of the research), but also helped instil a realistic sense of ‘profitability’ that would normally characterise personal investments. As a result, in spite of the small ‘capital’ participants were given to invest in the experiment, the line between the game and real-life could be partially blurred given that participants’ decisions did actually have an impact on their pockets at the end of the game [257].

5.2.1.3 *Questions*

The core part of the game consisted of 8 different investment scenarios comprising simple multiple-choice questions. In each of these proposed scenarios, subjects were asked to choose between traditional investment options and impact fund options. Each of these binary choices probed into different aspects of our research questions.

- (Q1) Simple "traditional versus impact" investment scenario: Participants were asked to choose between a TIF yielding a 5% return, or an IIF yielding a 4% return and helping provide access to clean water in developing countries.
- (Q2) "Traditional versus impact" investment scenario with different social return options: Participants were proposed the same scenario as (Q1) with a fixed TIF yielding a 5% return or an IIF with a 4%, 3%, 2% or 1% return. This scenario was aimed at understanding how much financial return investors are willing to sacrifice for social impact as they are progressively proposed different return options.
- (Q3) & (Q4) "Traditional versus impact" investment scenarios with additional details on social impact and with/without different social return options: These questions were set up in the same way as (Q1) and (Q2) respectively; however, participants were also provided with more information on the actual impact achieved by the IIF, whereby drawing a more concrete picture in the mind of the investor on the impact he/she could make.
- (Q5) & (Q6) "Traditional versus impact" investment scenario with a risk factor and with/without different social return options: Following the logic of the previous questions, (Q5) and (Q6) investigated how decisions between TIF and IIF are affected by the risk factor of having no returns at all, with the traditional investment option being more probable to generate economic returns (and no social impact) than the IIF (90% chance of yielding a return versus 80%).
- (Q7) "Traditional versus impact" investment scenario with a fiscal benefit: In this scenario, participants were asked to choose between a TIF yielding a 5% return or an IIF yielding a 4% with a tax deduction of 20% of the invested amount.

- (Q8) "Traditional versus impact" investment scenario with additional details on social impact and visual aid: Finally, in this scenario, participants were presented with a more detailed description of the social impact generated by the IIF supported by an illustration as well as additional geographical coordinates. The scenario was presented as follows:

  > *"According to WaterAid, Papua New Guinea has the world's worst access to clean water, with 60% of the population living without a safe water supply. For the poorest population section, getting ill or even dying from drinking dirty water is normal."*

  This question, along the lines of (Q3) and (Q4), was aimed at exploring the relationship between a potential empathy factor given by an additional visual incentive. The photo was intentionally chosen as it does not depict the beneficiary of the social investment as a victim (the community in the photo actively reacts to a problem – i.e. access to clean water - rather than passively bearing the consequences), while picturing human figures that can more easily trigger empathy or a fellow-feeling of solidarity [258].

In summary, each of the questions presented above has been designed within a given research framing as summarised in Table 5.9.

Framing	Question	Research Question
TIF vs IIF	Q1 & Q2	**RQ1**
Impact description	Q3 & Q4	**RQ2**
Risk factor	Q5 & Q6	**RQ3**
Tax deduction	Q7	**RQ4**
Visual aid	Q8	**RQ2**

Table 5.9: **Questions and Framings.** Summary of framings proposed to participants and its corresponding research question.

5.2.2 *Results*

To analyse the impact of each control factor on IIF investments, we performed a series of regressions (tables 5.10 - 5.12). Table 5.10 shows the effect of each question framing. Here, the intercept (*i.e.*, the reference value) represents the willingness to invest in IIFs over TIFs with no additional framing. The difference between IIF's and TIF's decisions is explicitly addressed by Question 1 - 85.2% of the participants chose the IIF over TIF. The rest of the coefficients, namely impact description, risk factor, tax deduction, and visual aid, correspond to the respective effects of each framing, more explicitly their difference in effect with respect to the intercept term. Let us begin by the results summarized on Table 5.10. As it can be observed, there is a positive effect of the impact description on the IIF ($p < 0.05$), and this effect is higher and more significant ($p < 0.001$) when a visual aid is added to the information provided to participants - that is adding a picture of the people helped by the impact investment (Question 8). This visual aid makes people more likely to choose the IIF option. On the other hand, people are less likely to choose IIF option when this is associated with a higher risk ($p < 0.001$). Regarding tax incentives, deductions were not found to have a significant effect on the preferability of impact investing options.

Regarding the outcome of proposing different return options on impact investing, Table 5.11 shows the results of three different logistic regressions, each column corresponding to a different regression analysis. The first column (TIF vs IIF) shows the results for the simplest case without additional information (questions Q1 and Q2). The second column corresponds to the scenario in which participants where provided with additional details on their social impact (Q3, Q4). The third column refers to scenarios where a risk factor was introduced, with TIF options being more probable to generate returns than IIF options (Q5, Q6). significant negative Delta coefficients indicate that, if IIF provides lower returns than TIF, the higher the difference in returns, the lower the investments in IIFs ($p < 0.001$). This effect is robust against information and risk framing (both $p < 0.001$).

	Model 1
Intercept	**1.75** (0.11)***
Impact description	**0.43** (0.18)*
Risk factor	−**0.67** (0.15)***
Tax deduction	−0.29 (0.16)
Visual aid	**0.63** (0.19)***
AIC	2520.36
BIC	2550.41
Log likelihood	-1255.18
Deviance	2510.36
Num. obs.	3010

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5.10: **Framing of questions.** Log-odds ratio of coefficients obtained by logistic regression. Questions coefficients correspond to their difference in effect with respect to the intercept (*i.e.*, the reference value). The intercept corresponds to the log-odds ratio which chose the IIF over the TIF (Question 1), computed over all the participants. Observations correspond to all participants' responses to the two-choice questions (Q1, Q3, Q5, Q7, Q8).

The results for the impact of the demographics variables on IIF investments are shown in Table 5.12. Here, each column corresponds to a different regression analysis, namely simpler scenarios without additional information (column TIF vs IIF - questions Q1, Q2); scenarios with additional details on social impact (Impact Description - Q3, Q4); scenarions with higher risk associated with IIF (Risk Factor - Q5, Q6); scenarios with fiscal benefit associated to the IIF (Tax Deduction - Q7); and scenarios with visual aid together with additional information on social impact (Visual Aid - Q8). The variables considered in this study are the following: gender; expertise (*i.e.*, belonging or not to the group of experts); age, educational level, prior knowledge on impact investing (based on the way participants answered to the question on whether they held any knowledge on impact investment); display of video on impact

	TIF vs IIF	Impact Description	Risk Factor
Intercept	**1.05** (0.02)***	**1.08** (0.02)***	**0.91** (0.02)***
Multiple Question	−0.03 (0.02)	−0.01 (0.02)	0.04 (0.02)
Delta	−**0.20** (0.01)***	−**0.19** (0.01)***	−**0.16** (0.01)***
AIC	3481.80	3270.87	3843.07
BIC	3505.83	3294.91	3867.11
Log likelihood	-1736.90	-1631.44	-1917.53
Deviance	558.86	521.04	630.13
Num. obs.	3010	3010	3010

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5.11: **Framing of multiple questions and fund profitability.** Log-odds ratio of coefficients obtained by logistic regression. Each column corresponds to a different regression analysis: TIF vs IIF (Q1 and Q2), impact description (Q3 and Q4), and risk factor (Q5 and Q6). The intercept corresponds to the log-odds ratio which chose the IIF over the TIF (Q1, Q3, Q5). Multiple Question coefficients correspond to the difference in effect when proposing different options (Q2, Q4, Q6) with respect to the intercept (Q1, Q3, Q5, respectively). Delta coefficients account for effect IIFs return (Q2, Q4, Q6), *i.e.*, they refer to the return differences between TIF and IIF. Observations correspond to all participants' responses to the two-choice questions and their corresponding multiple question (Q1 and Q2, Q3 and Q4, Q5 and Q6).

investment (50% among those without previous knowledge); and Delta (return differences between TIF and IIF).

Intercept coefficients show that participants are more likely to choose the IIF option, except, surprisingly, when this is associated with a fiscal benefit. In this scenario, we do not observe a significant gender difference, while being an expert appears to make a larger difference. Conversely, experts do not behave differently from the rest of participants when either risk or visual aid is taken into account. Age also affects responses distinctively in each question, as we discuss further in the next paragraphs.

The first column of Table 5.12 shows that: i) Women are more likely to invest in an IIF than men ($p < 0.001$), ii) experts are more likely to invest in an IIF than non-experts ($p < 0.001$), and iii) the willingness to invest in IIF increases with age ($p < 0.001$). On the other hand, the education level does not make a substantial difference in explaining the behaviour of investors. Furthermore, although previous knowledge on impact investing (according to the self-assessment of participants) does not influence the investment decision, the informative video played a positive role: participants who were shown a tutorial video on impact investing displayed a higher tendency to invest in IIF than those who did not watch it ($p < 0.001$).

When additional information on the actual impact achieved by the IIF was given to participants (second column of Table 5.12) i) gender differences persist, with women being more likely to invest in IIFs than men ($p < 0.001$), ii) experts are also more likely to invest in IIFs than non-experts ($p < 0.001$). Conversely, neither age nor education has a significant influence on IIF investment choices. Note that, although without additional information on social impact IIF investments increase with age (first column regression), this determinant disappears when additional information is provided (second column).

The effects of associating a higher risk with the IIF option (20% chance of not yielding a return with IIF versus a 10% chance with TIF) are shown on the third column of Table 5.12. It is shown that: i) Women are more likely to invest in higher risk IIF options than men ($p < 0.001$), ii) opting for higher risk IIF increases with age ($p < 0.001$). On the other hand, when a higher risk is associated with the IIF, the higher tendency of experts to invest in IIFs vanishes.

Regarding tax deductions (fourth column of Table 5.12), surprisingly, a significant effect of tax deductions on the impact investing option was not found, except for experts and older subjects, who display a positive response to tax benefits. It is observed that, when a tax incentive is included in the scenario, experts ($p < 0.05$) and older subjects ($p < 0.01$) show a higher tendency to invest in IIFs. As in previous cases, although prior knowledge on impact investing does not show a significant influence on impact investing choices, participants who watched the informative video showed a higher tendency to invest in IIFs than those who did not see it ($p < 0.001$). This tendency is stronger when a tax deduction is included.

Finally, regarding the effect of an additional visual incentive, the fourth column of Table 5.12 shows the logistic regression for the scenario in which additional details on social impact supported by an image were showed to participants. As explained before, the visual aid has a significant positive influence opting for IIFs. In this scenario, gender is the only demographic variable that plays a significant role in the willingness to invest in IIFs - women showed were more likely to opt for IIFs over TIFs ($p < 0.01$). Neither expertise, age, education level, prior knowledge showed a significant influence on investment choices. Although women display a higher probability to opt for IIFs than men in the presence of a visual incentive, it cannot be stated that visual aids affect more women than men, since the difference in its influence is not significant according to logistic regression.

5.2.3 *Discussion*

Our results indicate that in most scenarios experts are more likely than non-experts to choose the impact investment option. This does not really come as a surprise: it is likely that experts entered the impact investing field driven by personal principles and moral considerations [259, 260], as working in the world of social finance may already reflect personal compromises between a less lucrative career and an ethical professional path[261, 262].

Our findings show that older people have a higher tendency to choose impact investment options than younger people. This is somewhat surprising given the current momentum of narratives such as "Millennials Will Bring Impact Investing Mainstream" [263], whereby young generations are expected to shift large capitals towards social causes, as well as prioritising socially meaningful careers and thus focus on social entrepreneurship [264, 265]. Nevertheless, some studies have also shown that senior citizens are more prone to contribute to the common good [266], due to their willingness to leave a positive legacy behind. The explanation for such a result may be that the younger generations are interested in impact investment but do not have enough expertise or do not feel confident enough to take part in it. Indeed, the Financial Times [267] reports that "while 64% of the younger generation Credit

Suisse surveyed were interested in impact investing, only 24% had actually invested". Numbers even decrease when looking at high net worth families. A research from Morgan Stanley [268] shows that only 4% of Next Gen family members consider themselves fully-active participants spending "a great deal" of time engaged in impact investing, although the majority (60%) of Next Gens consider "important" to use their family's wealth to make a positive social or environmental impact.

For almost all the questions, we can observe that women are also more willing to choose an IIF than men, except for the tax reduction question. This is well in line with the abundant literature on philanthropy and charity-giving that shows that women are more likely to engage in altruistic behaviour [235, 269]. Even when a risk factor is introduced, more women prefer an IIF compared to male participants despite they are generally considered to exhibit a risk-averse behaviour. The tendency of women to prefer an IIF over a TIF is in line with existing research from the industry. Stephanie Luedke of Citi Investment Management, who works on the front lines of asset allocation, confirmed in a recent interview on Forbes [270] that "90% of women surveyed have indicated that they want to invest at least a portion of their wealth in a manner that aligns with their values". On the top of that, women are becoming wealthier, thanks to a more gender-equal intergenerational transfer of wealth [271], and are proving to have entered a traditionally "male" environment as capable investors, as showed in a research from Fidelity in which women tended to outperform men in generating a return on their investments [272].

When considering participants' prior knowledge on the topic, leaving expertise on the side, the difference between control groups is not significant. Yet, the experiment reveals that showing the video had a positive impact in prompting socially oriented decisions, whereby signalling a wider scope for promoting and raising awareness about impact investing. This is confirmed by our logistic regressions and represents one of the most important findings of our research in line with recent studies on the same topic [265]. Public administration bodies and civil society organisations have already started to put efforts in raising awareness about impact investment. Organisations such as Big Society Capital, the social investment "wholesaler" set up in 2012 by David Cameron together with his Big Society agenda, or the Social Impact Agenda promoted by the Portuguese Government are an example of this. International political bodies, such as the European Union, did not adopt a "wait and see" approach; on the contrary, they took significant, active steps forward, such as the creation of the Expert Group on Social Entrepreneurship (GECES) in 2011 and the consequent report in 2016 – "Social Enterprises and Social Economy going forward" [273] – advocating for a greater visibility and enhanced understanding of social enterprises and impact investments. This kind of initiatives, however, generated mixed results; more needs to be done not only by coordinating efforts between governments and international institutions but also by encouraging inter-sectorial collaborations between researchers, the private sector and practitioners from the social economy and the social enterprise world, who could work together to gather stronger evidence on the added value of impact investment and better communicate their main results through institutional channels. In this regard, media outlets are currently missing an opportunity, especially in light of the positive general attitude towards the topic in public narratives [265]. Furthermore, impact investing is not currently part of the curriculum of finance degrees and is not part of the formal training of a financier or corporate investor. Top universities are taking new steps in making innovative finance part of the mainstream and are increasingly engaged in the impact investing debate, knowledge-sharing and training. For instance, the Said Business School at the University of Oxford has recently launched a programme entitled "Oxford Impact Investing Programme: Build your investment skills to deliver maximum social return", directed at professionals and businesses that aim to enter the field - this integrates the work already undertook by the Skoll Centre for Social Entrepreneurship. In the same way, the Cambridge Institute for Sustainability Leadership (CISL) greatly focuses on sustainable business and leadership. Yet, these standarn university degrees hardly cover impact investing. As a result, whilst universities are increasingly treating topics related to management and innovation for social good, there is still a long way to go in shifting the way we approach mainstream financial training and education, which could be a great starting point to radically change mainstream finance.

Another reflection point is about tax incentives, usually seen as a strong market builder. In our experiment, the tax incentive is the only case in which gender does not play a significant role, and both men and women do not see it as an incentive. This was a somewhat surprising, key finding of our research. As a matter of fact, despite what academic evidence suggests and our experiment confirms, public bodies still put a great emphasis on the tax benefits of giving. The UK Government, for instance, has introduced the Social Investment Tax Relief (SITR) scheme in 2014 - yet, the results have not been as positive as expected. In 2016-17, 25 social enterprises received new investments through the SITR scheme and £1.8 million of funds were raised. Since SITR was launched in 2014-15, 50 social enterprises raised funds of £5.1 million through the scheme [274]. These figures are far from the 300,000 social enterprises and charities that could potentially benefit from SITR, according to Big Society Capital [275]. What is causing such a big difference? In a recent call for evidence launched by the British Government, organisations advocated for several changes suggesting that such incentives were not fit for purpose [276]. In this regard, our study confirms that tax incentives are not a game changer for people who are not experts in the field. One may wonder whether the problem lies in the design of incentive schemes or in the fact that tax incentives themselves are simply not a major determinant of investors' decisions. Other countries have launched similar tax incentive schemes in the past (i.e. France) and others (*i.e.*, Italy) have just followed. In a few years from now, it would be interesting to see the impact of these recently implemented incentives and run further research to understand whether fiscal incentives can still be considered as a main driver for investors' behaviour or are just a nice add-on impacting the decision of 'only' a few.

The experiment also brought about the lingering scepticism about impact investing. Indeed, impact investing is still perceived by some as a suspicious hybrid where money-driven actors, philanthropists and practitioners (*i.e.*, social entrepreneurs) are culturally polarised and still struggle to speak the same language [265]. By way of example, one of the participants - and more specifically a participant from the expert pool – reported feeling “almost angry” at the built-in reward mechanism of the game. He contested not feeling included in the scope of the experiment, which according to the participant implicitly assumed that people can only be incentivised by money; consequently, in this view, the experiment was meant for profit-oriented “venture capitalists” only. While the design of the game was merely aimed at resembling real-life circumstances, we did not predict that offering a reward could have triggered negative reactions. In the same way, another expert participant never claimed the prize, thus showing his ’pure’ willingness to engage in the debate and lack of responsiveness to monetary incentives.

5.2.3.1 *Conclusions*

Impact investing aims to generate social and environmental impact alongside a financial return. Here, we have run an experiment with 602 participants to understand what ’makes’ impact investors and what are the drivers for their decisions. We apply logistic regression analysis on the acquired data-set. One of the main weaknesses of the study is the sample limitation for experts. However, we must note that the process of finding experts and get them to run the experiment requires considerable resources. There is not such a thing like a pre-defined available data-set for this, and therefore having access to experts and ensuring their participation to the experiment is a challenge in itself, also due to their time limitation.

The main contribution of this work is the domain insight: our study shows that participants are generally favourable to invest in IIF, especially if they are women, older people, or individuals who were already familiar with the impact investing field (*i.e.*, “experts”). With reference to this last point, while prior experience in the field has an impact on choices (**RQ1**), there was no significant difference between non-experts who reported some or no previous knowledge on impact investing. This might lead to two competing explanations: i) non-experts who declared to have some knowledge on the field knew about it only vaguely; ii) simply knowing about impact investing is not enough, and prior experience rather than mere knowledge is a more significant determinant of choices.

Surprisingly, external incentives such as tax breaks do not appear to be a game-changer (**RQ4**), and future research might determine when and why they might affect investors' decisions. On the other hand, when participants are informed of the risks attached to their investment, the likelihood to invest in an IIF decreases (**RQ3**), but it increases when more information about the impact of their investment is made available (**RQ2**). Particularly, we have seen that visual aids further increase the investors' willingness to choose an IIF across all categories analysed in this work. We note that additional efforts should be made in raising awareness about impact investment, especially by policymakers and media outlets. Inter-sectorial collaboration between the public, private and third sector and academia (quadruple helix) should be encouraged, as well as the introduction of impact investing in the curriculum in financial training and education.

Future researches could benefit from a broader dataset. Tax incentives deserve special attention and researchers could focus on those countries that have already designed and implemented policies on this topic. An interesting twist to the research could be investigating how behaviour changes if the choice of the participants is made public, as an interest in reputation-building and positive self-branding may significantly drive people's choices.

	TIF vs IIF	Impact Desc.	Risk Factor	Tax Ded.	Visual Aid
Intercept	**2.09*****	**2.87*****	**1.49*****	-0.17	**2.38*****
	(0.20)	(0.22)	(0.19)	(0.48)	(0.64)
Male	$-$**0.47*****	$-$**0.50*****	$-$**0.65*****	0.11	$-$**0.78****
	(0.09)	(0.09)	(0.09)	(0.23)	(0.30)
Expert	**0.74*****	**0.65*****	0.34	**2.23***	1.76
	(0.19)	(0.20)	(0.18)	(1.05)	(1.09)
Age	**0.02*****	0.01	**0.02*****	**0.04****	0.01
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)
H.Ed.	-0.11	-0.16	$-$**0.25***	-0.15	-0.06
	(0.12)	(0.12)	(0.11)	(0.29)	(0.38)
P.Ed.	0.05	0.24	0.11	0.20	0.35
	(0.14)	(0.14)	(0.13)	(0.36)	(0.50)
Other	0.31	0.20	0.38	0.30	-0.29
	(0.23)	(0.24)	(0.22)	(0.61)	(0.70)
P.K.	0.15	0.06	0.06	0.22	-0.17
	(0.12)	(0.12)	(0.11)	(0.28)	(0.39)
V. Disp.	**0.35*****	**0.27***	**0.27****	**0.98*****	0.00
	(0.10)	(0.10)	(0.10)	(0.26)	(0.34)
Delta	$-$**0.97*****	$-$**0.96*****	$-$**0.71*****		
	(0.04)	(0.04)	(0.04)		
AIC	3262.16	3084.17	3545.96	551.29	352.44
BIC	3322.26	3144.27	3606.06	590.89	392.04
Log Likelihood	-1621.08	-1532.08	-1762.98	-266.64	-167.22
Deviance	3242.16	3064.17	3525.96	533.29	334.44
Num. obs.	3010	3010	3010	602	602

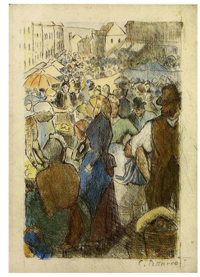
$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5.12: **Demographic variables impact.** Each model (column) corresponds to a regression for each framing type. *TIF vs IIF* considers data from Q1 and Q2; *Impact Desc., Impact Description* from Q3 and Q4; *Risk Factor* from Q5 and Q6; *Tax Ded., Tax deduction* from Q7; *Visual aid* from Q8. We consider three three different levels for education: *higher education (H. Ed.), postgraduate education (P.Ed.)*, and *other* (*i.e.*, non-curricular education besides basic education programs). *P.K.* is a dummy controlling for previous knowledge about impact investing and *V. Disp.* (Video displayed) is a dummy indicating if participants have watched the video. Delta coefficients correspond to the return differences between TIF and IIF. Observations correspond to all participants' responses to questions according to the framing type, columns from left to right: Q1 and Q2, Q3 and Q4, Q5 and Q6, Q7, and Q8.

6 TRADING IN COMPLEX NETWORKS

Whoever offers to another a bargain of any kind, proposes to do this. Give me that which I want, and you shall have this which you want, is the meaning of every such offer; and it is in this manner that we obtain from one another the far greater part of those good offices which we stand in need of.

Adam Smith, The Wealth of Nations



Market at Gisors, Rue Cappeville, Camille Pissarro

Up until now, we have dealt with scenarios wherein the manifestation of humans' pro-social tendencies provides the most efficient outcomes for the group. Nevertheless, in some situations, people or companies coalescing can have harmful effects to social welfare, such as in the case of oligopolies and cartels. They are detrimental by going against one important feature of markets: *free competition*. In an ideal market, the existence of many buyers and sellers would guarantee long-run efficiency in a barrier-free environment. As pointed by Adam Smith, competition would make market prices converge to their natural level and also push the economic system to higher levels of productivity and innovations [29]. Thus, without requiring a Leviathan [134], decentralized trade would work for the social welfare of the majority of people [277].

"buyers and sellers are in such free intercourse with each other that the prices of the same goods tend to equality easily and quickly." A. A. Cournot [278]

Consequently, concerning market interactions, competition can be seen as beneficial to the social good as cooperation is in social dilemmas. Nonetheless, in the real world, factors such as information asymmetries and cumulative advantage can lead to market inefficiency [279]. Specifically, structures centralizing market power would undermine competition, leading to monopolies and monopsonies determining prices [280]. Therefore, it is in the public interest to identify such structures in order to ensure free competition endures.

In this regard, most markets show an underlying network structure which has to be taken into account if we are to understand market outcomes [55, 281, 282]. Indeed, economic interactions are influenced by geographic proximity and individuals' relationships [283]. Moreover, global supply networks in agriculture, manufacturing, and services are a defining feature of the modern world, and trading outcomes are affected by all sorts of middlemen connecting producers to buyers [284]. Accordingly, the efficiency and the distribution of surpluses across different parts of these networks depend on the decisions of their intermediaries. In particular, their position can make them extract a large fraction of the trade surplus, and they can be positioned in such a way as to make trade inefficient [72].

Consequently, it is crucial to identify the principles governing intermediaries behaviour, especially if they decide simultaneously [279]. Furthermore, there is a non-trivial interplay between decisions and economic agents' links if the underlying network exhibits a complex topology [62], such is common in real systems [55, 285]. Thus, to improve our insights about these type of markets, in this Chapter, we present results of price formation experiments performed with human subjects located in large complex networks. Moreover, the observed behaviour leads us to create an agent-based model yielding macroscopic patterns consistent with the experimental findings. In sum, the results presented in this Chapter show that network topology is a chief determinant of pricing and efficiency.

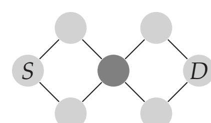
6.1 EFFECT OF NETWORK TOPOLOGY AND NODE CENTRALITY ON TRADING

Effect of network topology and node centrality on trading [286].

F. Maciel Cardoso, C. Gracia-Lázaro, F. Moisan, S. Goyal, A. Sánchez, & Y. Moreno

Globalization is a prominent feature of the modern economy [287]. Nowadays, supply, service and trading chains [288–292] play a central role in different contexts such as agriculture [293–296], transport and communication networks [297, 298], international trade [299] and finance [300, 301]. One key question on these systems is how pricing dynamics by intermediaries of the economy impacts both efficiency and surpluses. Our purpose here is to develop a better understanding of the forces that shape intermediary pricing behaviour in such complex networks.

Game theory constitutes a useful framework to study competition among trading agents [302]. In this context, the Nash Bargaining Game [303] studies how two agents share a surplus that they can jointly generate. In the Nash Bargaining Game, two players demand a portion of some good. If the total amount requested by both players is less than the total value of the good, both players get their request; otherwise, no player gets their request. There are many Nash equilibria in this game: any combination of demands whose sum is equal to the total value of the good constitutes a Nash equilibrium. There is also a Nash equilibrium where each player demands the entire value of the good [304].



*The dark grey vertex is a **critical node**.*

As a generalization of Nash demand game to n players, Choi et al. [72] proposed and tested in the laboratory a model of intermediation pricing. In this model, a good is supposed to go from a source S to a destination D. Intermediaries, which are located in the nodes of a network, may post a price for the passage of the good. Trading occurs if there exists a path between S and D on which the sum of prices is smaller than or equal to the value of the good. The key finding was that the pricing and the surpluses of the intermediaries depends on the presence of "critical" nodes: a node is said to be critical if it lies on all possible paths between S and D. Condorelli and Galeotti [284] provide an overview of the literature on intermediation and argue that the criticality of a node is an important determinant of pricing behaviour, intermediation rents, and the efficiency of trading, in a wide class of models of auctions and bargaining. The goal of the present work is to investigate this claim in large scale networks, so as to develop a better understanding of the role of network topology in commerce.

We conduct experiments with human subjects embedded in complex networks: specifically, we consider a random network and a small-world network each with 26 subjects (and the same level of average connectivity). In these networks there are *no* critical nodes: the results of Choi et al. would suggest that intermediary prices must be close to zero and that their surpluses must also be close to zero. As we will show below, our *first* finding is that, in all the networks studied, when there is not total but partial criticality, intermediaries set positive prices and they make large profits. Moreover, network topology has powerful effects: in particular, in the random network, intermediaries set lower prices as compared to a small-world network. As a consequence, there is full trading efficiency in the random network, but trade breaks down in almost one third of the cases in the small-world network.

This striking difference leads us to an examination of how location within a network affects pricing: our *second* finding is that within a given network, standard measures of network centrality appear to have no significant effect on pricing behaviour. As network location does not matter for prices, the presence on the cheapest and active path must be crucial for profits. And indeed, this is what we observe: intermediaries' earnings are positively related to their betweenness weighted by the path length.

Turning to the dynamics of price setting, we observe that traders raise prices if they lie on the successful trade path (*i.e.*, the least-cost path), and that they lower prices when they are off the least-cost path. Based on these observations, we build an agent based (ABM) model that reproduces qualitatively the experimental results. We then use simulations to extrapolate our findings to larger networks: our *third* finding is that network topology continues to matter and that random networks exhibit lower prices and higher level of efficiency even when there are 100 traders. Finally, our *forth* finding

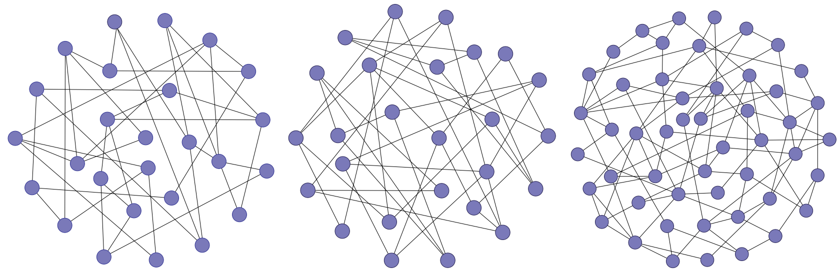


Figure 6.1: **Networks used in the experiments**: 26-nodes Random Network (left), 26-nodes Small-World-Like network (centre) and 50-nodes Random Network (right).

uncovers the role of node-disjoint paths − two paths are disjoint if and only if they do not share any node−and of the average path length in shaping level of pricing in the simulations: networks that have a larger number of node-disjoint paths exhibit lower prices and higher efficiency. Among networks with the same number of node-disjoint paths, average path length is an important driver of costs.

6.1.1 *Experimental setup*

We consider a simple game of price setting in networks to study supply, service and trading chains taken from Choi *et al* [72]. Let $\mathcal{N}$ be a set of nodes $\mathcal{N} = \{S, D, 1, 2, \ldots, n\}$, where S is a source and D a destination; and $\mathcal{L}$ a set of pairs of elements of $\mathcal{N}$. $\mathcal{N}$ and $\mathcal{L}$ define a trading network where the elements of $\mathcal{L}$ are the links. A path between S and D is a sequence of distinct nodes $\{i_1, i_2, ..., i_l\}$ such that $\{(S, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{l-1}, i_l), (i_l, D)\} \subset \mathcal{L}$.

Each experiment consists of 4 series of 15 rounds each, and it involves n human subjects that will play the role of intermediaries. Before starting the first round of a series, each subject is randomly assigned to a node in $\{1, 2, \ldots, n\}$. The positions of S and D are also assigned at random. These positions (players, S and D) remain constant over the 15 rounds. Subjects are always informed about the network and their position in it, that is, they can see the whole network including S and D. At each round, every subject has to make a decision; namely, she has to post a price from 0 to 100 tokens for the passage of a good by her node. The prices determine a total cost for every path between S and D. A path is feasible if its cost is not greater than a given threshold (100 tokens) that represents the value of the economic good generated by the path. After all players have made their choices, the cheapest path is selected: if it is feasible, each player located in this path receives her proposed price as a payoff. Otherwise, no trade takes place and payoffs are zero. Players who are out of the selected path do not get any payoff in that round. In the case of more than one cheapest path, the tie is resolved through a random choice among cheapest paths. From the second round onward, players are informed about the existence of a trade in the previous round, about the previously selected path, and about the prices and payoffs of all the players in the previous round together with their positions in the network.

We have conducted two experimental sessions in a random network of 26 nodes with $\langle k \rangle = 3$ and two more sessions in a small-world-like network of 26 nodes with $\langle k \rangle = 3$. All the networks used in the experiment are generated through the Watts-Strogatz [59] algorithm with different probabilities p of rewiring ($p = 0.1$ for the small-world-like network and $p = 1$ for the random networks). In any given treatment, the same network was used across all series and sessions. However, the selection of source-destination pairs is generated randomly at the beginning of each series such that the shortest path between the two nodes is of distance at least diameter - 2 (as a means to prevent uninteresting scenarios with very short distance between source and destination nodes).

Additionally, we have conducted another experimental session in a random network of 50 nodes with $\langle k \rangle = 4$ that will allow us to check the robustness of the results against the size and connectivity of the network [1]. Plots of the three networks are shown in Fig. 6.1. It depicts the structural representation of the network as viewed by the subjects in each corresponding treatment: Random Network of 26 nodes (left), Small-World-Like Network of 26 nodes (center) and Random Network of 50 nodes (right).

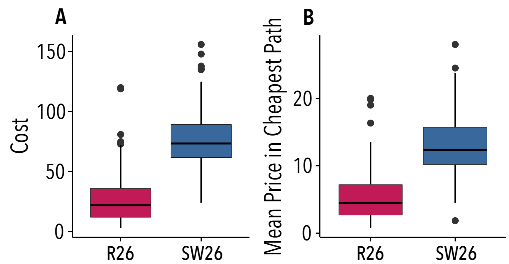


Figure 6.2: **Network topology affects trading costs and prices. A**: Cost of the cheapest path for the random network of 26 nodes (R26), and for the small-world network of 26 nodes (SW26). **B**: Mean price of participants in cheapest path for the same networks. Lines in the boxes denote the medians, whereas boxes extend to the lower- and upper-quartile values. Whiskers extend to the most extreme values within 1.5 interquartile range (IQR).

6.1.2 *Experimental Results*

The networks used in the experiment allow for coexisting paths with a different number of intermediaries, where theory predicts both efficient and inefficient (Nash) equilibria. Furthermore, these networks present different characteristics, such as degree and centrality distributions, that may affect the bargaining power of the intermediaries. These facts motivate our first question: How does the network topology affect costs and prices? Fig. 6.2A shows the cheapest path cost in each of the networks considered. As shown, the small-world networks exhibit higher costs than random networks (t(232.41)=15.5, $p < 0.001$). Fig. 6.2B instead displays the costs of the cheapest path normalized by the number of nodes on it, *i.e.*, the mean price of nodes along the cheapest path. The differences between networks persist, indicating that prices and costs strongly depend on the topology of the network. These results, separated by rounds, are shown in Fig. b.3 of the Appendix. Table 1 shows that there is a very large effect of topology on efficiency: in the random network trade is realized in practically all the cases, while in the small-world network trade breaks down in almost one third of the cases (binomial-test, 0.95 CI=(0.76, 0.90), $p < 0.001$). However, small-world networks involve higher costs and profits

1 These results can be found in the Appendix Section b.1 and Table b.1.

than random networks, since the higher posted prices compensate for the lower efficiency. Therefore, we conclude that the topology of the network matters for intermediation: the surpluses of the intermediaries vary significantly from one network class to the other in the experiments.

network	efficiency	price	price in CP	cost	profit	length
R 26	0.97	11.34	5.49	28.33	1.10	6.26
SW 26	0.68	18.10	13.16	76.52	2.38	7.00

Table 6.1: **Experimental results.** Efficiency (fraction of rounds in which the cheapest path cost was equal to or less than the threshold), and mean values of the price, price in the cheapest path, cost of the cheapest path, profit, and cheapest path length for the random network with 26 nodes (R 26) and for the small-world network with 26 nodes (SW 26).

Profit is only obtained when the subjects are on the cheapest path, i.e., when they are on the path through which the trading is realized. Thus, it is of interest to examine what is the role of the location of intermediaries in the network in shaping their behaviour, which we do next. First, we observe that the networks in our experiment do not contain any critical nodes and yet they generate large rents. So the results from the small scale experiments by Choi *et al* [72] do not apply to complex larger scale networks. It seems likely then that nodes that are present on more paths have greater market power. This motivates a generalization of the notion of criticality as follows:

$$sd(v) = \frac{|P_{SD}(v)|}{|P_{SD}|} \quad , \tag{6.1}$$

where $sd(v)$ is the *partial* criticality of node v, $|P_{SD}(v)|$ stands for the number of paths between the source and destination containing a given node v, and $|P_{SD}|$ for the total number of paths between the source and the destination. Following this line of thought, a higher partial criticality may indicate a potential for greater bargaining power and therefore nodes with a higher partial criticality should show higher prices and profits. Fig. 6.3A shows the accumulated prices of the intermediaries as a function of their partial criticality.

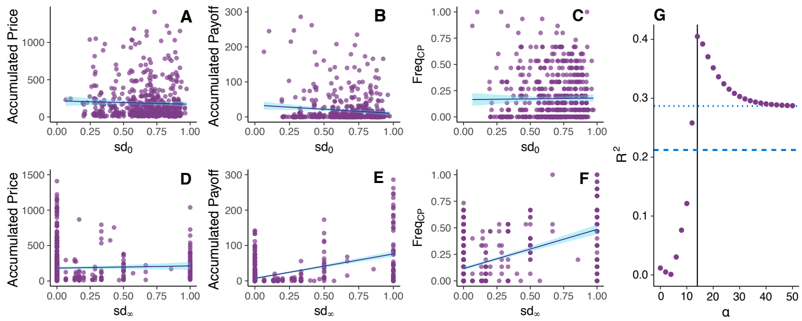


Figure 6.3: **SD-betweenness determines payoffs but not posted prices. A-F**: Accumulated price (**A,D**), accumulated payoff (**B,E**) and frequency in the cheapest path (**C,F**) of participants during a series of 15 rounds as a function of the node criticality sd_0 (**A,B,C**) and of the SD-betweenness sd_∞ (**D,E,F**). **G:** R^2 of the regression of participants accumulated payoff on sd_α versus α, where α modulates the weight of the length of the paths in the S-D centrality measure. Dashed (points) line show the value of R^2 for correlation of payoffs on the betweenness (SD-betweenness). Data is pooled across any series of 15 rounds in any experimental session. For similar analyses within each experimental network, see Appendix Fig. b.1.

There is no significant relation between partial criticality and the prices posted by participants. Even more strikingly, as illustrated in Fig. 6.3B, there is no relationship between the accumulated payoff obtained and a node's partial criticality. Fig. 6.3C shows the frequency that each player is on the cheapest path versus her partial criticality. Again, there is no relation between these variables.

This lack of correlation may be due to the equal weighting of paths with different length. In order to address this point, we refine our generalized notion of partial criticality to take path length into account:

$$sd_\alpha(v) = \frac{\sum_{[S,v,D]} l(p)^{-\alpha}}{\sum_{[S,D]} l(p)^{-\alpha}} \quad , \tag{6.2}$$

where the summations are over all the paths between S and D containing v (numerator) and over all the paths between S and D (denominator). $l(p)$ represents the length of path p and α stands for an arbitrary weight: as α increases, more importance is given to shorter paths. Specifically, when $\alpha \to \infty$ it will consider only the shortest paths, $sd_\infty(v)$ being a measure of the source-destination betweenness of node v (SD-betweenness(v)). On the opposite side, for $\alpha = 0$ the partial criticality of Equation 6.1 is recovered, that is, $sd_0(v) = sd(v)$.

Fig. 6.3D shows the accumulated prices of the intermediaries as a function of their SD-betweenness. Again, there is no relation observed between pricing behaviour and betweenness. However, as shown in Fig. 6.3E, there is a positive correlation between the accumulated payoff obtained by intermediaries v and their $sd_\infty(v)$. That is, although pricing is uncorrelated with SD-betweenness centrality, profits are positively correlated with it. The reason behind this difference must therefore lie in how the presence of v on the least-cost path is correlated with $sd_\infty(v)$. This is displayed in Fig. 6.3F, which represents the fraction of times that an intermediary is on the cheapest path versus her SD-betweenness. As shown, there is a positive correlation between these measures, which explains why – in a situation where prices are largely insensitive to network location – profits will be correlated with $sd_\infty(v)$. The robustness of these results against the size and connectivity of the network is discussed in the Appendix Section b.1.

So far, we have seen that node centrality does not influence earnings when we equally consider all the paths from S to D to compute it, but it does when we consider only the shortest paths. This fact indicates that the weight given to paths length is important to study the capacity of the nodes to extract surpluses. In order to verify this hypothesis, Fig. 6.3G shows the coefficient of determination R^2 of the regression of intermediaries payoffs on sd_α as a function of α. The best fit is obtained for $\alpha \sim 12$, which indicates that longer paths should have significantly smaller weight than shorter ones. As the number of paths grows exponentially with network size, SD-betweenness seems to be a feasible and good descriptor of participants' earnings.

6.1.3 *Behavioural rules*

We have noted that participants' behaviour is not determined by network position: criticality and classical measures of central-

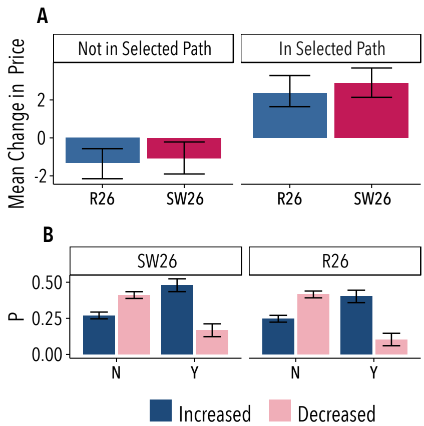


Figure 6.4: **Being or not in the cheapest path determines the intermediaries price increases. A:** Mean changes in the posted price conditioned to have been (right) or not (left) in the selected cheapest path in the previous round for the random networks of 26 nodes (R26), and for the small-world network of 26 nodes (SW26). **B:** Probability to increase (blue) and to decrease (pink) the posted price conditioned to have been (Y) or not (N) in the selected cheapest path, for each one of the studied networks. The error bars represent the 95% C.I. An extension of these results including the 50-nodes Random Network is displayed in Appendix Fig. b.2.

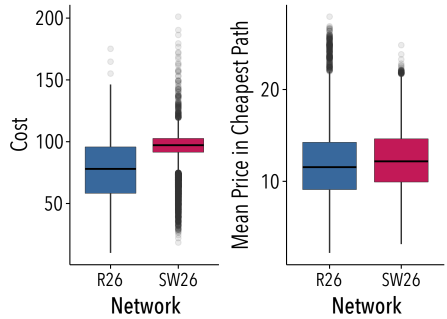


Figure 6.5: **Numerical results of the model executed over the networks, source and destination from the experiments.** Results shown are for 100 executions with 15 rounds for each network and source-destination pair, excluding the first round. Initial prices are bootstrapped from the experimental values. Values of σ and ρ are fixed and correspond to mean values of the experiment, respectively, 2.60 and 1.2.

ity are not good predictors of the prices posted by intermediaries. Nonetheless, results show differences in the prices posted by traders across different networks. Even if these networks might seem relatively small and similar, they are not. The environment (defined as the set of all the information that the individuals need to factor in their decisions) is very complex: there are many different paths passing through most of the traders, they need to take into account their price as well as those of other players, etc. It is thus reasonable to assume that the traders confronting such a complex and dynamic environment use rules of thumb, which on the other hand, should not depend on the network. In what follows, we develop a model that accounts for individual behaviour and for the differences observed experimentally.

Together with the network information, the other information shown to subjects is whether they were on the selected trading path. Fig. 6.4A shows, for each one of the networks considered, the mean change in price for the cases when the participant was or was not along the cheapest path in the previous round. In the same way, Fig. 6.4B shows the probabilities to increase and to decrease the posted price conditioned to have been (Y) or not (N) in the cheapest path. Players appear to follow a simple rule, namely, to increase their price if they were on the cheapest path in the previous round and to decrease it otherwise. Furthermore, the expected values shown in Fig. 6.4A point out that successful intermediaries keep increasing their prices and therefore, without sufficient competition, costs and prices would always grow.

We now build a simple agent based model (ABM) [305], as described below:

i) If node u belongs to a cheapest path at time t, it will change its posted price on time $t+1$ by σ;

ii) If node u does not belong to a cheapest path at time t, it will change its posted price on time $t+1$ by $-\rho$;

iii) The minimum price a node can post is 0.

network	efficiency	price	price in CP	cost	length
R 26	0.87	14.71	11.87	75.68	7.62
SW 26	0.66	15.14	12.56	94.32	8.97

Table 6.2: **Numerical results in experimental networks.** Efficiency (fraction of rounds in which the cheapest path cost was equal to or less than the threshold), and mean values of the price, price in the cheapest path, cost of the cheapest path, and cheapest path length. Results obtained from numerical simulations with each one of the two studied networks with their corresponding source and destinations.

To validate this model we executed it by bootstrapping the initial prices, the value of changes if on the cheapest path (σ) and the value of changes if not (ρ). The results, shown in Fig. 6.5, indicate that costs from simulations (resp. efficiency) are higher (resp. lower) in small-world networks than in random networks (t(9659.3)=68.33, $p < 0.001$), in agreement with our experimental results. Costs reached relatively high values in some rounds, as the model does not incorporate participants direct response to the maximum cost threshold. Table 6.2 also confirms that topological differences between the networks are driving the differences in cost.

Once we have shown that the model captures very well the experimental observations, we verify if the same phenomena are

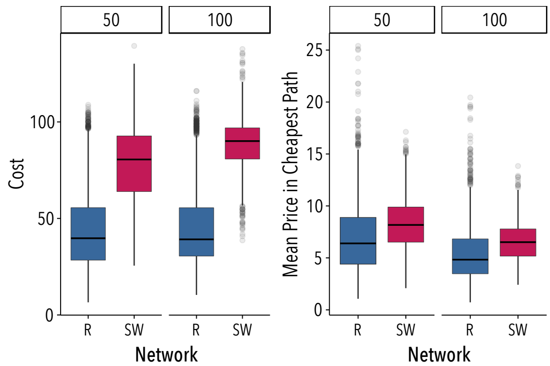


Figure 6.6: **Numerical results of the model for networks with 50 and 100 nodes.** Results shown are for 100 executions with 15 rounds for each network and source-destination pair, excluding the first round. Initial prices are bootstrapped from the experimental values. Values of σ and ρ are fixed and correspond to mean values of the experiment, respectively, 2.60 and 1.2. For similar analyses with random initial prices, see Appendix Fig. b.6.

observed in larger networks. Results for networks of size 50 and 100, shown in Fig. 6.6 and Table 6.3, are also consistent with the experimental data, confirming that the network topology has a significant effect on trading outcomes: small-worlds lead to higher costs and lower efficiency. A similar analysis with random initial prices, thus unlinking numerical results from those obtained from the experiments, can be found in Appendix section b.3.3 and Fig. b.6. Results in Fig. b.6 are compatible with those shown in Fig. 6.6, providing more evidence about the effects of the network structure on prices and costs.

6.1.4 *Topological properties behind the differences in cost*

Finally, we go one step further in order to explain what lies behind the differences found in costs. One possible theoretical hypothesis could be that costs depend on competition between paths. In our setup, this would be equivalent to assume that costs should decrease with the number of possible ways to reach the destination, *i.e.*, the number of independent (sets of) paths from S to D. Specif-

network	efficiency	price	price in CP	cost	length
R 50	0.98	13.12	7.03	44.76	7.68
SW 50	0.91	14.32	8.28	77.44	10.74
R 100	0.97	12.65	5.53	46.50	10.05
SW 100	0.82	13.20	6.56	88.56	15.41

Table 6.3: **Numerical results for larger networks.** Efficiency (fraction of rounds in which the cheapest path cost was equal to or less than the threshold), and mean values of the price, price in the cheapest path, cost of the cheapest path, and cheapest path length. Results obtained from numerical simulations with random networks with 50 and 100 nodes (R 50, R 100) for the small-world network with 50 and 100 nodes (R50, R 100).

ically, we expect competition to be proportional to the number M of node-disjoint paths [306], as it captures the possible number of simultaneous independent trades (see Appendix Section b.3.1 for a deeper discussion on this subject). According to this hypothesis, the larger the value of M, the lower the cost. Another possible explanation for the dependency of costs with the networks could be the structural differences between the latter. It is well known that clustering coefficients and average path lengths differ for the SW and the random networks considered in our experiments ($p \in \{0.1, 1\}$ [59]), and therefore the observed differences in cost could be tied to variations in those properties.

In order to verify the previous hypotheses, we executed a version of the model without the maximum cost threshold. With this setup, we can study long-term effects after a sufficiently large number of rounds and uncover the cost tendency. In this regime, we cannot analyse network efficiency, however, networks yielding higher cost should be more inefficient. Note that the proposed model allows extrapolating the observed behaviour to larger networks with a large range of values of M. Then, we can generalize the observed experimental results to larger networks, which allow us to find the (theoretically conjectured) influence of M on prices. We ran the algorithm for 10^4 rounds and then we considered the final cost of the trade for each configuration. Results for networks of size 26, 50 and 1000 nodes are shown in Figures 6.7A, 6.7B, and 6.7C, respectively. Simulations of trading dynamics on the aforementioned networks indicate that the number of node-disjoint paths (M) between S and D is the best indicator of final cost. Fig. 6.7D shows that as M grows, the costs are reduced so drastically that they go to 0 for $M > 3$. Moreover, the numerical results also reveal that for networks with the same value of M, the cost grows with the average path length. Indeed, this dependency explains why costs on small-world networks tend to be larger: these networks have a larger average path-length. To show that this finding is not a consequence of differences in the length of the cheapest paths, Fig. b.4 of the Appendix displays, for the same simulations, the costs of the cheapest path normalized by the number of nodes on it versus the average path length of the network. It can be seen that the mean price of nodes in the cheapest path also correlates with the average path length. Interestingly, even though in this regime the difference in the clustering is larger than the difference in average path length, the former is not as good as an indicator of costs ($R^2 = 0.57$ vs $R^2 = 0.79$, see Section b.3.2, Table b.2, and Fig. b.5 in the Appendix). In summary, these results provide two stylized facts that may guide future inquiries in this line, namely, trading costs will be null in setups with a relatively large number of node-disjoint paths and costs should be larger in networks with larger average path length.

6.2 CONCLUSIONS

Our experimental results indicate that the trading network has a powerful effect on both the pricing behaviour of intermediaries and the overall efficiency of the system, random networks being more efficient and showing significantly lower prices than small-world networks. However, within a network, prices are relatively insensitive to node location, but intermediaries with greater betweenness make larger profits. Informed by the experimental results, we introduced an ABM of pricing behaviour to understand traders' pricing. The key input of the model is the experimental observation that intermediaries raise prices when they lie on the cheapest path and lower their prices otherwise. The model successfully reproduced qualitatively the experimental results and allowed us to extrapolate

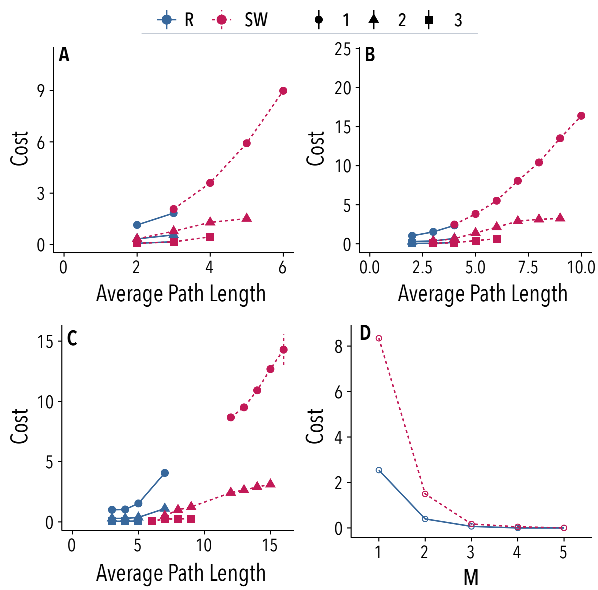


Figure 6.7: **Numerical results of the model. A,B,C**: Average final cost (in 10^4) of the cheapest path after a period of 10^4 rounds as a function of the average path length of the network. Different panels correspond to different network sizes: 26 (**A**), 50 (**B**), and 1000 (**C**) nodes; colours correspond to different network models: random (blue) and small-world (magenta); and different shapes correspond to different values of the number M of disjoint paths. For each configuration, there were generated 10000 networks of size 26, 50, and 1000, according to the Watts-Strogatz algorithm [59] with $p = 0.1, 1$ and average degree from 2 to 10. The initial cost was set to 0 and the increment/decrement ratio was fixed to the experimental value ($\sigma/\rho = 2.4$). Results for $M > 5$ are not shown as costs converge fast to 0. **D:** Mean value of the cost of the cheapest path versus M for the same networks.

and anticipate outcomes of pricing and efficiency to scenarios involving larger networks and longer timescales. Important enough, the model also enabled the discovery of what are the key determinants of cost, namely, the number of node-disjoint paths from source to destination and the network average path length. Ultimately, this explained the differences in our experimental results: in a small-world network, the average path length tends to be larger and this leads to higher costs and lower efficiency of trading in these networks as compared to random networks.

Overall, our work reveals that the topology of trading networks is key to determine their efficiency and cost. It would be interesting to further test our conclusions using real data on trading, in particular, the finding that the availability of node-disjoint paths takes trading costs down. On the other hand, our insights may be useful for the design of competition-improved networks for goods currently overpriced due to intermediation. Further research on the role of information provided to intermediaries and on other network topologies will be also relevant to address these issues.

7 MODELLING THE EVOLUTION OF COOPERATION

In all these scenes of animal life which passed before my eyes, I saw Mutual Aid and Mutual Support carried on to an extent which made me suspect in it a feature of the greatest importance for the maintenance of life, the preservation of each species, and its further evolution.

Mutual Aid: A Factor of Evolution, Kropotkin

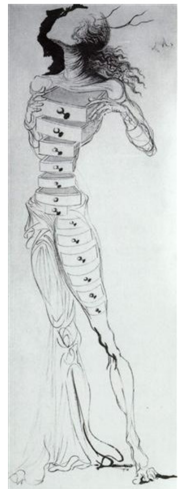


Figure with Drawers for a Four-part Screen, Salvador Dali

The results presented in the last chapters have shown how experiments can be used to obtain knowledge of human behaviour. A next step in the scientific dynamics is formulating new theories that can be tested by new experiments [30], or looking at the implications that the observed behaviour can have in systems that cannot be reproduced in the laboratory, such as we have done in Section 6.1.3. This especially important when studying social systems, given that humans usually live in large groups [2, 3], a condition that is unfeasible of being reproduced in a controlled experiment. Even though large systems can be replicated to some extent [17], some scenarios are just impossible, such as reproducing the evolutionary process of humans or other animals. In these situations, computer simulations are probably researchers' best tool [307], capturing emergent phenomena in situations that closed-form solutions cannot be obtained [115, 308].

Agent-based models (ABM) are especially important in this regard, [305], belonging to a third way of doing science according to Robert Axelrod, having strong assumptions as *deduction*, but coming to conclusions of the simulated data via *induction* [307]. Simulations of ABM allow exploring answers to complex questions, such as one underlying this thesis: how the cooperative behaviour observed in humans have originated? This question is not free from controversies, as we discuss in Sections 2.4.4 and 8.1. We attempt to contribute to this discussion in this Chapter by using evolutionary

"This is the essence of intuitive heuristics: when faced with a difficult question, we often answer an easier one instead, usually without noticing the substitution." Daniel Kahneman [309]

game theory, but without the hard assumptions of pure strategies. Instead, we rely on heuristics, which correspond to the method used by humans when making decisions [192]. Moreover, we use evolutionary algorithms [310] to model the dynamics of strategies selection and, thus, uncover the emerging heuristics. As we present in the next section, our findings resulted to be very insightful, shedding some light in how the evolutionary process might differ between humans and other non-human animals.

7.1 DYNAMICS OF HEURISTICS SELECTION FOR COOPERATIVE BEHAVIOUR

Dynamics of heuristics selection for cooperative behaviour, *under review.*

F. Maciel Cardoso, C. Gracia-Lázaro, & Y. Moreno

Game theory constitutes a powerful framework for the mathematical study of social dilemmas [19, 20]. Within this framework, the most representative and widely used game to model cooperation, the Prisoner's Dilemma, has become a paradigm for modelling the evolution of cooperative behaviour [23]. The Prisoner's Dilemma mimics the worst possible scenario for cooperation in which selfishness always provides a higher individual benefit than cooperative behaviour. Initial predictions indicated the social optimum would not be reachable by rational selfish individuals if the temptation for defecting (T) exceeded the reward for cooperating R. Nonetheless, cooperation is pervasive in human and animal societies [311–313], and a vast literature has demonstrated how cooperation can thrive in the presence of an appropriate evolutionary process [24, 26, 91, 111, 113, 115, 314–316]. The possible situations where cooperation might flourish are endless, and we are just beginning to uncover the ingredients behind the complexity observed in real systems [40, 155]. Consequently, theoretical studies usually focus on simplifications, such as individuals behaving according to fixed pure strategies [111, 115] or some arbitrary set of them [317, 318]. Yet, the reasoning and motivations of humans are more sophisticated and complex than pure strategies and decisions are usually taken factoring in many ingredients, weighting them differently [192]. In other words, generally speaking, the selection of strategies takes place in complex systems wherein imprecise behaviour and the environment are inputs of each other in a perpetual feedback loop [319].

In this line, behavioural economics has shown that humans respond in unexpected ways [33, 193] and often seem to possess hardwired heuristics while acting in experimental situations [320, 321]. Experiments have also shown that humans automatic responses are modelled by experiences from daily-life, building heuristics or intuitions which tend to favour cooperation [150, 154]. Therefore, it is plausible that cooperative societies are sustained by existent heuristics, maintained by norms [2, 101] or biological factors [82, 155, 320], that have resulted from a selection dynamics. It is thus imperative to understand how such possible heuristics have evolved, which will allow explaining the ingrained mechanisms behind the behaviour observed in living beings.

Here, we investigate the evolution of cooperative strategies through an agent-based model of heuristics selection inspired by evolutionary algorithms [310]. The ultimate goal is to obtain a description of the evolutionary process that could lead to different strategies. Explicitly, we consider agents composed of a chromosome and memory to store information of other players' previous actions (Fig. 7.1a). Their actions are responses, according to what is coded in their genes, to other players' history. The strategy space is given, thus, by all the possible genes' combinations. This does not mean that we model behaviours defined by real genomes: decision making, especially in humans, has entangled layers of complexity, and such an approach would be misguided. Rather, we use chromosomes as a tool to model heuristics formed through cultural or biological evolution [319, 322].

In our framework, the fitness of agents corresponds to the payoff obtained in iterated games, and it determines the agent's reproduction rates. Offspring will inherit its parents' chromosomes while being susceptible to mutation. Note that our approach differs from elementary evolutionary algorithms: they optimize functions in a constant fitness landscape, but in evolutionary games changes in the population imply changes in the fitness landscape [323], which can be easily seen in any form of the rock-paper-scissors game [24].

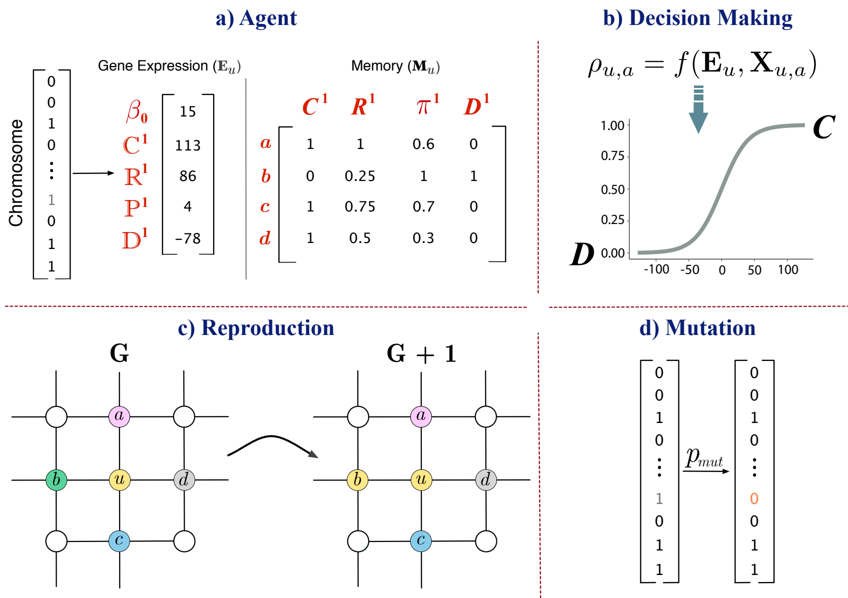


Figure 7.1: **Illustration of the model for memory 1. a)** Agents are composed of a chromosome and memory. Their memories store their experiences with their neighbours, and their chromosomes determine what will be their responses to the variables stored. **b)** Agent u cooperates probabilistically with agent a according to what is coded in its gene and the history of agent a. **c)** Reproduction takes place synchronously at the end of a generation (G): for each site u, a new agent is chosen proportionally to its fitness from the set $\{u \cup N(u)\}$ (coloured nodes), wherein $N(u)$ are u's neighbours. In the example, each colour corresponds to a different chromosome and, at generation $G + 1$, the chromosome of agent u happens to have reproduced in sites u and b, while its other neighbours by chance maintained the same chromosome. **d)** When an agent reproduces, with probability p_{mut} a bit will be flipped.

The use of evolutionary algorithms to explore the adaptation of agents is not new [314, 315, 324], and previous works have studied the evolution of automata-like strategies, though aiming at answering specific situations [325, 326]. In these studies, the equivalent of a chromosome is a tool to encode an extensive set of memory-based strategies used to understand when cooperation may thrive. Unfortunately, these types of strategies are hardly realistic and do not correspond to the optimal model for understanding the mechanisms behind human or animal responses. A model of heuristics should resemble more closely automatic responses based on intuition and past experiences [192], namely, by considering that intuitive responses are no more than stochastic processes which take as inputs the variables observed by the individual.

Here, we develop a modelling approach in which agents can evaluate different variables at the same time, thus resembling real situations wherein different factors interact and affect actions. Agents decisions are determined by an activation function taking as input their chromosome and the information to which they have access. Given their theoretical and practical importance, we focus on the evolution of cooperation in social dilemmas. For this case, therefore, we selected a set of variables based on the history of the players with whom they are playing. Nonetheless, our modelling framework is generic, and any arbitrary set of variables can be added or removed according to the question of interest. Our results show that the specified heuristics can evolve to cooperative equilibria for low mutation rates. An analysis of agents chromosome reveals that cooperation endures by reciprocity, indicating that the evolution drives heuristics to reproduce a fundamental mechanism underlying cooperation in nature, especially in humans [79, 89]. In this case, emerging strategies of conditional cooperators dominate, permitting cooperation to prosper. Finally, we provide an extension wherein agents can evaluate their genetic relatedness with others. The population in this scenario evolves to similar equilibria. However, the agents' chromosomes differ significantly from the first model. Kin identification becomes the main mechanism of cooperative heuristics. Nonetheless, agents still need to have a memory of their past actions for cooperation to endure.

Undoubtedly, varied environmental or perception variables affect the resulting behaviour in humans and other animals. Unfortunately, it is not straightforward to capture which variables guided evolution to the emerged behaviour in each particular scenario. In this line, our proposal provides one generic approach for the modelling of such processes. In particular, the model here presented also contribute some insightful results with the current specifications. Namely, we observe that cooperation can spread spontaneously when memory is available, and that mutation is essential to ensure this outcome. Moreover, although the same behaviour might be observed in distinct populations, the underlying causes might be significantly different, as we observe with our kin and non-kin models. These insights suggest that our method can be a useful tool to uncover the ultimate causes behind the evolution of pro-social behaviour.

7.1.1 *The model*

7.1.1.1 *Population Dynamics*

We consider a virtual environment inhabited by n haploid agents in a zero population growth condition, each one of them (u) containing a chromosome $\mathbf{A}_u$ defining the heuristic which will guide its decision. Each agent interacts with each other through links defined by a static contact structure, in which L is the set of edges connecting the two pairs of agents. In real systems, a generation embodies repeated interactions between individuals, and it is known that fast selection fluctuations can suppress cooperation even in the cases in which it is the only rational choice [113]. In our model, in each generation, there is a finite number of $s = 100$ time steps and, therefore, $s|L|$ dyadic interactions take place, *i.e.*, one for each edge at each time step. Thus, at each time step t, connected agents u and v interact in a game and obtain the payoffs π_u^t and π_v^t, respectively.

The generation reaches its end after the s time steps, and each agent u will have accumulated a total payoff of Π_u, corresponding to its fitness in a strong selection pressure process [113, 327]. Agents reproduce by a localized *death-birth* process [328]: at the end of each generation, each node u will be replaced by a node u' in the set $N_2(u) = N(u) \cup \{u\}$, which is composed by the neighbourhood of

u $(N(u))$ and u itself (Fig. 7.1c). Node u' is chosen probabilistically according to the fitness $(\Pi_{u'})$ of nodes in N_2. Thus, on one hand, the nodes which accumulate more payoff are more likely to be chosen, on the other hand, most adapted agents can reproduce up to sites of distance one. Finally, some fluctuations might affect offspring. Specifically, there is a probability p_{mut} of a newborn having a bit flipped in their chromosome (Fig. 7.1d).

7.1.1.2 *Game*

We are interested in the evolution of cooperation in a population of agents facing a social dilemma. Strictly speaking, we want to check if cooperative heuristics are the most adapted in conditions wherein pure strategies equilibria would be of full defection. We consider that at each interaction, agents play a round of a Prisoners' Dilemma (PD) with their neighbours. The PD game is a 2x2 game in which only two actions are available to the players, either cooperate or defect. If two players cooperate, they both get a reward R, if one cooperates and the other defects, the cooperator earns S and the defector gets a payoff T (the temptation to defect). Finally, if both defect, both of them obtain P. The PD occurs when the elements of the payoff matrix are such that $T > R > P > S$, which implies that a rational player should defect because, whatever your opponent does, the best (in terms of having larger payoff) is to defect. Henceforth, we consider that the values of each entry are a normalized version of Axelrod's tournament [90] values. Namely: $T = 1/\langle k \rangle ; R = 0.6/\langle k \rangle ; P = 0.2/\langle k \rangle ; S = 0$. As mentioned before, for these values, the prediction is that under a replicator dynamics, the system ends up in full defection [113].

7.1.1.3 *Agents*

Agents are hardwired, and their heuristics do not change in the course of one generation, which corresponds to their lifetime. Their heuristics are determined by their chromosomes and constitute a stochastic way to evaluate the variables stored in their memory and make a decision on whether to cooperate or not. Agents' memory stores the variables from previous interactions, and we assume their working memory is limited [329]. Hence, agents can only store

Variable	Gene	Description
	β_u^0	**Constant response.**
$C_{v,u}^l$	$\mathbb{C}_u^l$	**Direct Reciprocity:** 1 if v cooperated with u in round $t-l$, 0 otherwise.
$R_{v,u}^l$	$\mathbb{R}_u^l$	**Indirect Reciprocity:** Fraction of times agent v cooperates in round $t-l$ with players other than u. [1]
π_v^l	$\mathbb{P}_u^l$	**Payoff** obtained by agent v in round $t-l$.
$D_{v,u}^l$	$\mathbb{D}_u^l$	**Punishment:** 0 if v cooperated with u in round $t-l$, 1 otherwise.

Table 7.1: **Memory variables and genes.** Genes determine agents actions, and they are responses to the variables stored in their memory. Here we show the variables considered and their corresponding gene for a previous round l. The description indicates how the information is stored and the role of each gene.

.

a finite set of variables from the previous m rounds. Specifically, an agent u with the set of neighbours $N(u)$, has stored in its memory $\mathbf{M_u}$ variables for all $v \in N(u)$ and for all $l \in [1, m]$. Therefore, $\mathbf{M_u}$ is a matrix wherein each row contains the values stored for one neighbour, as shown in Fig. 7.1a.

The heuristics evaluate each stored variable according to a specific gene in the chromosome. Therefore, the expression of each gene is a weight given to a variable containing some information influencing agents' decision making. The vector $\mathbf{E}_u$ carries the responses of an agent u, i.e., its expressed genes values. They are given by a two complement representation of the gene bits, therefore, they are integers from -128 to 127[2]. The vector E_u contains the responses to the variables plus a constant response (β^0). Table 7.1 shows the set of variables stored and their corresponding genes. They are a basic set of external characteristics that an elementary agent can observe. Thus, they constitute a reasonable set of vari-

2 Hence, when a mutation occurs in a gene, from its expressed value it can be added/subtracted a random power of 2, or have its sign and value changed.

ables to be taken into account by a somewhat minimal heuristic. Finally, whether or not an agent will cooperate is determined by the sigmoid function specified by

$$\rho_{u,v} = f(\mathbf{E}_u, \mathbf{X}_{u,v}) = \frac{1}{1 + e^{-\kappa(\mathbf{X}_{u,v} \cdot \mathbf{E}_u)}}, \tag{7.1}$$

where $\rho_{u,v}$ corresponds to the probability of agent u cooperating with agent v. $\mathbf{X}_{u,v} = (1) \oplus \mathbf{M}_{u,v}$ corresponds to a vector composed by the number 1 in the first position, followed by the memory variables as specified in Table 7.1. κ (henceforth set to 0.05) provides the steepness of the curve, and it is chosen in such a way that if the dot product of both vectors is greater (*resp. lesser*) than 100, the probability should be approximately 1 (*resp. 0*), as illustrated by Fig. 7.1b.

7.1.2 *Results*

We ran simulations for populations of 1024 agents connected on a lattice (*LTT*) with a von Neumann neighbourhood and on Random Regular Networks (*RRN*) with the same nodes' degree ($k = 4$). We evolved the model for $5 \cdot 10^5$ generations, each with 100 rounds, for different values of the p_{mut} parameter. Results for memory between 0 and 5 are shown in Fig. 7.2. When $m = 0$, the agents' chromosome is composed of only the constant response (β^0) and strategies are reducible to mixed strategies. In this case, when no mutation is available the system quickly goes to full defection, as expected, and mutation increases the possibility of adding cooperative strategies by drift. Conversely, when agents have access to memory, cooperation is predominant in the regime of low mutation. Furthermore, cooperation is larger and more resilient to higher mutation when agents have access to a bigger memory. With more memory, agents should be able to construct more complex heuristics which seem to favour cooperation.

Figure 7.3 shows time evolution curves of individual realizations for $m = 1$. When $p_{mut} = 0$, the final fraction of cooperative fraction is highly dependent on the initial conditions, reaching a multitude of equilibria, some being fully cooperative and others showing a rather small level of cooperation, specially in the RRN network.

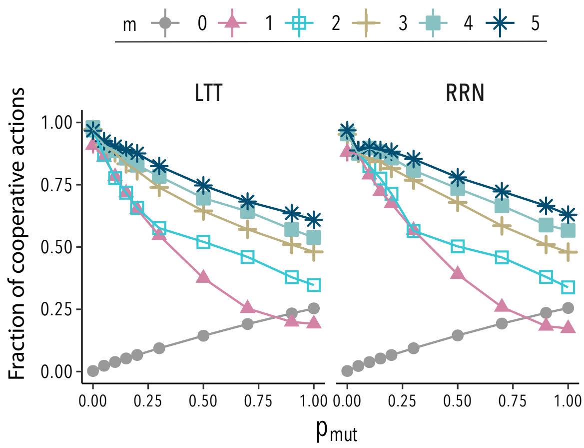


Figure 7.2: **Cooperation thrives at low mutation values**. Fraction of cooperative actions at the steady-state according to the mutation probability. Colours and shapes correspond for different memory (m) values. Averages plus .95 confidence interval of 100 realizations are presented for each mutation (p_{mut}) value.

In the regime of small mutation rates, fluctuations increase significantly. However, for some small values of the mutation rate, all realizations converge to highly cooperative equilibria, as can be seen when $p_{mut} = 0.05$. Note, additionally, that as the probability of mutation increases, the fraction of cooperative actions decreases. For the limiting value $p_{mut} = 1$, every new player is born with a mutation and the system evolves into a negligible average level of cooperation. Interestingly, this is a demonstration that a small noise can foster cooperation in the process of evolution. With more mutation, it gets harder for cooperative strategies to prevail and defection tends to increase, however, a sufficiently small mutation probability will guarantee that the system evolves to a cooperative equilibrium.

7.1.2.1 *Other payoff values*

To ensure that our results are robust with respect to differences in the payoff values, we ran simulations for different values of the temptation parameter T. To make our results comparable to previous work, we used the one-dimensional parametrization of

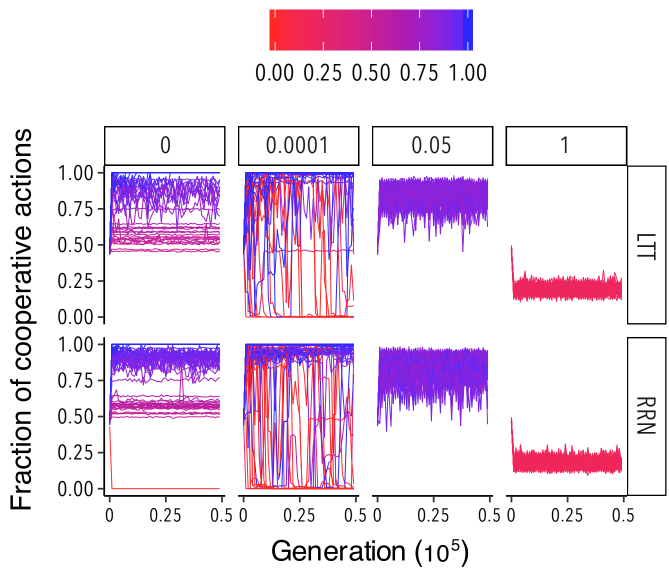


Figure 7.3: **Fraction of cooperative actions at the end of each generation**. Columns correspond to different mutation values (0, 0.0001, 0.05, 1) and horizontal panels to different networks (LTT, RRN). Agents have memory $m = 1$ and 100 realizations are performed for each mutation value and network. The colours of the lines correspond to the average of the last 1000 generations.

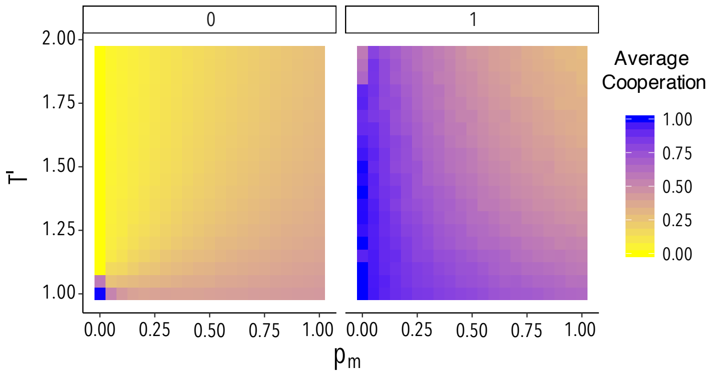


Figure 7.4: **Average cooperation at the stationary state**. The left panel shows results for $m = 0$ and the right panel for $m = 1$. 100 simulations were done for each p_m and T' combination. Colour coding corresponds to the average over all realizations and varies from blue (1) to yellow (0).

payoffs used by Nowak et. al [111]. In this version, $R = 1$, $P = \epsilon$, $S = 0$, and T varies from 1 to 2, with ϵ being a value close to zero. As we consider normalized versions, the payoff here is defined by $T = T'/\langle k\rangle; R = 1/\langle k\rangle; P = 0.01/\langle k\rangle; S = 0$, with T' varying from 1 to 2. Results for memory 0 and 1 are shown in Fig. 7.4. The results show that without memory, cooperation is only attainable when $T' = 1$ and low mutation. However, when agents have memory of their last interaction, cooperation endures even when the temptation to defect is around 2.

7.1.2.2 *Heuristics and Strategies*

In this section, we focus on the composition of the populations in the different regimes. It is not straightforward to evaluate how genes and variables interact, hence, it is hard to determine if agents are going to cooperate or not in a specific situation. A first step is to investigate what are the gene values in cooperative and non-cooperative equilibria. Fig. 7.5 show the distributions of genes for two mutation values: $p_{mut} = 0.05$ and $p_{mut} = 1$, wherein evolution leads to mostly cooperation and to mostly defection, respectively. Simulations in both LTT and RRN networks yielded similar distributions, indicating the presence of a common evolutionary pattern.

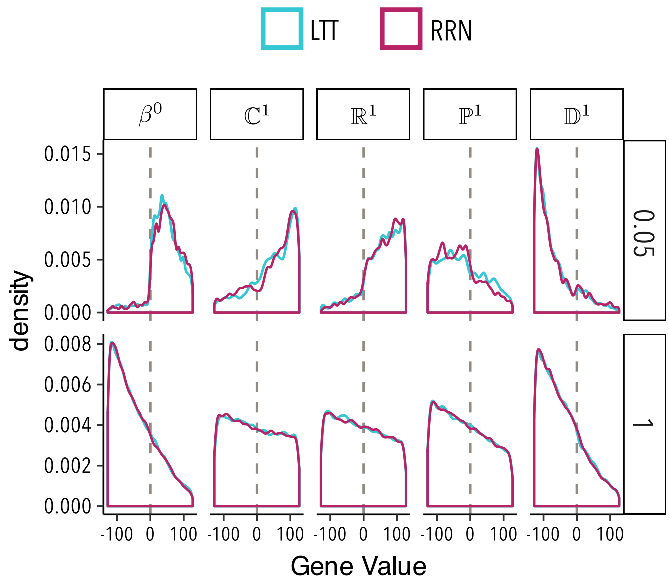


Figure 7.5: **Distribution of genes' expressed values.** Densities of genes values for simulations on LTT and RRN graphs for $m = 1$. Top panels show distributions for $p_{mut} = 0.05$ and bottom panels for $p_{mut} = 1$. The vertical dashed line indicates separate regions wherein the marginal probability to cooperate would be smaller (negative gene values) and greater (positive gene value) than 0.5.

When the majority of the population cooperates ($p_{mut} = 0.05$), β^0, $\mathbb{C}^1$, and $\mathbb{R}^1$ have a clear right-modality with most of these values being higher than 0. Conversely, $\mathbb{D}^1$ is left-modal with a clear peak at extreme negative values, while $\mathbb{P}^1$ shows a softer trend towards negative values. This implies that when cooperation thrives, agents have a baseline cooperative response and tend to reciprocate cooperation both directly and indirectly. On the other hand, the agents punish defectors rigorously and have a mild negative response to other agents' payoff, probably as a means to punish defectors, as only defectors can attain the highest payoffs. Interestingly, the distributions of β^0 indicate that the emerging strategies are willing to cooperate even in a one-shot game with an unknown player as show in Fig. 7.7, albeit this is not the expected behaviour for $m = 0$. In the other extreme, for $p_{mut} = 1$, defection prevails, and genes values indicate the underpinnings of this trend. All distributions are right-skewed, with β^0 and $\mathbb{D}^1$ having a noticeable peak at the lowest possible values. Thus, when mutations are too frequent agents are much more likely to exploit and punish, leading defection to be the default strategy. Too much drift will make it impossible for cooperative heuristics to be selected, and they will vanish in the population.

These last results provide a picture of the genotype space. However, there is still the need to identify which strategies have emerged. When studying evolutionary games, it is always challenging to bridge the gap between the genotype and phenotype spaces [323]. In our model, the profile of agents' actions would correspond to observable phenotypes, yet it is not straightforward to specify a method for heuristics classification. An unsupervised procedure would fall into the problem of how to identify the groups encountered, i.e., how to determine to which known strategies they correspond. Therefore, here we adopted an approach that consisted of classifying agents by looking at what would be their responses to the most basic strategies: a pure defector and a pure cooperator. Namely, we looked at whether agents were likely to cooperate or defect with agents having a history corresponding to each of the two pure strategies. For instance, a full defector v would always have defected with u ($C^1_{v,u} = 0$, $D^1_{v,u} = 1$), with its other neighbours

	Pure Cooperator	Pure Defector
$C^1_{v,u}$	1	0
$R^1_{v,u}$	1	0
π^1_v	$\langle \pi_C \rangle$	$\langle \pi_D \rangle$
$D^1_{v,u}$	0	1

Table 7.2: **Pure strategies memory.** Past history of the two basic strategies according to what would have been played by them for $m = 1$.

($R^1_{v,u} = 0$), and would have an expected payoff (π^1_v) corresponding to these actions.

Table 7.2 illustrates the variables contained in the memory of agent u with respect to a player v, corresponding to the two pure strategies for $m = 1$. All the values are given straightforwardly, expect for π^1_v. Payoffs values are more complicated, as they depend on the players with whom they are playing with, which we cannot define a priori. We decided to use the average payoff of individuals which cooperated and defected with all their neighbours for the pure cooperator and pure defector, respectively. Therefore, $\langle \pi_C \rangle = \langle \pi^t_i \rangle \, \forall i \in R_1, \forall t \in [1, 100]$ and $\langle \pi_D \rangle = \langle \pi^t_i \rangle \, \forall i \in R_0, \forall t \in [1, 100]$, wherein R_1 (resp. R_0) corresponds to the set of agents which cooperated with all (resp. none) of their neighbours in the last time step.

We then, use the threshold σ to divide the plane (ρ_C, ρ_D). Namely, we designate as cooperation when $\rho > (1 - \sigma)$, defection as $\rho < \sigma$, and random (-) when $\sigma \leq \rho \leq (1 - \sigma)$. This process results in the proposed classification is shown in Table 7.3. We considered strategies analogous to known ones, namely: *Full Cooperator (FC)*, cooperates with both pure cooperators and pure defectors; *Full Defector (FC)*, defects with both; *Conditional Cooperator(CC)*, reciprocates cooperation and defects otherwise; *Generous Conditional Cooperator(GCC)*, reciprocates cooperation and can cooperate randomly with defectors; *Conditional Defector(CD)*, cooperates randomly with cooperators and always defects with defectors; *Bully*, defects with cooperators, but cooperates with defectors; *Random*, behave randomly with both pure strategies. We labelled agents that could not be classified by this process as *Undefined*.

	Pure Cooperator	**Pure Defector**
FC	C	C
FD	D	D
CC	C	D
GCC	C	-
CD	-	D
Bully	D	C
Random	-	-

Table 7.3: **Classification of heuristics according to their responses to the two pure strategies:** pure defector and pure cooperator. We consider that agents cooperate (C) or defect (D) if their probability to cooperate is greater than $(1-\sigma)$ or smaller than σ, respectively.

In Fig. 7.6, we show the frequencies of each strategy from simulations of the heuristics selection dynamics. Top panels (**A**) show strategies for a lattice and bottom panels (**B**) for RRN networks. Results in both networks types are very consistent: when the mutation is low ($p_{mut} = 0.05$), most of the agents tend to be cooperators or conditional cooperators (mean fraction is 0.9 with a standard deviation of 0.07): CC constitutes most of the strategies, followed by a small fraction of GCC and FC players. In contrast, when mutation is high ($p_{mut} = 1$), FD and CD constitute the majority (mean=0.66, sd=0.038) of agents. However, a minority of CC players can persist (mean=0.17, sd=0.022), which explains the existence of a small fraction of cooperative actions even in this regime.

7.1.2.3 *Exploring kin discrimination: a first extension.*

It is known that cooperative behaviour can emerge and be sustained by factors that do not depend on players history of decisions. Namely, genetic relatedness or kinship plays a key role in the evolution of cooperation in nature [79, 82, 83, 89]. Kin selection is pervasive [311, 312], despite controversies over its role in particular phenomena [84, 107, 108, 110, 330, 331]. Indeed, these disagreements indicate the need to investigate the role played by genetic

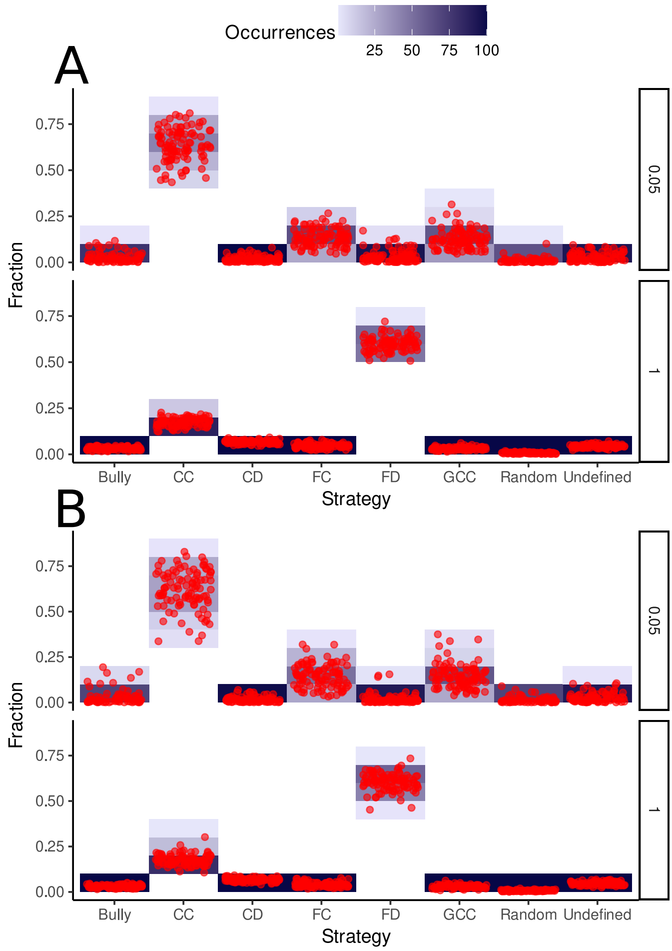


Figure 7.6: **Emerging Strategies.** Frequency of each strategy in executions in LTT (top panels **A**) and RRN (bottom panel **B**) networks for $p_{mut} = 0.05$, $\sigma = 0.3$ and $m = 1$. Each red dot correspond to the fraction of the strategy in a simulation and the histogram of fractions for each strategy is shown vertically, with darkest colours representing a higher number of occurrences.

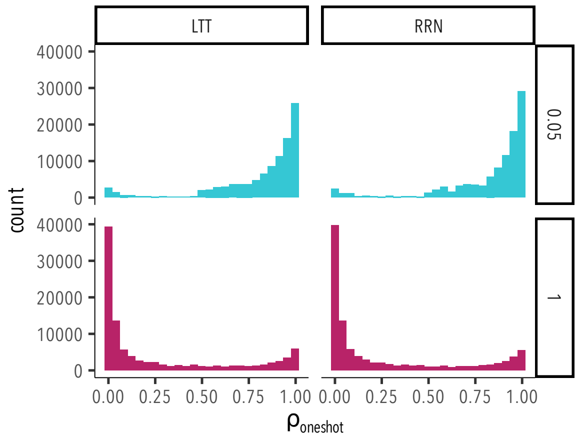


Figure 7.7: **Cooperation probability in one-shot games**. Distributions are calculated over agents of the final generation for $m = 1$ in a lattice (left panel) and in random regular networks (right panel). Probability is calculated considering that agents do not have access to other participants information, thus, only β^0 is used in the sigmoid. Top panels show distributions for $p_{mut} = 0.05$ and bottom panels for $p_{mut} = 1$.

relatedness in each specific scenario [84]. Therefore, to address this question, we take such mechanisms into account in the evolutionary dynamics of heuristics selection. Namely, we have extended the previous analysis and considered that agents could evaluate an additional variable that accounts whom they are interacting with, specifically, genetic proximity, which is one main mechanism ensuring interactions occur among related individuals [79].

We added to the agents' chromosome a gene $\mathbb{K}$ to account for genetic relatedness with the interacting agent. Operationally, we consider that this kinship relation is given by the Jaccard index of pairs of agents' chromosomes. Note that we are not specifying a method for kin selection, but allowing the heuristics to take into consideration agents similarity when deciding to cooperate or not. Enabling, thus, an estimation of the relevance of genetic relatedness by evaluating the weight organically given to the heuristics' new gene.

Results of simulations on a lattice are presented in Figure 7.8. Figure 7.8A shows the fraction of cooperation at the steady-state both

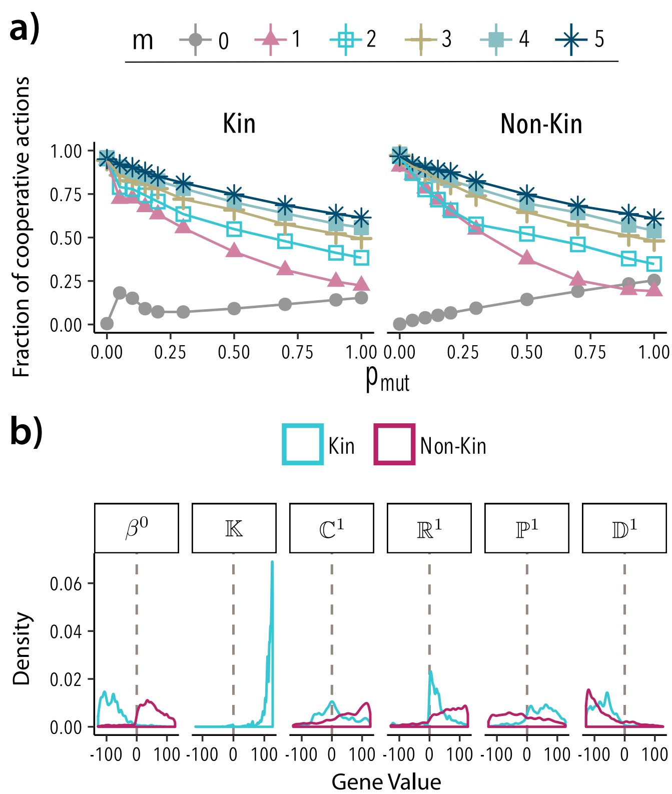


Figure 7.8: **Evolution of heuristics with kin identification on LTT. A** Fraction of cooperative actions at the steady state as a function of the mutation probability. Colours and shapes correspond for different memory (m) values. Averages plus .95 confidence interval of 100 realizations are presented for each mutation (p_{mut}) value. **B** Densities of genes values for simulations on LTT graphs for $p_{mut} = 0.05$ and $m = 1$. The vertical dashed lines separate the regions wherein the probability to cooperate would be smaller (gene value smaller than 0) and greater (gene value smaller than 0) than 0.5.

for our previous model (*Non-Kin*) and for the extended model (*Kin*). The evolution leads to similar scenarios in both cases, indicating that the presence of the ($\mathbb{K}$) gene did not enhance nor undermine cooperation significantly, though there is one modest exception. For heuristics without memory ($m = 0$) and low mutation, there is a modest increase in the level of cooperation. At variance with the model in a lattice, when there is no mutation, the fraction of cooperative actions can be different from zero in an RRN network, as shown in Fig. 7.9A. This demonstrates how important the kin identification mechanism can be in an adequate environment.

Despite the negligible differences in outcomes, there is a substantial effect on agents' chromosomes. Fig. 7.8B (resp. 7.9B) shows that including the possibility to weigh gene similarity changes the values of all other genes significantly in lattices (resp. RRN networks). For $m = 1$, cooperation is strongly determined by the ($\mathbb{K}$) gene, and genes for direct reciprocity and constant response becomes negative or neutral. The latter implies that most agents will not cooperate in one-shot interactions with unrelated individuals, as shown in Fig. 7.10, demonstrating a significant difference from the agents without the $\mathbb{K}$ gene. There still is a mostly positive response for indirect reciprocity and a negative for punishment, while the weight given to participants payoff inverts. This result points to a compelling message: when heuristics can evaluate genetic relatedness, the ones that do that will have a higher reproduction, therefore resulting in more adapted heuristics. Nonetheless, information from past interactions is still required, with punishment and reciprocity playing a role.

7.2 CONCLUSIONS

Natural selection has shaped the evolution of all sort of life forms. Advantageous strategies endure while others dwindle in a never-ending process of adaptation. Fundamental questions regarding the emergence of cooperative behaviour in social dilemmas have to be studied in the light of evolutionary mechanisms. Undoubtedly, emerging behaviour is intrinsically dependant on the individuals under study, e.g., humans commonly cooperate in large societies composed of unrelated individuals, while groups of animals are

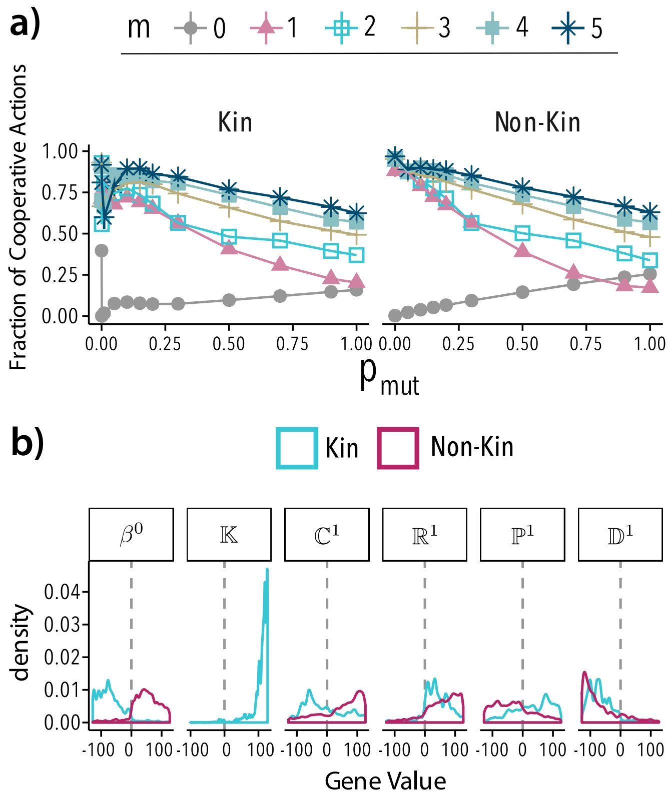


Figure 7.9: **Evolution of heuristics with kin identification on RRN. A** Fraction of cooperative actions at the steady state as a function of the mutation probability. Colours and shapes correspond for different memory (m) values. Averages plus .95 confidence interval of 100 realizations are presented for each mutation (p_{mut}) value. **B** Densities of genes values for simulations on RRN graphs for $p_{mut} = 0.05$ and $m = 1$. The vertical dashed lines separate the regions wherein the probability to cooperate would be smaller (gene value smaller than 0) and greater (gene value smaller than 0) than 0.5.

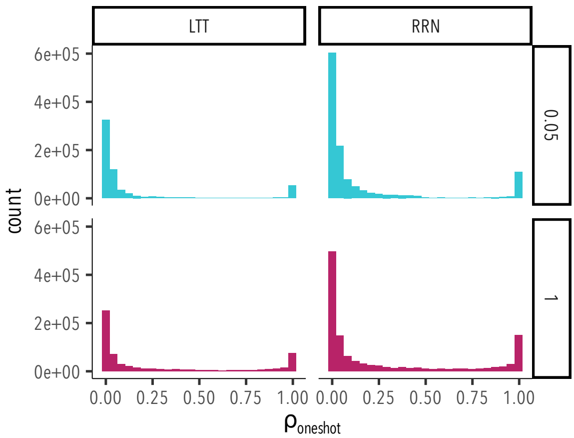


Figure 7.10: **Cooperation probability in one-shot games for the model with the kin identification gene**. Distributions are calculated over agents of the final generation for $m = 1$ in a lattice (left panel) and in random regular networks (right panel). Probability is calculated considering that agents do not have access to other participants information, thus, only β^0 is used in the sigmoid. Top panels show distributions for $p_{mut} = 0.05$ and bottom panels for $p_{mut} = 1$.

hardly greater than a few hundred [3]. In particular, variance in humans is especially relevant, as behaviour is deeply affected by the specifics of the interactions and the culture of the individuals [15, 40]. Moreover, given that it is an emergent phenomenon, behaviour can be deeply affected by the complex topology of interactions [117]. In an attempt to provide a framework for such scenarios, here we explore a model that allows unravelling what could be the drivers of cooperation by a heuristics selection process.

By exploring heuristics that make use of agents behavioural information to stochastically determine their decisions in iterated prisoners' dilemma games across generations, we have shown that, in a feasible environment, evolution will drive heuristics towards cooperation even when defection is expected for pure strategies. In these scenarios, reciprocity and punishment are the main ingredients of cooperators' decision-making, and most strategies will follow conditional cooperation. The fraction of cooperative decisions decreases with an increase in the mutation rate, nonetheless, for small mutation rates the system reaches a cooperative equilibrium. Without mutation, the configuration of the initial state is critical and the system can get trapped in equilibria of meagre cooperation. Increasing the memory of individuals also increases the fraction of cooperation, suggesting that heuristics with more resources are more cooperative. These aggregate results are indistinguishable from a version of the model wherein agents have, in addition to behavioural information, access to their similarity with others (which mimics genetic relatedness). For this latter scenario, the level of cooperation at the macroscopic level remains roughly the same. Important enough, however, at the level of individuals, chromosomes change significantly and cooperation is given through a kin identification process.

Therefore, when agents discriminate their kin, reciprocity loses much of its importance, which is especially insightful given the behaviour observed in nature. Kin selection is arguably the most important mechanism behind cooperation in non-human animals, while reciprocity is uncommon [79, 89]. Our result suggests that in order for reciprocity to be dominant, perfect kin discrimination cannot exist, which suggest that figuring out the interplay between both mechanisms is crucial for understanding human evolution.

Moreover, agents evolved in each condition presented a different expected response in one-shot games with unrelated individuals: cooperation is likely without the kin discrimination gene, while the majority of agent will defect when they can discriminate their genetic similarity.

To round off, we note that heuristics will adapt according to the information that they have access to, and they can change significantly according to the variables available. Surprisingly, despite changes in methods, cooperation is more likely than exploitation, due to reciprocity [87, 95] or to kin selection [82]. This suggests that even if individuals have limited cognitive capacities (a small memory weighed by a rather inexpensive function), cooperative heuristics can have higher reproduction rates and be pervasive. However, extrapolations have to be made with caution. As it is often the case of works in evolutionary game theory, our model sidesteps important details from biology and cognitive sciences [332]. Future work should explore the intersection between moral and material values and how it influences heuristics [313], and how selection works in more complex scenarios, for instance, when higher cognition has higher associated costs [333]. Moreover, our approach could be used to understand how cultural characteristics [2, 40] drive cooperation in different directions by modelling proper environmental variables, and whether costly punishment could sustain large scale cooperation [47]. We plan to explore this and similar questions next.

Part III

DISCUSSION

8 CONCLUSIONS

There is no real ending. It's just the place where you stop the story.

Frank Herbert

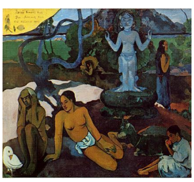
Grey Havens, Alan Lee

This thesis has focused on the study of people's interactions predominantly from an experimental perspective. Accordingly, we have presented our approach to observing human behaviour in controlled experiments, given their advantages and limitations (Chapter 3). Our results generally confirm that people are not entirely self-interested, joining the large body of evidence demonstrating that we are a cooperative species [313]. We show that people's behaviour is significantly varied: some can contribute more than others to public goods (Section 4.2 and 5.1), while others can be more willing to free ride and overexploit common resources (Section 4.1). Moreover, people not only coordinate and cooperate, they might also engage in costly altruism. This behaviour, nonetheless, is highly dependant on how information is presented to them, with some framings and contexts generating more altruistic responses than others (Chapter 5). Critically, however, people walk a thin line while in a dangerous scenario (Chapter 4) and their current pro-social response, despite existent, might not be sufficient for the challenges humanity still have to face [6, 179].

These results indicate that people, although relatively cooperative, also can compete significantly with others, which might hamper the unprecedented levels of cooperation we will inevitably need [6]. Nonetheless, both types of behaviour are the elementary constituents of groups' social psychology [11]. They are necessary elements of groups formation, as groups get cohesive with intra-group cooperation and inter-group competition. Competition might also be the driver for our unique level of cooperation, given that competition between groups required a higher level of in-group mutual-aid [4, 334]. Besides, competition is also behind the perfect functioning of markets and, consequently, the welfare of modern societies. In this line, our results from Chapter 6 show that human behaviour alone is not sufficient to understand the functioning of markets. In fact, in a complex environment, people might rely on rules of thumb and market efficiency will chiefly be determined by the underlying network topology.

Empirical findings indicate that rules of thumb or heuristics are the way humans act towards complex problems [32, 192], being them a result of biological and cultural evolution [319]. To understand how this process unfolded, we proposed a model for the evolution of cooperation in structured populations. Our findings show that heuristics will evolve towards conditional cooperation. However, if they can discriminate others by similarity, kin-discriminating strategies will outcompete conditional cooperators. This agrees with the general evidence that reciprocity is rare in non-human animals [89], and possibly reciprocity only outcompetes kin-discrimination when distinguishable characteristics of humans evolutionary history are at place [4].

Therefore, given that specific contributions are detailed in each chapter, the general contributions of this thesis can be summarised in four main points: *i)* incrementing our body of knowledge with respect to how humans contribute to public goods and manage *"public bads"* [12], such as common-pool resources; *ii)* a study of the interplay between framing and altruistic behaviour considering people's socio-demographic characteristics; *iii)* an understanding of human behaviour while bargaining in networks and how trade efficiency can be affected by network topology; *iv)* a novel approach to model strategies through evolutionary heuristics, illustrated by an analysis of the resulting equilibrium strategies in different scenarios.



Detail from ***Where do we come from? What are we? Where are we going?***, `Paul Gauguin`

8.1 FURTHER & BEYOND

There is still much to investigate and advance in our understanding of the behaviour of humans and other animals, especially with respect to the origins of pro-sociality and altruism. Despite being intensively researched (Section 2.4), determining the causes behind the emergence of cooperation still has controversies, even when focusing on non-human animals. For instance, recently, it has been under scrutiny whether inclusive fitness [82, 83] is an essential driver behind the evolution of biological altruistic behaviour (Section 2.4.1). Attackers argue that the inclusive fitness concept is not that useful for understanding the emergence of cooperation in eusocial species [107, 330], and it should be understood as a result of group (or multilevel) selection [335, 336]. On the other hand, defenders declare that inclusive fitness is relevant to the point of being inseparable from the concept of natural selection: *"Inclusive fitness is as general as the genetical theory of natural selection itself"* [108]. Ultimately, both sides seem to recognize the existence of kin selection [84], and this dialectical discussion appears to be reaching a synthesis [337]. Consensus has gone from approaches being somewhat equivalent [338], to recognizing that they make different causal hypothesis [84, 110]. Besides, this discussion has shown that better approaches to testing Hamilton's rule could be devised [332]. Thus, theory advances at last, yet this process demonstrates that we have to consider current findings prudently. Only by small increments we will understand the mechanisms behind the high cooperation level observed in humans.

Moreover, to evaluate humans pro-sociality, we should not neglect the inquiry of whom is benefiting from cooperation. In this regard, cooperating with kin [82] and cooperating with peers [87] are mechanisms prone to sustain corruption in modern societies, as they can lead to nepotism and cronyism, respectively. Indeed, tight kinship structures are correlated with high corruption indexes, while non-kin practices are associated with higher levels of cooperation [339]. Therefore, proposals to undermine corrupt practices should consider the underlying motivations behind them. For instance, while punishment towards defectors increases overall cooperation level [47], increasing the visibility of corrupt practices can be detrimental to cooperation, contradicting the idea that transparency will work as a punishment to corruptors [340]. In general, treating corruption as defection is misguided, and it is more appropriate to picture it as it is: the downside of cooperation. Only in this way, public policies can function properly to ensure impartial and generalized cooperation.

"As with Tolstoy's happy families, in this and other games there seems to be just one way to be self-interested ... but many ways to depart from the standard economic model"
`Samuel Bowles` *[12]*

Ultimately, there is no general law applying to all contexts and all populations. On the one hand, competition might be useful and make trade more efficient, as we saw in Chapter 6. On the other hand, competing for resources might deplete our natural resources, as we saw in Chapter 4, and solutions for this have to rely upon cooperation. Furthermore, humans vary not only across cultures but within countries themselves, and cooperative behaviour is highly dependant on context and local ecology [341]. As we have shown here, people from different countries (Chapter 4), gender and background (Chapter 5) vary significantly in their behaviour. Therefore, broader analyses from different scenarios are needed, as also thorough investigations on the causal mechanisms in each specific context.

Without proper consideration, however, the focus on data acquisition [342] can intensify the already vast collection of non-cohesive results in the social sciences, with some of them being not reproducible [343]. Consequently, improvements in experimental practices as also sharing standardized methodologies among researchers are mandatory and urgent. Still, solely increasing our availability of data and improving our empirical methodology will not be enough to solve current issues. As we have seen in Section 2.2.1.1, it is impossible to disentangle all experimental effects, which impedes a proper assessment of results' validity. In this regard, having sound foundations to evaluate the likelihood of empirical findings is a pressing need, which will be only possible if we devise an overarching theory of human behaviour [344].

"You step into the road, and if you don't keep your feet, there is no knowing where you might be swept off to."
Bilbo Baggins in ***The Fellowship of the Ring***, `J.R.R. Tolkien`

We hope this thesis provides basis and motivations for these quests and, luckily, soon we may better scrutinize human behaviour in all its varieties. Then the odds of building more pro-social societies and successfully overcoming the impending challenges should be in our favour.

9 CONCLUSIONES

Esta tesis se ha centrado en el estudio de las interacciones de las personas, predominantemente desde una perspectiva experimental. Por consiguiente, hemos presentado nuestro enfoque para observar el comportamiento humano a través experimentos controlados, considerando sus ventajas y limitaciones (Capítulo 3). Nuestros resultados generalmente confirman que las personas no son completamente egoístas, consonancia con la gran evidencia demostrando que somos una especie cooperativa [313]. Demostramos que el comportamiento de las personas es significativamente variado: algunos pueden contribuir más que otros a los bienes públicos (Sección 4.2 y 5.1), mientras otros pueden estar más dispuestos a aprovechar gratuitamente y sobreexplotar los recursos comunes. (Sección 4.1). Además, las personas no solo se coordinan y cooperan, también pueden actuar altruísticamente, suponiendoles un coste. Este comportamiento, sin embargo, depende en gran medida de cómo se les presenta la información, con algunos encuadres y contextos generando respuestas más altruistas que otros (Capítulo 5). No obstante, las personas caminan por una línea tenue en escenarios temerarios (Capítulo 4) y su respuesta prosocial vigente, a pesar de existir, podría no ser suficiente para los desafíos que la humanidad aún debe enfrentar [6, 179].

Estos resultados indican que las personas, aunque sean relativamente cooperativas, también pueden competir significativamente con otras, lo que podría obstaculizar los niveles sin precedentes de cooperación que inevitablemente necesitaremos [6]. Sin embargo, ambos tipos de comportamiento son los componentes elementales de la psicología social de grupos [11]. Son elementos necesarios para la formación de grupos, ya que los grupos se cohesionan con la cooperación intragrupal y la competencia entre grupos. La competencia también podría ser el motor de nuestro nivel único de cooperación, dado que la competencia entre grupos requiere un mayor nivel de ayuda mutua dentro del grupo [4, 334]. Además, la competencia también está detrás del perfecto funcionamiento de los mercados y, en consecuencia, del bienestar de las sociedades modernas. En esta línea, nuestros resultados del Capítulo 6 muestran que el comportamiento humano por sí solo no es suficiente para comprender el funcionamiento de los mercados. De hecho, en un entorno complejo, las personas pueden confiar en reglas generales y la eficiencia del mercado estará determinada principalmente por la topología de la red subyacente.

Los hallazgos empíricos indican que los seres humanos actúan a través de reglas de oro o heurísticas frente a problemas complejos [32, 192], siendo estas el resultado de la evolución biológica y cultural [319]. Para comprender cómo se desarrolló este proceso, propusimos un modelo estudiando la evolución de la cooperación en poblaciones estructuradas. Nuestros hallazgos muestran que las heurísticas evolucionarán hacia la cooperación condicional. Sin embargo, si pueden discriminar a otros por similitud, las estrategias basadas en la discriminación por parentesco superarán a los cooperadores condicionales. Esto coincide con la evidencia general de que la reciprocidad es poco común en animales no humanos [89], y posiblemente la reciprocidad sólo supera la discriminación por parentesco cuando tienen lugar características distinguibles de la historia evolutiva de los humanos [4].

Por lo tanto, dado que las contribuciones específicas se detallan en cada capítulo, las contribuciones generales de esta tesis se pueden resumir en cuatro puntos principales: *i)* incrementar nuestra base de conocimiento con respecto a cómo los humanos contribuyen a los bienes públicos y gestionan los *"malos públicos"* [12], tales como los recursos comunes; *ii)* un estudio de la interacción entre el encuadre y el comportamiento altruista considerando las características sociodemográficas de las personas; *iii)* la comprensión del comportamiento humano al negociar en redes y cómo la topología de la red puede afectar la eficiencia del comercio; *iv)* un enfoque novedoso para modelar estrategias a través de heurísticas evolutivas, ilustrado por un análisis de las estrategias de equilibrio resultantes en diferentes escenarios.

9.1 ADEMÁS & MÁS ALLÁ

Aún queda mucho por investigar y avanzar en nuestra comprensión del comportamiento de humanos y otros animales, especialmente con respecto a los orígenes de la pro-socialidad y del altruismo. A pesar de haber sido investigado intensamente (Sección 2.4), determinar las causas subyacentes al surgimiento de la cooperación todavía conlleva controversias, incluso cuando se enfoca en animales no humanos. Por ejemplo, recientemente, ha estado bajo escrutinio si la aptitud inclusiva [82, 83] es un impulsor esencial de la evolución del comportamiento altruista biológico (Sección 2.4.1). Los atacantes argumentan que el concepto de aptitud inclusiva no es tan útil para comprender el surgimiento de la cooperación en especies eusociales [107, 330], y esta debe entenderse como resultado de la selección de grupo (o multinivel) [335, 336] . Por otro lado, los defensores declaran que la aptitud inclusiva es relevante hasta el punto de ser inseparable del concepto de selección natural: *"Inclusive fitness is as general as the genetical theory of natural selection itself"* [108]. En última instancia, ambas partes parecen reconocer la existencia de la selección por parentesco [84], y esta discusión dialéctica parece estar llegando a una síntesis [337]. El consenso ha pasado de enfoques algo equivalentes [338], a reconocer que hacen diferentes hipótesis causales [84, 110]. Además, esta discusión ha demostrado que se podrían idear mejores enfoques para verificar la regla de Hamilton [332]. Por lo tanto, finalmente la teoría avanza, pero este proceso demuestra que tenemos que considerar los hallazgos actuales con prudencia. Solo mediante pequeños incrementos entenderemos los mecanismos por debajo del alto nivel de cooperación observado en los humanos.

Además, para evaluar la pro-socialidad de los seres humanos, no debemos ignorar la pregunta de quién se beneficia de la cooperación. En este sentido, cooperar con emparentados [82] y cooperar con pares [87] son mecanismos propensos a mantener la corrupción en las sociedades modernas, ya que pueden conducir al nepotismo y al amiguismo, respectivamente. De hecho, estructuras de parentesco estrechas se correlacionan con altos índices de corrupción, mientras que prácticas no familiares se asocian con niveles más altos de cooperación [339]. Por lo tanto, propuestas para socavar prácticas corruptas deben considerar las motivaciones subyacentes están por detrás. Por ejemplo, si bien el castigo hacia los desertores aumenta el nivel general de cooperación [47], aumentar la visibilidad de las prácticas corruptas puede ser perjudicial para la cooperación, lo que contradice la idea de que la transparencia funcionará como un castigo para los corruptores [340]. En general, tratar la corrupción como deserción es un error, y es más apropiado imaginarlo como lo que es: el lado negativo de la cooperación. Solo así las políticas públicas pueden funcionar adecuadamente para asegurar una cooperación imparcial y generalizada.

En última instancia, no existe una ley general que se aplique a todos los contextos y a todas las poblaciones. Por un lado, la competencia puede ser útil y hacer que el comercio sea más eficiente, como vimos en el Capítulo 6. Por otro lado, competir por los recursos podría agotar nuestros recursos naturales, como vimos en el Capítulo 4, y las soluciones para esto deben depender de la cooperación. Además, los seres humanos varían no solo entre culturas sino también dentro de cada país, y el comportamiento cooperativo dependerá en gran medida del contexto y la ecología local [341]. Como hemos mostrado aquí, las personas de diferentes países (Capítulo 4), género y antecedentes (Capítulo 5) varían significativamente en su comportamiento. Por lo tanto, se necesitan análisis más amplios de diferentes escenarios, así como también investigaciones exhaustivas sobre los mecanismos causales en cada contexto específico.

Sin embargo, sin la debida consideración, el enfoque en la adquisición de datos [342] puede intensificar la ya amplia colección de resultados no cohesivos en las ciencias sociales, algunos de ellos no siendo reproducibles [343]. En consecuencia, es obligatorio y urgente tanto mejorar las prácticas experimentales como el intercambio de metodologías estandarizadas entre investigadores. Aún así, aumentar únicamente nuestra disponibilidad de datos y mejorar nuestra metodología empírica no será suficiente para resolver los problemas actuales. Como hemos visto en la Sección 2.2.1.1, es imposible desentrañar todos los efectos experimentales, lo que impide una evaluación adecuada de la validez de los resultados. En este sentido, tener bases sólidas para evaluar la probabilidad de hallazgos empíricos es una necesidad apremiante, que solo será posible si ideamos una teoría global del comportamiento humano [344].

Esperamos que esta tesis proporcione la base y motivaciones para estas búsquedas y; esperemos que en breve podamos escudriñar mejor el comportamiento humano en todas sus variedades. De esta forma, las probabilidades de construir sociedades más prosociales y superar con éxito los desafíos inminentes deberían estar a nuestro favor.

Part IV

APPENDIX

a

A COMMON POOL OF DYNAMIC RESOURCES

A.1 BEHAVIOURAL MODEL DETAILS

To estimate the model parameters appearing in Eq. (4.6), i.e., β_R, β_T^{-s}, $\beta_{\langle T \rangle}^{-s}$, and β_i, as well as residual variances σ_i^2, we resorted to an appropriate variant of the maximum likelihood method. Specifically, the log-likelihood function to maximize was

$$\begin{aligned}\log L_{NT}\left(\boldsymbol{\beta}, \boldsymbol{\sigma}^2\right) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\log l_{NT}\left(\boldsymbol{\beta}, \boldsymbol{\sigma}^2\right) = \\ &= -\frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\log\left(2\pi\sigma_i^2\right) + \frac{\epsilon_i^2(t)}{\sigma_i^2}\right],\end{aligned} \tag{a.1}$$

where $\boldsymbol{\beta} = \left(\beta_R, \beta_T^{-s}, \beta_{\langle T \rangle}^{-s}, \beta_1, \ldots, \beta_N\right)$ and $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \ldots, \sigma_N^2\right)$ are symbolic notation for all model parameters and participant-specific residual variances, respectively, N is the number of players in a given country, and T is the number of rounds played. This type of log-likelihood function arises naturally when working with panel data [345], i.e., observations of multiple individuals over a period of time, as was the case herein. We assumed that the residual variance was constant for each player, but that players differed from one another, thus giving rise to a continuous spectrum of possible values. An additional assumption implied by log-likelihood function (a.1) was that residual covariances have the form $\mathrm{E}\epsilon_i\epsilon_j = \delta_{ij}\sigma_i\sigma_j$, where δ_{ij} is the Kronecker delta, i.e., $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. Similarly as with autocorrelations, this assumption was reasonable given that the lagged average efforts of others in Eq. (4.6) should account for potential cross-correlations in player decisions.

To estimate parameter values $\hat{\boldsymbol{\beta}}$ that maximize log-likelihood function (a.1), we used a generalized least squares estimator described by Hayashi [346], but only after casting the regression problem into a suitable form (see *Regression analysis* below). This estimator is implicit in the sense that residual variances to be estimated, $\hat{\boldsymbol{\sigma}}^2$, appear on the right side of the estimation equation. We therefore implemented an iterative numerical algorithm proposed by Amemiya [347]. Furthermore, to calculate robust confidence intervals, we approximated the covariance matrix of pair $\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2\right)$ with the sandwich estimator [348]. For the purpose of regression diagnostics, we confirmed the validity of these confidence intervals with a bootstrap procedure, as well as tested parameter estimability under model misspecification (see Appendix section a.1.2).

Among goodness-of-fit measures, the coefficient of determination R^2 is ubiquitous and offers intuitive appeal. We relied on a generalized definition of R^2 due to Nagelkerke [349]

$$R^2 = 1 - \exp\left[-\frac{2}{NT}\left(\log L_{NT}\left(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\sigma}}^2\right) - \log L_{NT}\left(\hat{\boldsymbol{\beta}}_\mathbf{0}, \hat{\boldsymbol{\sigma}}_\mathbf{0}^2\right)\right)\right], \tag{a.2}$$

where $\log L_{NT}\left(\hat{\boldsymbol{\beta}}_\mathbf{0}, \hat{\boldsymbol{\sigma}}_\mathbf{0}^2\right)$ is the log-likelihood of the null model for which $\hat{\boldsymbol{\beta}}_\mathbf{0} = (0, 0, 0, \hat{\beta}_0, \ldots, \hat{\beta}_0)$, while $\hat{\boldsymbol{\sigma}}_\mathbf{0}^2$ are the corresponding participant-specific residual variances. The said intuitive appeal of this definition stems from the fact that (i) $R^2 > 0$ for any model with free parameters that fits the data better than the null model, (ii) $R^2 = 1$ only if the fit is perfect, and (iii) R^2 is the proportion of explained variance in the data [349]. Furthermore, Eq. (a.2) explicitly incorporates the ratio of likelihood functions, thus showing that R^2 is closely related to the likelihood ratio test for the significance of regression [350], which we exploited to demonstrate that our results are indeed significant (see Appendix section a.1.2).

Here, we detail the elements of the regression analysis needed to estimate the model's parameters. Specifically, we describe how to (i) cast the regression problem into the most convenient form, (ii) numerically calculate the parameter values, and (iii) find robust confidence intervals using the sandwich estimator.

Let $\{i \in \mathbb{N} : 1 \leq i \leq N\}$ be an index set for players and $\{t \in \mathbb{N} : 1 \leq t \leq T\}$ an index set for time. We collected dependent variables $y_{it}\ (= T_{it})$, each of which represents the ith player's effort at moment t, in a T-vector of observations $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})^{\mathrm{tr}}$, where "tr" denotes the transpose operation, i.e., the ith player is represented by column-vector $\mathbf{y}_i$ containing T observations. Similarly, we collected the predictor variables at moment t in vector $\mathbf{X}_{it} = (R_{it}, T_{i,t-1}, T_{i,t-2}, \ldots, \langle T_{i,t-1}\rangle, 1)^{\mathrm{tr}}$, where R_{it} is the virtual forest's state, $T_{i,t-1}$, $T_{i,t-2}$, etc. are the past efforts, $\langle T_{i,t-1}\rangle$ is the past average effort of players excluding the ith player, and 1 is included to capture the fixed effects. We subsequently created a matrix of observations for the ith player with T rows $\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT})^{\mathrm{tr}}$. Finally, we set the vector of parameters to $\boldsymbol{\beta}_i = \left(\beta_{i1}, \beta_{i2}, \ldots, \beta_{ip}, \beta_i\right)^{\mathrm{tr}}$. With these definitions, the regression equation for the ith player became $\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i$. We assumed that residual vectors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT})^{\mathrm{tr}}$ are normally distributed with covariance matrices $V\boldsymbol{\epsilon}_i = E\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i^{\mathrm{tr}} = \sigma_i^2\mathbf{I}$. Here, $\mathbf{I}$ is the $T \times T$ identity matrix. We had no basis on which to assume constancy of the variance across different players; it generally holds that $\sigma_i \neq \sigma_j$. Defining $\mathbf{y} = (\mathbf{y}_1^{\mathrm{tr}}, \ldots, \mathbf{y}_N^{\mathrm{tr}})^{\mathrm{tr}}$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathrm{tr}}, \ldots, \boldsymbol{\beta}_N^{\mathrm{tr}})^{\mathrm{tr}}$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1^{\mathrm{tr}}, \ldots, \boldsymbol{\epsilon}_N^{\mathrm{tr}})^{\mathrm{tr}}$, and $\mathbf{X} = \mathbf{X}_1 \oplus \ldots \oplus \mathbf{X}_N$, the regression equations for all players turned into $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Here, $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are column vectors with TN entries, $\boldsymbol{\beta}$ is a column vector with $N(p+1)$ entries, and matrix $\mathbf{X}$ is a direct sum of $\mathbf{X}_1, \ldots, \mathbf{X}_N$, i.e., $\mathbf{X}$ is a diagonal block matrix with the first block being $\mathbf{X}_1$, the second block being $\mathbf{X}_2$, etc. The covariance matrix of $\boldsymbol{\epsilon}$ is given by the Kronecker product, $V\boldsymbol{\epsilon} = E\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{tr}} = \boldsymbol{\sigma}^2 \otimes \mathbf{I}$, where $\boldsymbol{\sigma}^2 = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_N^2\right)$, i.e., $\boldsymbol{\epsilon}$ is a diagonal matrix in which entries σ_1^2, σ_2^2, etc. repeat T times each. In general, the regression model we have just described belongs to a class of seemingly unrelated regression (SUR) models [347].

We appended the SUR model with constraints

$$
\begin{aligned}
\beta_{11} = \ldots = \beta_{i1} = \ldots = \beta_{N1} &= \beta_R, \\
\beta_{12} = \ldots = \beta_{i2} = \ldots = \beta_{N2} &= \beta_T^{-1}, \\
\beta_{13} = \ldots = \beta_{i3} = \ldots = \beta_{N3} &= \beta_T^{-2}, \\
&\vdots \\
\beta_{1p} = \ldots = \beta_{ip} = \ldots = \beta_{Np} &= \beta_{\langle T \rangle}^{-1},
\end{aligned}
$$

where β_R, β_T^{-1}, β_T^{-2}, etc. and $\beta_{\langle T \rangle}^{-1}$ are constant parameters as defined in Eq. (4.6). Although somewhat counter-intuitive at first, we cast these constraints in matrix form $\mathbf{Q}^{\mathrm{tr}}\boldsymbol{\beta} = \mathbf{0}$ to make the implementation easier. Matrix $\mathbf{Q}$ is an $N(p+1) \times p(N-1)$ matrix. One way to set up this matrix was for the first column to indicate that the first parameter for player 1 equals the first parameter for player 2 ($\beta_{11} = \beta_{21}$), the second column to indicate that the first parameter for player 2 equals the first parameter for player 3 ($\beta_{21} = \beta_{31}$), and so on until column $N-1$ ($\beta_{N-1,1} = \beta_{N1}$). Columns N to $2(N-1)$ did the same for the second parameter (β_{*2}), and so on until columns $(p-1)N$ to $p(N-1)$ for the pth parameter (β_{*p}). To finally implement the described constraints we had to define another $N(p+1) \times (p+N)$ matrix $\mathbf{R}$ such that the block-matrix $\mathbf{A} = [\mathbf{QR}]^{\mathrm{tr}}$ is non-singular and $\mathbf{R}^{\mathrm{tr}}\mathbf{Q} = \mathbf{0}$. Putting $\gamma = \mathbf{A}\boldsymbol{\beta}$, the SUR regression model turned into $\mathbf{y} = \mathbf{X}\mathbf{A}^{-1}\gamma + \boldsymbol{\epsilon}$. In this equation, the parameter vector γ by definition consists of two parts, $\gamma_1 = \mathbf{Q}^{\mathrm{tr}}\boldsymbol{\beta} = \mathbf{0}$ and $\gamma_2 = \mathbf{R}^{\mathrm{tr}}\boldsymbol{\beta}$. Only the latter part, which is an $(p+N)$-vector, remained unspecified. The number of entries in γ_2 reflected the fact that our model had p parameters and N fixed effects, i.e., one fixed effect per player. Using property $\mathbf{A}^{-1} = [\mathbf{Q}(\mathbf{Q}^{\mathrm{tr}}\mathbf{Q})^{-1}\mathbf{R}(\mathbf{R}^{\mathrm{tr}}\mathbf{R})^{-1}]$, we made additional transformation $\mathbf{X}' = \mathbf{XR}(\mathbf{R}^{\mathrm{tr}}\mathbf{R})^{-1}$ to ultimately reformulate the initial regression problem as $\mathbf{y} = \mathbf{X}'\gamma_2 + \boldsymbol{\epsilon}$. The last equation was no longer a SUR model, because matrix $\mathbf{X}'$ could not be expressed as a direct sum of other matrices. Revisiting the definitions of models based on covariance matrices revealed that we were facing a heteroscedastic model with a constant variance within subsets of the sample [347].

A log-likelihood function corresponding to the described regression problem is specified in Eq. (a.1). To maximize this log-likelihood function, we first defined a generalized least squares estimator [346] in the form

$$\hat{\gamma}_2^{\mathrm{GLS}} = \left[\mathbf{X}'^{\,\mathrm{tr}}(V\boldsymbol{\epsilon})^{-1}\mathbf{X}'\right]^{-1}\mathbf{X}'^{\,\mathrm{tr}}(V\boldsymbol{\epsilon})^{-1}\mathbf{y}. \tag{a.3}$$

This expression, unlike the expression for the more common ordinary least squares estimator, is not an explicit equation whose evaluation immediately results in parameter estimates. Instead, covariance matrix $V\boldsymbol{\epsilon}$ is unknown and needs to be estimated from data alongside parameter vector γ_2. To this end, we employed an iterative numerical algorithm as follows [347]. The algorithm is initialized with the parameter estimates obtained from the ordinary least squares estimators, $\hat{\gamma}_2^{\mathrm{OLS}} = [\mathbf{X}'^{\,\mathrm{tr}}\mathbf{X}']^{-1}\mathbf{X}'^{\,\mathrm{tr}}\mathbf{y}$ and $\hat{\sigma}_i^2 = \hat{\boldsymbol{\epsilon}}_i^{\mathrm{tr}}\hat{\boldsymbol{\epsilon}}_i$. Here, the residual vector estimator is $\hat{\boldsymbol{\epsilon}}_i = \mathbf{y}_i - \mathbf{X}_i'\hat{\gamma}_{2,i}^{\mathrm{OLS}}$. Vector $\mathbf{y}_i$ (resp., $\hat{\gamma}_{2,i}^{\mathrm{OLS}}$) contains the elements of $\mathbf{y}$ (resp., $\hat{\gamma}_2^{\mathrm{OLS}}$) that correspond to the ith player, while similarly matrix $\mathbf{X}_i'$ contains the rows of $\mathbf{X}'$ that correspond to the ith player. Estimates $\hat{\sigma}_i^2$ are then used to construct the covariance matrix according to $V\epsilon = \sigma^2 \otimes \mathbf{I}$, which inserted into Eq. (a.3) yields the first iteration value of $\hat{\gamma}_2^{\text{GLS}}$. Subsequent iterations differ from the first one only in that $\hat{\epsilon}_i = \mathbf{y}_i - \mathbf{X}_i' \hat{\gamma}_{2,i}^{\text{GLS}}$, where $\hat{\gamma}_{2,i}^{\text{GLS}}$ follows from a preceding iteration. The algorithm stops when all the elements of $\hat{\gamma}_2^{\text{GLS}}$ change less than desired precision δ between two consecutive iterations. We used $\delta = 10^{-12}$ throughout.

Once we had estimated the parameter values, the first among the remaining tasks was to estimate the corresponding 95% confidence intervals. We relied on a theorem of the maximum likelihood theory stating that estimator $\hat{\gamma}_2$ asymptotically has a normal distribution with mean γ_2 and covariance matrix $V\hat{\gamma}_2$, the elements of which can be approximated by the sandwich estimator, $\mathbf{S}_{NT}$, defined as follows [348, 351]. Let $\boldsymbol{\theta} = \left(\gamma_2^{\text{tr}}, (\sigma^2)^{\text{tr}}\right)^{\text{tr}}$, then

$$\mathbf{S}_{NT}\left(\hat{\boldsymbol{\theta}}\right) = \mathbf{A}_{NT}^{-1}\left(\hat{\boldsymbol{\theta}}\right) \mathbf{B}_{NT}\left(\hat{\boldsymbol{\theta}}\right) \mathbf{A}_{NT}^{-1}\left(\hat{\boldsymbol{\theta}}\right),$$

where matrices $\mathbf{A}_{NT}\left(\hat{\boldsymbol{\theta}}\right)$ and $\mathbf{B}_{NT}\left(\hat{\boldsymbol{\theta}}\right)$ are symbolically given by

$$\mathbf{A}_{NT}\left(\hat{\boldsymbol{\theta}}\right) = \left[\sum_{i=1}^{N}\sum_{t=1}^{T} \left.\frac{\partial^2 \log l_{NT}}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\text{tr}}}\right|_{\hat{\boldsymbol{\theta}}}\right],$$

$$\mathbf{B}_{NT}\left(\hat{\boldsymbol{\theta}}\right) = \left[\sum_{i=1}^{N}\sum_{t=1}^{T} \left.\frac{\partial \log l_{NT}}{\partial\boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} \left.\frac{\partial \log l_{NT}}{\partial\boldsymbol{\theta}^{\text{tr}}}\right|_{\hat{\boldsymbol{\theta}}}\right],$$

with l_{NT} being the probability density function for model residuals appearing in Eq. (a.1). Remembering that our main concern was parameter vector $\boldsymbol{\beta}$, its estimator $\hat{\boldsymbol{\beta}}$, and the corresponding covariance matrix $V\hat{\boldsymbol{\beta}}$, we proceeded to find the link between $V\hat{\boldsymbol{\beta}}$ and $V\hat{\gamma}_2$. From the definition of γ_2, it followed that $\hat{\boldsymbol{\beta}} = \mathbf{R}\left(\mathbf{R}^{\text{tr}}\mathbf{R}\right)^{-1}\hat{\gamma}_2$. This equality shows that estimator $\hat{\boldsymbol{\beta}}$ is also asymptotically normal with covariance matrix $V\hat{\boldsymbol{\beta}} = \mathbf{R}\left(\mathbf{R}^{\text{tr}}\mathbf{R}\right)^{-1} V\hat{\gamma}_2 \left(\mathbf{R}^{\text{tr}}\mathbf{R}\right)^{-1}\mathbf{R}^{\text{tr}}$. The diagonal elements of $V\hat{\boldsymbol{\beta}}$ contained information on the standard errors of the estimated parameters. We used this information to calculate the corresponding z-scores and thereafter the 95% parameter confidence intervals.

A.1.1 *Fixed Effects*

With the same set of parameter values applied to all participants from a given country, our model was implicitly designed to cap-

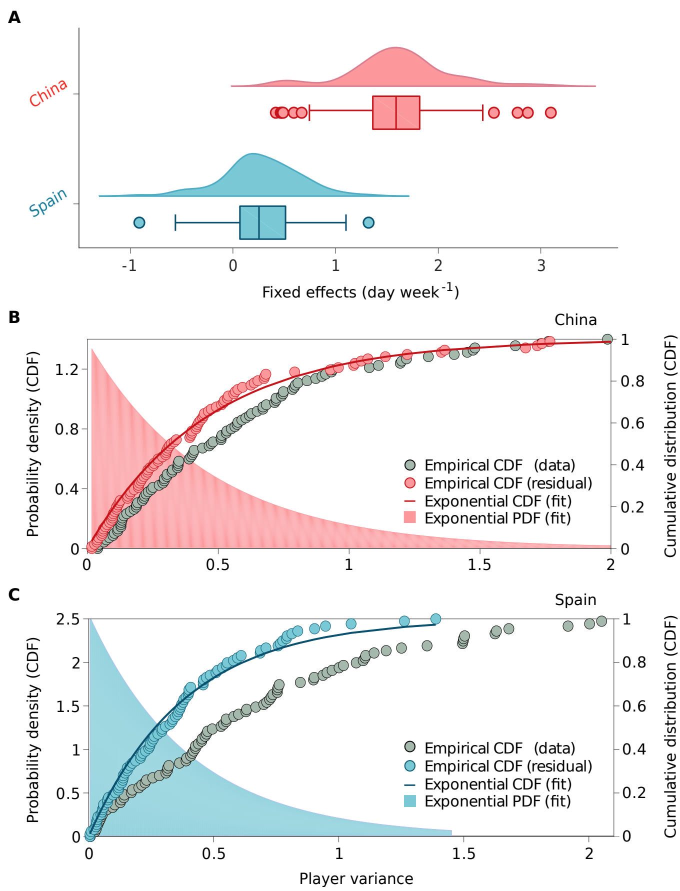


Figure a.1: **Chinese participants manifested stronger individualistic propensities than their Spanish counterparts.** Individual differences entered the model through constant terms called fixed effects and residual variances, both specific to each participant. **A,** Fixed effects have considerably larger absolute values in China than in Spain. This reveals an individualistic propensity to exert effort by the Chinese participants irrespective of the state of the explanatory variables, including the virtual forest's state or efforts posted by others. Here, shown are the kernel-smoothed probability density for fixed effects and the corresponding box plots in which medians, interquartile ranges, limits that would encompass 99.3% of normally distributed data, and outliers are respectively represented by the central vertical line, boxes, whiskers, and circles. **B, C,** In both countries, residual and data variances appear to be drawn from exponential distributions. The means of these distributions for the Chinese participants are 0.4655 (95% CI $[0.3848, 0.5746]$) and 0.5710 (95% CI $[0.4720, 0.7050]$), respectively, indicating a variance reduction of ≈20%. The same quantities for the Spanish participants are 0.3617 (95% CI $[0.2970, 0.4505]$) and 0.6635 (95% CI $[0.5447, 0.8262]$), respectively, indicating a reduction of ≈45%.

ture the collective behavior. Individual differences in the form of propensities to exert and vary effort at random respectively entered the model via participant-specific constant terms called fixed effects and participant-specific residual variances. Fixed effects have considerably larger absolute values among the Chinese participants (Fig. a.1A), thus revealing their stronger propensity to exert effort irrespective of the state of the explanatory variables, including the virtual forest's state or efforts posted by others. This result means that non-zero effort is more likely in China than Spain when the number of trees left for cutting is low. Consequently, six Chinese, but none of the Spanish groups kept depleting the resource until the last round of the game (Fig. 4.3B, C).

A straightforward interpretation of participant-specific residual variances is that they represent individualistic propensities to randomly vary effort. We find that these variances approximately follow the exponential distribution (Fig. a.1B, C), thus reflecting a spectrum of individual behaviors. Some participants stick with previous decisions (smaller residual variance), while others tend to explore all possibilities (larger residual variance). Overall, the Chinese participants were more inclined to randomly vary effort, as evidenced by the mean residual variance of 0.4655 as opposed to 0.3617 for the Spanish participants. Aside from this, there is another interesting way to interpret residual variances.

The total data variance comprises contributions from data variances specific to each participant and from mutual covariances. A model without individualized parameters prioritizes collectiveness and is bound to capture the latter contribution. The former contribution is only captured to the extent that individuals mimic the collective, which we quantified by contrasting participant-specific residual variances with the corresponding data variances. We find that residuals carry an average of 20% and 45% less variance than the corresponding data for the Chinese (Fig. a.1B) and the Spanish (Fig. a.1C) participants, respectively. A greater variance reduction in the latter case indicates that the Spanish participants better mimicked the collective behavior. This result additionally helps to explain the faster resource depletion in Spain than in China (Fig. 4.3B, C). The stronger the collective, the weaker is the individual resistance to a dominant trend.

A.1.2 *Model Robustness*

The maximum likelihood theory makes fairly strong assumptions that are often violated in practice when the theory is applied to statistical regression modeling. To cope with this problem, the theory has been amended with estimators that are robust to "reasonable" assumption violations. One such example is the sandwich estimator [348] used herein to estimate the covariance matrix and ultimately the 95% confidence intervals for the model parameters. However, when the theoretical assumptions are severely violated, even the robust estimators fail. Gauging the severity with which assumptions are violated is possible via statistical tests [348], but these tests are fairly technical and demand rather elaborate calculations. Bootstrapping is a set of statistical techniques that offer workarounds for these difficulties [352] at a computational cost that is almost negligible with modern computing power.

We resorted to the resampling of residuals in order to probe the probability distributions of the parameter estimators arising in the context of our behavioural regression model. By estimating these distributions, it is possible to detect potential biases in parameter estimates, as well as situations in which relying on asymptotic normality leads to the wrong estimates of confidence intervals. Our datasets, for example, consisted of one time series (50 rounds) for each participant (96 and 90 in China and Spain, respectively), in which case both strong autocorrelations within a time series or strong cross-correlations between any two time series would violate the model assumptions.

Because simple random resampling would erase potential autocorrelations or cross-correlations in the data, we performed the moving block bootstrap [353]. This approach to bootstrapping preserves the potential structure in residuals by dividing them into $T - B + 1$ overlapping blocks of length B, where T is the length of the original time series. The first block then covers residuals 1 to B, the second 2 to $B + 1$, and so on. The resampling is performed by randomly drawing T/B blocks with replacement. To obtain a sufficiently detailed picture of the probed distributions, we performed 1000 bootstrap simulations with $B = 5$ and, due to the lagged predictors, $T/B = 9$ instead of 10.

The described bootstrapping procedure confirms the validity of the behavioral regression model (Fig. a.2). We find no evidence of biased parameter estimates. The bootstrap 95% confidence intervals correspond reasonably well to the 95% confidence intervals obtained by means of the sandwich estimator, thus indicating that the model assumptions are not violated for asymptotic normality to become inapplicable.

We additionally tested the statistical significance of the overall model performance. The purpose was to estimate the probability of obtaining a coefficient of determination as large as $R^2 = 0.689$ (in the case of the Spanish data) by pure chance when in fact the null model was true. The null model presumed constant effort supplemented with noise (see the definition of R^2 in the Supporting Methods section above). Even when the null model is true, the behavioral regression model with its multiple degrees of freedom should capture some data variance, but the key question is how much. If the captured data variance as measured by R^2 were on par with the R^2 obtained by fitting the regression model to the original data, then the significance of this regression model would be questionable.

The performance of the behavioral regression model is highly statistically significant (Fig. a.3). As intuitively expected given the coefficient of determination as high as $R^2 = 0.689$ for the Spanish data, we find that the probability of obtaining such a large value is minuscule when the null model is true. It is therefore highly improbable that the results of the behavioral regression model (see Fig. 4.5) are a product of pure chance, thus firmly establishing the statistical significance of the model's performance.

In the context of regression diagnostics, we also performed a numerical experiment to test the employed estimator's robustness to misspecification. Namely, we had no a priori way of knowing what a full set of predictors for our behavioral regression model might look like. It was therefore important to establish that estimating the correct parameter values would still be likely in the absence of a valid or the presence of an invalid predictor. To this end, we created 1000 synthetic datasets using (i) the seven-parameter model as shown in Fig. 4.5 in conjunction with (ii) the resampling of residuals as described above. We then attempted to fit to these synthetic datasets an eight-parameter version of the model containing a spurious lag-six predictor. If the estimator employed in our behavioral regression model was truly robust to misspecification, the most likely outcome of such regression attempts would be that the estimated value of the spurious parameter was close to zero, while the estimated values of other parameters were close to their true values.

Numerical simulation strongly suggest that the behavioral regression model is robust to misspecification (Fig. a.4). The estimate of the spurious parameter is indeed likely to be close to zero, while the estimates of other parameters are likely to be close to their true values (Fig. a.4). There is no evidence of bias. A detail to be aware of is that a non-zero estimate of the spurious parameter seemingly explains some fraction of the data variance, thus reducing the uncertainty in other estimated parameter values. For this reason, it is a good practice to keep in mind the influence of non-significant, non-zero parameters on regression results.

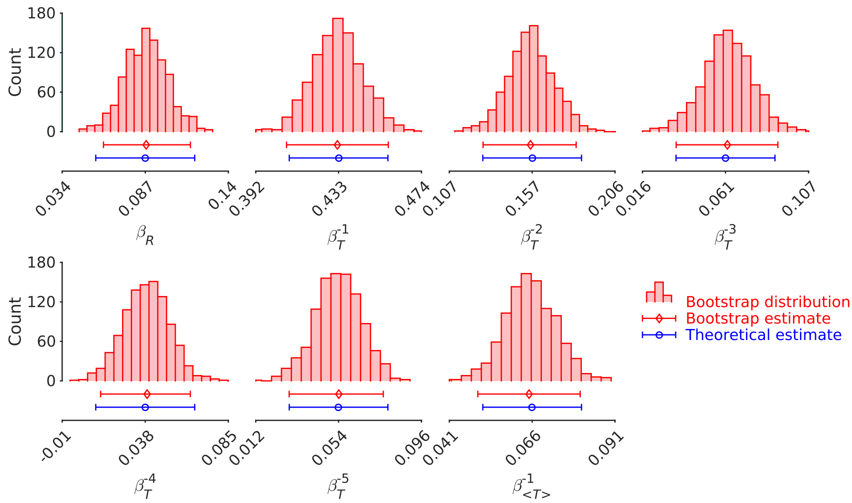


Figure a.2: **Bootstrapping confirms the validity of the behavioral regression model.** Strong assumptions of the maximum likelihood theory are often violated in practice, especially in the context of statistical regression modeling. Although robust estimators circumvent some problems, severe assumption violations may result in biased parameter estimates or overstated parameter significance. An adequate resampling of model residuals, however, allows probing the probability distributions of parameter estimators, thus possibly detecting biases (i.e., a distribution's expectation differs from the true parameter value) or failures of asymptotic normality in constructing confidence intervals (i.e., a distribution's p% interquantile range differs from the p% confidence interval obtained via the covariance matrix). Fortunately, no such problems plague our model as evidenced by a fairly good agreement of bootstrap and theoretical estimates using the Spanish data. Shown are the bootstrap distributions for all parameter estimators (histograms), as well as the corresponding means (red diamonds) and the 95% interquantile ranges (red error bars). Also shown are the maximum likelihood parameter estimates (blue circles) and the corresponding 95% confidence intervals (blue error bars) for an easy comparison.

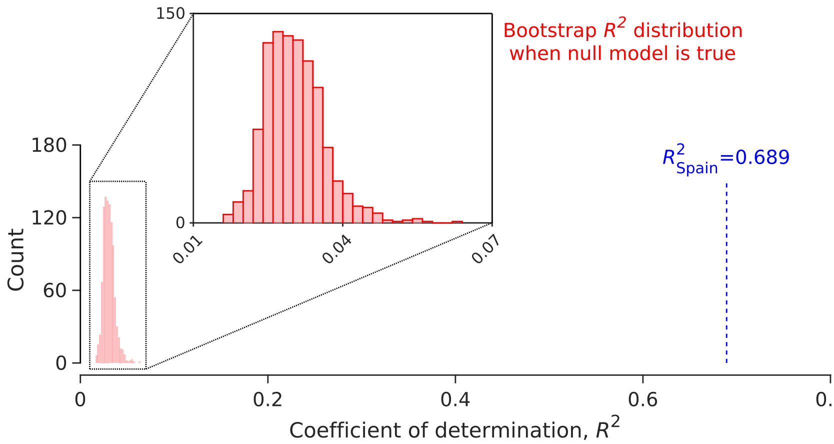


Figure a.3: **Performance of the behavioral regression model is statistically significant.** The histogram represents a bootstrap R^2 distribution for the behavioral regression model when, in fact, the null model (i.e., constant effort with random noise) is true. This distribution shows that the former model with its multiple degrees of freedom explains some data variance even if the data is just noise, yet the performance is poor as intuitively expected. Comparing the 99% quantile of the bootstrap R^2 distribution, $R^2_{99\%}$=0.049, to the value of $R^2 = 0.689$ obtained with the behavioral regression model fitted to the Spanish data, indicates that the model's performance is highly significant. Put alternatively, there would be a negligible probability of obtaining $R^2 = 0.689$ if the null model were indeed true.

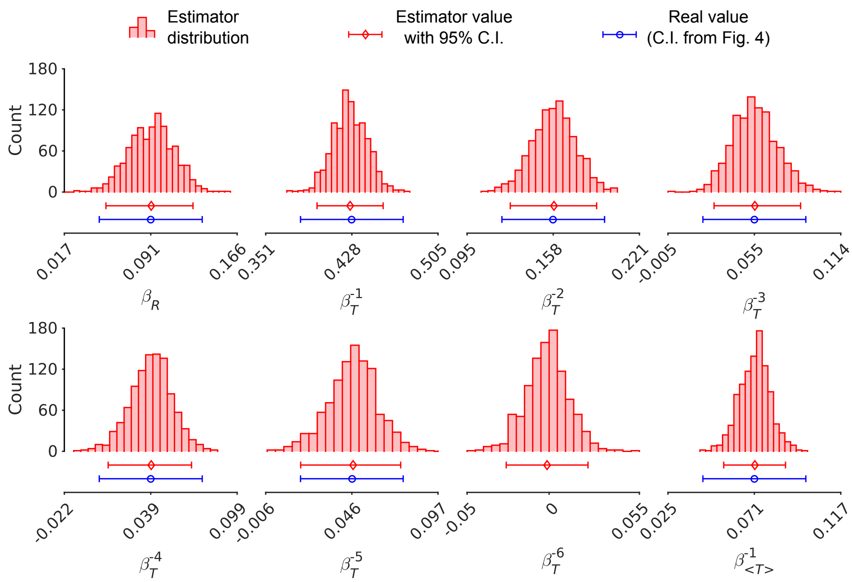


Figure a.4: **Behavioral regression model is robust to misspecification.** Using the same bootstrapping algorithm as in Fig. a.2, we generated 1000 synthetic datasets to which we fitted a version of the behavioral regression model with a spurious parameter, β_T^{-6}. The zero value of this parameter is, on average, correctly identified, as are the true values of other parameters. Because the estimator of the spurious parameter is not identically equal to zero, the corresponding predictor seemingly "explains" some of the data variance, thus reducing the uncertainty in other parameters, as evidenced by the upper (red) error bars that are narrower than the lower (blue) ones.

b

EFFECT OF NETWORK TOPOLOGY AND NODE CENTRALITY ON TRADING

Here, we present additional results aimed to support the findings shown in Chapter 6.

B.1 50 NODES RANDOM NETWORK

The robustness of the findings shown in Chapter 6 against the size and connectivity of the network are explored through an additional experimental session in a random network of 50-nodes with $\langle k \rangle = 4$. At variance with the smaller networks, the selection of the randomly source-destination pairs is generated such that the shortest path between the two nodes is of distance at least diameter - 1. Table b.1 puts together the results corresponding to the random network with 50 nodes and those corresponding to 26-nodes networks. Although the larger network shows lower costs than the smaller ones (actually, for the 50-nodes network intermediation rents are close to zero), both the correlation between payoffs and SD-betweenness (Fig. b.1) and the behavioral rule (Fig. b.2) are verified in the network of 50 nodes, as will be discussed in the next paragraphs.

Fig. 6.3 of Chapter 6 shows the prices, payoffs and frequencies on the cheapest path for all realizations. This result indicates a correlation between profits and the source-destination betweenness (sd_∞). To ensure that this result is not spurious, Fig. b.1 shows how those variables correlate with the centrality in each network. It is clear that the same behavior is maintained when we look at the results for each network.

As an extension of Fig. 6.4 of Chapter 6, Fig. b.2A shows the mean change in price for the cases where the participant was or was not on the cheapest path in the previous round, while Fig. b.2B shows the probabilities to increase and to decrease the posted price conditioned to have been (Y) or not (N) on the cheapest path. In

network	efficiency	price	price in CP	cost	profit	length
R 50	1	7.85	1.18	5.67	0.12	5.78
R 26	0.97	11.34	5.49	28.33	1.10	6.26
SW 26	0.68	18.10	13.16	76.52	2.38	7

Table b.1: **Experimental results.** Efficiency (fraction of rounds in which the cheapest path cost was equal to or less than the threshold), and mean values of the price, price on the cheapest path, cost of the cheapest path, profit, and cheapest path length for each one of the three studied networks: random networks with 50 and 26 nodes (R 50, R 26) and small-world network with 26 nodes (SW 26).

this figure, the results corresponding to the random network with 50 nodes have been added showing that, within the limitations of the current experiment, the behavioral rule according to which players increase their price if they were on the cheapest path in the previous round and decrease it otherwise is robust against the size and connectivity of the network.

b.2 ADDITIONAL EXPERIMENTAL RESULTS: EVOLUTION OF COSTS AND PRICES

Fig. b.3A shows the evolution of the cheapest path cost for each network. As described in Chapter 6, the costs are significantly higher in the Small-World network than in the Random Network. Among the random networks considered, the costs are lower in the larger network (50 nodes) than in the smaller one (26 nodes). These higher costs in the SW network entail a lower efficiency, as can be seen in table b.1. To deepen this issue, Fig. b.3B represents the evolution of the mean price of participants on the cheapest path for each network. Note that the pattern is consistent with Fig. b.3A, pointing out that the previous finding was not a consequence of different cheapest path lengths. To complement these results, Fig. b.3C and b.3D show, respectively, the evolution of the mean and median price posted by all the participants. As has been discussed in Chapter 6, the effect of the network topology on prices (Fig. b.3C, b.3D) is not as clear as it is on costs and payoffs (Fig. b.3A, b.3B)

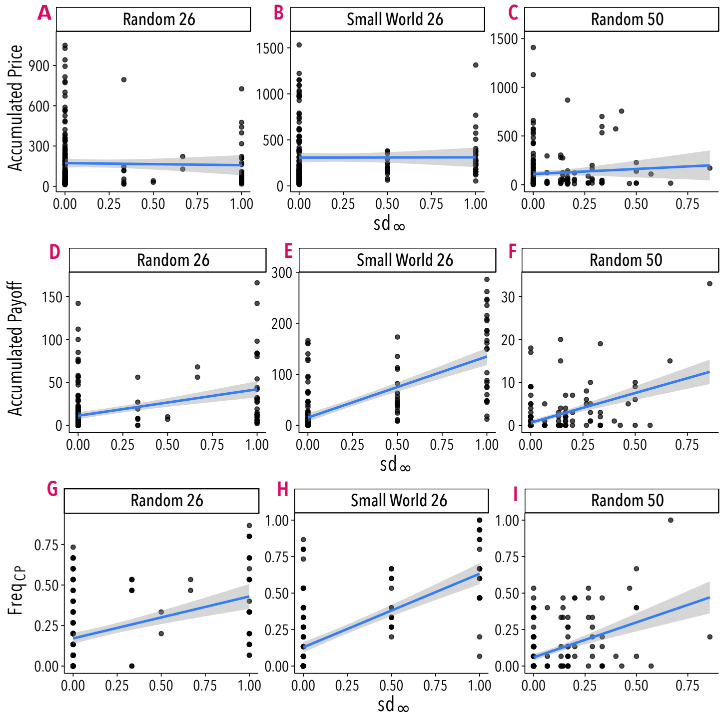


Figure B.1: **SD-betweenness determines payoffs but not posted prices. A-I**: Accumulated price (**A-C**), accumulated payoff (**D-F**) and frequency on the cheapest path (**G, I**) of participants during a series of 15 rounds as a function of the SD-betweenness sd_∞ for the Random 50 (**A,D,G**), Small World 26 (**B,E,H**), and Random 50 networks(**C,F,I**).

since the topology mainly affects the probability for the nodes to be on the cheapest path more than the posted prices.

B.3 ADDITIONAL MODEL RESULTS: FINAL COSTS AND NETWORK METRICS. ROBUSTNESS.

B.3.1 *Node-disjoint paths*

In the game, nodes in a path between the source and the destination constitute a group, and all such groups can be seen as competing to be the cheapest path. At the same time, individuals within those groups try to maximize their own profit, and as such are competing

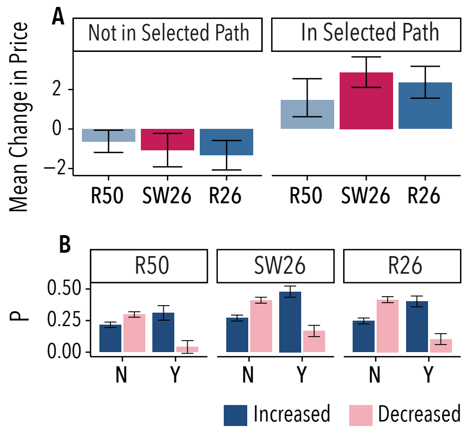


Figure b.2: **Extension of Figure 6.4 of Chapter 6 with the 50-nodes random network. A:** Mean changes in the posted price for the studied networks: random network of 50 nodes (R50) and 26 nodes (R26), and small-world network of 26 nodes (SW26). The panel discriminates the cases in which the participant was (right) from those in which she was not (left) on the selected cheapest path in the previous round. **B:** Probability to increase (blue) and to decrease (pink) the posted price conditioned to have been (Y) or not (N) on the selected cheapest path, for each one of the studied networks. The error bars represent the 95% C.I.

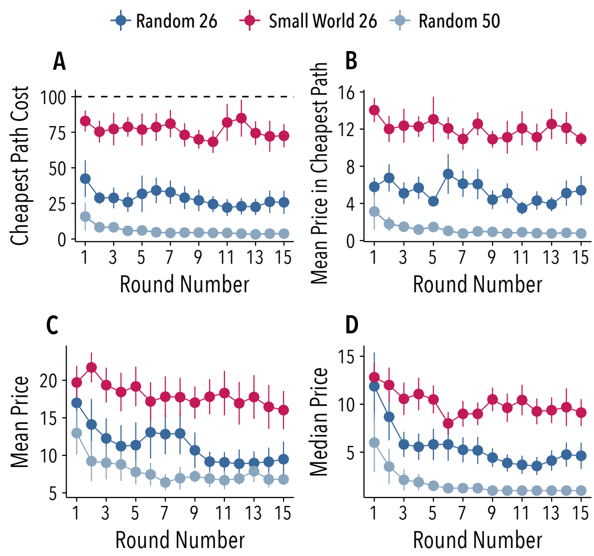


Figure b.3: **Evolution of costs and prices for each experimental network.** Cheapest path cost (**A**), mean price of nodes on the cheapest path (**B**), mean price (**C**), and median price (**D**) as a function of the round number. Each series of points corresponds to a given network: Random Network with 50 nodes (dark blue), Small-World with 26 nodes (magenta), and Random Network with 26 nodes (light blue). The error bars represent $1.96 \times$ SEM.

with other group members. Therefore, with a large number of paths there will be a large number of groups competing to be the cheapest path, with one caveat: paths do not constitute disjoint sets of nodes, which implies that nodes compete concurrently in a large number of groups. These conditions make the assessment of a good descriptor of the competition in a network a non-trivial task. Nonetheless, we can provide a lower bound estimator based on the minimum number of independent groups that can compete with each other using the number of node-disjoint paths[1].

1 There is also an important practical fact to take into consideration: metrics based on counting all paths are unfeasible for relatively large and dense networks, as the number of paths grows exponentially. Fortunately, computing the maximum number of independent paths is reducible to the maximum flow problem, thus it can be computed in polynomial time.

The results shown in Fig. 6.7 of the Chapter 6 show that, when considering the behavioral rule found, the number M of node-disjoint paths is a particularly good indicator of the final trading costs. With more independent paths, more groups of different nodes can coordinate resulting in a cheaper cost. To illustrate the mechanism behind this relationship, we propose the following simple scenario wherein a formal relationship between the two variables can be demonstrated: i) all the node-disjoint paths are shortest paths and ii) all the nodes start with the same price. For this kind of scenario, we can see a clear relation between costs and M, as shown by Lemma 1. Indubitably, this lemma does not provide us a rule for how costs will change in all the possible networks. Nonetheless, it provides a useful insight into how the competition between paths for the cheapest path should be related to M.

Lemma 1. *Let us consider a graph G, a source S, a destination D, and identical initial prices across all nodes. Let us assume that nodes increase their posted prices by σ if they were located in the previous round in the cheapest path, otherwise decrease it by ρ. If there are M node-disjoint shortest paths of the same length between S and D, the cheapest path cost will increase indefinitely if and only if*

$$\frac{\sigma}{\rho} > (M-1)$$

Proof. Let us consider an enumeration of the disjoint paths from S to D: $p_1, p_2, \ldots, p_M$. Let x be the initial posted price for all the nodes. Initially, a path (without loss of generality, p_1) will be selected, and nodes belonging to path p_1 will increase their price from x to $x+\sigma$, while nodes belonging to the rest of paths $(p_2, p_3, \ldots, p_M)$ will decrease their price to $x-\rho$. In the subsequent steps, the rest of the paths $(p_2, p_3, \ldots, p_M)$ will be selected until all the M paths will have been selected. At step M, the selected path will have a cost per node of $x-(M-1)\rho$, thus, its nodes will increase their price to $x+\sigma-(M-1)\rho$, which will be higher than x only if $\sigma > (M-1)\rho$. Note that, by the same reasoning, at step $M+1$ all the nodes located in $p_1, \ldots, p_M$ will have a cost of $x+\sigma-(M-1)\rho$. Therefore, the cost will increase indefinitely if and only if $\sigma > (M-1)\rho$. □

B.3.2 *Average Path Length and Clustering Coefficient*

Fig. 6.7 from Chapter 6 shows how trade costs scale with the average path length of the network. This result is not a consequence of cheapest paths length differences: the mean price of nodes on the cheapest path also correlates with the average path length, as shown in Fig. b.4.

It is well known that small world networks differ from random networks with respect to clustering and average path length in different ways. Therefore, the clustering coefficient is also a natural candidate for capturing the differences in the cheapest path cost, as it is indeed the case as shown in Fig. b.5. To check which of both observables is more connected with the network properties driving the differences in cost, we executed two linear regressions with C_i, the final trade cost (i.e., on the cheapest path), as the dependent variable: one having the clustering coefficient as independent variables and one with the average path length as independent variables. The data considered are the results from simulations executed without the threshold, which are shown in Fig. 6.7 of Chapter 6. As the slopes of both properties change with respect to M, we added an individual coefficient for each value of M obtained by multiplying the network property by a Kronecker delta (δ) dummy variable. The regression model with clustering coefficient (CC) is shown in Eq. b.1, and that with average path length ($\langle d \rangle$) in Eq. b.2. We restricted the data to $M \leq 3$, as for larger values final costs are mostly zero.

$$C_i = \beta_1 CC_i \delta_{M_i 1} + \beta_2 CC_i \delta_{M_i 2} + \beta_3 CC_i \delta_{M_i 3} \tag{b.1}$$

$$C_i = \beta_1 \langle d \rangle_i \delta_{M_i 1} + \beta_2 \langle d \rangle_i \delta_{M_i 2} + \beta_3 \langle d \rangle_i \delta_{M_i 3} \tag{b.2}$$

The results from the regression are displayed in Table b.2 which shows that the model with average path length as regressors provides a better description of the final trade costs: the coefficient of determination when considering the average path length is $R^2(\langle d \rangle) = 0.79$, while the model with the clustering coefficient achieves $R^2(CC) = 0.57$.

	Model with CC	Model with $\langle d \rangle$
CC for $M = 1$	0.74^{***}	
	(0.00)	
CC for $M = 2$	0.04^{***}	
	(0.00)	
CC for $M = 3$	-0.09^{***}	
	(0.00)	
$\langle d \rangle$ for $M = 1$		0.97^{***}
		(0.00)
$\langle d \rangle$ for $M = 2$		0.20^{***}
		(0.00)
$\langle d \rangle$ for $M = 3$		0.07^{***}
		(0.00)
R^2	0.57	0.79
Adj. R^2	0.57	0.79
Num. obs.	131148	131148
RMSE	0.66	0.46

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table b.2: **Coefficients of the statistical models.** Coefficients of the two statistical models, having the clustering coefficient (CC) as the independent variable (*left*), and having the average path length ($\langle d \rangle$) as the independent variable (*right*). The coefficients are standardized (centered and divided by their standard deviation). As the coefficients of clustering and average path length seem to vary with M, we consider it as a dummy variable. Slopes (clustering coefficient and average path length) are M-specific.

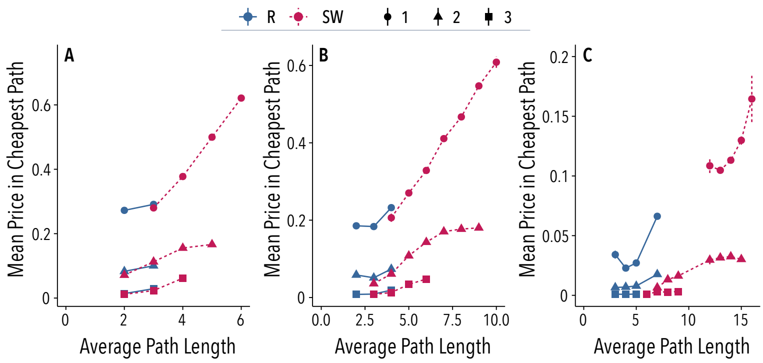


Figure b.4: **Numerical results of the model.** Average price of nodes on the cheapest path after a period of 10^4 rounds as a function of the mean path length of the network. Each panel corresponds to a different network size: 26 (**A**), 50 (**B**), and 1000 (**C**) nodes. Different colors correspond to different network models: random (blue) and small-world (magenta). Different simbols correspond to different values of the number M of disjoint paths: $M = 1$ (circles), 2 (triangles), and 3 (squares). For each configuration, there were generated 10000 networks of each size according to the Watts-Strogatz algorithm [59] with $p = 0.1$ (SW) and $p = 1$ (R), and average degree from 2 to 10. The increment/decrement ratio was fixed to the experimental value ($\sigma/\rho = 2.4$).

B.3.3 *Robustness of the model against the choice of initial conditions*

To validate the proposed model, we executed it by bootstrapping the initial prices from those obtained from the experiments. The results, shown in Fig. 6.6 of Chapter 6, indicate that costs and mean prices in the cheapest path are higher in small-world networks than in random networks, in agreement with the experimental results.

Furthermore, to study the effect of initial prices on the model, we have executed the algorithm with random initial prices, disregarding the data from experiments and then making the model self-consistent. To this end, we have taken the initial prices according to a Poissonian distribution ($\lambda = 10$). The results for the Poissonian distribution for the initial prices are displayed in Fig. b.6, being compatible with those corresponding to bootstrapping displayed in Fig. 6.6 of Chapter 6, therefore highlighting the ro-

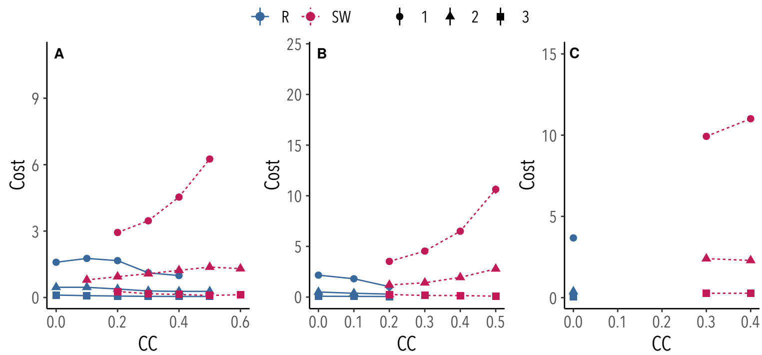


Figure b.5: **Numerical results of the model.** Average cost of the cheapest path after a period of 10^4 rounds as a function of the network clustering coefficient. Each panel corresponds to a different network size: 26 (**A**), 50 (**B**), and 1000 (**C**) nodes. Different colors correspond to different network models: random (blue) and small-world (magenta); while different symbols correspond to different values of the number M of disjoint paths: $M = 1$ (circles), 2 (triangles), and 3 (squares). For each configuration, there were generated 10000 networks of each size, according to the Watts-Strogatz algorithm [59] with $p = 0.1$ (SW), $p = 1$ (R), and average degree from 2 to 10. The increment/decrement ratio was fixed to the experimental value ($\sigma/\rho = 2.4$).

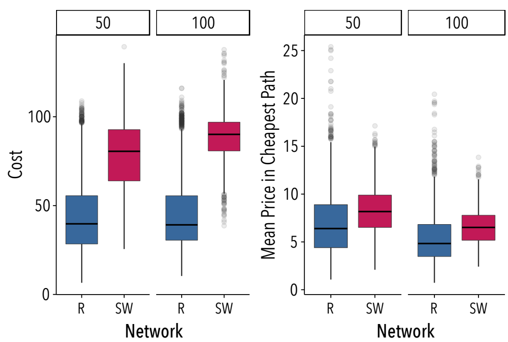


Figure b.6: **Numerical results of the model for networks with 50 and 100 nodes.** Results shown are for 100 executions with 15 rounds for each network and source-destination pair, excluding the first round. Initial prices are taken from a Poissonian distribution ($\lambda = 10$). Other values are $\sigma = 2.60$ and $\rho = 1.2$.

bustness of the model against the choice of initial conditions and validating the effects discussed in Chapter 6 about the network structure influence on prices and costs.

C

FRAMING EFFECTS IN CONTRIBUTIONS AND DONATIONS

C.1 CONDITIONAL CONTRIBUTION

In order to study the response of participants to other players' behaviour, we have performed a dynamic panel data model having as a regressor previous contributions from the other participants in the group, such as specified by the following equation:

$$C_{itf} = \beta_0 + \beta_1 ID_i + \beta_2 FKP_i + \beta_3 ID_i * FKP_i + \beta_4 W_i + \beta_5 O_{i,t-1} + u_{it} \tag{C.1}$$

where $O_{i,t-1}$ corresponds to the total amount contributed by the other players of player i's group at round $t-1$. Thus β_5 measures the conditional contribution of participants, the rest of the terms correspond to the contributions of: ID treatment (β_1), FKP order (β_2), interaction term between the order and treatment (β_3), gender (β_5). To check if men and women have a significantly different conditional response, we can add a regression coefficient $\beta_6 O$ for the interaction between gender (W_i) and $O_{i,t-1}$ such as specified by the equation below:

$$C_{itf} = \beta_0 + \beta_1 ID_i + \beta_2 FKP_i + \beta_3 ID_i * FKP_i + \beta_4 W_i + \beta_5 O_{i,t-1} + \beta_6 O_{i,t-1} * W_i + u_{it} \tag{C.2}$$

The results of the regression are shown in results columns (1) and (2) of Table C.1. They indicate that participants do not condition their contribution with respect to the contribution of others, and there is no evidence that men or women would react differently to this general trend.

In order to uncover a possible influence of framing on conditional contributions to the funds, we have performed a similar analysis to study the individuals' response to other players' contributions to the different funds. To this end, we have introduced in

Equation (c.1) the regression term β_6 that accounts for the interaction between the treatment (ID) and the contributions by the other players of player i's group at round $t-1$:

$$C_{itf} = \beta_0 + \beta_1 ID_i + \beta_2 FKP_i + \beta_3 ID_i * FKP_i + \beta_4 W_i + \beta_5 O_{i,t-1} + \beta_6 O_{i,t-1} * ID_i + u_{it} \tag{c.3}$$

Results displayed in column (3) of Table c.1 indicate that conditional cooperation plays no role in explaining differences in contributions between framing.

	Dependent variable:		
	Total Contribution to the PGG		
	(1)	(2)	(3)
ID	-1.039	-1.075	4.430
	(7.361)	(7.379)	(26.568)
FKP	8.350	8.364	8.049
	(6.707)	(6.701)	(6.564)
Women	12.287^{**}	9.936	12.358^{**}
	(5.232)	(24.860)	(5.271)
$O_{i,t-1}$	0.010	0.007	0.014
	(0.022)	(0.037)	(0.027)
ID x FKP	18.881^{*}	18.928^{*}	20.653^{*}
	(10.086)	(10.097)	(12.093)
Women x $O_{i,t-1}$		0.004	
		(0.037)	
ID x $O_{i,t-1}$			-0.010
			(0.045)
Constant	48.320^{***}	49.725^{**}	45.921^{***}
	(14.466)	(23.815)	(17.429)
Observations	2,223	2,223	2,223
R^2	0.145	0.145	0.145
Adjusted R^2	0.143	0.142	0.143
F Statistic	374.484^{***}	374.748^{***}	375.159^{***}
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Table c.1: **Dynamic panel model regression with cluster robust standard errors at the individual level.** Columns (1), (2), and (3) refer to the regressions described in formulas (c.1), (c.2), and (c.3) respectively.

BIBLIOGRAPHY

[1] J. Henrich. *The Secret of Our Success*. Princeton: Princeton University Press, 2016. ISBN: 9781400873296.

[2] R. Boyd. *A different kind of animal: how culture transformed our species*. Princeton, New Jersey: Princeton University Press, 2018, p. 229. ISBN: 9780691177731.

[3] M. W. Moffett. “Human Identity and the Evolution of Societies.” In: *Human Nature* 24.3 (2013), pp. 219–267. ISSN: 10456767. DOI: 10.1007/s12110-013-9170-3.

[4] M. Tomasello. “The ultra-social animal.” In: *European Journal of Social Psychology* 44.3 (2014), pp. 187–194. ISSN: 00462772. DOI: 10.1002/ejsp.2015.

[5] S. L. Lewis and M. A. Maslin. “Defining the Anthropocene.” In: *Nature* 519.7542 (2015), pp. 171–180. ISSN: 0028-0836. DOI: 10.1038/nature14258.

[6] Core Writing Team. *IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by R. Pachauri and L. Meyer. Geneva, Switzerland: IPCC, 2014, p. 151.

[7] E. C. Ellis. *Anthropocene: A Very Short Introduction*. Oxford: Oxford University Press, 2018, p. 208.

[8] J. A. Tucker, Y. Theocharis, M. E. Roberts, and P. Barberá. “From Liberation to Turmoil: Social Media And Democracy.” In: *Journal of Democracy* 28.4 (2017), pp. 46–59. ISSN: 1086-3214. DOI: 10.1353/jod.2017.0064.

[9] E. O. Wilson. *The Social Conquest of Earth*. Liveright, 2012, p. 352. ISBN: 978-0871403636.

[10] R. Boyd and J. Silk. *How Humans Evolved*. 5th. W. W. Norton & Company, 2009, p. 541.

[11] M. A. Hogg and J. M. Levine. "Cooperation and Competition." In: *Encyclopedia of Group Processes & Intergroup Relations*. Ed. by J. M. Levine and M. A. Hogg. 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc., 2010, pp. 150–160. DOI: 10.4135/9781412972017.n49.

[12] S. Bowles. *The Moral Economy: Why Good Incentives Are No Substitute for Good Citizens*. Yale University Press, 2016. ISBN: 978-0-300-16380-3.

[13] M. B. Brewer. "The Psychology of Prejudice: Ingroup Love and Outgroup Hate?" In: *Journal of Social Issues* 55.3 (1999), pp. 429–444. ISSN: 0022-4537. DOI: 10.1111/0022-4537.00126.

[14] H. Bernhard, U. Fischbacher, and E. Fehr. "Parochial altruism in humans." In: *Nature* 442.7105 (2006), pp. 912–915. ISSN: 0028-0836. DOI: 10.1038/nature04981.

[15] C. F. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton, New Jersey: Princeton University Press, 2003, p. 568.

[16] D. G. Rand, M. A. Nowak, J. H. Fowler, and N. A. Christakis. "Static network structure can stabilize human cooperation." In: *Proceedings of the National Academy of Sciences* 111.48 (2014), pp. 17093–17098. ISSN: 0027-8424. DOI: 10.1073/pnas.1400406111.

[17] C. Gracia-Lazaro, A. Ferrer, G. Ruiz, A. Tarancon, J. A. Cuesta, A. Sanchez, and Y. Moreno. "Heterogeneous networks do not promote cooperation when humans play a Prisoner's Dilemma." In: *Proceedings of the National Academy of Sciences* 109.32 (2012), pp. 12922–12926. ISSN: 0027-8424. DOI: 10.1073/pnas.1206681109.

[18] L. Shi, I. Romić, Y. Ma, Z. Wang, B. Podobnik, H. E. Stanley, P. Holme, and M. Jusup. "Freedom of choice adds value to public goods." In: *Proceedings of the National Academy of Sciences* 117.30 (2020), pp. 17516–17521. ISSN: 0027-8424. DOI: 10.1073/pnas.1921806117.

[19] J. von Neumann, O. Morgenstern, H. W. Kuhn, and A. Rubinstein. "Theory of Games and Economic Behavior." In: *Princeton University, Princeton* (1944), p. 776.

[20] R. B. Myerson. *Game Theory: Analysis of Conflict*. 1st ed. Harvard University Press, 1997, p. 600. ISBN: 0-674-34115-5.

[21] K. Steele and H. O. Stefánsson. "Decision Theory." In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 201. Metaphysics Research Lab, Stanford University, 2016.

[22] Eric Rasmusen. *Games and Information: An Introduction to Game Theory*. 4th. Wiley-Blackwell, 2006, p. 558. ISBN: 978-1-405-13666-2.

[23] A. Rapoport and A. M. Chammah. *Prisoner's dilemma: a study in conflict and cooperation*. Ann Arbor: University of Michigan Press, 1965. ISBN: 0472061658.

[24] K. Sigmund. *The Calculus of selfishness*. Princeton, New Jersey: Princeton University Press, 2010, p. 173. ISBN: 9780691142753. DOI: 10.5860/choice.47-6917.

[25] R. D. Luce and H. Raiffa. *Games and Decisions: Introduction and Critical Survey*. John Wiley & Sons, Inc., 1957, p. 509.

[26] M. A. Nowak. "Five rules for the evolution of cooperation." In: *Science* 314.5805 (2006), pp. 1560–1563. ISSN: 00368075. DOI: 10.1126/science.1133755.

[27] K. Leyton-Brown and Y. Shoham. "Essentials of Game Theory: A Concise Multidisciplinary Introduction." In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 2.1 (2008), pp. 1–88. ISSN: 1939-4608. DOI: 10.2200/S00108ED1V01Y200802AIM003. arXiv: arXiv:1011.1669v3.

[28] J. F. Nash. "Equilibrium points in n-person games." In: *Proceedings of the National Academy of Sciences* 36.1 (1950), pp. 48–49. ISSN: 0027-8424. DOI: 10.1073/pnas.36.1.48.

[29] A. Smith. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Ed. by E. Cannan. University of Chicago Press, 1977. ISBN: 9780226763743.

[30] F. Guala. *The Methodology of Experimental Economics*. New York, NY: Cambridge University Press, 2005, p. 286.

[31] G. Hardin. "The tragedy of the commons." In: *Science* 162.3859 (1968), pp. 1243–1248. ISSN: 00368075. DOI: 10.1126/science.162.3859.1243.

[32] T. Gilovitch, D. Griffin, and D. Kahneman, eds. *Heuristics and biases: the psychology of intuitive judgment*. Cambridge University Press, 2002. ISBN: 9780511808098.

[33] R. H. Thaler. "Behavioral Economics: Past, Present and Future." In: *SSRN Electronic Journal* 20.38 (2016), p. 9. ISSN: 1556-5068. DOI: 10.2139/ssrn.2790606.

[34] J. S. Mill. "On the Definition and Method of Political Economy." In: *The Philosophy of Economics*. Ed. by D. M. Hausman. Cambridge: Cambridge University Press, 2007, pp. 41–58. ISBN: 9780511819025. DOI: 10.1017/CBO9780511819025.003.

[35] D. Kahneman, J. L. Knetsch, and R. H. Thaler. "Fairness and the Assumptions of Economics." In: *The Journal of Business* 59.S4 (1986), S285. ISSN: 0021-9398. DOI: 10.1086/296367.

[36] R. Forsythe, J. L. Horowitz, N. E. Savin, and M. Sefton. "Fairness in simple bargaining experiments." In: *Games and Economic Behavior* 6.3 (1994), pp. 347–369. ISSN: 10902473. DOI: 10.1006/game.1994.1021.

[37] F. Guala and L. Mittone. "Paradigmatic experiments: The Dictator Game." In: *Journal of Socio-Economics* 39.5 (2010), pp. 578–584. ISSN: 10535357. DOI: 10.1016/j.socec.2009.05.007.

[38] E. Fehr and U. Fischbacher. "Why Social Preferences Matter – the Impact of non-Selfish Motives on Competition, Cooperation and Incentives." In: *The Economic Journal* 112.478 (2002), pp. C1–C33. ISSN: 0013-0133. DOI: 10.1111/1468-0297.00027.

[39] I. Gilboa, A. Postlewaite, L. Samuelson, and D. Schmeidler. "What are axiomatizations good for?" In: *Theory and Decision* 86.3-4 (2019), pp. 339–359. ISSN: 0040-5833. DOI: 10.1007/s11238-018-09685-1.

[40] J. Henrich et al. ""Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies." In: *Behavioral and Brain Sciences* 28.6 (2005), pp. 795–815. ISSN: 0140-525X. DOI: 10.1017/S0140525X05000142.

[41] J. H. Kagel and A. E. Roth, eds. *Handbook of Experimental Economics*. Princeton University Press, 1995.

[42] C. Plott and V. Smith, eds. *Handbook of Experimental Economics Results*. Vol. Volume 1. 2008, p. 1097.

[43] A. E. Roth. "Introduction to Experimental Economics." In: *Handbook of Experimental Economics*. Ed. by J. H. Kagel and A. E. Roth. Princeton University Press, 1995, pp. 3–109.

[44] P. Lipton. *Inference to the best explanation*. 2nd ed. Routledge, 2004. ISBN: 0-203-47085-0.

[45] F. Guala. "Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate." In: *Behavioral and Brain Sciences* 35.1 (2012), pp. 1–15. ISSN: 0140-525X. DOI: 10.1017/S0140525X11000069.

[46] J. O. Ledyard. "Public Goods: A Survey of Experimental Research." In: *Public Economics*. University Library of Munich, Germany, 1994.

[47] E. Fehr and S. Gächter. "Altruistic punishment in humans." In: *Nature* 415.6868 (2002), pp. 137–140. ISSN: 0028-0836. DOI: 10.1038/415137a.

[48] R. M. Isaac, J. M. Walker, and S. H. Thomas. "Divergent evidence on free riding: An experimental examination of possible explanations." In: *Public Choice* 43.2 (1984), pp. 113–149. ISSN: 00485829. DOI: 10.1007/BF00140829.

[49] S. Bowles and H. Gintis. "Homo reciprocans." In: *Nature* 415.6868 (2002), pp. 125–127. ISSN: 0028-0836. DOI: 10.1038/415125a.

[50] S. Bowles and H. Gintis. "The evolution of strong reciprocity: cooperation in heterogeneous populations." In: *Theoretical Population Biology* 65.1 (2004), pp. 17–28. ISSN: 00405809. DOI: 10.1016/j.tpb.2003.07.001.

[51] A.-L. Barabási. *Network science*. Cambridge University Press, 2014. ISBN: 0309100267. DOI: 10.17226/11516.

[52] F. Menczer, S. Fortunato, and C. A. Davis. *A First Course in Network Science*. Cambridge University Press, 2020. ISBN: 9781108653947. DOI: 10.1017/9781108653947.

[53] Jacob Moreno. *Who Shall Survive: A New Approach to the Problem of Human Interrelations*. Nervous and Mental Disease Publishing Co, 1934, p. 466.

[54] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994, p. 825. ISBN: 9780521387071.

[55] M. O. Jackson. *Social and Economic Networks*. Princeton, New Jersey: Princeton University Press, 2010, p. 504. DOI: 10.1093/acprof:oso/9780199591756.003.0019.

[56] M. Faloutsos, P. Faloutsos, and C. Faloutsos. "On power-law relationships of the Internet topology." In: *ACM SIGCOMM Computer Communication Review* 29.4 (1999), pp. 251–262. ISSN: 01464833. DOI: 10.1145/316194.316229.

[57] R. Albert, I. Albert, and G. L. Nakarado. "Structural vulnerability of the North American power grid." In: *Physical Review E* 69.2 (2004), p. 025103. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.69.025103. arXiv: 0401084 [cond-mat].

[58] B. Vogelstein, D. Lane, and A. J. Levine. "Surfing the p53 network." In: *Nature* 408.6810 (2000), pp. 307–310. ISSN: 0028-0836. DOI: 10.1038/35042675.

[59] D. J. Watts and S. H. Strogatz. "Collective dynamics of 'small-world' networks." In: *Nature* 393.6684 (1998), pp. 440–442. ISSN: 0028-0836. DOI: 10.1038/30918.

[60] A.-L. Barabási and R. Albert. "Emergence of Scaling in Random Networks." In: *Science* 286.5439 (1999), pp. 509–512. ISSN: 0036-8075. DOI: 10.1126/science.286.5439.509.

[61] M. E. J. Newman. "The Structure and Function of Complex Networks." en. In: *SIAM Review* 45.2 (2003), pp. 167–256. ISSN: 0036-1445. DOI: 10.1137/S003614450342480. arXiv: 0303516 [cond-mat].

[62] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang. "Complex networks: Structure and dynamics." In: *Physics Reports* 424.4-5 (2006), pp. 175–308. ISSN: 03701573. DOI: 10.1016/j.physrep.2005.10.009.

[63] M. Newman. *Networks: An Introduction*. Oxford: Oxford University Press, 2010. DOI: 10.1093/acprof:oso/9780199206650.001.0001.

[64] R. van der Hofstad. *Random graphs and complex networks*. Vol. 1. Cambridge Series in Statistical and Probabilistic Mathematics, 2017. ISBN: 978-1-107-17287-6. DOI: 10.1017/9781316779422.

[65] R. Guimerà, S. Mossa, A. Turtschi, and L. A. Amaral. "The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles." In: *Proceedings of the National Academy of Sciences of the United States of America* 102.22 (2005), pp. 7794–7799. ISSN: 00278424. DOI: 10.1073/pnas.0407994102. arXiv: 0312535 [cond-mat].

[66] A. Aleta, S. Meloni, M. Perc, and Y. Moreno. "From degree-correlated to payoff-correlated activity for an optimal resolution of social dilemmas." In: *Physical Review E* 94.6 (2016), pp. 1–8. ISSN: 24700053. DOI: 10.1103/PhysRevE.94.062315.

[67] A. Aleta, S. Meloni, and Y. Moreno. "A Multilayer perspective for the analysis of urban transportation systems." In: *Scientific Reports* 7.1 (2017), p. 44359. ISSN: 2045-2322. DOI: 10.1038/srep44359. arXiv: 1607.00072.

[68] L. G. Alves, A. Aleta, F. A. Rodrigues, Y. Moreno, and L. A. Nunes Amaral. "Centrality anomalies in complex networks as a result of model over-simplification." In: *New Journal of Physics* 22.1 (2020), p. 13043. ISSN: 13672630. DOI: 10.1088/1367-2630/ab687c.

[69] H. Tong, C. Faloutsos, and J.-y. Pan. "Fast Random Walk with Restart and Its Applications." In: *Sixth International Conference on Data Mining (ICDM'06)*. IEEE, 2006, pp. 613–622. DOI: 10.1109/ICDM.2006.70.

[70] M. Miller, C. Sathi, D. Wiesenthal, J. Leskovec, and C. Potts. "Sentiment Flow Through Hyperlink Networks." In: *Fifth International AAAI Conference on Weblogs and Social Media*. 2011, pp. 550–553.

[71] S. A. Myers, A. Sharma, P. Gupta, and J. Lin. "Information network or social network? The structure of the twitter follow graph." In: *WWW 2014 Companion - Proceedings of the 23rd International Conference on World Wide Web*. New York, New York, USA: ACM Press, 2014, pp. 493–498. ISBN: 9781450327459. DOI: 10.1145/2567948.2576939.

[72] S. Choi, A. Galeotti, and S. Goyal. "Trading in networks: Theory and experiments." In: *Journal of the European Economic Association* 15.4 (2017), pp. 784–817.

[73] Stanley Milgram. "The Small-World Problem." In: *Psychology Today* 1.1 (1967), pp. 61–67.

[74] M. E. Newman. "Power laws, Pareto distributions and Zipf's law." In: *Contemporary Physics* 46.5 (2005), pp. 323–351. ISSN: 00107514. DOI: 10.1080/00107510500052444. arXiv: 0412004 [cond-mat].

[75] M. Perc. "The Matthew effect in empirical data." In: *Journal of the Royal Society, Interface / the Royal Society* 11.98 (2014), p. 20140378. ISSN: 1742-5662. DOI: 10.1098/rsif.2014.0378. arXiv: 1408.5124.

[76] D. D. S. Price. "A general theory of bibliometric and other cumulative advantage processes." In: *Journal of the American Society for Information Science* 27.5 (1976), pp. 292–306. ISSN: 10974571. DOI: 10.1002/asi.4630270505.

[77] B. Bollobás. "A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs." In: *European Journal of Combinatorics* 1.4 (1980), pp. 311–316. ISSN: 01956698. DOI: 10.1016/S0195-6698(80)80030-8.

[78] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras. "Generation of uncorrelated random scale-free networks." In: *Physical Review E* 71.2 (2005), p. 027103. ISSN: 1539-3755. DOI: 10.1103/PhysRevE.71.027103. arXiv: 0408110 [cond-mat].

[79] S. A. West, A. S. Griffin, and A. Gardner. "Evolutionary Explanations for Cooperation." In: *Current Biology* 17.16 (2007), R661–R672. ISSN: 09609822. DOI: 10.1016/j.cub.2007.06.004.

[80] S. A. West, A. S. Griffin, and A. Gardner. "Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection." In: *Journal of Evolutionary Biology* 20.2 (2007), pp. 415–432. ISSN: 1010061X. DOI: 10.1111/j.1420-9101.2006.01258.x.

[81] J. M. Smith. "Group Selection and Kin Selection." In: *Nature* 201.4924 (1964), pp. 1145–1147. ISSN: 00280836. DOI: 10.1038/2011145a0.

[82] W. Hamilton. "The genetical evolution of social behaviour. I & II." In: *Journal of Theoretical Biology* 7.1 (1964), pp. 1–16. ISSN: 00225193. DOI: 10.1016/0022-5193(64)90038-4.

[83] D. C. Queller. "A general model for kin selection." In: *Evolution* 46.2 (1992), pp. 376–380. ISSN: 00143820. DOI: 10.1111/j.1558-5646.1992.tb02045.x.

[84] J. Birch. "The inclusive fitness controversy: finding a way forward." In: *Royal Society Open Science* 4.7 (2017), p. 170335. ISSN: 2054-5703. DOI: 10.1098/rsos.170335.

[85] A. Grafen. "Do animals really recognize kin?" In: *Animal Behaviour* 39.1 (1990), pp. 42–54. ISSN: 00033472. DOI: 10.1016/S0003-3472(05)80724-9.

[86] W. D. Hamilton. "Altruism and Related Phenomena, Mainly in Social Insects." In: *Annual Review of Ecology and Systematics* 3.1 (1972), pp. 193–232. ISSN: 0066-4162. DOI: 10.1146/annurev.es.03.110172.001205.

[87] R. L. Trivers. "The Evolution of Reciprocal Altruism." In: *The Quarterly Review of Biology* 46.1 (1971), pp. 35–57. ISSN: 0033-5770. DOI: 10.1086/406755.

[88] R. Dawkins. *The Selfish Gene: 30th Anniversary Edition*. 2006, p. 384. ISBN: 0199291152.

[89] T. Clutton-Brock. "Cooperation between non-kin in animal societies." In: *Nature* 462.7269 (2009), pp. 51–57. ISSN: 0028-0836. DOI: 10.1038/nature08366.

[90] R. Axelrod and W. Hamilton. "The evolution of cooperation." In: *Science* 211.4489 (1981), pp. 1390–1396. ISSN: 0036-8075. DOI: 10.1126/science.7466396.

[91] M. A. Nowak and K. Sigmund. "Tit for tat in heterogeneous populations." In: *Nature* 355.6357 (1992), pp. 250–253. ISSN: 0028-0836. DOI: 10.1038/355250a0.

[92] M. Nowak, K. Sigmund, M. Noway, and K. Sigmund. "A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game." In: *Nature* 364.6432 (1993), pp. 56–58. ISSN: 0028-0836. DOI: 10.1038/364056a0.

[93] P. Dal Bó and G. R. Fréchette. "The Evolution of Cooperation in Infinitely Repeated Games: Experimental Evidence." In: *American Economic Review* 101.1 (2011), pp. 411–429. ISSN: 0002-8282. DOI: 10.1257/aer.101.1.411.

[94] D. Fudenberg, D. G. Rand, and A. Dreber. "Slow to Anger and Fast to Forgive: Cooperation in an Uncertain World." In: *American Economic Review* 102.2 (2012), pp. 720–749. ISSN: 0002-8282. DOI: 10.1257/aer.102.2.720.

[95] R. D. Alexander. *The Biology of Moral Systems*. Routledge, 1987, p. 323. ISBN: 9780203700976. DOI: 10.4324/9780203700976.

[96] M. A. Nowak and K. Sigmund. "Evolution of indirect reciprocity." In: *Nature* 437.7063 (2005), pp. 1291–1298. ISSN: 14764687. DOI: 10.1038/nature04131.

[97] M. A. Nowak and K. Sigmund. "Evolution of indirect reciprocity by image scoring." In: *Nature* 393.6685 (1998), pp. 573–577. ISSN: 0028-0836. DOI: 10.1038/31225.

[98] K. Panchanathan and R. Boyd. "Indirect reciprocity can stabilize cooperation without the second-order free rider problem." In: *Nature* 432.7016 (2004), pp. 499–502. ISSN: 0028-0836. DOI: 10.1038/nature02978.

[99] C. Wedekind and M. Milinski. "Cooperation through image scoring in humans." In: *Science* 288.5467 (2000), pp. 850–852. ISSN: 00368075. DOI: 10.1126/science.288.5467.850.

[100] H. Gintis, S. Bowles, R. Boyd, and E. Fehr. "Moral Sentiments and Material Interests." In: *Moral Sentiments and Material Interests*. Ed. by H. Gintis, S. Bowles, R. Boyd, and E. Fehr. The MIT Press, 2005. ISBN: 0262072521. DOI: 10.7551/mitpress/4771.003.0004.

[101] P. Richerson et al. "Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence." In: *Behavioral and Brain Sciences* 39 (2016), e30. ISSN: 0140-525X. DOI: 10.1017/S0140525X1400106X.

[102] E. Ostrom. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press, 1990.

[103] E. Ostrom, J. Walker, and R. Gardner. "Covenants with and without a Sword: Self-Governance Is Possible." In: *American Political Science Review* 86.02 (1992), pp. 404–417. ISSN: 0003-0554. DOI: 10.2307/1964229.

[104] E. Ostrom. "Collective action and the evolution of social norms." In: *Journal of Economic Perspectives* 14.3 (2000), pp. 137–158. ISSN: 08953309. DOI: 10.1257/jep.14.3.137.

[105] S. Gächter. "In the lab and the field: Punishment is rare in equilibrium." In: *Behavioral and Brain Sciences* 35.1 (2012), pp. 26–28. ISSN: 0140-525X. DOI: 10.1017/S0140525X11001415.

[106] E. Wilson and D. Wilson. "Evolution "for the Good of the Group"." In: *American Scientist* 96.5 (2008), p. 380. ISSN: 0003-0996. DOI: 10.1511/2008.74.380.

[107] M. A. Nowak, C. E. Tarnita, and E. O. Wilson. "The evolution of eusociality." In: *Nature* 466.7310 (2010), pp. 1057–1062. ISSN: 0028-0836. DOI: 10.1038/nature09205.

[108] P. Abbot et al. "Inclusive fitness theory and eusociality." In: *Nature* 471.7339 (2011), E1–E4. ISSN: 0028-0836. DOI: 10.1038/nature09831.

[109] S. A. West, C. El Mouden, and A. Gardner. "Sixteen common misconceptions about the evolution of cooperation in humans." In: *Evolution and Human Behavior* 32.4 (2011), pp. 231–262. ISSN: 10905138. DOI: 10.1016/j.evolhumbehav.2010.08.001.

[110] J. Birch. "Are kin and group selection rivals or friends?" In: *Current Biology* 29.11 (2019), R433–R438. ISSN: 09609822. DOI: 10.1016/j.cub.2019.01.065.

[111] M. A. Nowak and R. M. May. "Evolutionary games and spatial chaos." In: *Nature* 359.6398 (1992), pp. 826–829. ISSN: 0028-0836. DOI: 10.1038/359826a0.

[112] B. A. Huberman and N. S. Glance. "Evolutionary games and computer simulations." In: *Proceedings of the National Academy of Sciences of the United States of America* 90.16 (1993), pp. 7716–7718. ISSN: 00278424. DOI: 10.1073/pnas.90.16.7716.

[113] C. P. Roca, J. A. Cuesta, and A. Sánchez. "Evolutionary game theory: Temporal and spatial effects beyond replicator dynamics." In: *Physics of Life Reviews* 6.4 (2009), pp. 208–249. ISSN: 15710645. DOI: 10.1016/j.plrev.2009.08.001. arXiv: 0911.1720.

[114] E. Lieberman, C. Hauert, and M. A. Nowak. "Evolutionary dynamics on graphs." In: *Nature* 433.7023 (2005), pp. 312–316. ISSN: 0028-0836. DOI: 10.1038/nature03204.

[115] F. C. Santos, J. M. Pacheco, and T. Lenaerts. "Evolutionary dynamics of social dilemmas in structured heterogeneous populations." In: *Proceedings of the National Academy of Sciences* 103.9 (2006), pp. 3490–3494. ISSN: 0027-8424. DOI: 10.1073/pnas.0508201103.

[116] H. Ohtsuki, C. Hauert, E. Lieberman, and M. A. Nowak. "A simple rule for the evolution of cooperation on graphs and social networks." In: *Nature* 441.7092 (2006), pp. 502–505. ISSN: 14764687. DOI: 10.1038/nature04605.

[117] G. Szabó and G. Fáth. "Evolutionary games on graphs." In: *Physics Reports* 446.4-6 (2007), pp. 97–216. ISSN: 03701573. DOI: 10.1016/j.physrep.2007.04.004. arXiv: 0607344 [cond-mat].

[118] C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, and M. A. Nowak. "Strategy selection in structured populations." In: *Journal of Theoretical Biology* 259.3 (2009), pp. 570–581. ISSN: 00225193. DOI: 10.1016/j.jtbi.2009.03.035.

[119] M. A. Nowak, C. E. Tarnita, and T. Antal. "Evolutionary dynamics in structured populations." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 365.1537 (2010), pp. 19–30. ISSN: 0962-8436. DOI: 10.1098/rstb.2009.0215.

[120] J. A. Cuesta, C. Gracia-Lázaro, A. Ferrer, Y. Moreno, and A. Sánchez. "Reputation drives cooperative behaviour and network formation in human groups." In: *Scientific Reports* 5.1 (2015), p. 7843. ISSN: 2045-2322. DOI: 10.1038/srep07843.

[121] B. Allen, G. Lippner, Y.-T. T. Chen, B. Fotouhi, N. Momeni, S.-T. T. Yau, and M. A. Nowak. "Evolutionary dynamics on any population structure." In: *Nature* 544.7649 (2017), pp. 227–230. ISSN: 0028-0836. DOI: 10.1038/nature21723. arXiv: 1605.06530.

[122] X. Li, M. Jusup, Z. Wang, H. Li, L. Shi, B. Podobnik, H. E. Stanley, S. Havlin, and S. Boccaletti. "Punishment diminishes the benefits of network reciprocity in social dilemma experiments." In: *Proceedings of the National Academy of Sciences* 115.1 (2018), pp. 30–35. ISSN: 0027-8424. DOI: 10.1073/pnas.1707505115.

[123] F. C. Santos and J. M. Pacheco. "A new route to the evolution of cooperation." In: *Journal of Evolutionary Biology* 19.3 (2006), pp. 726–733. ISSN: 1010061X. DOI: 10.1111/j.1420-9101.2005.01063.x.

[124] J. Gómez-Gardeñes, M. Campillo, L. M. Floría, and Y. Moreno. "Dynamical Organization of Cooperation in Complex Topologies." In: *Physical Review Letters* 98.10 (2007), p. 108103. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.98.108103.

[125] C. Gracia-Lázaro, J. A. Cuesta, A. Sánchez, and Y. Moreno. "Human behavior in Prisoner's Dilemma experiments suppresses network reciprocity." In: *Scientific Reports* 2.1 (2012), p. 325. ISSN: 2045-2322. DOI: 10.1038/srep00325. arXiv: 1201.3019v1.

[126] J. Grujić, C. Gracia-Lázaro, M. Milinski, D. Semmann, A. Traulsen, J. A. Cuesta, Y. Moreno, and A. Sánchez. "A comparative analysis of spatial Prisoner's Dilemma experiments: Conditional cooperation and payoff irrelevance."

In: *Scientific Reports* 4.1 (2015), p. 4615. ISSN: 2045-2322. DOI: 10.1038/srep04615.

[127] C. Schill, T. Lindahl, and A. S. Crépin. "Collective action and the risk of ecosystem regime shifts: Insights from a laboratory experiment." In: *Ecology and Society* 20.1 (2015), art48. ISSN: 17083087. DOI: 10.5751/ES-07318-200148.

[128] J.-C. Cardenas, M. Janssen, and F. Bousquet. "Dynamics of rules and resources: three new field experiments on water, forests and fisheries." In: *Handbook on Experimental Economics and the Environment*. Ed. by J. A. L. Price and M. K. Edward Elgar, Cheltenham, UK: Edward Elgar Publishing, 2013, pp. 319–345. DOI: 10.4337/9781781009079.00020.

[129] G. Knapp and J. J. Murphy. "Voluntary approaches to transitioning from competitive fisheries to rights-based management: bringing the field into the lab." In: *Agric. Resour. Econ. Rev.* 39.2 (2010), pp. 245–261.

[130] G. R. Fréchette. "Session-effects in the laboratory." In: *Experimental Economics* 15.3 (2012), pp. 485–498. ISSN: 1386-4157. DOI: 10.1007/s10683-011-9309-1.

[131] E. Gamma, R. Helm, R. Johnson, and J. M. Vlissides. *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1994, p. 416. ISBN: 0201633612.

[132] D. L. Chen, M. Schonger, and C. Wickens. "oTree—An open-source platform for laboratory, online, and field experiments." In: *J. Behav. Exp. Finance* 9 (2016), pp. 88–97.

[133] M. Pereda, I. Tamarit, A. Antonioni, J. A. Cuesta, P. Hernández, and A. Sánchez. "Large scale and information effects on cooperation in public good games." In: *Scientific Reports* 9.1 (2019), p. 15023. ISSN: 2045-2322. DOI: 10.1038/s41598-019-50964-w.

[134] T. Hobbes. *Leviathan*. Oxford University Press, 2008. ISBN: 9780199537280.

[135] R. M. Hilton. "Institutional Incentives for Resource Mobilization: An Analysis of Irrigation Systems in Nepal." In: *Journal of Theoretical Politics* 4.3 (1992), pp. 283–308. ISSN: 14603667. DOI: 10.1177/0951692892004003003.

[136] S. J. Cox. "No Tragedy on the Commons." In: *Environmental Ethics* 7 (1985), pp 49–61.

[137] C. J. Dahlman. "The tragedy of the commons that wasn't: On technical solutions to the institutions game." In: *Population and Environment* 12.3 (1991), pp. 285–296. ISSN: 0199-0039. DOI: 10.1007/BF01357919.

[138] E. Ostrom, J. Burger, C. B. Field, R. B. Norgaard, and D. Policansky. "Revisiting the commons: Local lessons, global challenges." In: *Science* 284.5412 (1999), pp. 278–282. ISSN: 00368075. DOI: 10.1126/science.284.5412.278.

[139] E. Ostrom. "A general framework for analyzing sustainability of social-ecological systems." In: *Science* 325.5939 (2009), pp. 419–422.

[140] J. Pretty. "Social capital and the collective management of resources." In: *Science* 302.5652 (2003), pp. 1912–1914.

[141] J. C. Cardenas, J. Stranlund, and C. Willis. "Local Environmental Control and Institutional Crowding-Out." In: *World Development* 28.10 (2000), pp. 1719–1733. ISSN: 0305750X. DOI: 10.1016/S0305-750X(00)00055-3.

[142] S. Battersby. "News Feature: Can humankind escape the tragedy of the commons?" In: *Proc. Natl. Acad. Sci. U.S.A.* 114.1 (2017), pp. 7–10.

[143] L. Boerema and J. Gulland. "Stock assessment of the Peruvian anchovy (*Engraulis ringens*) and management of the fishery." In: *J. Fish. Res. Board. Can.* 30.12 (1973), pp. 2226–2235.

[144] L. C. Hamilton and M. J. Butler. "Outport adaptations: Social indicators through Newfoundland's cod crisis." In: *Hum. Ecol. Rev.* 8.2 (2001), pp. 1–11.

[145] C. Mullon, P. Fréon, and P. Cury. "The dynamics of collapse in world fisheries." In: *Fish Fish.* 6.2 (2005), pp. 111–120.

[146] L. M. Curran, S. N. Trigg, A. K. McDonald, D. Astiani, Y. Hardiono, P. Siregar, I. Caniago, and E. Kasischke. "Lowland forest loss in protected areas of Indonesian Borneo." In: *Science* 303.5660 (2004), pp. 1000–1003.

[147] D. Brack. "Illegal logging and the illegal trade in forest and timber products." In: *Int. Forest. Rev.* 5.3 (2003), pp. 195–198.

[148] M. A. Janssen. "Introducing ecological dynamics into common-pool resource experiments." In: *Ecol. Soc.* 15.2 (2010).

[149] M. A. Janssen, R. Holahan, A. Lee, and E. Ostrom. "Lab experiments for the study of social-ecological systems." In: *Science* 328.5978 (2010), pp. 613–617.

[150] D. G. Rand, J. D. Greene, and M. A. Nowak. "Spontaneous giving and calculated greed." In: *Nature* 489.7416 (2012), pp. 427–430. ISSN: 00280836. DOI: 10.1038/nature11467. arXiv: NIHMS150003.

[151] Z. Wang, M. Jusup, L. Shi, J.-H. Lee, Y. Iwasa, and S. Boccaletti. "Exploiting a cognitive bias promotes cooperation in social dilemma experiments." In: *Nat. Commun.* 9.1 (2018), p. 2954.

[152] Z. Wang, M. Jusup, R.-W. Wang, L. Shi, Y. Iwasa, Y. Moreno, and J. Kurths. "Onymity promotes cooperation in social dilemma experiments." In: *Sci. Adv.* 3.3 (2017), e1601444.

[153] O. P. Hauser, D. G. Rand, A. Peysakhovich, and M. A. Nowak. "Cooperating with the future." In: *Nature* 511.7508 (2014), p. 220.

[154] D. G. Rand, A. Peysakhovich, G. T. Kraft-Todd, G. E. Newman, O. Wurzbacher, M. A. Nowak, and J. D. Greene. "Social heuristics shape intuitive cooperation." In: *Nature Communications* 5.1 (2014), p. 3677. ISSN: 2041-1723. DOI: 10.1038/ncomms4677.

[155] K. R. Foster and F. L. Ratnieks. "A new eusocial vertebrate?" In: *Trends in Ecology and Evolution* 20.7 (2005), pp. 363–364. ISSN: 01695347. DOI: 10.1016/j.tree.2005.05.005.

[156] E. Stokstad. "Global efforts to protect biodiversity fall short." In: *Science* 369.6510 (2020), pp. 1418–1418. ISSN: 0036-8075. DOI: 10.1126/science.369.6510.1418.

[157] T. R. Karl and K. E. Trenberth. "Modern global climate change." In: *Science* 302.5651 (2003), pp. 1719–1723.

[158] R. Kline, N. Seltzer, E. Lukinova, and A. Bynum. “Differentiated Responsibilities and Prosocial Behavior in Climate Change Mitigation: Behavioral Evidence from the United States and China.” In: *SSRN Electronic Journal* (2016). ISSN: 1556-5068. DOI: 10.2139/ssrn.2822093.

[159] R. Kline, N. Seltzer, E. Lukinova, and A. Bynum. “Differentiated responsibilities and prosocial behaviour in climate change mitigation.” In: *Nature Human Behaviour* 2.9 (2018), pp. 653–661. ISSN: 2397-3374. DOI: 10.1038/s41562-018-0418-0.

[160] M. Milinski, R. D. Sommerfeld, H.-J. Krambeck, F. A. Reed, and J. Marotzke. “The collective-risk social dilemma and the prevention of simulated dangerous climate change.” In: *Proceedings of the National Academy of Sciences* 105.7 (2008), pp. 2291–2294. ISSN: 0027-8424. DOI: 10.1073/pnas.0709546105.

[161] M. Jusup, F. Maciel-Cardoso, C. Gracia-Lázaro, C. Liu, Z. Wang, and Y. Moreno. “Behavioural patterns behind the demise of the commons across different cultures.” In: *Royal Society Open Science* 7.7 (2020), p. 201026. ISSN: 2054-5703. DOI: 10.1098/rsos.201026.

[162] E. Ostrom. “How types of goods and property rights jointly affect collective action.” In: *J. Theor. Politics* 15.3 (2003), pp. 239–270.

[163] T. Dietz, N. Dolšak, E. Ostrom, and P. C. Stern. “The drama of the commons.” In: *The drama of the commons*. Ed. by E. Ostrom, T. Dietz, N. Dolšak, P. C. Stern, S. Stonich, and E. U. Weber. National Academy Press, 2002, pp. 3–35.

[164] J. Tanimoto. “Environmental dilemma game to establish a sustainable society dealing with an emergent value system.” In: *Physica D* 200.1-2 (2005), pp. 1–24.

[165] S. A. Levin, S. Barrett, S. Aniyar, W. Baumol, C. Bliss, B. Bolin, P. Dasgupta, P. Ehrlich, C. Folke, I.-M. Gren, et al. “Resilience in natural and socioeconomic systems.” In: *Environ. Dev. Econ.* 3.2 (1998), pp. 221–262.

[166] V. Dakos, B. Matthews, A. P. Hendry, J. Levine, N. Loeuille, J. Norberg, P. Nosil, M. Scheffer, and L. De Meester. "Ecosystem tipping points in an evolving world." In: *Nat. Ecol. Evol.* 3 (2019), pp. 355–362.

[167] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. Van Nes, M. Rietkerk, and G. Sugihara. "Early-warning signals for critical transitions." In: *Nature* 461.7260 (2009), p. 53.

[168] S. A. Levin. "Self-organization and the emergence of complexity in ecological systems." In: *Bioscience* 55.12 (2005), pp. 1075–1079.

[169] P. A. Stephens, W. J. Sutherland, and R. P. Freckleton. "What is the Allee effect?" In: *Oikos* 87.1 (1999), pp. 185–190.

[170] Y. Liu, Z. Li, H. Xiong, X. Gao, and J. Wu. "Understanding of internal clustering validation measures." In: *2010 IEEE 10th International Conference on Data Mining (ICDM)*. IEEE. 2010, pp. 911–916.

[171] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data*. John Wiley & Sons, Inc., 1990.

[172] J. Vicens, N. Bueno-Guerra, M. Gutiérrez-Roig, C. Gracia-Lázaro, J. Gómez-Gardeñes, J. Perelló, A. Sánchez, Y. Moreno, and J. Duch. "Resource heterogeneity leads to unjust effort distribution in climate change mitigation." In: *PLOS ONE* 13.10 (2018), e0204369.

[173] J. Poncela-Casasnovas, M. Gutiérrez-Roig, C. Gracia-Lázaro, J. Vicens, J. Gómez-Gardeñes, J. Perelló, Y. Moreno, J. Duch, and A. Sánchez. "Humans display a reduced set of consistent behavioral phenotypes in dyadic games." In: *Sci. Adv.* 2.8 (2016), e1600451.

[174] A. Peysakhovich, M. A. Nowak, and D. G. Rand. "Humans display a 'cooperative phenotype' that is domain general and temporally stable." In: *Nat. Commun.* 5 (2014), p. 4939.

[175] J.-C. Cardenas, M. Janssen, F. Bousquet, et al. "Dynamics of rules and resources: three new field experiments on water, forests and fisheries." In: *Handbook on experimental economics*

and the environment. Ed. by J. List and M. Price. Edward Elgar Publishing, 2013, pp. 319–345.

[176] W. Steffen, Å. Persson, L. Deutsch, J. Zalasiewicz, M. Williams, K. Richardson, C. Crumley, P. Crutzen, C. Folke, L. Gordon, et al. "The Anthropocene: From global change to planetary stewardship." In: *Ambio* 40.7 (2011), p. 739.

[177] W. A. Jury and H. J. Vaux Jr. "The emerging global water crisis: managing scarcity and conflict between water users." In: *Adv. Agron.* 95 (2007), pp. 1–76.

[178] W. N. Adger, T. P. Hughes, C. Folke, S. R. Carpenter, and J. Rockström. "Social-ecological resilience to coastal disasters." In: *Science* 309.5737 (2005), pp. 1036–1039.

[179] K. Anderson, J. F. Broderick, and I. Stoddard. "A factor of two: how the mitigation plans of 'climate progressive' nations fall far short of Paris-compliant pathways." In: *Climate Policy* (2020), pp. 1–15. ISSN: 1469-3062. DOI: 10.1080/14693062.2020.1728209.

[180] J. I. Allan. "Dangerous Incrementalism of the Paris Agreement." In: *Global Environmental Politics* 19.1 (2019), pp. 4–11. ISSN: 1526-3800. DOI: 10.1162/glep_a_00488.

[181] A. Tavoni, A. Dannenberg, G. Kallis, and A. Löschel. "Inequality, communication, and the avoidance of disastrous climate change in a public goods game." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.29 (2011), pp. 11825–11829. ISSN: 00278424. DOI: 10.1073/pnas.1102493108.

[182] A. Lange, A. Löschel, C. Vogt, and A. Ziegler. "On the self-interested use of equity in international climate negotiations." In: *European Economic Review* 54.3 (2010), pp. 359–375. ISSN: 00142921. DOI: 10.1016/j.euroecorev.2009.08.006.

[183] J. T. Roberts and B. C. Parks. *A Climate of Injustice*. The MIT Press, 2007, xi, 404 p. ISBN: 9780262681612.

[184] U. T. Srinivasan, S. P. Carey, E. Hallstein, P. A. T. Higgins, A. C. Kerr, L. E. Koteen, A. B. Smith, R. Watson, J. Harte, and R. B. Norgaard. "The debt of nations and the distribution of ecological impacts from human activities." In: *Proceedings of*

the National Academy of Sciences 105.5 (2008), pp. 1768–1773. ISSN: 0027-8424. DOI: 10.1073/pnas.0709562104.

[185] K. Hyndman and M. Embrey. “Econometrics for Experiments.” In: *The Handbook of Behavioral Operations*. Ed. by K. Donohue. John Wiley & Sons, 2019, pp. 35–88.

[186] J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. MIT Press, 2010. ISBN: 9780262232586.

[187] F. C. Santos and J. M. Pacheco. “Risk of collective failure provides an escape from the tragedy of the commons.” In: *Proceedings of the National Academy of Sciences of the United States of America* 108.26 (2011), pp. 10421–10425. ISSN: 00278424. DOI: 10.1073/pnas.1015648108.

[188] J.-C. Cardenas. “Norms from outside and from inside: an experimental analysis on the governance of local ecosystems.” In: *Forest Policy and Economics* 6.3-4 (2004), pp. 229–241. ISSN: 13899341. DOI: 10.1016/j.forpol.2004.03.006.

[189] U. Gneezy and A. Rustichini. “A Fine is a Price.” In: *The Journal of Legal Studies* 29.1 (2000), pp. 1–17. ISSN: 0047-2530. DOI: 10.1086/468061.

[190] N. Machiavelli, J. C. Bondanella, and P. Bondanella. *Discourses on Livy*. Oxford University Press, 2008, p. 448. ISBN: 9780199555550.

[191] R. H. Thaler. *Misbehaving: The Making of Behavioral Economics*. W. W. Norton & Company, 2015. ISBN: 978-0-393-08094-0.

[192] A. Tversky and D. Kahneman. “Judgment under Uncertainty: Heuristics and Biases.” In: *Science* 185.4157 (1974), pp. 1124–1131. ISSN: 0036-8075. DOI: 10.1126/science.185.4157.1124.

[193] R. J. Aumann. “A synthesis of behavioural and mainstream economics.” In: *Nature Human Behaviour* (2019), p. 1. ISSN: 2397-3374. DOI: 10.1038/s41562-019-0617-3.

[194] C. A. Cox and B. Stoddard. “Framing and feedback in social dilemmas with partners and strangers.” In: *Games* 6 (2015), pp. 394–412.

[195] Giving USA Foundation. *Giving USA 2018: The Annual Report on Philanthropy for the Year 2017*. Giving USA Foundation, 2018.

[196] NCVO (The National Council for Voluntary Organisations). *UK Civil Society Almanac 2019*. NCVO, 2019.

[197] B. I. Bittker. "Charitable contributions: Tax deductions or matching grants." In: *Tax. L. Rev.* 28 (1972), p. 37.

[198] M. Feldstein and A. Taylor. "The income tax and charitable contributions." In: *Econometrica: Journal of the Econometric Society* (1976), pp. 1201–1222.

[199] J. Peloza and P. Steel. "The price elasticities of charitable contributions: a meta-analysis." In: *Journal of Public Policy & Marketing* 24.2 (2005), pp. 260–272.

[200] C. T. Clotfelter. *Federal tax policy and charitable giving*. University of Chicago Press, 2007.

[201] R. Bradley, S. Holden, and R. McClelland. "A robust estimation of the effects of taxation on charitable contributions." In: *Contemporary Economic Policy* 23.4 (2005), pp. 545–554.

[202] P. Kollock. "Social dilemmas: the anatomy of cooperation." In: *Annu. Rev. Sociol.* 24 (1998), pp. 183–214.

[203] I. Smillie, H. Helmich, J. Randel, and T. German. *Stakeholders: Government-NGO partnerships for international development*. Routledge, 2013.

[204] K. Sansom. "Complementary Roles? Ngo–Government Relations For Community-Based Sanitation In South Asia." In: *Public Administration and Development* 31.4 (2011), pp. 282–293.

[205] Charities Aid Foundation. *World Giving Index 2016*. Charities Aid Foundation, 2016.

[206] J. Andreoni. "Chapter 18 Philanthropy." In: *Handbook of the Economics of Giving, Altruism and Reciprocity*. Elsevier, 2006, pp. 1201–1269. DOI: 10.1016/s1574-0714(06)02018-5.

[207] R. Bekkers and P. Wiepking. "A literature review of empirical studies of philanthropy: Eight mechanisms that drive charitable giving." In: *Nonprofit and voluntary sector quarterly* 40.5 (2011), pp. 924–973.

[208] S. Rose-Ackerman. "Charitable giving and "excessive" fundraising." In: *The Quarterly Journal of Economics* 97.2 (1982), pp. 193–212.

[209] P. Hankinson. "Brand orientation in charity organisations: qualitative research into key charity sectors." In: *International Journal of Nonprofit and Voluntary Sector Marketing* 5.3 (2000), pp. 207–219.

[210] D. Karlan and J. A. List. "Does price matter in charitable giving? Evidence from a large-scale natural field experiment." In: *American Economic Review* 97.5 (2007), pp. 1774–1793.

[211] D. Webber. "Understanding charity fundraising events." In: *International Journal of Nonprofit and Voluntary Sector Marketing* 9.2 (2004), pp. 122–134.

[212] A. Sargeant and J. Kähler. "Returns on fundraising expenditures in the voluntary sector." In: *Nonprofit Management and Leadership* 10.1 (1999), pp. 5–19.

[213] C. S. Kumru and L. Vesterlund. "The effect of status on charitable giving." In: *Journal of Public Economic Theory* 12.4 (2010), pp. 709–735.

[214] S. Huck and I. Rasul. "Matched fundraising: Evidence from a natural field experiment." In: *Journal of Public Economics* 95.5-6 (2011), pp. 351–362.

[215] C. C. Eckel and P. J. Grossman. "Rebate versus matching: does how we subsidize charitable contributions matter?" In: *Journal of Public Economics* 87.3-4 (2003), pp. 681–701.

[216] C. C. Eckel and P. J. Grossman. "Subsidizing charitable contributions: a natural field experiment comparing matching and rebate subsidies." In: *Experimental Economics* 11.3 (2008), pp. 234–252.

[217] E. Das, P. Kerkhof, and J. Kuiper. "Improving the effectiveness of fundraising messages: The impact of charity goal attainment, message framing, and evidence on persuasion." In: *Journal of Applied Communication Research* 36.2 (2008), pp. 161–175.

[218] D. J. Mesch, M. S. Brown, Z. I. Moore, and A. D. Hayat. "Gender differences in charitable giving." In: *Int. J. Nonprofit Volunt. Sect. Mark.* 16 (2011), pp. 342–355.

[219] D. Balliet, N. P. Li, S. J. Macfarlan, and M. van Vugt. "Sex differences in cooperation: A meta-analytic review of social dylemmas." In: *Psycol. Bull.* 137 (2011), pp. 881–909.

[220] A. H. Eagly. "The His and Hers of Prosocial Behavior: An Examination of the Social Psychology of Gender." In: *Am. Psychol.* 64 (2009), pp. 644–658.

[221] D. T. Kenrick, M. R. Trost, and J. M. Sundie. "Sex Roles as Adaptations: An Evolutionary Perspective on Gender Differences and Similarities." In: *The psychology of gender (2nd ed.)* Ed. by A. H. Eagly, A. E. Beall, and R. J. Sternberg. New York: Guilford Press, 2004, pp. 65–91.

[222] G. E. Bolton and E. Katok. "An experimental test for gender differences in beneficent behavior." In: *Econ. Lett.* 48 (1995), pp. 287–292.

[223] M. Dufwenberg and A. Muren. "Gender composition in teams." In: *J. Econ. Behav. Org.* 61 (2006), pp. 50–54.

[224] J. A. Molina, C. Gracia-Lázaro, J. I. Jiménez-Nadal, Y. Moreno, J. A. Cuesta, and A. Sánchez. "Gender differences in cooperation: experimental evidence on high school students." In: *PLoS ONE* 8 (2013), e83700.

[225] P. Kuhn and M. C. Villeval. "Are Women More Attracted to Co-operation Than Men?" In: *Econ. J.* 125 (2015), pp. 115–140.

[226] J. Andreoni and L. Vesterlund. "Which is the fair sex? Gender differences in altruism." In: *Q. J. Econ.* 116 (2001), pp. 293–312.

[227] J. L. Solow and N. Kirkwood. "Group identity and gender in public goods experiments." In: *J. Econ. Behav. Organ.* 88.3 (2013), p. 032807.

[228] J. Andreoni and R. Petrie. "Public goods experiments without confidentiality: a glimpse into fund-raising." In: *J. Pub. Econ.* 88 (2004), pp. 1605–1623.

[229] C. E. Landry, A. Lange, J. List, M. Price, and N. Rupp. "Toward an understanding of the economics of charity: evidence from a field experiment." In: *Q. J. Econ.* 121 (2006), pp. 747–782.

[230] Y. Chen, M. Harper, J. Konstan, and S. Li. "Social comparisons and contributions to online communities: a field experiment on movielens." In: *Am. Econ. Rev* 100 (2010), pp. 1358–1398.

[231] C. Blackwell and M. McKee. "Only for my own neighborhood? Preferences and voluntary provision of local and global public goods." In: *J. Econ. Beh. ORg.* 52 (2003), pp. 115–131.

[232] N. R. Buchan, G. Grimalda, R. Wilson, M. Brewer, E. Fatás, and M. Foddy. "Globalization and Human Cooperation." In: *Proc. Natl. Acad. Sci. USA* 106 (2009), pp. 4138–4142.

[233] L. Corazzini, C. Cotton, and P. Valbonesi. "Donor coordination in project funding: Evidence from a threshold public goods experiment." In: *J. Pub. Econ.* 128 (2015), pp. 16–29.

[234] J. Krieg and A. Samek. "When charities compete: A laboratory experiment with simultaneous public goods." In: *J. Beh. Exp. Econ.* 66 (2017), pp. 40–57.

[235] P. Wiepking and R. Bekkers. "Who gives? A literature review of predictors of charitable giving. Part Two: Gender, family composition and income." In: *Voluntary Sector Review* 3.2 (2012), pp. 217–245.

[236] L. De Amicis, S. Binenti, F. Maciel Cardoso, C. Gracia-Lázaro, Á. Sánchez, and Y. Moreno. "Understanding drivers when investing for impact: an experimental study." In: *Palgrave Communications* 6.1 (2020), p. 86. ISSN: 2055-1045. DOI: 10.1057/s41599-020-0447-y.

[237] GIIN. *Annual Impact Investor Survey*. \https://thegiin.org/research/publication/annualsurvey2018. Accessed: 2019-09-02. 2019.

[238] M. Bowes. *Future Trends in Impact Investing*. \https://www.ampcapital.com/au/en/insights-hub/articles/2019/\may/future-trends-in-impact-investing. Accessed: 2019-09-02. 2019.

[239] GIIN. *Impact Investing Trends Evidence of a Growing Industry*. \https://thegiin.org/research/publication/impact-investing-trends. Accessed: 2019-09-02. 2016.

[240] A. M. Esteves, D. Franks, and F. Vanclay. "Social impact assessment: the state of the art." In: *Impact Assessment and Project Appraisal* 30.1 (2012), pp. 34–42.

[241] E. T. Jackson. "Evaluating social impact bonds: questions, challenges, innovations, and possibilities in measuring outcomes in impact investing." In: *Community Development* 44.5 (2013), pp. 608–616.

[242] M. J. Epstein and K. Yuthas. *Measuring and improving social impacts: A guide for nonprofits, companies and impact investors*. Routledge, 2017.

[243] S. Karytsas, D. Mendrinos, and C. Karytsas. "Measurement methods of socioeconomic impacts of renewable energy projects." In: *IOP Conference Series: Earth and Environmental Science*. Vol. 410. 2020, p. 012087. DOI: 10.1088/1755-1315/410/1/012087.

[244] K. Tsotsotso. "Is Programme Evaluation the Same as Social Impact Measurement?" In: *Journal of Social Entrepreneurship* (2020). ISSN: 19420684. DOI: 10.1080/19420676.2019.1683877.

[245] F. Findler. "Scoping What Matters: An Introduction to Impact Mapping." In: *Business and the Sustainable Development Goals*. Springer International Publishing, 2019, pp. 55–74.

[246] L. Portales. "Social Impact in Social Innovations: Definition, Design, and Evaluation." In: *Social Innovation and Social Entrepreneurship*. Springer International Publishing, 2019, pp. 161–176.

[247] C. Hervieux and A. Voltan. "Toward a systems approach to social impact assessment." In: *Social Enterprise Journal* 15.2 (2019), pp. 264–286. ISSN: 17508533. DOI: 10.1108/SEJ-09-2018-0060.

[248] C. O'Connor and S. Labowitz. *Measuring Human rights performance for Investors*. Tech. rep. March. NYU Stern Center for Business and Human Rights, 2017.

[249] G. Serafeim, K. Trinh, and R. Zochowski. "A Preliminary Framework for Product Impact-Weighted Accounts." In: *SSRN Electronic Journal* (2020). ISSN: 1556-5068. DOI: 10.2139/ssrn.3532472.

[250] A. McWilliams and D. Siegel. "Corporate social responsibility and financial performance: correlation or misspecification?" In: *Strategic management journal* 21.5 (2000), pp. 603–609.

[251] M.-D. P. Lee. "A review of the theories of corporate social responsibility: Its evolutionary path and the road ahead." In: *International journal of management reviews* 10.1 (2008), pp. 53–73.

[252] J. L. Campbell. "Why would corporations behave in socially responsible ways? An institutional theory of corporate social responsibility." In: *Academy of management Review* 32.3 (2007), pp. 946–967.

[253] J. N. Druckman and C. D. Kam. "Students as experimental participants." In: *Cambridge handbook of experimental political science* 1 (2011), pp. 41–57.

[254] D. MBA. *Three Things: What Is Impact Investing?* \https://www.youtube.com/watch?v=WBNAsvlnERs. Accessed: 2019-09-02. 2017.

[255] M. Argyle and A. Furnham. *The psychology of money*. Routledge, 2013.

[256] U. Gneezy, E. A. Keenan, and A. Gneezy. "Avoiding overhead aversion in charity." In: *Science* 346.6209 (2014), pp. 632–635.

[257] G. Charness, U. Gneezy, and B. Halladay. "Experimental methods: Pay one or pay all." In: *Journal of Economic Behavior & Organization* 131 (2016), pp. 141–150.

[258] C. D. Burt and K. Strongman. "Use of images in charity advertising: Improving donations and compliance rates." In: *International Journal of Organisational Behaviour* 8.8 (2005), pp. 571–580.

[259] N. Carter and M. Huby. "Ecological citizenship and ethical investment." In: *Environmental Politics* 14.2 (2005), pp. 255–272. ISSN: 09644016. DOI: 10.1080/09644010500055159.

[260] M. Buttle. "'I'm not in it for the money': Constructing and mediating ethical reconnections in UK social banking." In: *Geoforum* 38.6 (2007), pp. 1076–1088. ISSN: 00167185. DOI: 10.1016/j.geoforum.2006.12.011.

[261] A. Lewis. *Morals, markets and money : ethical, green and socially responsible investing*. Harlow: Prentice Hall, 2002, p. 214. ISBN: 0273653237.

[262] A. Wins and B. Zwergel. "Private ethical fund investors across countries and time: a survey-based review." In: *Qualitative Research in Financial Markets* 7.4 (2015), pp. 379–411. ISSN: 1755-4179. DOI: 10.1108/QRFM-10-2014-0030.

[263] J. Emerson and L. Norcott. *Millennials Will Bring Impact Investing Mainstream*. 2016.

[264] M. J. Stecker. "Revolutionizing the nonprofit sector through social entrepreneurship." In: *Journal of Economic Issues* 48.2 (2014), pp. 349–358.

[265] P. K. Novak, L. De Amicis, and I. Mozetič. "Impact investing market on Twitter: influential users and communities." In: *Applied network science* 3.1 (2018), p. 40.

[266] J. A. Molina, A. Ferrer, J. I. Giménez-Nadal, C. Gracia-Lázaro, Y. Moreno, and A. Sánchez. "Intergenerational cooperation within the household: a Public Good game with three generations." In: *Review of Economics of the Household* 17.2 (2019), pp. 535–552.

[267] FT. *The younger generation needs better advice on impact investing*. \https://www.ft.com/content/bebd2b92-605c-11e9-9300-0becfc937c37. Accessed: 2019-09-02. 2019.

[268] M. Stanley. *Morgan Stanley Private Wealth Management & Campden Wealth (2014)*. \http://www.campdenresearch.com/content/\next-generation-wealth-report-2014. Accessed: 2019-09-02. 2014.

[269] B. Gunstone and E. Gavin. *Insights into charity fundraising*. https://www.institute-of-fundraising.org.uk/\library/insights-into-charity-fundraising-final-report/. Accessed: 2019-09-02. 2017.

[270] P. Doerr. *Changing The Status Quo: How Women Are Leading The Charge On Impact Investing*. https://www.forbes.com/sites/patsydoerr/2018/10/23/\changing-the-status-quo-how-women-are-leading-\the-charge-on-impact-investing/#5fe92cab5bbe. Accessed: 2019-09-02. 2018.

[271] A. Brown. *Six Reasons Why Women Are Shaking Up Impact Investing*. https://www.triplepundit.com/story/2019/\six-reasons-why-women-are-shaking-impact-investing/82571/. Accessed: 2019-09-02. 2019.

[272] M. S. Ficher. *Women Are Savvier Investors Than They Think*. https://www.thinkadvisor.com/2017/05/18/women-are-savvier-investors-\than-they-think/?slreturn=20190806061128. Accessed: 2019-09-02. 2017.

[273] GECES. *Social Enterprises and Social Economy going forward*. https://ec.europa.eu/growth/content/\social-enterprises-and-social-economy-going-forward-0_en. Accessed: 2019-09-02. 2016.

[274] HMRC. *Enterprise Investment Scheme, Seed Enterprise Investment Scheme and Social Investment Tax Relief – Statistics on companies raising funds*. https://www.gov.uk/government/statistics/\enterprise-investment-scheme-seed-enterprise-investment-\scheme-and-social-investment-tax-relief-statistics-may-2018. Accessed: 2019-09-02. 2018.

[275] BSC. *Social Investment Tax Relief critical for social enterprises and charities, argues Big Society Capital*. https://www.bigsocietycapital.com/latest/type/news/social-investment-\tax-relief-critical-social-enterprises-and-charities. Accessed: 2019-09-02. 2019.

[276] D. Floyd. *What a relief! A review of Social Investment Tax Relief for charities and social enterprises*. https://www.sibgroup.org.uk/sites/default/files/files/\What%20A%20Relief%20-%20SITR%20research%20report.pdf. Accessed: 2019-09-02. 2019.

[277] H. D. Kurz. "Adam Smith on markets, competition and violations of natural liberty." In: *Cambridge Journal of Economics* 40.2 (2016), pp. 615–638. ISSN: 14643545. DOI: 10.1093/cje/bev011.

[278] A. Cournot. *Researches into the mathematical principles of the theory of wealth*. 1838.

[279] D. Condorelli, A. Galeotti, and L. Renou. "Bilateral Trading In Networks." In: *Review of Economic Studies* 84.1 (2017), pp. 82–105. ISSN: 1467937X. DOI: 10.1093/restud/rdw034.

[280] A. P. Lerner. "The Concept of Monopoly and the Measurement of Monopoly Power." In: *The Review of Economic Studies* 1.3 (1934), pp. 157–175. ISSN: 0034-6527. DOI: 10.2307/2967480.

[281] A. Kirman. *Networks: A Paradigm Shift for Economics?* April 2018. 2016, pp. 13–46. ISBN: 9780199948277. DOI: 10.1093/oxfordhb/9780199948277.013.4.

[282] Y. Bramoullé, A. Galeotti, and B. Rogers, eds. *The Oxford Handbook of the Economics of Networks*. Oxford University Press, 2016. ISBN: 9780199948277. DOI: 10.1093/oxfordhb/9780199948277.001.0001.

[283] J. E. Rauch. "Networks versus markets in international trade." In: *Journal of International Economics* 48.1 (1999), pp. 7–35. ISSN: 00221996. DOI: 10.1016/S0022-1996(98)00009-9.

[284] A. Galeotti and D. Condorelli. "Strategic Models of Intermediation Networks." In: *The Oxford Handbook on The Economics of Networks*. Ed. by A. G. Y. Bramoulle and B. Rogers. Vol. 1. Oxford: Oxford University Press, 2016. DOI: 10.1093/oxfordhb/9780199948277.013.16.

[285] F. Vega-Redondo. *Complex Social Networks*. Cambridge: Cambridge University Press, 2007, pp. 1–295. ISBN: 9780511804052. DOI: 10.1017/CBO9780511804052.

[286] F. M. Cardoso, C. Gracia-Lázaro, F. Moisan, S. Goyal, Á. Sánchez, and Y. Moreno. "Effect of network topology and node centrality on trading." In: *Scientific Reports* 10.1 (2020), p. 11113. ISSN: 2045-2322. DOI: 10.1038/s41598-020-68094-z.

[287] D. Rodrik. *The globalization paradox: democracy and the future of the world economy*. WW Norton & Company, 2011.

[288] J. E. Anderson and E. Van Wincoop. "Trade costs." In: *Journal of Economic literature* 42.3 (2004), pp. 691–751.

[289] P. Antràs and D. Chor. "Organizing the global value chain." In: *Econometrica* 81.6 (2013), pp. 2127–2204.

[290] D. Hummels, J. Ishii, and K.-M. Yi. "The nature and growth of vertical specialization in world trade." In: *Journal of international Economics* 54.1 (2001), pp. 75–96.

[291] P. Antras and A. Costinot. "Intermediated trade." In: *The Quarterly Journal of Economics* 126.3 (2011), pp. 1319–1374.

[292] A. Chaabane, A. Ramudhin, and M. Paquet. "Design of sustainable supply chains under the emission trading scheme." In: *International Journal of Production Economics* 135.1 (2012), pp. 37–49.

[293] M. Fafchamps and B. Minten. "Relationships and traders in Madagascar." In: *The Journal of Development Studies* 35.6 (1999), pp. 1–35.

[294] B. Bayley et al. *A revolution in the market: the deregulation of South African agriculture.* Oxford Policy Management, 2000.

[295] J. Meerman. *Reforming agriculture: The World Bank goes to market*. The World Bank, 1997.

[296] L. N. Traub and T. S. Jayne. "The effects of price deregulation on maize marketing margins in South Africa." In: *Food Policy* 33.3 (2008), pp. 224–236.

[297] A. D'Ignazio and E. Giovannetti. "Antitrust analysis for the Internet upstream market: A border gateway protocol approach." In: *Journal of Competition Law and Economics* 2.1 (2006), pp. 43–69.

[298] A. D'Ignazio and E. Giovannetti. "Asymmetry and discrimination in Internet peering: Evidence from the LINX." In: *International Journal of Industrial Organization* 27.3 (2009), pp. 441–448.

[299] P. Dicken, P. F. Kelly, K. Olds, and H. Wai-Chung Yeung. "Chains and networks, territories and scales: towards a relational framework for analysing the global economy." In: *Global networks* 1.2 (2001), pp. 89–112.

[300] D. Li and N. Schürhoff. *Dealer networks: market quality in over-the-counter markets*. Tech. rep. Mimeo University of Lausanne, 2013.

[301] D. M. Gale and S. Kariv. "Trading in networks: A normal form game experiment." In: *American Economic Journal: Microeconomics* 1.2 (2009), pp. 114–32.

[302] M. J. Osborne and A. Rubinstein. *Bargaining and markets*. Academic press, 1990.

[303] J. F. Nash Jr. "The bargaining problem." In: *Econometrica: Journal of the Econometric Society* (1950), pp. 155–162.

[304] J. Nash. "Two-person cooperative games." In: *Econometrica: Journal of the Econometric Society* (1953), pp. 128–140.

[305] R. L. Goldstone and M. A. Janssen. "Computational models of collective behavior." In: *Trends in Cognitive Sciences* 9.9 (2005), pp. 424–430. ISSN: 1364-6613. DOI: https://doi.org/10.1016/j.tics.2005.07.009.

[306] M. van Steen. *Graph Theory and Complex Networks*. 2010. ISBN: 978-90-815406-1-2.

[307] R. M. Axelrod. *The complexity of cooperation: Agents-based models of competition and collaboration*. Vol. 1. 1. Princeton University Press, 1997, pp. 10–68. ISBN: 0691015686.

[308] C. Cioffi-Revilla. *Introduction to Computational Social Science*. Texts in Computer Science. London: Springer London, 2014, p. 320. ISBN: 978-1-4471-5660-4. DOI: 10.1007/978-1-4471-5661-1.

[309] D. Kahneman. *Thinking , Fast and Slow*. Farrar, Straus and Giroux, 2011, p. 499. ISBN: 978-0374275631.

[310] A. Eiben and J. Smith. *Introduction to Evolutionary Computing*. Berlin, Heidelberg: Springer-Verlag, 2003.

[311] L. A. Dugatkin. *Cooperation among animals: an evolutionary perspective*. Oxford University Press, 1997. ISBN: 019508621X. DOI: 10.5860/choice.35-0904.

[312] A. F. Bourke. *Principles of Social Evolution*. Oxford University Press, 2011. ISBN: 9780199231157.

[313] S. Bowles and H. Gintis. *A cooperative species: Human reciprocity and its evolution*. Princeton University Press, 2011, pp. 1–262. ISBN: 9781400838837.

[314] R. Axelrod. "The Evolution of Strategies in the Iterated Prisoner's Dilemma." In: *Genetic Algorithms and Simulated Annealing*. Ed. by L. Davis. Morgan Kaufman, Los Altos, CA, 1987, pp. 32–41.

[315] Lindgren. "Evolutionary Phenomena in Simple Dynamics." In: *Artificial Life II* X (1991), pp. 295–312.

[316] A. M. Colman, L. Browning, and B. D. Pulford. "Spontaneous similarity discrimination in the evolution of cooperation." In: *Journal of Theoretical Biology* 299 (2012), pp. 162–171. ISSN: 00225193. DOI: 10.1016/j.jtbi.2011.05.022.

[317] A. Traulsen, C. Hauert, H. De Silva, M. A. Nowak, and K. Sigmund. "Exploration dynamics in evolutionary games." In: *Proceedings of the National Academy of Sciences* 106.3 (2009), pp. 709–712. ISSN: 0027-8424. DOI: 10.1073/pnas.0808450106.

[318] Q. Su, A. Li, L. Zhou, and L. Wang. "Interactive diversity promotes the evolution of cooperation in structured populations." In: *New Journal of Physics* 18.10 (2016), p. 103007. ISSN: 13672630. DOI: 10.1088/1367-2630/18/10/103007.

[319] H. Gintis. "Gene–culture coevolution and the nature of human sociality." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1566 (2011), pp. 878–888. ISSN: 0962-8436. DOI: 10.1098/rstb.2010.0310.

[320] A. Knafo et al. "Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promoter region and correlation between RS3 length and hippocampal mRNA." In: *Genes, Brain and Behavior* 7.3 (2008), pp. 266–275. ISSN: 1601-1848. DOI: 10.1111/j.1601-183X.2007.00341.x.

[321] M. Bateson, D. Nettle, and G. Roberts. "Cues of being watched enhance cooperation in a real-world setting." In: *Biology Letters* 2.3 (2006), pp. 412–414. ISSN: 1744957X. DOI: 10.1098/rsbl.2006.0509.

[322] P. J. Richerson and R. Boyd. *Not by genes alone: How culture transformed human evolution*. Chigado University Press, 2005, p. 332. ISBN: 9780226712123.

[323] M. A. Nowak. "Evolutionary Dynamics of Biological Games." In: *Science* 303.5659 (2004), pp. 793–799. ISSN: 0036-8075. DOI: 10.1126/science.1093411.

[324] J. H. Miller. "The coevolution of automata in the repeated prisoner's dilemma." In: *Journal of Economic Behavior and Organization* 29.1 (1996), pp. 87–112. ISSN: 01672681. DOI: 10.1016/0167-2681(95)00052-6.

[325] M. Van Veelen, J. García, D. G. Rand, and M. A. Nowak. "Direct reciprocity in structured populations." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.25 (2012), pp. 9929–9934. ISSN: 00278424. DOI: 10.1073/pnas.1206694109.

[326] K. G. Binmore and L. Samuelson. "Evolutionary stability in repeated games played by finite automata." In: *Journal of Economic Theory* 57.2 (1992), pp. 278–305. ISSN: 00220531. DOI: 10.1016/0022-0531(92)90037-I.

[327] A. Traulsen, J. M. Pacheco, and M. A. Nowak. "Pairwise comparison and selection temperature in evolutionary game dynamics." In: *Journal of Theoretical Biology* 246.3 (2007),

pp. 522–529. ISSN: 00225193. DOI: 10.1016/j.jtbi.2007.01.002.

[328] K. Lindgren and M. G. Nordahl. "Evolutionary dynamics of spatial games." In: *Physica D: Nonlinear Phenomena* 75.1-3 (1994), pp. 292–309. ISSN: 01672789. DOI: 10.1016/0167-2789(94)90289-5.

[329] M. Milinski and C. Wedekind. "Working memory constrains human cooperation in the Prisoner's Dilemma." In: *Proceedings of the National Academy of Sciences* 95.23 (1998), pp. 13755–13758. ISSN: 0027-8424. DOI: 10.1073/pnas.95.23.13755.

[330] M. A. Nowak, A. McAvoy, B. Allen, and E. O. Wilson. "The general form of Hamilton's rule makes no predictions and cannot be tested empirically." In: *Proceedings of the National Academy of Sciences* 114.22 (2017), pp. 5665–5670. ISSN: 0027-8424. DOI: 10.1073/pnas.1701805114.

[331] M. van Veelen. "Can Hamilton's rule be violated?" In: *eLife* 7 (2018). ISSN: 2050-084X. DOI: 10.7554/eLife.41901.

[332] J. Van Cleve. "Building a synthetic basis for kin selection and evolutionary game theory using population genetics." In: *Theoretical Population Biology* (2020). ISSN: 00405809. DOI: 10.1016/j.tpb.2020.03.001.

[333] L. F. Seoane and R. V. Solé. "Information theory, predictability and the emergence of complex life." In: *Royal Society Open Science* 5.2 (2018), p. 172221. ISSN: 2054-5703. DOI: 10.1098/rsos.172221. arXiv: 1701.02389.

[334] C. Handley and S. Mathew. "Human large-scale cooperation as a product of competition between cultural groups." In: *Nature Communications* 11.1 (2020), p. 702. ISSN: 2041-1723. DOI: 10.1038/s41467-020-14416-8.

[335] D. S. Wilson. "A theory of group selection." In: *Proceedings of the National Academy of Sciences of the United States of America* 72.1 (1975), pp. 143–146. ISSN: 00278424. DOI: 10.1073/pnas.72.1.143.

[336] A. Traulsen and M. A. Nowak. "Evolution of cooperation by multilevel selection." In: *Proceedings of the National Academy of Sciences* 103.29 (2006), pp. 10952–10955. ISSN: 0027-8424. DOI: 10.1073/pnas.0602530103.

[337] J. Birch. "Kin Selection, Group Selection, and the Varieties of Population Structure." In: *The British Journal for the Philosophy of Science* 57.1 (2018), pp. 323–357. ISSN: 0007-0882. DOI: 10.1093/bjps/axx028.

[338] J. A. Marshall. "Group selection and kin selection: Formally equivalent approaches." In: *Trends in Ecology and Evolution* 26.7 (2011), pp. 325–332. ISSN: 01695347. DOI: 10.1016/j.tree.2011.04.008.

[339] M. Akbari, D. Bahrami-Rad, and E. O. Kimbrough. "Kinship, fractionalization and corruption." In: *Journal of Economic Behavior & Organization* 166 (2019), pp. 493–528. ISSN: 01672681. DOI: 10.1016/j.jebo.2019.07.015.

[340] M. Muthukrishna, P. Francois, S. Pourahmadi, and J. Henrich. "Corrupting cooperation and how anti-corruption strategies may backfire." In: *Nature Human Behaviour* 1.7 (2017), p. 0138. ISSN: 2397-3374. DOI: 10.1038/s41562-017-0138.

[341] T. Talhelm, X. Zhang, S. Oishi, C. Shimin, D. Duan, X. Lan, and S. Kitayama. "Large-Scale Psychological Differences Within China Explained by Rice Versus Wheat Agriculture." In: *Science* 344.6184 (2014), pp. 603–608. ISSN: 0036-8075. DOI: 10.1126/science.1246850.

[342] P. Raghavan. "It's time to scale the science in the social sciences." en. In: *Big Data & Society* 1.1 (2014), p. 205395171453224. ISSN: 2053-9517. DOI: 10.1177/2053951714532240.

[343] Open Science Collaboration. "Estimating the reproducibility of psychological science." In: *Science* 349.6251 (2015), aac4716–aac4716. ISSN: 0036-8075. DOI: 10.1126/science.aac4716.

[344] M. Muthukrishna and J. Henrich. "A problem in theory." In: *Nature Human Behaviour* 3.3 (2019), pp. 221–229. ISSN: 2397-3374. DOI: 10.1038/s41562-018-0522-1.

[345] B. Baltagi. *Econometric analysis of panel data*. 3rd. John Wiley & Sons, 2005.

[346] F. Hayashi. *Econometrics*. Princeton University Press, 2000.

[347] T. Amemiya. *Advanced econometrics*. Harvard University Press, 1985.

[348] H. White. "Maximum likelihood estimation of misspecified models." In: *Econometrica* 57.1 (1982), pp. 1–25.

[349] N. J. Nagelkerke. "A note on a general definition of the coefficient of determination." In: *Biometrika* 78.3 (1991), pp. 691–692.

[350] Q. H. Vuong. "Likelihood ratio tests for model selection and non-nested hypotheses." In: *Econometrica* 57.2 (1989), pp. 307–333.

[351] D. A. Freedman. "On the so-called "Huber sandwich estimator" and "robust standard errors"." In: *Am. Stat.* 60.4 (2006), pp. 299–302.

[352] B. Efron. "Better bootstrap confidence intervals." In: *J. Am. Stat. Assoc.* 82.397 (1987), pp. 171–185.

[353] H. R. Kunsch. "The jackknife and the bootstrap for general stationary observations." In: *Ann. Stat.* (1989), pp. 1217–1241.