Games **2015**, *6*, 214-230; doi:10.3390/g6030214

OPEN ACCESS
games
ISSN 2073-4336
www.mdpi.com/journal/games

Article

Fairness and Trust in Structured Populations

Corina E. Tarnita

Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA;
E-Mail: ctarnita@princeton.edu; Tel.: +1-609-258-3896

Academic Editor: Martin A. Nowak and Christian Hilbe

Received: 5 June 2015 / Accepted: 13 July 2015 / Published: 20 July 2015

Abstract: Classical economic theory assumes that people are rational and selfish, but behavioral experiments often point to inconsistent behavior, typically attributed to "other regarding preferences." The Ultimatum Game, used to study fairness, and the Trust Game, used to study trust and trustworthiness, have been two of the most influential and well-studied examples of inconsistent behavior. Recently, evolutionary biologists have attempted to explain the evolution of such preferences using evolutionary game theoretic models. While deterministic evolutionary game theoretic models agree with the classical economics predictions, recent stochastic approaches that include uncertainty and the possibility of mistakes have been successful in accounting for both the evolution of fairness and the evolution of trust. Here I explore the role of population structure by generalizing and expanding these existing results to the case of non-random interactions. This is a natural extension since such interactions do not occur randomly in the daily lives of individuals. I find that, in the limit of weak selection, population structure increases the space of fair strategies that are selected for but it has little-to-no effect on the optimum strategy played in the Ultimatum Game. In the Trust Game, in the limit of weak selection, I find that some amount of trust and trustworthiness can evolve even in a well-mixed population; however, the optimal strategy, although trusting if the return on investment is sufficiently high, is never trustworthy. Population structure biases selection towards strategies that are both trusting and trustworthy trustworthy and reduces the critical return threshold, but, much like in the case of fairness, it does not affect the winning strategy. Further considering the effects of reputation and structure, I find that they act synergistically to promote the evolution of trustworthiness.

Keywords: Ultimatum Game; Trust Game; reputation; evolutionary game theory; population structure

1. Introduction

Game theorists have traditionally assumed that people act fully rationally to maximize their own financial gains. However, behavioral data has been inconsistent with these expectations. The Ultimatum Game (UG) and the Trust Game (TG) are two of the most influential examples of such inconsistent behavior, usually attributed to "other-regarding preferences" [1]. In the UG, two players have to divide a certain sum of money between them. One player (the proposer) makes an offer. The other player (the responder) can either accept the offer, in which case each receives the money as proposed, or reject the offer, in which case neither player receives anything. In a one-shot anonymous UG, a rational self-interested proposer will offer the minimum amount that she believes will be acceptable to the responder. A rational self-interested responder will accept any nonzero offer. Thus, the rational decision is for the proposer to make the minimum possible offer, and for the responder to accept it. To evaluate these predictions, many behavioral experiments have been conducted using the UG (see [1] for a review). Although there is considerable quantitative variation across studies, two clear qualitative deviations from rational self-interest are always observed: (i) many responders choose to reject low (but nonzero) offers; and (ii) many proposers offer more than the minimum amount required to avoid rejection.

In the TG, the investor begins with an initial amount of one monetary unit and can either keep it or transfer it (or some part of it) to the trustee. To represent the value created by interactions based on trust, whatever gets transferred is multiplied by a factor $b > 1$. The trustee then chooses how much to return to the investor. In a one-shot anonymous trust game, there is no reason for a self-interested trustee to return anything [2]. Hence, there is no reason for a self-interested investor to make the transfer, and the potential gains of trust and exchange are lost. In all behavioral experiments with the trust game, however, investors do make transfers and trustees return significant amounts [3]. These inconsistencies have similarly been attributed to "other-regarding preferences".

While these are perfectly acceptable proximate explanations, the question that still remains is: what is the source of these other-regarding preferences? [4] Evolutionary biologists have attempted to explain these phenomena from an evolutionary game theoretic perspective in which agents are not assumed to be rational; instead, they reproduce proportional to how well they do in the specific games, their offspring inherit their strategies, and natural selection chooses the winning strategies [5]. Traditional, deterministic models of evolutionary game theory agree with the classical game theoretic predictions: in the one-shot anonymous UG and TG, natural selection favors rationality: low offers and demands in the UG and no investment or return in the TG [6,7]. However, more complex evolutionary models have showed promising results. Rand and collaborators [8] showed using stochastic evolutionary game theory, where agents make mistakes when judging the payoffs and strategies of others, that natural selection can favor fairness in the UG. Gale and collaborators [9] have shown that assuming that learning processes are subject to constant perturbations and noise can lead to fairness. Page and collaborators [10] found that fairness can evolve in the UG in a spatial setting, in which interactions are non-random. Manapat and collaborators [7] showed that when individuals' strategies evolve in a context in which investors sometimes have knowledge about trustees' reputations before transactions, natural selection can favor both trust and trustworthiness.

Here I examine the same stochastic evolutionary setup proposed by [8], but extended to apply to a structured population playing the UG and the TG respectively. An underlying population structure simply means that individuals are not equally likely to interact with each other, but that interactions are more likely occurring with one's "neighbors" [11]. The neighbors could be geographical but they could also be individuals with similar strategies, preferences, physical features, cultural or genetic backgrounds etc. Population structure has been very powerful to explain the evolution of cooperation and social behavior both theoretically [12–14] and empirically [15,16]. However, its effects on fairness and trust have not been much studied with the exception of one promising experimental study in which a behavioral clustering mechanism—pairing trusting individuals with trustworthy ones—leads to an increase in the levels of cooperation compared to random pairings of investors and trustees [17]. I begin by developing a general theory and analytical results for the study of a continuous strategy space in structured populations, in the limit of weak selection. I then apply these general results to the study of fairness and the study of trust and trustworthiness with and without access to reputation.

2. General Model Description and Results

I consider a structured population of N players able to choose from a continuum of strategies situated on the n-dimensional hypercube. A strategy $\mathbf{p} \in [0, 1]^n$ is an n-dimensional vector of numbers between 0 and 1. The underlying population structure determines who interacts with whom to accumulate payoff and who competes with whom for reproduction. The expected payoff for an interaction between any two strategies is given by a function E determined by the specific game played. Individuals that accumulate higher payoff are more likely to reproduce: the rate of reproduction of each individual is proportional to 1 + δE, where δ is a constant that measures the intensity of selection. The higher the intensity of selection, the more likely agents with higher payoffs are to be imitated (to reproduce). At the extreme of $\delta \to \infty$, only those who obtain the highest payoff are imitated (strong selection). At the other extreme, $\delta \to 0$ selection is weak; in this case, all strategies have almost the same effective payoff and the dynamics is dominated by neutral drift. Weak selection is a natural situation that can arise in different ways: (i) payoff differences are small; (ii) strategies are similar; or (iii) individuals are confused about payoffs when updating their strategies. In such situations, the particular game makes only a small contribution to the overall reproductive success of an individual. Finally, reproduction is subject to symmetric mutation: with probability $1 - u$ the offspring inherits the strategy of the parent, but with probability u a random strategy is chosen. This process leads to a stationary distribution characterizing the mutation-selection equilibrium.

I am interested in the dynamics of this process for large but finite population size and weak selection, which here will mean δ $\ll 1/N$. In this case, although the frequencies of the strategies can widely fluctuate in time, all strategies have approximately the same abundance on average in the stationary distribution of the mutation-selection process. I am interested in the small deviations from this uniform distribution. I say that strategy $\mathbf{p}$ is favored on average in the mutation-selection equilibrium, if its abundance exceeds the mean. To calculate this deviation I use a perturbation theory in the selection strength, δ and, for a strategy to be selected, I require that the first order term with respect to δ is positive [8,18,19]. I show in Appendix A that this is equivalent to:

$$E_{\mathbf{p}} = \tilde{L}_{\mathbf{p}} + \sigma_2 \tilde{H}_{\mathbf{p}} > 0 \tag{1}$$

where

$$\begin{aligned} \tilde{L}_{\mathbf{p}} &= \int_{[0,1]^n} [\sigma_1 E(\mathbf{p},\mathbf{p}) + E(\mathbf{p},\mathbf{p}') - E(\mathbf{p}',\mathbf{p}) - \sigma_1 E(\mathbf{p}',\mathbf{p}')] \, d\mathbf{p}' \\ \tilde{H}_{\mathbf{p}} &= \int_{[0,1]^n} \int_{[0,1]^n} [E(\mathbf{p},\mathbf{p}'') - E(\mathbf{p}',\mathbf{p}'')] \, d\mathbf{p}' d\mathbf{p}'', \end{aligned} \tag{2}$$

and $E(\mathbf{p}',\mathbf{p}'')$ is the expected payoff that strategy $\mathbf{p}'$ receives from strategy $\mathbf{p}''$ in a given game. The parameters σ_1 and σ_2 are structural coefficients that need to be calculated for the specific evolutionary process under investigation [20]. These parameters depend on: the population structure that determines who interacts with whom; the update rule that determines whose strategy gets imitated and how; and the mutation rate. The first term, $\tilde{L}_{\mathbf{p}}$, integrates over all pairwise competitions that involve strategy $\mathbf{p}$, each pairwise comparison including the first structural coefficient, σ_1. Thus, σ_1 encapsulates how much more likely a strategy is to play its own kind than the opposite kind in a pairwise encounter; henceforth I will call σ_1 the pairwise structural coefficient and I will consider only structures for which individuals are more likely to interact with their own kind, *i.e.*, $\sigma_1 > 1$. The second term, $\sigma_2 \tilde{H}_{\mathbf{p}}$, evaluates the competition between strategy $\mathbf{p}$ and all other strategies simultaneously, weighted by the second structural coefficient σ_2. Thus, σ_2 captures the interaction in the other extreme case, when all strategies are simultaneously present in the population; henceforth I will call σ_2 the mean structural coefficient and I will consider only structures for which $\sigma_2 > 0$.

In the limit of low mutation only one structural coefficient is needed and Condition (1) becomes:

$$E^0_{\mathbf{p}} = \int_{[0,1]^n} [\sigma_0 E(\mathbf{p},\mathbf{p}) + E(\mathbf{p},\mathbf{p}') - E(\mathbf{p}',\mathbf{p}) - \sigma_0 E(\mathbf{p}',\mathbf{p}')] \, d\mathbf{p}' > 0 \tag{3}$$

where σ_0 is the low mutation limit of $\sigma = (2\sigma_1 + \sigma_2)/(2 + \sigma_2)$, the structure coefficient for games with two strategies [20,21].

Finally, the most favored (optimal) strategy is determined by maximizing $E_{\mathbf{p}}$ (or $E^0_{\mathbf{p}}$ in the limit of low mutation).

3. Evolution of Fairness: Ultimatum Game

I model the ultimatum game by imagining two players who have to split an amount summing to unity. In any given interaction, players are randomly assigned to the roles of proposer and responder. An agent's strategy is given by the two dimensional vector $S = (p, r) \in [0, 1]^2$, where p is the amount offered when acting as proposer, and r is the minimum amount demanded when acting as responder, or the "rejection threshold." An offer p is accepted by a responder with the rejection threshold r if and only if $p \geq r$. Let $U(S_1, S_2)$ be the expected payoff that strategy $S_1 = (p_1, r_1)$ gets from strategy $S_2 = (p_2, r_2)$ in the ultimatum game. Since I assume that in the interaction between a player using strategy S_1 and a player using strategy S_2, each player can be in the role of the proposer with equal probability, $U(S_1, S_2)$ is given (up to a $1/2$ factor which I henceforth omit) by the function:

$$U(S_1,S_2)=\begin{cases} 1-p_1+p_2 & \text{if } p_1 \geq r_2 \text{ and } p_2 \geq r_1 \\ 1-p_1 & \text{if } p_1 \geq r_2 \text{ and } p_2 < r_1 \\ p_2 & \text{if } p_1 < r_2 \text{ and } p_2 \geq r_1 \\ 0 & \text{if } p_1 < r_2 \text{ and } p_2 < r_1 \end{cases} \tag{4}$$

Depending on whether $p \geq r$ or $p < r$, the payoff function $U(S,S)$ takes two different values (1 and 0 respectively) and hence I find the condition for strategy S to be favored by selection to be

$$\begin{aligned} E_S &= \tilde{L}_S + \sigma_2 \tilde{H}_S = \sigma_1 \left(I(p \geq r) - \frac{1}{2} \right) + p - 2p^2 + r - r^2 + \sigma_2 (p - p^2 - \frac{r^2}{2}) > 0 \\ E_S^0 &= \sigma_0 \left(I(p \geq r) - \frac{1}{2} \right) + p - 2p^2 + r - r^2 > 0 \end{aligned} \tag{5}$$

where $I(\mathit{condition})$ is one if *condition* is *true* and is zero if *condition* is *false*. In [8] it was shown that stochasticity and uncertainty can select for fairness in a well-mixed population. To recover their results I simply use the values of the structure coefficients for a well-mixed population, $\sigma_1 = 1$ and $\sigma_2 = Nu$. Adding population structure expands the region of strategies selected for to include increasingly more fair strategies (Figure 1). Already $\sigma \geq 2$ is sufficient to select for the entire $p \geq r$ region. However, not all strategies selected for have the same abundance in the stationary distribution, so next I focus on the optimum strategy (most abundant in the stationary distribution and hence, by the measure above, most favored by selection). Maximizing E_S and E_S^0 I conclude that the optimal strategy is achieved when $p \geq r$ and is given by:

$$(p_{\text{opt}}, r_{\text{opt}}) = \begin{cases} (\frac{1}{3}, \frac{1}{3}) & \text{if } u \to 0 \text{ or } 0 \leq \sigma_2 \leq 1 \\ \left(\frac{1+\sigma_2}{4+2\sigma_2}, \frac{1}{2+\sigma_2}\right) & \text{if } u >> 0 \text{ and } \sigma_2 > 1 \end{cases}$$

Note that for low mutation the pairwise structural coefficient σ_0 has no effect on the optimum strategy. Thus, in the limit of weak selection and low mutation the optimum strategy is $(1/3, 1/3)$ – one that offers 33% and also rejects any offer lower than 33% – which is the same optimum obtained for the well-mixed population [8]. As mutation increases, the second structural coefficient, σ_2, that captures the effect of mutation and structure in the center of the hypercube, starts to influence the optimal strategy, but only when σ_2 is larger than 1. However, the pairwise structural coefficient σ_1 continues to have no effect on the outcome. As σ_2 increases, the proposal increases and the rejection threshold decreases. For high σ_2, the most frequent strategy is $(1/2, 0)$; thus, the proposal is 50% and the rejection threshold is 0. This outcome is achieved in the high mutation limit, where all strategies are present in the population simultaneously with approximately equal frequency. Hence, the optimum strategy is the one that maximizes its expected absolute payoff against a randomly chosen opposing strategy. As has been shown previously [6,8], it is intuitive that this strategy is $(1/2, 0)$ since the offer $p = 1/2$ maximizes the proposer's expected payoff of $p(1-p)$ when playing against a randomly chosen opponent; the demand $r = 0$ maximizes the expected payoff as responder because any nonzero demand results in lost profit.

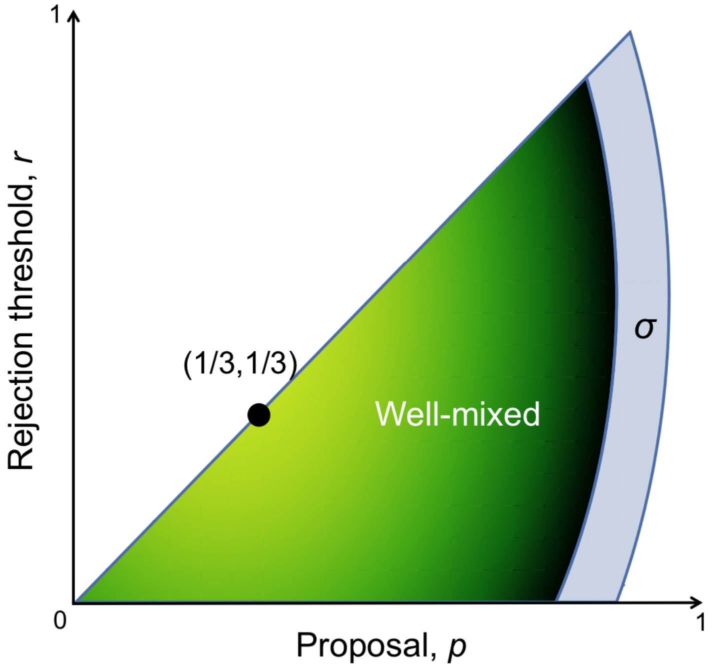

Figure 1. Structure increases the region of strategies that are selected for: light blue region gets added to the yellow-green gradient region corresponding to a well-mixed population. Results are shown for weak selection with $\sigma_0 = 1.5$; here the optimum strategy remains the same $(1/3, 1/3)$.

The important conclusion for the ultimatum game is that, in the limit of weak selection, the population structure has an impact on the space of strategies that get selected, increasing the density of fair strategies, but has little-to-no impact on the optimum strategy. The optimal strategies observed are very close to those selected for in the well-mixed population and depend only on the mean structural coefficient σ_2. This is because the pairwise σ_1 affects the overall effect of selection on a strategy $S = (p, r)$ via two terms: $U(S, S)$ which is either 0 if $p < r$ or 1 if $p \geq r$ (independent of the actual values of p and r) and $\int_{[0,1]^2} U(S', S')$ which is always equal to $1/2$. Thus the effect of structure on pairwise interactions is the same for all strategies above and below the diagonal, respectively. This therefore can not affect the relative ordering of the strategies; however, it can change the region of strategies that are selected for.

4. Evolution of Trust: Trust Game

I model the trust game by imagining two players that have an investor-trustee transaction. In any given interaction, players are randomly assigned to the roles of investor and trustee. An agent's strategy is given by the two dimensional vector $S = (p, r) \in [0, 1]^2$, where p is the amount invested when the agent acts as investor and r is the fraction returned when the agent acts as trustee. The investor begins with an initial stake of one unit and can choose to either keep the stake or transfer some fraction $0 \leq p \leq 1$ of it to the trustee. To represent the value created by interactions based on trust, the transferred amount is multiplied by a factor $b > 1$, which I will refer to as the return on investment. The trustee then chooses what fraction r of the enhanced transfer amount pb to return to the investor. Let $T(S_1, S_2)$ be the expected payoff that strategy $S_1 = (p_1, r_1)$ gets from strategy $S_2 = (p_2, r_2)$. Since I assume that in the interaction between a player using strategy S_1 and a player using strategy S_2, each player can be in the role of the investor with equal probability, $T(S_1, S_2)$ is given (up to a $1/2$ factor which I henceforth omit) by the function:

$$T(S_1, S_2) = 1 - p_1 + p_1 b r_2 + p_2 b(1 - r_1) \tag{6}$$

4.1. Well-Mixed Population

Here it is worth first investigating the results obtained in this stochastic game theoretic framework, in the limit of weak selection, in a well mixed population. This is the analog of the study conducted in [8] for the evolution of fairness. Although a very similar framework has been employed in [7], the analysis was not for weak selection. Using the payoff function in Equation (6) I find for any mutation:

$$E_S = \frac{2 + Nu}{2}\big(-br + p(b-2) + 1\big) \tag{7}$$

which is positive for the same set of strategies for which its low mutation version (obtained for $Nu \to 0$) is positive. Thus, mutation does not affect the selective outcome for a well-mixed population at weak selection. From this I conclude that any strategy with $0 \leq p \leq 1$ and $0 < r < 1/b + (1 - 2/b)p$ is selected for (Figure 2).

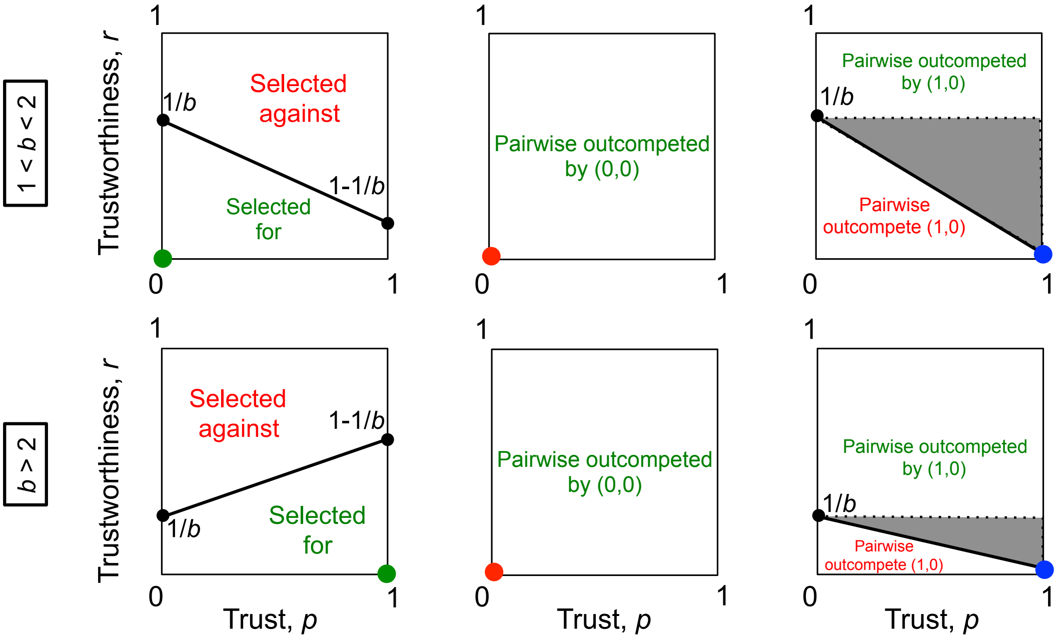
1 < b < 2
b > 2
Trustworthiness, r
Trust, p
1
0
1/b
1-1/b
Selected against
Selected for
Pairwise outcompeted by (0,0)
Pairwise outcompeted by (1,0)
Pairwise outcompete (1,0)

Figure 2. In a well-mixed population, in the limit of weak selection, both trust and trustworthiness are selected for and, for sufficient returns, the trusting but not trustworthy strategy, $(1, 0)$, is the optimal strategy. Upper panels: $1 < b < 2$. Lower panels: $b > 2$. Left panels: the most abundant strategy (green dot). Middle panels: strategy $(0, 0)$ outcompetes all other strategies. Right panels: strategy $(1, 0)$ (blue dot) outcompetes a large number of strategies in pairwise interactions; for a subset of these strategies, $(1, 0)$ wins more in pairwise interactions than does $(0, 0)$ (grey shaded area).

Thus, as was the case for fairness [8], stochasticity and mistakes alone can promote the evolution of some amount of trust and trustworthiness even in a well-mixed population. In particular, strategy $(1,0)$ that trusts but is not trustworthy is always selected for. This is because at weak selection all strategies have similar abundances and no strategies of the hypercube, even seemingly non-sensical ones – are a priori excluded from the game. Only natural selection eventually removes the strategies that do not perform well. In this case, strategies that display some trust and trustworthiness can win many pairwise encounters with strategies that are more trusting and trustworthy than they are (Figure 2). And although the $(0,0)$ strategy that neither trusts nor is trustworthy outcompetes all other strategies pairwise (Figure 2 middle panels), the $(1,0)$ strategy that trusts but is not trustworthy gains more from certain pairwise interactions than does $(0,0)$ (Figure 2 right panels). Therefore, on average, as long as the return is sufficiently high ($b > 2$), strategy $(1,0)$ can be the optimum strategy.

To conclude, for weak selection, stochasticity and mistakes can lead to the evolution of trust and low-to-intermediate levels of trustworthiness.

Finally, here I used a one-population formulation where individuals are equally likely to be found in the two roles. However, a two-population formulation in which one population is made of investors and the other of trustees can also be employed when both populations are well-mixed. Ohtsuki *et al.* [22] derived analytical conditions for a pair of strategies to be favored in the limit of weak selection in bimatrix well-mixed games and [8] found that the one-population and two-populations formulations yield identical results for the well-mixed ultimatum game. Applying the approach in [22] to the trust game between two well-mixed populations I similarly find that the two-populations formulation yields identical results to the one population formulation (Appendix B). However, no extensions have been made for the study of bimatrix games with population structure.

4.2. Structured Populations

Next, I move on to the study of population structure. Using the same payoff Function (6), I find for any mutation:

$$E_S = \frac{1}{2}\Big(- br(2+\sigma_2) + p(2b\sigma_1 + b\sigma_2 - 2(1+\sigma_1+\sigma_2)) + 1 + b - (b-1)\sigma_1 + \sigma_2\Big) \tag{8}$$

and in the limit of low mutation

$$E_S^0 = \frac{1}{2}\Big(- 2br + p(-2 + 2(b-1)\sigma_0) + 1 + b - (b-1)\sigma_0\Big) \tag{9}$$

Here σ_0 is the low mutation limit of the structural coefficient for a game with only two strategies $\sigma = (2\sigma_1 + \sigma_2)/(2+\sigma_2)$. From these I conclude that any strategy (p, r) with

$$\begin{aligned} &0 \le p \le 1 \\ &0 < r < \frac{1}{b} - \frac{(b-1)(\sigma-1)}{2b} + p\sigma\left(1 - \frac{2 - \frac{\sigma-1}{\sigma}}{b}\right) \end{aligned} \tag{10}$$

is selected for. For low mutation, one simply needs to replace σ by σ_0 in the above. Note that σ_1 and σ_2 do not influence the selection outcome independently. They only influence it together via σ. Population structure affects the well-mixed results in two ways. First of all, it changes the region of strategy space that is selected for: it selects for fewer trustworthy strategies that do not trust and it selects for more strategies that are both trusting and trustworthy (Figure 3). Overall, it is easy to check that the same total area gets selected for so that the density of strategies selected for does not change, only their types.

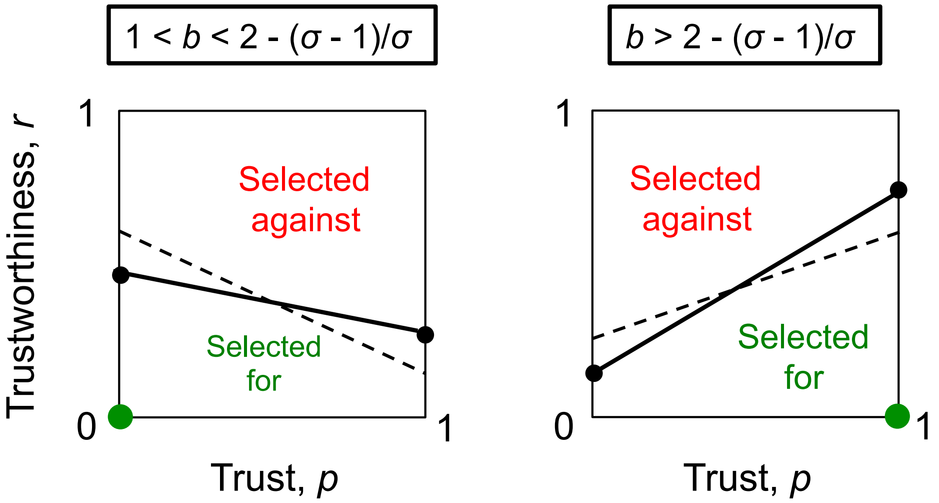

Figure 3. Population structure reduces the density of trustworthy strategies that do not trust and increases the density of trusting and trustworthy ones. Left panel: $1 < b < 2-(\sigma-1)\sigma$. The most abundant strategy (green dot) is $(0, 0)$ – no trust and no trustworthiness. Right panel: $b > 2 - (\sigma - 1)\sigma$. The most abundant strategy (green dot) is $(1, 0)$ – complete trust and no trustworthiness. In both panels, compared to well-mixed populations (region under dashed line) population structure selects for fewer trustworthy strategies that do not trust and for more trustworthy strategies that also trust (region under solid line).

Second, population structure affects the winning strategy. When maximizing Equation (8), the optimum strategy is

$$(p_{\text{opt}}, r_{\text{opt}}) = \begin{cases} (0,0) & \text{if } 1 < b < 2 - \frac{\sigma-1}{\sigma} \\ (1,0) & \text{if } b \geq 2 - \frac{\sigma-1}{\sigma} \end{cases}$$

provided that $\sigma_1 > 1$ and $\sigma_2 > 0$ (which are the general assumptions I make throughout this paper). To obtain the analog of the above for the limit of low mutation, one simply needs to replace σ with its low mutation limit, σ_0. Thus, although the optimum strategy is still either the non-trusting and non-trustworthy $(0, 0)$ or the trusting but non-trustworthy $(1, 0)$, as was the case in the well-mixed population, the population structure helps by reducing the critical threshold for the return b. Furthermore, it is worth noting that the threshold for evolving trust is likely to be smaller for low mutation since typically mutation weakens the effect of population structure (*i.e.*, $\sigma < \sigma_0$), as shown in the several cases studied in [21].

To sum up, population structure selects for more trust and trustworthiness and reduces the return needed for strategy $(1, 0)$ to be the optimal strategy. However, population structure does not lead to an optimal strategy that has some level of trustworthiness.

5. Evolution of Trust with Reputation

King-Casal *et al.* [23] and Manapat *et al.* [7] proposed to study a modification of the above trust game that includes reputation. In this context, at least some of the time, individuals know the reputation of their trustees and can choose to invest with them or not, depending on whether they are trustworthy or not. To capture this, I use a probability q that the investor knows the reputation of the trustee. This game can also be interpreted as a mix between the trust and the ultimatum games. Thus, with probability $1-q$ they play the same simple trust game as above. However, with probability q, an investor transfers money to a trustee only if their return fraction r is sufficiently high so that the investor is at least no worse off by making the transfer (*i.e.*, $r > 1/b$). Therefore, up to a factor $1/2$, the payoff function between two players using strategies $S_1 = (p_1, r_1)$ and $S_2 = (p_2, r_2)$ is:

$$\hat{T}(S_1, S_2) = (1-q)(1 - p_1 + p_1 b r_2 + p_2 b(1-r_1)) + \\ + q \begin{cases} 1 & \text{if } r_1 \leq 1/b \text{ and } r_2 \leq 1/b \\ br_2 & \text{if } r_1 \leq 1/b \text{ and } r_2 > 1/b \\ 1 + b(1-r_1) & \text{if } r_1 > 1/b \text{ and } r_2 \leq 1/b \\ br_2 + b(1-r_1) & \text{if } r_1 > 1/b \text{ and } r_2 > 1/b \end{cases} \tag{11}$$

Here because $\hat{T}(S, S)$ depends on how r compares to $1/b$, there will be different strategy spaces selected for depending on whether $r \leq 1/b$ or $r > 1/b$ (Appendix C). Overall, I find that adding reputation decreases the density of strategies with low levels of trustworthiness ($r \leq 1/b$) and increases the density of strategies with high levels of trustworthiness ($r > 1/b$) (Figure 4). Furthermore, I find that the optimum strategy is: for $r \leq 1/b$

$$(p_{\text{opt}}, r_{\text{opt}}) = \begin{cases} (0, 0) & \text{if } 1 < b < 2 - \frac{\sigma - 1}{\sigma} \\ (1, 0) & \text{if } b \geq 2 - \frac{\sigma - 1}{\sigma} \end{cases} \tag{12}$$

for $r > 1/b$

$$(p_{\text{opt}}, r_{\text{opt}}) = \begin{cases} (0, 1/b + \epsilon) & \text{if } 1 < b < 2 - \frac{\sigma - 1}{\sigma} \\ (1, 1/b + \epsilon) & \text{if } b \geq 2 - \frac{\sigma - 1}{\sigma} \end{cases} \tag{13}$$

As above, $\sigma = (2\sigma_1 + \sigma_2)/(2 + \sigma_2)$ is the structural coefficient for a game with only two strategies. To obtain the analog of the above for the limit of low mutation, one simply needs to replace σ by its low mutation limit, σ_0. Thus, having access to accurate reputation some of the time increases the pressure for the evolution of trustworthiness. In fact, the trustworthy strategy can even be the optimal one if its expected payoff is higher than that of the trusting strategy that is not trustworthy, *i.e.*, if $E_{(1,0)} < E_{(1,1/b+\epsilon)}$. For this, a sufficient condition is:

$$\frac{2}{3b} - \frac{\sigma - 1}{3b(2\sigma + 1)} \leq q \leq 1 \tag{14}$$

For a well mixed population ($\sigma = 1$) the above condition reduces to $q \geq 2/3b$. Since for a structured population with $\sigma_1 > 1$ and $\sigma_2 > 0$ it follows that $\sigma > 1$, (14) shows that the structure decreases the amount of reputation knowledge needed for trustworthiness to evolve. When the probability of knowing the reputation surpasses the critical threshold in (14), the winning strategy is one that trusts fully, $p = 1$, and is also moderately trustworthy, $r = 1/b + \epsilon$ where

$$0 < \epsilon < \min\left\{\sigma\frac{b\sigma - (1+\sigma)}{b(1+2\sigma)+(1+\sigma)}, 1 - \frac{1}{b}\right\} \tag{15}$$

The upper bound on ϵ is derived from satisfying Equation (14) (see Appendix C for details) as well as $r = 1/b + \epsilon \leq 1$.

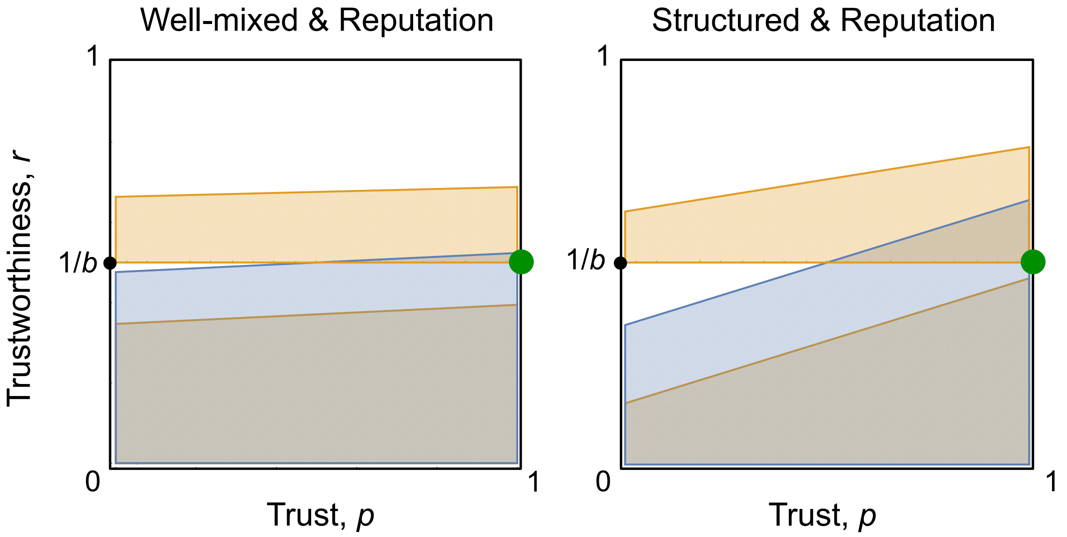

Figure 4. Reputation decreases the density of strategies with low trustworthiness ($r < 1/b$) that are selected for and increases the density of strategies with high trustworthiness ($r > 1/b$), both in a well-mixed and in a structured population. If there is sufficient knowledge of reputation then the winning strategy is one that is both trusting and trustworthy, $(1, 1/b + \epsilon)$. Light blue = strategies selected for in the absence of reputation; light orange = strategies selected for with reputation.

There are two interesting conclusions to note here: first, as for the simple TG, σ_1 and σ_2 do not influence the outcome independently. They only influence it together via σ. Second, the structure decreases both the lower bound for the necessary return b and, for a fixed b, it decreases the lower bound for the reputation recognition, q. So, for example, if the reward can be at most $b = 1.5$, one at least needs a structure with $\sigma = 2$ to allow for a trusting and trustworthy strategy to be the optimal one (according to Equations (12) and (13) above). Such a structure is the 3-regular graph in which every node has three neighbors, studied in the limit of low mutation. However, from Equation (14), $\sigma_0 = 2$ would require a reputation knowledge of at least 40% whereas a 2-regular graph (cycle) that has $\sigma_0 = 3$ for low mutation would only require 38% knowledge. So, for a fixed b, the higher the σ, the better. As was the case for the simple trust game, low mutation is most favorable since the existing examples suggest that $\sigma < \sigma_0$ [21].

6. Discussion

I have derived several main conclusions. First of all, in the limit of weak selection, stochasticity and mistakes can select for strategies with some amount of fairness, trust and trustworthiness in both the ultimatum and the trust games, with or without reputation, even in a well-mixed population. Intuitively this is because for weak selection all strategies have roughly the same abundances and fair or trusting strategies can sometimes do better in pairwise comparisons with other strategies than do unfair ones. Second, population structure improves the strategy space selected for: for the ultimatum game it extends the strategy space selected for to include increasingly more fair strategies, while for the trust game it selects for fewer strategies that are little-trusting and little-trustworthy and for more trusting and trustworthy strategies. Third, in the ultimatum game and trust game without reputation, structure does not affect the optimal strategy; in the trust game however, it decreases the critical threshold for the return on investment necessary for a trusting strategy to be the optimal one. The optimal strategy is however never trustworthy, regardless of the population structure. Finally, in the trust game with reputation, trustworthiness can be a feature of the optimal strategy. In this case, structure and reputation act synergistically such that having high enough σ requires less knowledge of reputation and *vice versa* (see Figure 5). For a well-mixed population, when $\sigma = 1$, I recover the results of [7].

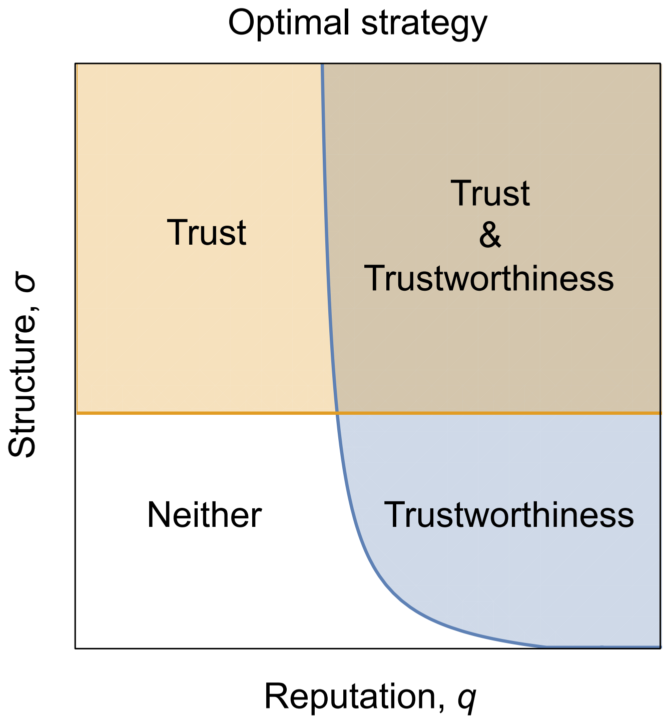

Figure 5. Structure and reputation act synergistically to allow the evolution of trust and trustworthiness. For a fixed return b, the optimum strategy can be both trusting and trustworthy only if both q and σ are large enough. The larger the σ, the less knowledge of reputation is necessary; the more likely to know the reputation, the less structure (smaller σ) is needed.

It is also worth noting that my final results for the trust game depend only on the σ, the structural coefficient for games with two strategies [21] and do not depend independently on the two structural coefficients σ_1 and σ_2 arising for games with at least three strategies. This is very convenient in terms of studying the evolution of trust in different structures, since it is much easier to calculate σ than σ_1 and σ_2 independently, and results already exist for many structures of interest [21].

In future work, it will be interesting to relax the present assumptions in several ways. First, the weak selection assumption can often obscure effects that are essential to the overall dynamics [24,25]. One interesting such possibility is the effect of structure for the evolution of fairness. While I found little-to-no effect in the limit of weak selection, [10] used simulations to find that structure with a type of local Birth-Death updating can indeed promote the evolution of fairness in the limit of strong selection. Reconciling these two findings would shed light on the mechanisms by which structure can play a significant role as well as on the interaction between structure and selection intensity.

Second, I assume that individuals are equally likely to be in both roles: proposer and responder in the ultimatum game and investor and trustee in the trust game. However, this is not always the case. Therefore, it would be interesting to study the effects of population structure on games between two populations, sometimes referred to as bimatrix games. When the two populations are well-mixed, [22] derived analytical conditions for a pair of strategies to be favored in the limit of weak selection and Rand *et al.* [8] applied these results to the study of the ultimatum game in a well-mixed population. They found identical results to the one-population approach I employed here. Furthermore, applying the approach in [22] to the trust game between two well-mixed populations I similarly find identical results here. However, no description exists for bimatrix games when the two populations are structured. In an experimental setting, [17] showed that considering a population of investors and one of trustees and exogenously introducing population structure by pairing investors and trustees with similar levels of trust and trustworthiness, can lead to the evolution of both traits. It is worthwhile considering an evolutionary update rule that would allow the interaction between two structured populations and inspecting whether the same results presented here will be obtained.

Finally, here the access to the trustees' reputation is assumed to be instantaneous. In [26] however Manapat and Rand introduce delays in information propagation and endow investors with memories, thereby allowing for the possibility that investors might face conflicting information about trustees. Similar directions can and should also be studied in a population structure context.

Acknowledgments

Funding was provided by the National Science Foundation grant EF-1137894.

Conflicts of Interest

The authors declare no conflict of interest.

A. Selection of Continuous Strategies in a Structured Population

To avoid very cumbersome notation I will show the proof for only two dimensions but the extension to the n-dimensional hypercube is a straightforward analogy. First I assume, as in [8], that the strategies do not cover the entire unit square, but in fact are only of the form $\boldsymbol{s} = (i/m, j/m)$ with $1 \leq i, j \leq m$ being integers. This discretizes the problem, making it possible to invoke previous results. To then go back to the continuous strategy space I simply take the limit $m \to \infty$.

Having turned the continuous problem into a discrete one, I am now interested in the stationary abundance of these discrete strategies. For this problem, I can use the result in [19] to conclude that, for large population size N, strategy $\boldsymbol{s}$ is favored by selection if $L_{\boldsymbol{s}} + \sigma_2 H_{\boldsymbol{s}} > 0$ where

$$\begin{aligned} L_{\boldsymbol{s}} &= \frac{1}{m^2} \sum_{i'=1}^{m} \sum_{j'=1}^{m} [\sigma_1 E(\boldsymbol{s}, \boldsymbol{s}) + E(\boldsymbol{s}, \boldsymbol{s}') - E(\boldsymbol{s}', \boldsymbol{s}) - \sigma_1 E(\boldsymbol{s}', \boldsymbol{s}')] \\ H_{\boldsymbol{s}} &= \frac{1}{m^4} \sum_{i'=1}^{m} \sum_{j'=1}^{m} \sum_{i''=1}^{m} \sum_{j''=1}^{m} [E(\boldsymbol{s}, \boldsymbol{s}'') - E(\boldsymbol{s}', \boldsymbol{s}'')] \end{aligned} \tag{16}$$

Here $\boldsymbol{s}' = (i'/m, j'/m)$ and $\boldsymbol{s}'' = (i''/m, j''/m)$ and $E(\mathrm{s}, \mathrm{s}')$ is the expected payoff that strategy s receives from strategy s$'$ in a given game. Moreover, [19] showed that the higher the quantity $L_{\boldsymbol{s}} + \sigma_2 H_{\boldsymbol{s}}$, the more the strategy s is favored by selection. Consequently, to determine which strategy is most favored by selection, one simply has to maximize $L_{\boldsymbol{s}} + \sigma_2 H_{\boldsymbol{s}}$.

Taking the limit $m \to \infty$ as in [18], the sums in Equation (16) converge to the integrals

$$\begin{aligned} \tilde{L}_S &= \int_0^1 \int_0^1 [\sigma_1 E(S, S) + E(S, S') - E(S', S) - \sigma_1 E(S', S')] \; dp' dr' \\ \tilde{H}_S &= \int_0^1 \int_0^1 \int_0^1 \int_0^1 [E(S, S'') - E(S', S'')] \; dp' dr' dp'' dr'', \end{aligned} \tag{17}$$

where $S' = (p', r')$ and $S'' = (p'', r'')$. It follows that the condition for strategy S to be favored by selection is $\tilde{L}_S + \sigma_2 \tilde{H}_S > 0$ and that the most favored strategy is determined by maximizing $E_S = \tilde{L}_S + \sigma_2 \tilde{H}_S$.

B. Trust Game: Two Well-Mixed Populations Formulation

For the ultimatum game in a well-mixed population, [8] showed that using a one-population formulation as above or a two-population formulation (in which half the population always plays the role of proposer and the other half always plays the role of responder) leads to identical results. Here I show that the same holds for the trust game. I only show the calculation for the simple trust game but the analysis is identical for the trust game with reputation. Consider two populations – one of investors, the other of trustees. Then, the payoff of an investor using strategy p from a trustee using strategy r is $E_I = 1 - p + pbr$ and the payoff of the trustee is $E_T = pb(1 - r)$. Using the extension of the result in [21] to continuous strategies employed by [8] I find that the condition for a pair of strategies (p, r) to be selected for is $L(p, r) + 2(N - 1)uH(p, r)$ where

$$\begin{aligned} L(p, r) &= \int_{[0,1]^2} \Big[E_I(p, r) + E_I(p, r') - E_I(p', r) - E_I(r', r')\Big] dp' dr' + \\ &+ \int_{[0,1]^2} \Big[E_T(p, r) - E_T(p, r') + E_T(p', r) - E_T(r', r')\Big] dp' dr' \\ H(p, r) &= \int_{[0,1]^2} \Big[E_I(p, r') - E_I(p', r') + E_T(p', r) - E_T(p', r')\Big] dp' dr' \end{aligned} \tag{18}$$

I find

$$L(p,r) + 2(N-1)uH(p,r) = (1+(N-1)u)(-br + p(b-2) + 1) \tag{19}$$

This is greater than zero for the same strategy space as derived in Section 4.1, Equation (7). Furthermore, optimizing, I obtain the same condition as for the one-population formulation:

$$(p_{\text{opt}}, r_{\text{opt}}) = \begin{cases} (0,0) & \text{if } 1 < b < 2 \\ (1,0) & \text{if } b \geq 2 \end{cases}$$

Thus, using the two population formulation for the trust game when both populations are well-mixed yields the same outcome as using the one population formulation.

C. Calculation of E_S for the Trust Game with Reputation

Because $\hat{T}(S,S)$ depends on whether $r > 1/b$, I find:

for $r \leq 1/b$

$$\begin{aligned} E_S &= \frac{1}{2}\Big(-b(1-q)(2+\sigma_2)r + p(1-q)(b(2\sigma_1+\sigma_2) - 2(1+\sigma_1+\sigma_2)) + C\Big) \\ E_S^0 &= \frac{1}{2}\Big(-2b(1-q)r + 2p(1-q)(b\sigma_0 - 2(1+\sigma_1)) + C_0\Big) \end{aligned}$$

and for $r > 1/b$

$$\begin{aligned} E_S &= \frac{1}{2}\Big(-b(1+q)(2+\sigma_2)r + p(1-q)(b(2\sigma_1+\sigma_2) - 2(1+\sigma_1+\sigma_2)) + C + \\ &\quad + q(1-\sigma_1 + b(1+\sigma_1+\sigma_2))\Big) \\ E_S^0 &= \frac{1}{2}\Big(-2b(1-q)r + 2p(1-q)(b\sigma_0 - 2(1+\sigma_1)) + C_0 + q(1-\sigma_1+b(1+\sigma_1))\Big) \end{aligned} \tag{20}$$

where C_0 and C are constants with respect to the strategy $S = (p,r)$. The trustful and trustworthy strategy can be the winning strategy if its expected payoff is higher than that of the trusting strategy that is not trustworthy, *i.e.*, if $E_{(1,0)} < E_{(1,1/b+\epsilon)}$. This is the case if and only if:

$$q \geq \frac{(1+b\epsilon)\frac{1}{\sigma}}{b(1+\frac{1}{\sigma}) - b\epsilon\frac{1}{\sigma} - 1} \tag{21}$$

which is less than 1 if

$$\epsilon \leq \frac{(b-1)(\sigma+1)}{2b} \tag{22}$$

A sufficient condition for q to satisfy Equation (21) is that

$$\frac{2}{3b} - \frac{\sigma-1}{3b(2\sigma+1)} \leq q \leq 1 \tag{23}$$

as long as

$$\epsilon < \sigma\frac{b\sigma - (1+\sigma)}{b(1+2\sigma)+(1+\sigma)} \tag{24}$$

References

1. Camerer, C.F. *Behavioral Game Theory: Experiments in Strategic Interaction*; Princeton University Press: Princeton, NJ, USA, 2003.
2. Berg, J.; Dickhaut, J.; McCabe, K. Trust, reciprocity, and social history. *Games Econ. Behav.* **1995**, *10*, 122–142.
3. Johnson, N.D.; Mislin, A.A. Trust games: A meta-analysis. *J. Econ. Psychol.* **2011**, *32*, 865–889.
4. Wilson, D.S.; Gowdy, J.M. Evolution as a general theoretical framework for economics and public policy. *J. Econ. Behav. Organ.* **2012**, *90*, S3–S10.
5. Maynard, S.J. *Evolution and the Theory of Games*; Cambridge University Press: Cambridge, UK, 1982.
6. Nowak, M.A.; Page, K.M.; Sigmund, K. Fairness versus reason in the ultimatum game. *Science* **2000**, *289*, 1773–1775.
7. Manapat, M.L.; Nowak, M.A.; Rand, D.G. Information, irrationality and the evolution of trust. *J. Econ. Behav. Organ.* **2013**, *90*, S57–S75.
8. Rand, D.G.; Tarnita, C.E.; Ohtsuki, H.; Nowak, M.A. Evolution of fairness in the one-shot anonymous Ultimatum Game. *Proc. Natl. Acad. Sci. USA* **2013**, *7*, 2581–2586.
9. Gale, J.; Binmore, K.G.; Samuelson, L. Learning to be imperfect: The ultimatum game. *Games Econ. Behav.* **1995**, *8*, 56–90.
10. Page, K.M.; Nowak, M.A.; Sigmund, K. The spatial ultimatum game. *Proc. Roy. Soc. B* **2000**, *267*, 2177–2182.
11. Lieberman, E.; Hauert, C.; Nowak, M.A. Evolutionary dynamics on graphs. *Nature* **2005**, *433*, 312–316.
12. Ohtsuki, H.; Hauert, C.; Lieberman, E.; Nowak, M.A. A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **2006**, *441*, 502–505.
13. Nowak, M.A. Five rules for the evolution of cooperation. *Science* **2006**, *314*, 1560–1563.
14. Nowak, M.A.; Tarnita, C.E.; Antal, T. Evolutionary dynamics in structured populations. *Phil. Trans. R. Soc. Lond. B* **2006**, *365*, 19–30.
15. Rand, D.G.; Arbesman, S.; Christakis, N.A. Dynamic social networks promote cooperation in experiments with humans. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 19193–19198.
16. Rand, D.G.; Nowak, M.A.; Fowler, J.H.; Christakis, N.A. Static network structure can stabilize human cooperation. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 17093–17098.
17. McCabe, K.A.; Rigdon, M.L.; Smith, V.L. Sustaining Cooperation in Trust Games. *Econ. J.* **2007**, *117*, 991–1007.
18. Antal, T.; Traulsen, A.; Ohtsuki, H.; Tarnita, C.E.; Nowak, M.A. Mutation-selection equilibrium in games with multiple strategies. *J. Theor. Biol.* **2009**, *258*, 614–622.
19. Tarnita, C.E.; Antal, T.; Nowak, M.A. Mutation-selection equilibrium in games with mixed strategies. *J. Theor. Biol.* **2009**, *261*, 50–57.
20. Tarnita, C.E.; Wage, N.; Nowak, M.A. Multiple strategies in structured populations. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 2334–2337.

21. Tarnita, C.E.; Ohtsuki, H.; Antal, T.; Fu, F.; Nowak, M.A. Strategy selection in structured populations. *J. Theor. Biol.* **2009**, *259*, 570–581.
22. Ohtsuki, H. Stochastic evolutionary dynamics of bimatrix games. *J. Theor. Biol.* **2010**, *264*, 136–142.
23. King-Casas, B.; Tomlin, D.; Anen, C.; Camerer, C.F.; Quartz, S.R.; Montague, P.R. Getting to know you: Reputation and trust in a two-person economic exchange. *Science* **2005**, *308*, 78–83.
24. Cavaliere, M.; Sedwards, S.; Tarnita, C.E.; Nowak, M.A.; Csikasz-Nagy, A. Prosperity is associated with instability in dynamical networks. *J. Theor. Biol.* **2012**, *299*, 126–138.
25. Wu, B.; Garcia, J.; Hauert, C.; Traulsen, A. Extrapolating weak selection in evolutionary games. *PLoS Comp. Biol.* **2013**, *9*, e1003381.
26. Manapat, M.L.; Rand, D.G. Delayed and Inconsistent Information and the Evolution of Trust. *Dyn. Games Appl.* **2012**, *2*, 401–410.