Engineering Trust in Complex Networks

Manuel Tavares de Sousa Rosa Galamba
Instituto Superior Técnico
Lisboa, Portugal
mrgalamba@gmail.com

***Abstract*—In a world where every day more and more social and economic transactions are done via the internet, eliminating the usual face to face component, Trust and Trustworthiness are pivotal for many services to work properly. To study both of these elements, through game theory, it is common to use the Trust Game. Even though game theory dictates that in one-shot interactions, namely by means of the game's unique Nash equilibrium, investors should not trust the trustees nor should these be trustworthy, behavioral data from several experiments shows the opposite for both cases. Hence, there is still the question of how trust can be stabilized in the original Trust Game in order for it to capture what happens in reality. Even though there are several studies addressing versions of the game, that consider reputation or are played in social networks, the effects of combining both of these components are not clear. In this work, we propose a new model, consisting of a Trust Game version with reputation, using Evolutionary Game Theory as a framework, played in both finite unstructured populations and in static social networks, where we introduce other variations with the objective of increasing Trust and Trustworthiness in the population. We concluded that taking into account players' reputation has a positive effect in both Trust and Trustworthiness. When played in a Social Network, the introduction of network based role and strategy assignment, namely based on individuals' degree in the network, may yield a considerable increase of Trust and Trustworthiness. The most successful variations were when considering the more connected individuals as Investors and the introduction of pathological players in the population.**

***Index Terms*—Trust, Evolutionary Game Theory, Social Networks, Reputation, Game Theory**

I. Introduction

Trust and trustworthiness are fundamental to a successful society. In today's world, many social and economic transactions occur through the internet between people that will never actually meet in real life. In order for these transactions to work, individuals must expect that their partner in the transaction will not behave opportunistically in an attempt to maximize their own payoff, regardless of the fact that their decisions could possibly cause prejudice to their counterpart. Such trust, however, is not easy to explain or sustain.

In economics, many times, one of the main assumptions is that individuals will act in their own self-interest in order to maximize their payoff. Therefore, in individual choice settings, an individual choosing an action that deviates from his self-interest is considered irrational (assuming that individuals act following a utility function that only accounts for their own gains). In group settings, however, there are situations where acting in self-interest will make all the individuals worse off [3]. In an attempt to understand human behaviour in this last situation, Berg et al. [3] developed an experimental setting named the Investment (or Trust) Game.

The Trust Game, in brief, is an interaction between an investor and a trustee. The investor initially has a certain amount of money that he can either keep or transfer to the trustee. This value is multiplied by a factor $b > 1$ before reaching the trustee which will then decide how much to return to the investor [3].

Based on the mathematical models that are used to study this kind of interactions, namely Game Theory, one would assume that people playing the Trust Game would act to maximize their own payoff, particularly when considering that the unique Nash Equilibrium [18] for this game is for the investors to transfer zero money [3]. Behavioral data resultant from real experiments with this game, however, reveals that investors do make transfers and trustees return considerable amounts, *e.g.* [3], [13].

Because traditional Game Theoretical models fail to predict the behavior of humans playing Trust Games, some open question remain: Why do high investment preferences arise? How are those preferences maintained? How is trust impacted by specific interventions in a population?

This Thesis proposes new computational models to study the Trust Game and approach the previous questions. To do this we propose our own model where we combine Trust Games with reputations played on networked populations. The results of these simulations, as detailed in section V, show that the introduction of both reputation and structure to the population playing the Trust Game, particularly for certain variations of the game that use some of the networks' properties, lead to an increase in the promotion of both trust and trustworthiness.

A. Objectives

As previously mentioned, real-world sharing economy transactions usually consider some kind of reputation. The model we propose consists of applying a reputation system to: firstly finite unstructured populations; secondly, a networked version of the Trust Game, namely with a static scale-free network, using evolutionary game theory as a framework.

Many times, considering infinite populations is more convenient from a mathematical point of view, however, real world populations are finite and considering these instead introduces considerable changes [28]. For the version using unstructured populations, our goal is to verify if the results regarding infinite populations in [29] are extendable to finite populations. To study this we first consider the case where there are two populations, one of each role (investors and trustees) and every individual from either of the populations interacts with all of the individuals from the other population. Additionally, we run the same simulations for the symmetric version of the game, *i.e.* only one population where every individual plays as both roles while still interacting with all of the other members of the population.

Furthermore, for the networked version, we explore three different scenarios:

1. Asymmetric role assignment (λ model)
2. Diversity in the reputation
3. Hybrid societies with pathological players

Regarding asymmetric role assignment, we consider a λ parameter that controls how much a player's role (either investor or trustee) depends on his degree in the network. The main goal here is to study network-based role assignment, *e.g.* where highly connected individuals may also be in a better position to play as investors.

The second scenario consists of assigning different reputation values to players according to their degree in the network. We do this by forcing players with a higher degree to have a higher reputation value.

Finally, the third scenario consists in having pathological players, *i.e.* a group of players that, regardless of the time step of the simulation and the player they are interacting with, will always act cooperatively during the experiment. Our main objective with this scenario is to see whether having the pathological players assignment dependent on the players' degree in the network will have an impact on the promotion of trust and trustworthiness.

II. Background Theory

A. Game Theory

The purpose of game theory is to help us understand strategic interactions between decision-makers through the use of mathematical models. These decision-makers are often referred to as players, who act rationally according to a set of rules, hence the usage of the word "game". The scope of game theory is however much larger, varying from economic, like the Trust Game, to biological or even political phenomena [22].

These mathematical models are composed of three elements: a set of players, a set of actions that are available for each player, and a specification of each player's preferences. Every player knows the set of actions for all the players, including himself, and the resultant payoff from all the combinations between their actions and the other players'. This is often represented by a payoff matrix or a decision tree. Based on this information, for every interaction, players must select the actions - or sequences of actions - that most likely will maximize their payoffs.

1) Nash Equilibrium: A Nash Equilibrium [18] corresponds to a set of actions with the property that, if every player adheres to this set, no individual player will do better by choosing a different one.

In a formal definition [22], let a_i be the strategy profile of player i and a_{-i} be the strategy profile of all the other players, then a^* is a Nash equilibrium if:

$$\forall i, a_i : u_i(a_i^*, a_{-i}^*) \geq u_i(a_i, a_{-i}^*) \tag{1}$$

and is a strict Nash equilibrium if:

$$\forall i, a_i : u_i(a_i^*, a_{-i}^*) > u_i(a_i, a_{-i}^*) \tag{2}$$

where u_i is the payoff function for player i

B. Evolutionary Game Theory

Evolutionary game theory (EGT) was first introduced in [31] and consists in the application of game theory to populations. It originated in a biological context and it comes from the realization that the average payoff, here named fitness, of particular phenotypes, meaning the observable characteristics or traits of an individual, depends on their frequencies in the population [30]. In recent times, however, EGT has been of interest in other fields like economics and sociology. Here instead of measuring the fitness of a phenotype, we measure the fitness of a strategy in terms of how successful (*e.g.* economically) it is.

Evolution, in this context, works by selecting the individuals that perform better than average, modeling Darwinian competition. In EGT, however, instead of the less apt individuals dying, evolution occurs due to social learning. Individuals with better fitness are imitated by the others, according to a certain update rule.

1) Well-mixed populations: Evolutionary games have traditionally dealt with infinite unstructured populations (well-mixed populations), in which each agent interacts with all other agents with equal probability. This setup can be conveniently described through the so-called replicator equation [10], [29], a deterministic equation which allows the study of fitness-based evolution in time. This equation may define both genetic evolution or a process of social learning in which, in the first case, individuals with higher fitness will reproduce more or, in the later, individuals with higher fitness will tend to be imitated more often. In any case, strategies that do better than average will grow, whereas those that do worse than average will diminish. As usual, fitness is here defined as the average return each agent gets from interacting with all the other members of the population.

2) Games on Graphs: Although the use of infinite unstructured populations may be more convenient from a mathematical point of view, in the sense that the replicator equation can be used to describe the dynamics of the populations, in real-world situations, populations are finite and individuals are constrained to interact with (and imitate) a subset of the population, an idea conveniently defined as a network: each agent is represented by a node that is constrained to play solely with its closest neighbours. The impact of topological constraints is known to induce profound evolutionary effects, as demonstrated experimentally in the study of the evolution of different strains of *Escherichia coli* (Kerr et al. [14]). In social settings, computational and mathematical models have also shown that cooperation is favoured on spatially structured populations [20]. This result has been recently demonstrated experimentally with humans [24].

In most social settings, and contrarily to spatially unstructured populations where all individuals (homogeneously) interact with the same number partners, some individuals engage in more interactions than others which, as a result, may potentially create conditions for a broad distribution of fitness values. Such heterogeneous scenarios often comprise a small number of nodes with many interaction links, called hubs, connecting the majority of nodes that contain fewer neighbours [27].

In this Thesis, we analyze both homogeneous and heterogeneous populations. For the latter, we adopt a paradigmatic example of such interaction structures: scale-free networks [2].

A network, also called an undirected graph, consists of a pair $G = (V, E)$, where V is a set of vertices, also named nodes, and E is a set of edges, *i.e.* the existing links between the network nodes. When two nodes are connected by an edge we consider them neighbours in the network.

Before defining scale-free networks we must first introduce the concept of degree distribution. When considering a network, the degree of a node corresponds to the number of connections it has with all the other nodes in the network. Consequently, the degree distribution of a network is the distribution of these degrees over the whole network. An SF network is a network whose degree distribution follows a power-law for large k, *i.e.* $P(k) \sim k^{-\gamma}$, where $P(k)$ is the fraction of nodes in the network with degree k and γ is the exponent of that specific power law. In order to generate scale-free networks, the Barabási-Albert model can be used, detailed in section III.

For the case of simple one-shot 2-player games cooperation as the prisoner's dilemma, scale-free interaction structures were shown to help cooperation to thrive [26] when compared with homogeneous interaction structures, as highly-connected nodes are promptly taken over by cooperators who can then influence the whole community into cooperating. This enhancement is grounded on the diverse nature of real interactions. However, there are still a reduced number of studies on the impact of such structures on the evolution of trust, a question aimed by this Thesis.

III. Related Work

The objective of this work is to study the importance of trust, reciprocity, and reputation in the context of money transactions, or, generally, situations that require trusting another person or entity in order to achieve a payoff maximizing outcome. To do this we will use as a starting point the Trust Game (also called investment game) suggested by Berg et al. [3] in 1995. Although this study does not take into consideration reputation, it is one of the earliest experiments in the field to use a simple game interaction to systematically study the dilemma of trust and reciprocity.

The experiment consists of a group of subjects that are placed in a room (room A) and receive an initial amount of money. Each subject in room A must then choose how much of this initial endowment to send (they can opt to keep all the money) to another anonymous individual located in a second room (room B), knowing that the amount they send will be tripled by the time it gets to room B. After receiving the money, each subject in room B then must decide on how much money to send back to room A and how much money to keep. When mentioning other experiments, subjects from room A, in this report, will be treated as Investors, and subjects from room B as Trustees.

The main reason we found this experiment of interest and consequently made us explore more of this field is the fact that assuming that all the subjects were rational, it was expected that the subjects from room A would not send any money to room B since this is the predicted and unique Nash equilibrium as it is shown in section IV and in [3], in the trust game section of [29], as well as in many other studies in very similar games like the peasant-dictator game for the discrete case [11] or even for the N-player version of the game [1]. Furthermore, if subjects from room A actually did send any money, one would think that none of the subjects from room B would send money back. The results from the experiment, however, show the complete opposite of this along with other behavioral experiments with the trust game [4], [13], [9], [15], [8].

In [3] the authors divide the experiment into two sessions, one with no history treatment in which subjects were not given any information about prior similar experiences, and one with a social history treatment in which new subjects were given the results of the first experiment before "playing" the game. Although this provides some interesting results, namely that the average amount returned by subjects in room B increases with a social history treatment, we wanted to see the effect it would have if subjects had prior information about the other subjects with whom they are actually "playing" the game, which would represent, in a sense, their reputation.

A. *Reputation in unstructured population Trust Games*

One of the first studies to introduce reputation to Trust Games was written by Sigmund in [29]. This analysis considered, however, infinite populations, an assumption that we relax in this Thesis. The author firstly introduces a similar version to the Trust Game proposed in [3], yielding the same results predicted by the Nash Equilibrium, after the initial state the population evolves so that all the players become defectors (*i.e.* refuse to offer, or return, anything to the other player). Later, another version was introduced considering reputation, showing its positive effects on trust and trustworthiness.

Another important work regarding reputation in unstructured populations, although this time with finite populations, was done by Manapat et al. [16]. In their simulations, through the use of an unstructured population of investors and trustees, *i.e.* every investor interacts with all the trustees, the authors simulate the behaviour of the agents in a Trust Game where the Investors have, sometimes, information about the Trustees.

For each interaction, randomly picking a pair of an investor and a trustee, the investor knows, with a probability q, the exact fraction r of the amount the trustee will return, q acts, this way, as a measure of the availability of information. An investor will always transfer money if $r > 1/b$, with $b > 1$ being the factor by which the stake is multiplied if the transfer is made.

We believe that this resembles our own model in the sense that the r rate of returning the money acts basically as a reputation system. The main difference lies in the fact that, because investors always act rationally, as long as $r > 1/b$ they will make the transfer. In our reputation system, however, defective investors will only transfer an amount proportional to the trustee's reputation.

It should also be noted the fact that in [16] there is a probability $1 - q$ (with $q < 1$) that the investor does not have any information on the trustee, which does not happen in our reputation system since we consider that Investors always have access to Trustees' reputation.

As far as the evolution of the population goes, in the study by Manapat et al. [16], an evolutionary process is used, which according to the authors can be interpreted as genetic evolution [19]. This evolutionary process causes higher payoff strategies to widespread and lower payoff strategies to eventually disappear.

Since the evolution of the population occurs according to the evolutionary process mentioned above, the highest payoff strategy tends to dominate. For the trustees this highest payoff corresponds to a value of r as close as possible to $1/b$ so that the investors still make the transfer and, simultaneously, the trustees keep the maximum amount of money possible, $r = 1/b + \epsilon$ is a Nash equilibrium [16].

B. Networked Trust Games

Abbass et al. [1] proposed an evolutionary N-player trust game with an unstructured population consisting of investors, trustees who are trustworthy, and trustees who are untrustworthy. In their study, they concluded that even though the optimal solution for the population includes investors as part of it, the evolutionary dynamics converge to a population with no investors and only trustees (of both kinds). The exception to this occurs when the initial population does not have any untrustworthy players. In [6] the authors use a population consisting of the same three types in an attempt to see whether trust can be promoted when the population is structured, namely a specific spatial topology or a social network.

Because in their study [6] populations have two different types of trustees, two different multipliers were also used, *i.e.* the factor by which the investors' money is multiplied before reaching the trustees, R_T for the trustworthy trustees and R_U for the untrustworthy, with $1 < R_T < R_U < 2R_T$. A temptation to defect investors' trust ratio r_{UT} is also introduced, with $r_{UT} = \frac{R_U - R_T}{R_T}$, which we will use in order to make it easier to analyze the results obtained.

Regarding the network topologies, Chica et al. [6] consider a regular lattice, scale-free (SF) networks with different densities as well as Erdös-Rényi (ER) random networks. The SF networks were chosen as the default network for the study since these have been widely used in studies of evolutionary games. The synthetic SF networks were generated with the Barabási–Albert algorithm [2] which allows the generation of networks with different densities.

The algorithm starts with a small number m_0 of vertices, every time step a new vertex is added with *m* edges, $m < m_0$, linked to *m* different vertices that already belong to the graph. These vertices are chosen with a probability proportional to their degree, this way assuring the preferential attachment component. After *t* time steps the model results in a random network with $t+m_0$ vertices and mt edges. To create networks with different densities and $\langle k \rangle$ in [6], the authors used different values of $m \in \{2, 3, 4, 6, 8\}$. Most of the times, the authors use the SF network with $m = 3$ since this generates a network with $\langle k \rangle = 4$, which is the same as the regular lattice and the ER network.

Concerning the experiments and results in [6], several values for the temptation to defect ratio r_{UT}, and various initial population conditions, were considered in order to study the evolution of trust and global net wealth in different setups. For the r_{UT}, since its value regulates the game difficulty, three different values were mainly used, creating three different versions of the game: the easier version with $r_{UT} = 0.11$, medium with $r_{UT} = 0.33$ and harder with $r_{UT} = 0.66$.

For the easier version of the game, we can observe that trust can be promoted in all the different networks tested, investors and trustworthy trustees only disappear when the initial population is clearly dominated by untrustworthy trustees. Consequently, this results in, not only high levels of trust, but also high levels of global net wealth. These values are particularly high for the regular lattice, followed by SF networks with high densities. When $r_{UT} = 0.33$, the initial population, in order to promote trust, needs to have a higher number of investors and trustworthy trustees. It is interesting to note that, conversely to the previous case, SF networks are better for promoting trust, specifically SF with lower densities. Lastly, for the harder ($r_{UT} = 0.66$) version of the game, trust can only be promoted when there are no untrustworthy trustees in the initial population.

IV. Model

The version of the Trust Game making use of reputation in [29] considers infinite unstructured populations. As mentioned in section I, many times, from a mathematical point of view considering infinite populations is more convenient, however, real world populations (namely those where individuals play the Trust Game) are finite and individuals interact over social networks. Considering finite populations can introduce considerable changes [28]; one that we should note is the fact that, in finite populations, strategies that form Nash Equilibria, or that are evolutionary stable, may not always be highly prevalent [28], [12].

Our model aims to implement an altered version of the Trust Game where we consider the effects of reputation in finite populations as in [16]. However, the reputation aspect will be added in a different way, similar to the one used in [29], *i.e.* while in [16] investors only have information on the trustees with a certain probability, in our model and in [29] Investors always have access to Trustees' reputation. Furthermore, instead of only using unstructured populations, we will also consider a networked version of the Trust Game similar to the one in [5].

We use this model to study the evolution of trust making use of computer simulations and evolutionary game theory, in the contexts previously described, by creating an evolutionary process where the strategies of the investors and trustees change through time. This change is done via a mechanism of social learning [21], where players can imitate the strategies of those performing better, causing higher payoff strategies to proliferate while lower payoff strategies diminish.

A. Game Payoffs

The Trust Game is played by all the players for a finite number of time steps. At each time step of the simulation, players interact with each other accordingly to the type of population structure considered. These interactions between the players result in a certain payoff for each player. Every time step of the simulation new payoffs are calculated for every player since players might have changed their strategy in the previous time step, hence resulting in new payoffs in the present one.

A player's (total) payoff for any time step of the simulation corresponds to the sum of the payoffs from every interaction that player was part of in that time step. The payoffs of any interaction are calculated according to the payoff matrix in Table I, which corresponds to Table 5.28 in [29] (when the parameter μ in the book is 0).

TABLE I
TRUST GAME PAYOFF MATRIX. CORRESPONDS TO TABLE 5.28 IN [29], WHEN $\mu = 0$

	$\mathbf{f}_1$	$\mathbf{f}_2$
$\mathbf{e}_1$	$(\beta - c, b - \gamma)$	$(-c, b)$
$\mathbf{e}_2$	$((\beta - c)v, (b - \gamma)v)$	$(0, 0)$

This payoff matrix can be understood as following: players can act as one of two roles in each interaction: either as an investor (row) or as a trustee (column). The first position of each cell corresponds to the investor's payoff, while the second position payoff corresponds to the trustee's. An investor may choose to make a transfer, *i.e.* cooperate ($\mathbf{e}_1$); or to defect ($\mathbf{e}_2$). The same applies to trustees who can either return a certain amount to the investor, *i.e.* cooperate ($\mathbf{f}_1$); or defect ($\mathbf{f}_2$), *i.e.* do not transfer back anything. In this version of the trust game, if an investor decides to cooperate, then he will donate a sum c to the trustee, which will be multiplied by a factor $r > 1$ resulting in $b > c$. The trustee, in turn, will return an amount β to the investor, costing him γ, if he decided to cooperate, and 0 otherwise (no cost in this case). We assume $0 < c < \beta$ and $0 < \gamma < b$. Lastly, the variable v corresponds to the likelihood that defective investors cooperate if they know that they will be rewarded. This means essentially that v corresponds to the trustees' reputation, *i.e.* the probability that a player with a strategy $\mathbf{f}_1$ becomes known as a cooperator, and consequently a player with a strategy $\mathbf{e}_2$ cooperates too. Thus, defective investors ($\mathbf{e}_2$) can either have payoff of $(\beta - c)v$ when interacting with a trustee with a strategy $\mathbf{f}_1$ or 0 if interacting with a trustee with a strategy $\mathbf{f}_2$, *i.e.* no costs or benefits.

B. Evolutionary Update

Each round, firstly one player a is randomly selected, then mutation occurs with a probability μ (parameter of the process), *i.e.* there is a small chance the chosen player will just change his strategy to a random one (might be the same he already had). If mutation did not happen, then a new player b from the same population is randomly selected (investors can only imitate other investors and trustees can only imitate other trustees) accordingly to the population structure considered. A pairwise comparison rule was then adopted in order to calculate the probability (p) of the first player (a) imitating the second player (b) based on both of their resultant payoffs in that time step. In our work, as in [16], we use the Fermi function as this pairwise comparison rule, as studied by Traulsen et al. [32]:

$$p = \frac{1}{1 + e^{-\beta(\pi_b - \pi_a)}} \tag{3}$$

The variables π_a and π_b correspond to the accumulated payoff of player a and player b respectively, calculated for each player as the sum of all his interactions' payoffs. Let us consider here the parameter β as the intensity of selection (not to be confused with β introduced in the context of the Trust Game payoff, as the Trustee return). This means imitation of strategies will occur with a probability proportional to the difference between both players' payoff (for $\beta > 0$), and that if β increases so does the dependence on this difference.

C. Unstructured Populations

We first consider the case where there are two populations, one of each role (investors and trustees); at every time step of the simulation, each individual from either of the populations interacts with all of the individuals from the population with a different role. In this case, the imitation process occurs only between individuals of the same population, and every player can imitate any of the others.

Secondly, we consider a symmetric version of the game, where there is only one population; all the individuals play as both roles and may have different strategies for each one (each individual has two independent strategies). Players interact with the entire population, playing in both roles, and the total payoff of each player corresponds to the sum of his total payoff as an investor with his total payoff as a trustee. Once again, every player can imitate any of the others, however, they can only imitate strategies of the same role, *i.e.* if a player is imitating another player's strategy, for instance as an investor, he may only change his strategy as an investor as well.

D. Structured Populations

For this part of our study, players are placed in a social network, which will have an impact on the way information is spread, and how the pairs to play games are formed. Regarding the network structure, we use scale-free networks since they capture important characteristics of real world social networks, such as a highly heterogeneous degree distribution. In fact these networks have been used extensively in studies related to evolutionary games (*e.g.* [25]), namely in the case of networked Trust Games like [6] and [7].

By using networks, interactions between players are constrained by the network topology, *i.e.* every player will only interact with his direct neighbours. For most of our experiences, players will play in both roles, one at a time, every time step; their total payoff will correspond to the sum of the payoffs when playing in both roles. The imitation process, analogous to the interactions, will be restrained to a player's direct neighbours. Like in the unstructured symmetric version of the game, each player will have two independent strategies (one as an investor and one as a trustee) and can only imitate strategies of the same role.

1) Asymmetric role assignment: As previously mentioned, our default setting for the networked version of the Trust Game is for all the players to play as both an Investor and a Trustee. The idea behind asymmetric role assignment version of the game, however, is to, for every interaction, force a dependency between the role a player has in that interaction and his characteristics, namely his degree in the network.

Every time any player a interacts with any player b, a will act as an investor with a probability p_i, calculated accordingly with the following equation:

$$p_i = \frac{k_a^\lambda}{k_a^\lambda + k_b^\lambda} \tag{4}$$

Here, k_a and k_b correspond to player a and player b's degree in the network respectively. The variable λ, which may take negative values, controls the dependency between the degree of a player and his role. If λ is 0 then the player's role is uniformly random, *i.e.* the role of investor is attributed to one of the agents with the same probability. The higher λ is, the more likely the player with the larger degree is to act as an investor. For negative λ values, the lower λ is, the more likely the player with the higher degree is to act as a trustee. With this experience, our main interest is to see the consequences of forcing most of the individuals that are hubs in the network to have a certain role - which happens with either considerably large or small values for λ - regarding the promotion of Trust and Trustworthiness.

2) Diversity in the reputation: In the previous versions of the networked Trust Game considered, all the players have the same reputation, *i.e.* the same v value in the payoff matrix (Table I). In this version, our aim is to determine the effects of varying the value of v, while always maintaining the average v in the population the same.

We shall then consider 3 scenarios to compare the effects of this diversity in reputation: the baseline where the whole population has the same v; a second one where we consider two different v values and assign the larger to the half of the population with a higher degree and the smaller to the remaining half; another one where we consider the same two different v values and assign each of them to half of the population, randomly selected.

3) Hybrid societies with pathological players: In order to avoid initial imbalances in terms of strategies in the population, the default setting before starting the simulation is to randomly assign strategies to each player, while making sure that 50% of the individuals have one strategy (either cooperative or defective) and the remaining the other. For this version of the Trust Game, however, we introduce pathological players, *i.e.* players that, regardless of what happens during the simulation and of the role they are assuming, will always cooperate for every interaction.

In order to compare the effects of the pathological players being network hubs or not, we consider two variations: Firstly, one where we define several thresholds for the degree of the nodes in the network. Players with a degree above these thresholds will be pathological players; Secondly, we count the number of pathological players assigned for each of the thresholds considered in the first scenario and assign the same number of pathological players but by selecting them randomly out of the population.

V. Results and Discussion

As mentioned in section IV, the experiments done consist of applying our model to computer simulations. To do this we wrote and simulated a program in Python 3.6.8.

A. Methods

In our simulations, depending on the experiment we are considering there are some small changes in its initial setup. In the non symmetric version with unstructured populations, we consider two populations of 500 individuals each. For the symmetric version with unstructured populations and all the versions with networks, we consider a single population with 500 individuals. The networks considered are scale-free networks with an average degree of 4.

Every individual is represented by his payoff, his strategy as an investor, and his strategy as a trustee (with the exception of the non-symmetric version where each individual only plays in one role and therefore only has one strategy).

Every simulation has 10^5 rounds, consisting of one run. Regarding the stream plots for the unstructured populations we consider the state of the population after a single run of the whole experience, however, transition probabilities correspond to an average of 10^3 runs for each of the possible combinations of states. All the other results correspond to an average of 200 runs for each experience.

The players' payoffs, as it was previously mentioned, are calculated according to the payoff matrix in **Table I**. The default values for the payoff matrix are: initial stake of a cooperative investor $c = 1$; amount returned by a cooperative trustee $\beta = 3$; initial stake multiplied by factor $r = 3$

corresponds to $b = 3 \times 1$; cost to a cooperative trustee $\gamma = 2$; and trustees' reputation $\upsilon = 0.5$. These values are the same for all the versions with the exception of: the asymmetric role assignment version, where we consider multiple values for the multiplication factor r (value by which the initial stake of an investor is multiplied before reaching a trustee); the diversity in reputation version where we consider different values for υ (the trustees' reputation).

Considering **Table I** with the values mentioned above, we can conclude that there is no Nash equilibrium in pure strategies. It is possible, however, to compute the Mixed-Strategy Nash Equilibria which corresponds to the mixed strategies $(0.2; 0.8)$, for the investors, and $(0.5; 0.5)$, for the trustees.

Lastly, concerning the evolutionary update, the probability of a random mutation occurring before imitation takes place is $\mu = 0.01$. The imitation occurs according to the Fermi function (equation 3), as defined in section IV, here, the default value for the intensity of selection is $\beta = 10$.

B. Results

We first start by trying to prove that the results in [29] are extendable to finite (unstructured) populations by comparing the evolutionary dynamics of both. Furthermore, we calculate the average values for both trust and trustworthiness, *i.e.* out of the 500 individuals in the population how many are cooperators when they play as an investor and how many are cooperators when they play as a trustee respectively, in order to compare the promotion of trust and trustworthiness between unstructured and structured populations.

1) Unstructured Populations: Firstly we considered the reduced Trust game in [29] which is very similar to the original version of the Trust Game introduced in [3] (does not take reputation into account). The results for this version show that, after the initial state, all the players become defectors, regardless of the role they are assuming.

Using our model we calculate the evolutionary dynamics for both the non-symmetric and symmetric versions of the reduced Trust Game, by simulating the game in all the possible states of the population, *i.e.* all the possible combinations of the players' strategies, and estimating for each one the most likely state it will transit to.

From the simulations with both the non symmetric and the symmetric version of the reduced Trust Game we concluded that regardless of the initial proportions of players with defective and cooperative strategies, the population tends to evolve to a state where all the players become defectors.

Next, we consider the Trust Game version that takes into account reputation. The results, for both the non-symmetric and the symmetric version in were similar to the ones in [29], showing that they are indeed extendable to finite populations and that reputation does have a positive effect regarding the promotion of trust and trustworthiness.

In order to verify the effects of introducing reputation into the model, we calculate the average number of individuals with cooperative strategies, over time, which allows to quantify the level of trust and trustworthiness in the population in these simulations. For the non symmetric version the average number of cooperative Investors and Trustees was 108.38 and 323.84 respectively. For the symmetric the results are rather similar, with an average 106.74 of cooperative investors and 322.49 cooperative trustees.

2) Structured Populations: Although results for Trust Game with reputation and unstructured populations, revealed a clear improvement regarding the promotion of trust and trustworthiness, the average value of trust is still low when compared to behavioral data resultant from experiments with real people like in [3]. In an attempt to more accurately emulate human behaviour in real life situations we did the same simulations just mentioned, however this time, using structured populations in a scale-free network.

a) Asymmetric role assignment: So far, in the unstructured version of the game, we considered that every player always plays as a certain role (or as both roles in the symmetric version), however, in real life situations that resemble the Trust Game, not every individual acts in a certain role with the same probability. One could say that this probability of playing as an Investor or Trustee depends on an individual's degree in the network considered. For instance, if we consider the network in which nodes are Uber's drivers and Uber's clients (let us consider Uber's drivers as trustees and the passengers as investors), where a driver is linked to every client he ever had (and vice versa), we may assume that the drivers are hubs in this network, *i.e.* the individuals with larger degrees in this network represent almost always trustees.

In this version, we try to apply the idea we just described to the trust game, Fig. 1 shows the results regarding the trust values and Fig. 2 the trustworthiness values. As detailed in section IV, for every interaction between two players we force a dependence between the players' role in that interaction and their degree in the network. This dependence is controlled by equation 4, namely through the value that the variable λ assumes. In our experiments we consider $\lambda = \{-2, -1, 0, 1, 2\}$. For $\lambda = 0$ the role assignment is attributed uniformly, therefore the game is played as up until now. Higher values (1 and 2) make the player with the higher degree more likely to be an investor and lower values(-2 and -1) more likely to be a trustee.

Additionally to the role assignment variation we just described we also consider three different levels of difficulty in the game, by varying the multiplication factor of the investors' initial stake.

b) Diversity in the reputation: In the previous version, we studied the influence of an individual's degree in the network on his most probable role. In regard to reputation, however, we considered that all the players have the same value, *i.e.* if we consider Table I all the individuals have the same υ. Yet, when picturing real life situations it is easy to think that individuals with a higher degree will have, most of the times, a higher probability of having a known reputation as well. For instance, if we consider the sellers in Alibaba retail store as nodes in a network, that are linked to every client they ever had, most likely, the ones with a higher degree in

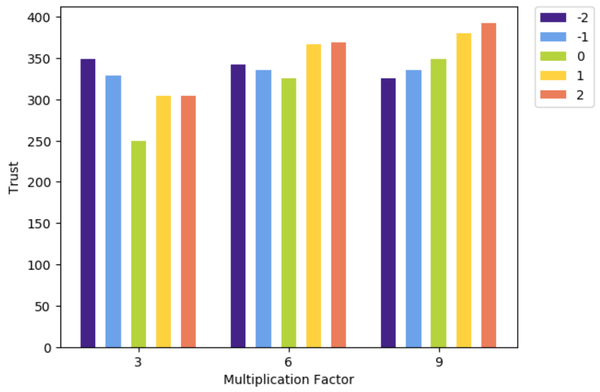

Fig. 1. Average number of cooperative Investors over time by varying the multiplication factor (r) and λ (equation 4), each colour represents a different λ value

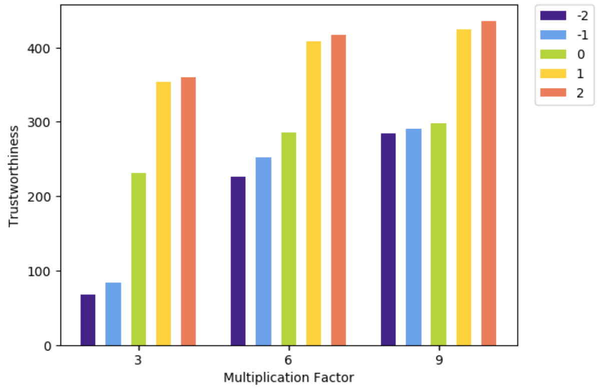

Fig. 2. [Average number of cooperative Trustees over time by varying the multiplication factor (r) and λ (equation 4), each colour represents a different λ value

the network will also have more reviews and, as such, a higher probability of having a public reputation.

In this version, we return to a symmetric game (players play as both roles and their payoff corresponds to a sum of the payoff as each role) and we consider three different situations regarding the reputation values distribution:

- a uniform distribution where all the players are assigned the exact same reputation value, $\upsilon = 0.5$

- a distribution taking into consideration players' degree in order to replicate what was just described, where we pick two different values for reputation, $\upsilon_1 = 0.1$ and $\upsilon_2 = 0.9$, assign υ_1 to the half of the individuals with the lowest degrees and υ_2 to the half with highest degrees, assuring this way that the average reputation value remains the same

- a random distribution of the same υ_1 and υ_2 values, while assuring again that the average reputation value stays the same, *i.e.* a random half of the population gets assigned υ_1 and the remaining individuals υ_2

c) Hybrid societies with pathological players: In this version we follow the main idea of the previous one, however, rather than assuming that players with a higher degree have a higher reputation as well, we will instead assume that they will always cooperate, regardless of the role they are playing as, and therefore are denominated pathological players [23] (also named resilient [17]). In a real life context, this means we presume that individuals with really large degrees in the network, *e.g.* the British online retailer Asos, most likely always cooperates (in this example it would mean, never deceive the clients and always send the product they asked for) when compared to individual sellers in smaller retail shops.

Fig. 3 shows the results for both the pathological players chosen in function of their degrees and the ones chosen randomly regarding the average trust values, and Fig. 4 regarding the average trustworthiness values. The thresholds for players' degrees are 10, 20, 30, and 40 which correspond to 29, 11, 3, and 2 pathological players assigned respectively. Lastly, we should note that, for comparison reasons, for the regular version with no pathological players the average trust was 254 and the average trustworthiness 208.

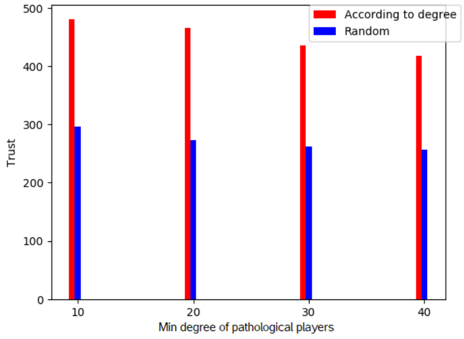

Fig. 3. Trust levels for pathological players either selected according to thresholds for degree or randomly

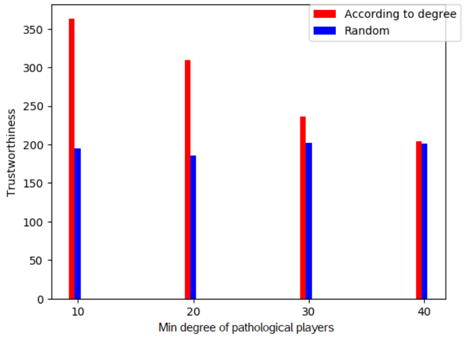

Fig. 4. Trustworthiness levels for pathological players either selected according to thresholds for degree or randomly

Since the results for this version of the Trust Game were the best ones regarding trust and trustworthiness (if we are only considering the regular multiplication factor of 3), particularly when looking at the smaller degree threshold considered, and therefore the most pathological players, we wanted to see how this would show in the stream plots for this version of the game, *i.e.* how it affects the most likely direction of evolution of the population.

In order to more precisely analyze the effects of the pathological players we also consider the stream plot for the regular version of the Trust Game with reputation played in a network, which can be seen in Fig. 5. Fig 6 shows the evolutionary dynamics for the Trust Game with pathological players selected in function of their degrees, considering the lowest degree threshold (29 pathological players).

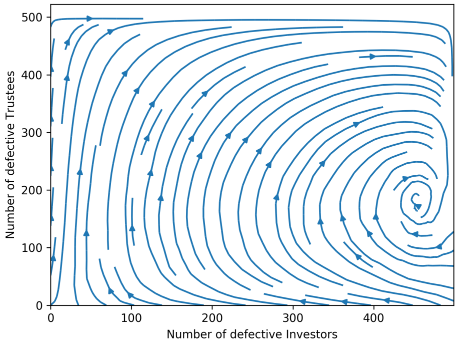

Fig. 5. Evolutionary Dynamics of the regular version of Trust Game with reputation for Structured Populations

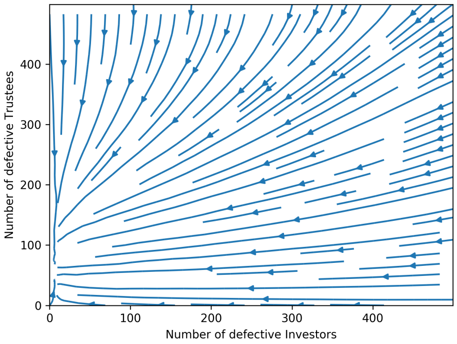

Fig. 6. Evolutionary Dynamics of the Trust Game version with (29) pathological players selected in function of their degree in the Network

C. Discussion

1) Unstructured Populations: Firstly, regarding the reduced Trust Game version, as predicted by the replicator dynamics and the Nash Equilibrium of the game presented in [29], for both the non-symmetric and the symmetric version of the game the results for trust and trustworthiness are 0 after the initial state of the game. This allows us to conclude that there are no major changes for this version when changing from infinite to finite populations.

Secondly, for the Trust Game version with reputation, introducing reputations had a positive result regarding the promotion of cooperation in both investors and trustees, particularly in the latter. If we only consider two possible strategies, *i.e.* either cooperate or defect, in the behavioral experiment in [3], namely with the version with social history (which is in a way mimicked by introducing reputation), we can see that about 89% of investors and 71% of trustees cooperated. In our simulations, results were 22% of the players cooperated when playing as investors and 65% when playing as trustees (on average, over time), showing there is still a considerable difference, mainly in regard to trust, and thus motivating us to once again add complexity to the model, namely by structuring the population in a SF network. Lastly, one should note that doing this comparison is not completely fair since: firstly Berg et al. [3] uses a relatively small number of subjects; secondly, in our simulations, we only consider two possible strategies, a player either cooperates or defects, however, in [3] investors can choose the precise amount they want as an initial stake and trustees can do the same for the return amount.

2) Structured Populations: As in the previous section, here we divide the discussion in the three versions considered with structured populations.

a) Asymmetric role assignment: Let us first start by analysing the effects of only structuring the population in an SF network, *i.e.* for a multiplication factor of 3 and a $\lambda = 0$ in equation 4. For these values the results were that on average, (over time) 50% of the players playing as investors and 46% playing as trustees acted cooperatively, meaning that trust levels increased considerably (more than double) and trustworthiness values decreased by 15%.

Regarding the multiplication factor, when its value increases, the game becomes ”easier”, since the payoff for cooperative trustees increases proportionally. As the multiplication factor increases so does the number of trustworthy trustees, which may be explained by the fact that the payoff difference between cooperating and defecting decreases. This happens regardless of the λ considered, however, for higher values of λ the trustworthiness results are considerably better (Fig. 2), *i.e.*, contrarily to the Uber’s example we gave before when the players with the larger degree in the network act as investors with a higher probability it favors the promotion of trustworthiness. Furthermore, increasing these players’ probability of playing as trustees seems to have a negative effect.

Lastly, one would think that for higher λ values the trust values would be higher as well since a larger number of cooperative trustees means investors will have a better payoff when adopting a cooperative strategy as well. Results for trust (Fig. 1), however, show that for lower multiplication factors, contrarily to intuition, results are the opposite of the trustworthiness results, *i.e.* players with larger degree acting as trustees with a higher probability (negative λ values) favors the promotion of trust. When the multiplication factor increases, however, having the larger degree players playing as investors with a higher probability increases the trust values, as we expected.

b) Diversity in the reputation: For this version, there were not any major improvements in terms of promotion of trust and trustworthiness. However, there are still some differences, namely, the variation where reputation is assigned while taking into account players’ degrees shows better results, in regard to both trust and trustworthiness values, than the other two.

One explanation for these differences would be that, even though we are keeping the average reputation in the network the same, the fact that we are assigning higher reputation values to the players with the larger degree increases the probability of these players cooperating when playing as trustees since now their payoff when interacting with cooperative and defective (by having higher reputation values, defective investors will still trust them with a significant amount) investors is considerably closer than before. Because these players have a larger degree, they also have a higher probability of being chosen in the imitation process, thus increasing the number of cooperative trustees. As a consequence, this increase may in turn influence positively the number of cooperative investors, since more cooperative trustees makes cooperating more profitable.

c) Hybrid societies with pathological players: This last version of the game was the one where we obtained the best results (if only considering the results for the regular multiplication factor of 3), namely for the variation where the pathological players are assigned whilst taking into account players' degree. Even though this is also the only version where we force certain players to always cooperate, the number of players chosen as pathological is always really small when compared to the population size, *e.g.* for a threshold of 10 only 29 players are selected and for a threshold of 40 this number decreases to only 2 players our of 500.

Despite the reduced number of pathological players selected, as we can see in Fig. 3 and in Fig. 4, when these players are chosen while taking into account their degree the trust and trustworthiness values increase considerably, particularly for the first two thresholds considered. This increase is rather noticeable not only when compared to previous versions, but also when compared to the variation with the exact same number of pathological players chosen randomly, for which essentially results are the same regardless of the number of pathological players selected.

In regard to the evolutionary dynamics of the population, for the regular version of the Trust Game, with reputation, played in a network (Fig. 5), the dynamics are similar to the unstructured version with reputation. For the version with pathological players Fig. 6 (considering the lowest degree threshold), however, the most likely direction of evolution of the population is extremely different, namely, the population tends to evolve to a state where almost all the investors and about 80% of the Trustees have a cooperative strategy, showing that the introduction of pathological players in the population, when taking into account their degree during their assignment, has an extremely positive effect concerning the promotion of trust and trustworthiness.

One explanation for this major increase in the number of cooperative players of both roles would be the fact that the players with the larger degree, as mentioned before, have a higher probability of being chosen in the imitation process, furthermore because they are also linked to more players they have the potential to have higher payoffs making them more likely to be imitated by other players. Due to both these reasons, and by forcing these players to always cooperate we are increasing the chances of other players (non pathological players) adopting a cooperative strategy as well.

VI. Conclusion

With the objective of more accurately modeling human behavior in the trust game, various studies were done where researchers add complexity to the original model in different ways. Some of the more important (and the ones we mainly focused on) were: considering altered payoff matrices to account for reputations, evolutionary game theory, and different populations structures. Thus, the main question that motivated this work was ”which mechanisms explain the promotion of trust and trustworthiness in the context of the trust game?”.

We propose a new model (detailed in section IV) where we use the Trust Game payoff matrix with reputations, introduced by Sigmund in [29], and apply it to a 2-player version of the game with finite populations, using evolutionary game theory as a framework.

We firstly studied our model when applied to unstructured populations, although other works using different models, as in [16], already consider reputation in an unstructured population context. Secondly, we introduce a new version where we consider reputation in a population structured in a scale-free network.

The structuring of populations allowed us to study some new components exclusive to populations structured in networks. Firstly, we studied the effects of forcing a dependence between a player role and his degree, controlled by equation 4, while varying the difficulty of the game. We concluded that the setup that is most favorable for the promotion of trust is: for a more difficult game having the higher degree players act as trustees; for a lower difficulty game having the higher degree players act as investors. Concerning trustworthiness, we concluded that having the higher degree players play as investors, regardless of the game difficulty, is always the most favorable. Secondly, we studied the effects of different distributions of the reputations values in the population, we concluded that the distribution where we take into account the players' degrees when assigning the reputation values is the one that most favors the promotion of both trust and trustworthiness. Lastly, we studied the effects of introducing pathological players in the population. We concluded that by selecting the higher degree players to be pathological players the values for both trust and trustworthiness increase considerably.

References

[1] Abbass, H., Greenwood, G., and Petraki, E. The n-player trust game and its replicator dynamics. *IEEE Transactions on Evolutionary Computation 20* (2015), 1–1.

[2] Barabási, A.-L., and Albert, R. Emergence of scaling in random networks. *Science 286*, 5439 (1999), 509–512.

[3] Berg, J., Dickhaut, J., and McCabe, K. Trust, reciprocity, and social history. *Games and Economic Behavior 10*, 1 (1995), 122 – 142.

[4] Bohnet, I., and Zeckhauser, R. Trust, risk and betrayal. *Journal of Economic Behavior & Organization 55*, 4 (2004), 467 – 484.

[5] Chica, M., Chiong, R., Adam, M. T. P., Damas, S., and Teubner, T. An evolutionary trust game for the sharing economy. In *2017 IEEE Congress on Evolutionary Computation (CEC)* (2017), pp. 2510–2517.

[6] Chica, M., Chiong, R., Kirley, M., and Ishibuchi, H. A networked$N$-player trust game and its evolutionary dynamics. *IEEE Transactions on Evolutionary Computation 22*, 6 (2018), 866–878.

[7] Chica, M., Chiong, R., Ramasco, J., and Abbass, H. A. Effects of update rules on networked n-player trust game dynamics. *CoRR abs/1712.06875* (2017).

[8] Fehr, E. On the economics and biology of trust. *Journal of the European Economic Association 7* (2009), 235–266.

[9] Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. Measuring trust*. *The Quarterly Journal of Economics 115*, 3 (2000), 811–846.

[10] Hofbauer, J., and Sigmund, K. *Evolutionary games and population dynamics*. Cambridge university press, 1998.

[11] Huyck, J. B. V., Battalio, R. C., and Walters, M. F. Commitment versus discretion in the peasant-dictator game. *Games and Economic Behavior 10*, 1 (1995), 143 – 170.

[12] Imhof, L. A., Fudenberg, D., and Nowak, M. A. Evolutionary cycles of cooperation and defection. *Proceedings of the National Academy of Sciences 102*, 31 (2005), 10797–10800.

[13] Johnson, N. D., and Mislin, A. A. Trust games: A meta-analysis. *Journal of Economic Psychology 32*, 5 (2011), 865 – 889.

[14] Kerr, B., Riley, M. A., Feldman, M. W., and Bohannan, B. J. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature 418*, 6894 (2002), 171.

[15] Malhotra, D. Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes 94*, 2 (2004), 61 – 73.

[16] Manapat, M. L., Nowak, M. A., and Rand, D. G. Information, irrationality, and the evolution of trust. *Journal of Economic Behavior & Organization 90* (2013), S57 – S75.

[17] Mao, A., Dworkin, L., Suri, S., and Watts, D. J. Resilient cooperators stabilize long-run cooperation in the finitely repeated prisoner's dilemma. *Nature communications 8*, 1 (2017), 1–10.

[18] Nash, J. F., et al. Equilibrium points in n-person games. *Proceedings of the national academy of sciences 36*, 1 (1950), 48–49.

[19] Nowak, Martin A., A. S. C. T., and Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature 428(6983)* (2004), 646–650.

[20] Nowak, M. A., and May, R. M. Evolutionary games and spatial chaos. *Nature 359*, 6398 (1992), 826.

[21] Nowak, M. A., and Sigmund, K. Evolutionary dynamics of biological games. *Science 303*, 5659 (2004), 793–799.

[22] Osborne, M. J., et al. *An introduction to game theory*, vol. 3. Oxford university press New York, 2004.

[23] Pacheco, J. M., and Santos, F. C. The messianic effect of pathological altruism. *Pathological Altruism* (2011), 300.

[24] Rand, D. G., Nowak, M. A., Fowler, J. H., and Christakis, N. A. Static network structure can stabilize human cooperation. *Proceedings of the National Academy of Sciences 111*, 48 (2014), 17093–17098.

[25] Santos, F. C., and Pacheco, J. M. Scale-free networks provide a unifying framework for the emergence of cooperation. *Physical Review Letters 95*, 9 (2005), 098104.

[26] Santos, F. C., Pacheco, J. M., and Lenaerts, T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences 103*, 9 (2006), 3490–3494.

[27] Santos, F. C., Santos, M. D., and Pacheco, J. M. Social diversity promotes the emergence of cooperation in public goods games. *Nature 454*, 7201 (2008), 213–216.

[28] Santos, F. P., Santos, F. C., and Pacheco, J. M. Social norms of cooperation in small-scale societies. *PLoS computational biology 12*, 1 (2016), e1004709.

[29] Sigmund, K. *The calculus of selfishness*. Princeton University Press, 2010.

[30] Smith, J. M. *Evolution and the Theory of Games*. Cambridge university press, 1982.

[31] Smith, J. M., and Price, G. R. The logic of animal conflict. *Nature 246*, 5427 (1973), 15.

[32] Traulsen, A., Nowak, M. A., and Pacheco, J. M. Stochastic dynamics of invasion and fixation. *Phys. Rev. E 74* (Jul 2006), 011909.