Not all Prisoner's Dilemma games are equal: Incentives, social preferences, and cooperation

Frederic Moisan, Robert ten Brincke, Ryan O. Murphy, Cleotilde Gonzalez

In Press: Decision

Not all Prisoner's Dilemma Games are Equal: Incentives, Social Preferences, and Cooperation

Frederic Moisan[1], Robert ten Brincke[2], Ryan O. Murphy[3], Cleotilde Gonzalez[1*]

[1]Dynamic Decision Making Laboratory
Department of Social and Decision Sciences
Carnegie Mellon University

[2] ETH Zürich

[3] University of Zürich

Author Note

This research was supported by the National Science Foundation Award number: 1530479 and by the Army Research Laboratory under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA) to Cleotilde Gonzalez. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.

[*]Correspondence concerning this article should be addressed to Cleotilde Gonzalez, Dynamic Decision Making Laboratory, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: coty@cmu.edu

Words: **5253 words**

Abstract

The Prisoner's Dilemma (PD) is a classic decision problem where two players simultaneously must decide whether to cooperate or to act in their own narrow self-interest. The PD game has been used to model many naturally occurring interactive situations, at the personal, organizational and social levels, in which there exists a tension between individual material gain and the common good. At least two factors may influence the emergence of cooperative behavior in this well known collective action problem: the incentive structure of the game itself, and the intrinsic social preferences of each of the players. We present a framework that integrates these two factors in an effort to account for patterns of high or low cooperation from repeated choice interactions. In an experiment using a collection of different PD games, and a measure of individual social preferences, we identify regions of PD games in which: (1) cooperation is independent of social preferences; (2) nice people can be exploited; and (3) being nice is consistently rewarded.

Keywords: prisoner's dilemma, cooperation, social preferences, social value orientation.

Introduction

Cooperation is an essential component in efficient social interactions. However, it requires that decision makers forgo some of their own potential gains in order to achieve better communal outcomes. This tension creates a social dilemma, a situation that pits an individual's self-interest against collective interests (Hardin, 1968; Dawes, 1980; Ostrom, Burger, Field, Norgaard & Policansky, 1999; Van Lange, Joireman, Parks & Van Dijk, 2013). The Prisoner's Dilemma (PD) is a social dilemma that archetypically represents this tension. Its simple structure makes clear that each player has a strictly dominant option, namely to choose defect (i.e., not to cooperate); and the logic of this "rational" choice, fueled by narrow self-interest, yields unambiguous but lamentable results: people may act in their own self-interest and this negates more efficient communal outcomes. This is the tension at the heart of collective action problems Ostrom (2014).

Fortunately however, universal defection is not observed when people make choices in social dilemmas and a sizable proportion of decision makers choose to cooperate instead (e.g., Axelrod & Hamilton, 1981; Dawes & Thaler, 1988; Camerer & Fehr, 2006; Rand, Dreber, Ellingsen, Fudenberg & Nowak, 2009; Martin, Gonzalez, Juvina & Lebiere, 2014). Two factors may influence the emergence of this cooperative behavior: (1) the particular payoffs in the PD game may reward cooperation in different ways (i.e., the "PD game" is not just *one* game but it is rather a constellation of potential 2-player games with particular properties) and (2) the social preferences that people hold may intrinsically compel them to care about the other's outcomes as well as their own, hence making cooperation more likely. Here, we present a coherent and integrated framework that brings together the incentive structures in the PD social dilemmas, and the intrinsic individual social preferences people have, to explain the emergence of cooperation

in repeated social interactions. The results of our empirical investigation reveal that not all PD games are equal in terms of elicited behavior. We identify patterns of high and low cooperation rates that emerge from different incentive structures in PD games, and also from individual social preferences (i.e., Social Value Orientation (SVO)) among decision makers. In particular we identify incentive structures in which decision makers with a high degree of SVO are exploited; and those incentives where high SVO players can thrive and earn greater payoffs through repeated cooperation.

Social preferences and cooperation

Social preferences are perhaps best explained when contrasted against Homo-economicus, an idealized decision maker posited to be narrowly self-interested and further believing that other decision makers are also narrowly self-interested (Aumann, 1976). A Homo-economicus' sole motivation is to maximize his/her own payoffs, indifferent to the payoffs of others, and further he/she expects the same from all other decision agents. Many economic models use Homo-economicus as a foundation, and game theoretic normative solutions are founded in part on this construct. In contrast, we consider here social decision agents who derive some personal pleasure from the positive payoffs that other decision makers receive. This "nicer" kind of decision maker is thus willing to make sacrifices for the benefit of others and hence promote the common good. A general utility framework for this kind of social decision maker can be expressed as:

$$u(\pi_s, \pi_o) = \pi_s + \alpha \times \pi_o \qquad (1)$$

Here a decision maker garners positive utility for the payoff to themselves π_s as well as some utility for the other player's payoff π_o. Homo-economicus is simply the special case where $\alpha = 0$. One can consider α as a representation a decision maker's degree of "niceness" or the

regard for other's outcomes as a result of his own actions (Gonzalez, Ben-Asher, Martin & Dutt, 2015). This construct is known as *Social Value Orientation* (SVO) in social psychology (Van Lange, 1999), *other regarding preferences* in economics (Cooper & Kagel, 2009), *welfare tradeoff ratios* in evolutionary psychology (Sell et al., 2009), and *altruism* in biology (Hamilton, 1964) as well as in economics (Becker, 1976; Simon, 1993; Fehr & Schmidt, 2006). Regardless of the moniker, widespread empirical work has established that the majority of human decision makers have some significant degree of positive social preferences and that α typically ranges between zero and one[1] (Messick & McClintock, 1968; Van Lange, De Cremer, Van Dijk & Van Vugt, 2007; Murphy & Ackermann, 2013). SVO can be measured as a reliable individual difference that captures these intrinsic social preferences (Murphy et al., 2011; Murphy & Ackerman, 2015).

The relationship between prosocial preferences (i.e., high SVO) and cooperation in PD games is sometimes (Balliet, Parks & Joireman, 2009), but not always (Kummerli, 2010; Kümmerli, Burton-Chellew, Ross-Gillespie & West, 2010; Burton-Chellew & West, 2013; Burton-Chellew, Nax & West, 2015), found to be positive. A complicating factor that may help explain the conflicting results is that "the" PD game actually represents an infinite set of potential payoff structures that may influence cooperation in different ways. Different PD games have different relative payoffs and different games yield different behavior from decision makers holding heterogeneous social preferences. In order to better understand how individuals with diverse intrinsic social preferences may influence each other when interacting repeatedly,

[1] In the literature SVO scores are sometimes reported in the metric of degrees, which is a consequence of legacy from the Ring Measure (Liebrand, 1984), which yielded output in terms of an angle. Transforming between the index α and an SVO angle score in degrees is straightforward as $\alpha = \tan(SVO°)$. The linear metric of α may be easier to understand than a trigonometric scaling of social preferences and that is in part why we use it here.

empirical investigation is needed and hence a major motivation for this paper. Beyond the theoretical view, it is pragmatically important and collectively beneficial to be able to create incentive structures that allow genuinely nice decision makers (i.e., those with high SVO) to promote the emergence and maintenance of cooperation. Similarly, it is important to identify the features of environments that preclude nice individuals from being exploited, and lastly to find ways in which those less inclined to cooperate may be encouraged to do so.

In the next section we define a comprehensive normalized space of PD games motivated by early theoretical work (Rapoport & Chammah, 1965; Axelrod, 1967). We also develop expectations regarding the relationship between SVO and the particular payoff structures of the games, and explicate how these factors would influence the emergence and sustainability of ongoing cooperation.

The normalized Prisoner's Dilemma space

In a standard 2x2 PD game (Figure 1), each player has two options and the game's payoffs must conform to the strict inequalities $T > R > P > S$. For each player, option D strictly dominates option C, but the outcome of both players choosing D leads to the payoff P for each player. This outcome is inefficient as both players could do better and receive R by each choosing C rather than D. A secondary condition of a PD game requires that $2R > T+S$ (Rapoport & Chammah, 1965), which ensures that the payoffs obtained from continued joint cooperation (CC) are greater than those obtained from alternating cooperation and defection (where T is earned on one round, then S on the next, both players alternate, and so on).

Early theoretical work (Rapoport & Chammah, 1965; Axelrod, 1967) proposed indices of anticipated cooperation that correspond to different payoffs in PD games (see Stivers, Murphy, & Kuhlman, 2017 for an overview). This theoretical approach is appealing because it reflects the

reality that people often actually choose to cooperate in PD games; and it is simple, as a prediction of cooperation is possible only from the relative differences in payoffs. Other approaches to rationalize cooperation could be more complicated. For example Homo-economicus would require a refinement of the assumption of "common knowledge of rationality" in order to cooperate in any finite PD game (Kreps et al., 1982). Also, specifying precisely what preferences and beliefs would replace narrow self-interest and the assumption of common knowledge in order to justify cooperation is tractable, but non-trivial (e.g., Murphy & Ackermann 2015).

Following a theoretical perspective of anticipated cooperation from normalized indices in the PD games, Figure 1 presents a configuration of all possible PD games using a straightforward transformation of PD games with scaled (normalized) payoffs as follows:

$$\boldsymbol{T' = \frac{T-S}{T-S} \equiv 1} \qquad \textbf{(2)}$$

$$\boldsymbol{R' = \frac{R-S}{T-S}} \qquad \textbf{(3)}$$

$$\boldsymbol{P' = \frac{P-S}{T-S}} \qquad \textbf{(4)}$$

$$\boldsymbol{S' = \frac{S-S}{T-S} \equiv 0} \qquad \textbf{(5)}$$

The best-known measure of anticipated cooperation is Rapoport's *K-index* (Rapoport, 1967), and in this notation it is defined as:

$$\boldsymbol{K = R' - P' = \frac{R-P}{T-S}} \qquad \textbf{(6)}$$

The K-index is designed to capture the severity of a game as it reflects the intuition that some PD games are "easier" than others, and hence cooperation is more likely to emerge in games with higher K values (Rapoport & Chammah, 1965; Roth & Murnighan, 1978) contrasted with games that have relatively low K values.

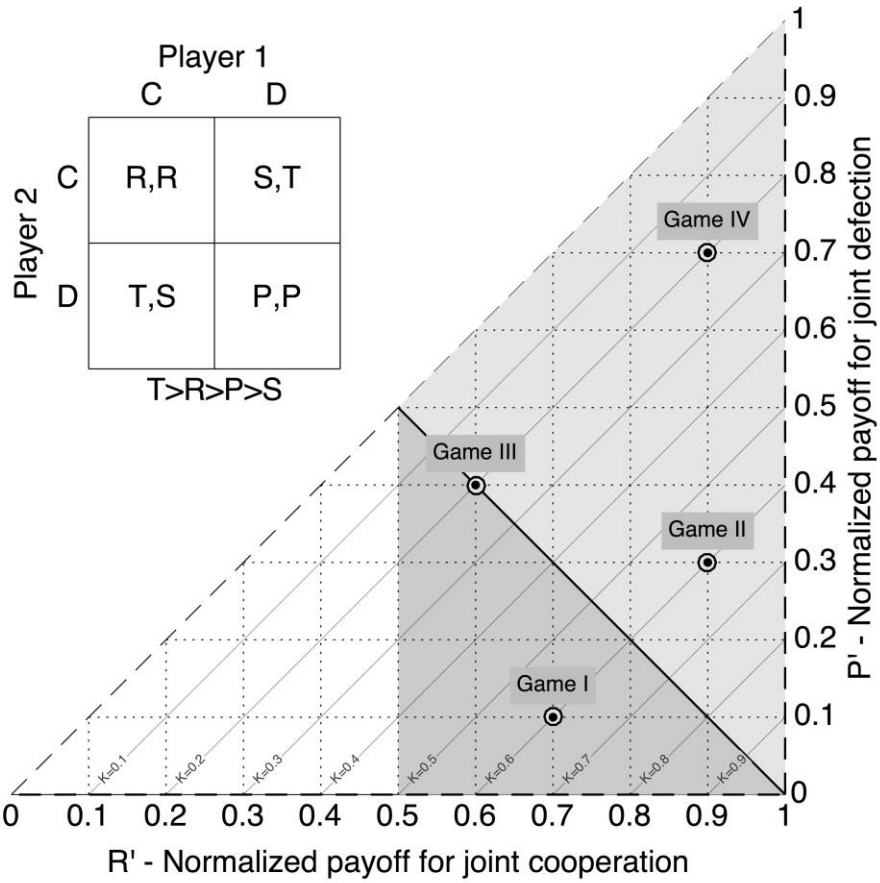

Figure 1: The normalized PD space. The set of PD games can be feature scaled (normalized) such that payoffs T = 1, S = 0, and R' and P' are the free parameters on the x and y axes respectively. Any point within the displayed triangle is a particular PD game. Games to the right of R' = 0.5, which correspond to any point within the shaded right trapezoid, conform to the additional payoff requirement that 2R > (T+S), an important restriction for iterated PD games. In any case, all PD games have the same defining characteristic of a Pareto-deficient pure strategy Nash Equilibrium. However some games are "easier" dilemmas than others, and indices like Rapoport's K-index reflect this. The four games used in the current experiment are indicated in the PD space and are labeled I-IV. Sustainable exploitation within heterogeneous pairs of players is more likely in games from the dark grey area (R+P < T+S) than in games from the light area (R+P > T+S). Behavior is expected to be constant over time in games located on the border between the dark and light gray areas (R+P = T+S).

One advantage of using these normalized payoffs is that *any* PD game can be defined with just two parameters (R' and P'), each bounded between 0 and 1, but not inclusive. A second advantage of this standardization is that it allows one to make theoretical predictions regarding the level of cooperation that are isomorphic to other well know indexes. From this theoretical

development we pose two questions. First, how is the exogenous payoff structure of the PD game related to emergent cooperation rates among players? Games with the same K-index (diagonals in Figure 1) are expected to result in the same level of cooperation between two players, and as the K-index increases, cooperation should also increase. Further, we can explore the separate potential effects of R' (which corresponds to the r_2 index from Rapoport & Chammah, 1965), on cooperation which, according to Rapoport & Chammah, is conjectured to be positive. Second, how are endogenous social preferences related to emergent cooperation rates among players? Previous work has produced mixed results and this could be in part due to the different payoff structures that were used. Through a laboratory experiment, we can isolate the independent effects of different incentive structures by randomly assigning dyads to different PD games. We can also examine what sort of incentive structures allow genuinely nice players to spread their prosocial influence and induce other players they interact with to cooperate more. The dynamics related to this question raise issues about reciprocity and learning that we address in the results section.

To address these questions we conducted an experiment that allows us to identify patterns of low and high cooperation rates across different PD games while concordantly measuring the players' social preferences. We selected four games across the normalized PD space, marked as Game I, II, III, and IV in Figure 1. These games allow us wide coverage of possible strategic interactions and do so in a way that allows us to disentangle the effects of the K-index from the effects of R' levels on cooperation. Game I and II share the same high K-index (K=0.6), while Games III and IV share the same low K-index (K=0.2). The expectation is that there would be more cooperation in Games I and II than in Games III and IV; and there would be similar levels of cooperation in Games I and II as well as in Games III and IV. Furthermore, Game III has the

lowest R' level (R' =0.6), followed by Game I (R' =0.7) and Games II and IV (R' =0.9). The expectation is that Games II and IV should yield the highest cooperation levels, followed by Game I and then Game III.

As highlighted above, incentives are not the sole driver of behavior and we expect intrinsic social preferences to matter as well. SVO reflects the extent to which a decision maker cares about the material welfare of others, which can be formally expressed through the index α in the utility function from equation (1). For any PD game, there exists a threshold $\bar{\alpha} = max(\frac{T-R}{R-S}, \frac{P-S}{T-P})$ such that, for any $\alpha > \bar{\alpha}$, choosing option C is always strictly preferred to option D, regardless of whatever beliefs a person has about the other's anticipated behavior[2]. Similarly, there exists a threshold $\underline{\alpha} = min(\frac{T-R}{R-S}, \frac{P-S}{T-P})$ such that, for any $\alpha < \underline{\alpha}$, choosing D is always strictly preferred to C, regardless of the beliefs about the other's anticipated behavior. However, whenever $\underline{\alpha} < \alpha < \bar{\alpha}$, SVO alone is not sufficient to determine the anticipated choice, and the behavioral prediction is contingent on a decision maker's expectations about the other player's behavior. Analytical work by Murphy & Ackermann, (2015) provides detailed predictions of such combined effects on cooperative choice in a number of PD games, including the ones selected for the current study. In particular, we have $\bar{\alpha} = \frac{3}{7}$ and $\underline{\alpha} = \frac{1}{9}$ in Games I and II (K=0.6), $\bar{\alpha} = \underline{\alpha} = \frac{2}{3}$ in Game III (K=0.2), and $\bar{\alpha} = \frac{7}{3}$ and $\underline{\alpha} = \frac{1}{9}$ in Game IV (K=0.2).

Assuming social preferences as in equation (1), three types of PD games can be distinguished:

---

[2] Option C is strictly preferred to option D whenever outcome CD is strictly preferred to outcome DD and outcome CC is strictly preferred to outcome DC. In these cases, preferences alone are sufficient to account for behavior as beliefs are inconsequential, assuming decision makers have social utilities consistent with equation 1.

1. If $R + P = T + S$ (the border line in Figure 1), then no dynamic behavior is expected. In fact, since $\underline{\alpha} = \bar{\alpha}$ in this case, people are expected to almost[3] always either strictly prefer C to D, or strictly prefer D to C across the repeated iterations of the game. In this case, mutual cooperation clearly requires that $\alpha > \bar{\alpha}$ for both players in the pair. If this condition is not satisfied, it may then lead to exploitation (CD or DC) in heterogeneous pairs (where players have different social preferences): the most prosocial player ($\alpha > \bar{\alpha}$) would keep cooperating while the most individualistic one ($\alpha < \underline{\alpha}$) would keep defecting. We test these predictions in our experiment via the Game III condition.
2. If $R + P > T + S$ (light grey area in Figure 1), social preferences cannot promote exploitation within a pair whenever $\underline{\alpha} < \alpha < \bar{\alpha}$ for at least one player whose preferred option directly depends on one's expectations of cooperation from the other player. More precisely, preferring to cooperate requires believing that the other player will also cooperate. As a result, mutual cooperation can be reached as long as R' is sufficiently large such that $\alpha > \underline{\alpha} = \frac{T-R}{R-S}$ for both players (e.g., if $\underline{\alpha} < \alpha_1 < \bar{\alpha}$ for Player 1 and $\alpha_2 > \bar{\alpha}$ for Player 2, then Player 1 will l$earn\ ov$er time that Player 2 always p$lays\ C, which\ w$ill lead Player 1 $to\ also\ sel$ect C). Similarly, mutual defection can be reached as long as P' is sufficiently large such that $\alpha < \bar{\alpha} = \frac{P-S}{T-P}$ for both players (e.g., if $\underline{\alpha} < \alpha_1 < \bar{\alpha}$ for Player 1 and $\alpha_2 < \underline{\alpha}$ for Player 2, then Player 1 will l$earn\ ov$er time that Player 2 always p$lays\ D, which\ w$ill lead Player 1 $to\ also\ sel$ect D). While mutual cooperation and mutual defection can be equally attractive in these games, we expect little effects of SVO whenever $R' + P'$ is large

[3] In the particular case where $\alpha = \underline{\alpha} = \bar{\alpha}$, one is indifferent between choosing either actions and may therefore play differently in different rounds.

(prosocials and individualists behaving alike since $\underline{\alpha} < \alpha < \bar{\alpha}$ is likely for both players)[4]. We test this hypothesis through games II and IV (both with a large R') in our experiment. The main difference between those two games is that Game IV is expected to be more belief dependent than Game II because of its larger gap between $\underline{\alpha}$ and $\bar{\alpha}$, which makes it more likely that $\underline{\alpha} < \alpha < \bar{\alpha}$ for both players than in Game II. One may therefore expect more convergence to outcome DD in game IV than in game II, the latter being less restrictive on unconditional cooperators.

3. If $R + P < T + S$ (dark grey area in Figure 1), then social preferences are expected to promote exploitation within heterogeneous pairs. As in the previous case, whenever $\underline{\alpha} < \alpha < \bar{\alpha}$, one's preferred option directly depends on one's expectations of cooperation from the other player. However, the difference with the previous case lies in that choosing to cooperate here requires believing that the other will defect[5]. Such a counterintuitive observation has significant consequences on the expected emergence of mutual cooperation within a pair: mutual cooperation indeed requires that $\alpha > \bar{\alpha} = \frac{T-R}{R-S}$ for both players here[6]. If instead $\alpha < \bar{\alpha}$ for at least one player, then exploitation outcomes CD and DC can be stable outcomes ($e.g.$, if $\alpha_1 < \bar{\alpha}$ for Player 1 $and\ \alpha_2 > \bar{\alpha}$ for Player 2, then Player 1 will learn over time that Player 2 always plays $C, which$ will lead Player 1 $to\ eventual$ly select D). As a result, whenever R' is not too large (such that $\alpha < \bar{\alpha}$ for at least one player in the pair), as in Game I, we expect a

[4] When instead $R' + P'$ are small (under the assumption that $R + P > T + S$), it becomes more likely that $\alpha < \underline{\alpha}$ or $\alpha > \bar{\alpha}$ for both players, and therefore constant exploitation can be expected in heterogeneous pairs.

[5] For more details, we refer to Table 3 in (Murphy & Ackermann, 2015).

[6] Similarly, mutual defection requires that $\alpha < \underline{\alpha} = \frac{P-S}{T-P}$ for both players.

significant influence of SVO on cooperation to occur where prosocial players would be exploited by individualistic players.

In summary, (1) no effect of SVO is expected whenever $R + P$ is sufficiently large and $R + P > T + S$ (e.g., Games II and IV), and (2) exploitation within heterogeneous pairs is predicted whenever R is sufficiently small[7] and $R + P \leq T + S$ (e.g., Games I and III). Note that the above theoretical analysis is consistent with the K-index conjecture previously described. Indeed, in all PD games, increasing K implies a decrease of both $\underline{\alpha}$ and $\bar{\alpha}$, thereby making unconditional cooperation more likely to emerge. Furthermore, our analysis reveals that *a priori* there is no PD game where heterogeneous pairs can benefit prosocial players more than they benefit individualistic players. Instead, the only predicted effect of SVO in this case is the exploitation of prosocials by individualists in some games. In other words, in theory, it never pays off to be a nice player when paired with an individualistic player. We therefore aim at testing these hypotheses in the following experiment.

Methods

A total of 220 American individuals were recruited through Amazon Mechanical Turk to voluntarily participate in an online study. A total of 172 people completed all parts of the study (78%) (Mean age= 33 years old, SD=11), and of those 74 participants were female (43%). The study was incentive compatible, with a minimum guaranteed payoff of $0.92 and a maximum attainable payoff of $3.60. This includes a show-up fee of $0.60. The actual average payoff was $2.17. Because the approximate duration to complete the study in full is approximately 20 minutes, these incentives are in line with the norm for online studies. Partial payment was awarded only to subjects whose matching partner left prematurely, and then only for those parts

[7] Note that the lower bound for R is 0.5 because of the assumption that $2R > T + S$.

of the study that these subjects had completed in full. All payoffs were displayed in points, where 100 points corresponds to 20 cents (a 1-to-5 ratio).

Procedure

An introductory screen informed participants of the general flow of the study. Participants were explicitly told that their choices have monetary consequences for both themselves as well the other participant with whom they would be matched. Basic demographic information was collected before the actual experiment began. Then participants completed the SVO Slider Measure that is comprised of 15 monetary allocations between the participant and an anonymous other party, an example shown in Figure 2. The measure yields an individual score that corresponds to α in Equation 1[8], and provides some insight into a decision maker's willingness to make costly tradeoffs between themselves and another person. Scores near zero indicate narrow self-interest whereas values closer to one indicate greater prosociality (i.e., niceness). As shown in the Appendix (Figure 6), the distribution of SVO scores in this experiment reveals a clear bimodal pattern that is routinely observed for this construct. Consistently with previous studies conducted in the laboratory, more extreme motivations such as competitiveness and altruism, which can be measured by the SVO Slider Measure, are rarely observed in our data (competitiveness corresponds to $SVO_o < 0$ whereas altruism corresponds to $SVO_o > 1$). We refer to Murphy et al. (2011) for further details regarding the SVO slider measure.

[8] As in Murphy et al., 2011, the SVO index can be computed based on the 6 primary money allocation tasks (the other 9 secondary slider items, which allow disentangling prosocial motivations of reducing inequality and maximizing joint payoff, have not been used in our analysis): $SVO_o = \frac{\boldsymbol{A_o - 50}}{\boldsymbol{A_s - 50}}$ where $\boldsymbol{A_s}$ and $\boldsymbol{A_o}$ represent the average allocations to the self and other respectively across the 6 primary slider items.

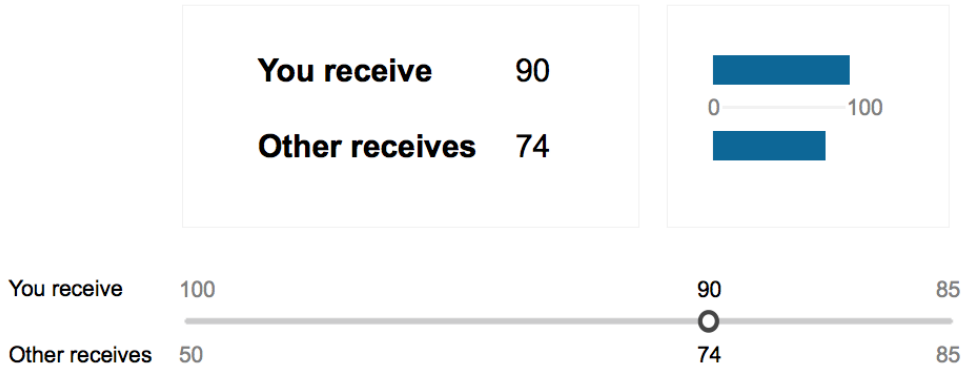

Figure 2. Example of one of the SVO Slider Measure items. A participant can move the slider to change the joint allocation of resources between herself and some other person, and then push a submit button to register the choice. In this decision task, the participant acts as a "dictator" as the other person has no choice in the matter of resource allocations. This technique provides a clean and reliable measure of how a decision maker weighs tradeoffs between her own payoffs and the payoffs of some other mutually anonymous person.

In part two of the experiment, participants were randomly matched in fixed dyads and each dyad was randomly assigned to one of the 4 previously specified PD games, repeatedly for 60 rounds. Payoffs were scaled such that the maximum range (the difference between T and S) is 100, where T=110 and S=10 in order to explicitly avoid a potential payoff of zero. These four games are shown in the matrices in Table 1.

Table 1. The four payoff scaled games used in the experiment, corresponding to the four games in Figure 1.

Game I (K=0.6)

	C	D
C	80, 80	10, 110
D	110, 10	20, 20

Game II (K=0.6)

	C	D
C	100, 100	10, 110
D	110, 10	40, 40

Game III (K=0.2)

	C	D
C	70, 70	10, 110
D	110, 10	50, 50

Game IV (K=0.2)

	C	D
C	100, 100	10, 110
D	110, 10	80, 80

Participants were not informed of the total number of rounds (60) to avoid end effects. However, participants knew the research session would last about 30 minutes in total. On each round, participants chose between options labeled A or B (the choice options were not explicitly

named “cooperate” or “defect”). The payoffs were displayed in matrix format, where the row of the matrix that corresponds to the respective option was highlighted when the mouse pointer hovered over the choice button. Once both participants had made a choice, the resulting payoffs, as well as the choices for both players, were displayed. However, no historical overview or summary of choices for earlier rounds was provided. Participants' earnings were allocated based on the sum of ten randomly selected rounds that were determined at the end of the experimental session.

Results

Rapoport's K-index predicts cooperation to be higher for Games I and II (each game as a K-index of 0.6) than for Games III and IV (each with a K-index of 0.2). Our results are consistent with this general prediction. The linear regression between the K-index and the mutual cooperation level is positive and significant (r=0.10, p=0.003)[9]. However, the K-index does not account for the more nuanced pattern of results. Table 2 shows the mean fraction of rounds in which players unilaterally choose to cooperate (*C rate*), and the mean fraction of rounds in which both players in a dyad cooperate bilaterally (*CC rate*). Note for example that Games I and IV have different K-values, but have similar levels of mutual cooperation (CC rate: t=0.17, p=0.86). Furthermore, games that have the same K-index (I and II; III and IV) have different cooperation levels (CC rate: t=2.07 and p=0.04; t=2.26 and p=0.03).

Table 2: Mean fraction of rounds in which players unilaterally choose to cooperate (C) and the mean fraction of rounds in which both players cooperate simultaneously (CC).

[9] Here the unit of analysis is the ratio of mutual cooperation over 60 rounds for each pair of matched subjects.

PD	K	R'	C rate	CC rate
I	0.6	0.7	0.55	0.43
II	0.6	0.9	0.75	0.68
III	0.2	0.6	0.30	0.17
IV	0.2	0.9	0.50	0.41

Using each player's choice in every round as the unit of analysis, we ran multilevel logistic regressions models that confirm the significant effect of K and R' on cooperative behaviour. See Table 3 for more details.

Table 3. Mixed-effect logistic regression models were fit to predict binary cooperation choices in terms of round, K and R as fixed effects (subject is the random effect) with likelihood ratio tests comparing models (M_B and M_C with M_A, M_D with M_C).

	Models							
	M_A		M_B		M_C		M_D	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Intercept	0.546	0.273	0.643	0.278	-1.736	0.568	-3.697	0.950
Round			-0.003	0.002				
K					5.600	1.259	35.907	1.247
R							1.284	0.505
df	2		3		3		4	
AIC	6860		6859		6844		6840	
BIC	6874		6880		6865		6868	
Log-likelihood	-3428		-3426		-3419		-3416	
χ^2			3.045		17.927		6.086	
df χ^2			1		1		1	
p			0.081		<0.0001		0.014	

There exists another organizing principle that helps explain the patterns of cooperation. Figure 3 shows that the correlation between SVO and the individual cooperation rate (over the whole 60 rounds) is positive for games with a lower R' and close to zero for games with a higher R'[10]. The normalized index R' moderates the regimes of the PD game where SVO is a useful explanatory variable. The interpretation is that for low R' games, the payoff for unilateral defection (T) is relatively larger than the payoff for mutual cooperation (R), whereas for high R'

[10] Using each player's choice in every round as the unit of analysis, running some multilevel logistic regressions similarly indicates that the probability of choosing to cooperate in any given round significantly increases with the SVO index only in games I and III. See Tables 11 and 12 in the appendix for more details.

games the payoff difference between mutual cooperation and unilateral defection is much smaller. A low R' game therefore offers a strong incentive to defect for self-regarding individuals, but this incentive is weaker for prosocial players. Meanwhile, even for almost perfectly self-regarding players, a high R' game offers a weak incentive to defect. The effect of this mechanism may be amplified by players considering the shadow of the future (Axelrod, 1984), knowing that their own defection may spawn the other player's defection (e.g., negative reciprocity) in subsequent rounds.

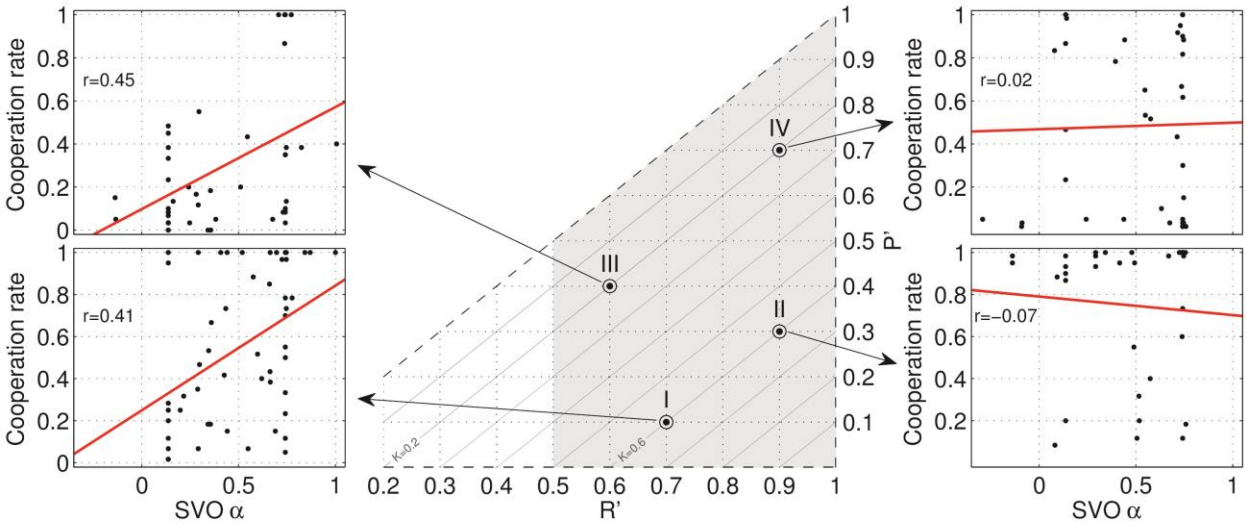

Figure 3: Effect of R' on the strength of SVO as an explanatory variable for cooperation. Scatter plots show SVO α versus individual cooperation rates (average over 60 rounds per participant), where each dot is a subject and the least squares fit is drawn in as a line. Note that only for Games I and III is there a significant relationship between SVO and cooperation. For the other games from the right side of the PD space there exists no such relationship between social preferences and bilateral cooperative choices. Detailed statistics are provided in Table 8 from the Appendix.

The above results show a relationship between SVO and cooperation for PD games with a lower R' but they do not shed light on whether cooperation is an effective strategy for prosocial players. When is it beneficial for prosocial players to act consistently with their intrinsic preferences? Figure 4 addresses this question by showing the correlations between SVO and

earnings, showing that it is only in Game I where prosocial players actually fare well. The average payoffs are higher for prosocials in Game I but not in Game III.

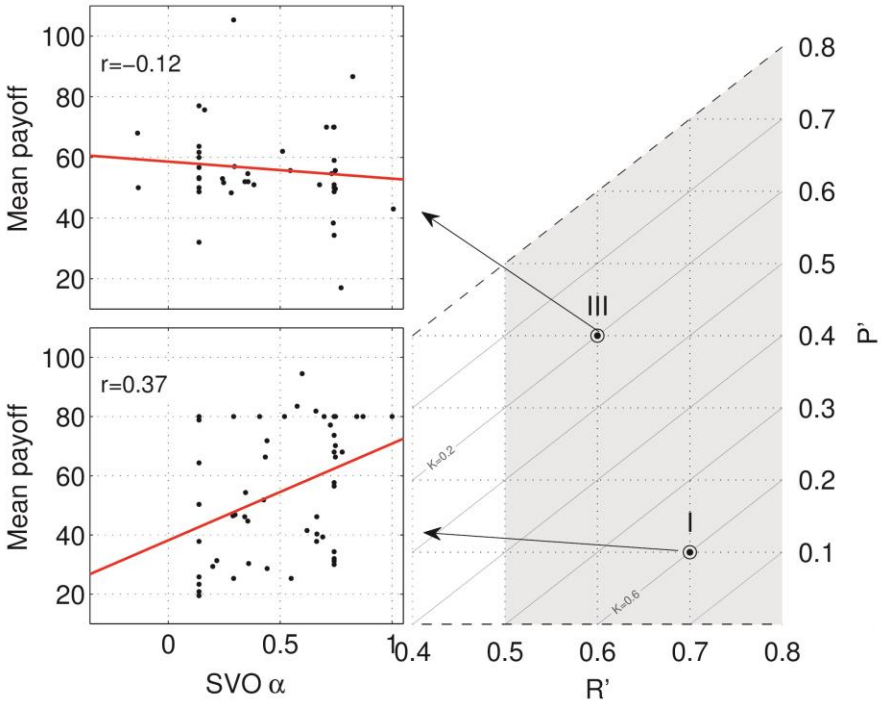

Figure 4: Relation between mean payoff (over 60 rounds per participant) and SVO. For Game I, subjects with higher SVO receive higher mean payoffs. For Game III, no such correspondence emerges. Detailed statistics are provided in Table 9 from the Appendix.

The combined analysis from Figures 3 and 4 suggests that prosocials can thrive in Game I—this is a case where the payoff for mutual defection (P) is relatively smaller than the payoff for unilateral defection (T). If the motives for prosocials lean towards cooperation, what does this tell us about the partner's social preferences? Do prosocials cooperate despite being paired to self-regarding players? Examining the combination of social preferences within pairs provides insights into this question. Table 4 reveals that in Game I the partner's SVO is as influential as one's own SVO (the β coefficient is similar for both variables), while in Game III cooperative behavior is only driven by each individual's own SVO (cooperation is largely independent of the partner's type).

Table 4. Multiple regressions predicting individual cooperation rate (over 60 rounds per participant) in terms of one's own SVO (SVO_o) and the partner's SVO (SVO_p) in Games I and III ($p<0.01$(**), $p<0.001$(***))

	Game I		Game III	
Variable	SE_β	β	SE_β	β
SVO_o	0.118	0.429***	0.142	0.441**
SVO_p	0.118	0.419***	0.142	0.210
R^2	0.349		0.263	
F	12.6***		6.61**	

Figure 5 shows the relationship between payoffs and the relative difference in SVO within pairs. The relative difference between own and partner's SVO has no relationship to payoffs in Game I but this relationship is negative in Game III. Payoffs in Game III are lower for prosocials when the difference between a player's SVO and the other player's SVO is large (r=0.36, p=0.024). This result indicates that participants with higher relative SVO compared to their partner tend to receive lower payoffs in Game III (i.e., a more prosocial member can be exploited by a more individualistic member of the pair), and the absence of this phenomenon in Game I is what allows prosocials to thrive (i.e., a more prosocial member in a dyad can induce an individualistic partner to cooperate).

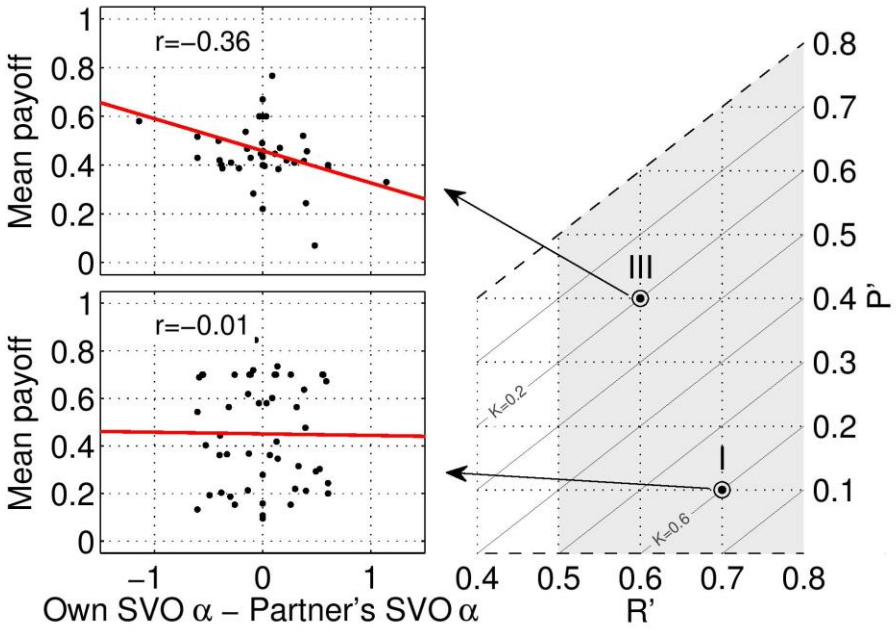

Figure 5. Relation between mean payoff (over 60 rounds per participant) and relative difference in SVO of Own and Partner's SVO in Games I and III. Detailed statistics are provided in Table 10 from the Appendix.

The above results suggest that in Games I and III prosocial people tend to be more generally cooperative than others. However, one may wonder whether prosociality improves all types of conditional cooperation. To answer this question, we consider two simple measures of reciprocal behavior: a person *positively reciprocates* by cooperating (choosing C) in the round following a cooperative move by the partner. Similarly, a person *negatively reciprocates* by defecting (choosing D) in the round following a defective move by the partner. If social preferences make people unconditionally more cooperative, then SVO should positively correlate with positive reciprocity and negatively correlate with negative reciprocity. Table 5 indeed indicates that positive reciprocity significantly increases with higher SVO in both Games I and III. However, Table 6 reveals that SVO has no significant effect on negative reciprocity. Instead, it suggests that people are more likely to negatively reciprocate in later rounds, regardless of their social preferences. This result simply means that although prosocial individuals appear willing to strive for mutual cooperation, they do not tolerate being exploited any more than individualistic people do.

Table 5. Mixed-effect logistic regression models fitted to predict positive reciprocity choices in Games I and III in terms of round and own SVO (SVO_o) as fixed effects (subject is the random effect) with likelihood ratio tests comparing models (PR_{Bx} and PR_{Cx} with PR_{Ax}).

	Models											
	Game I						Game III					
	PR_{A1}		PR_{B1}		PR_{C1}		PR_{A3}		PR_{B3}		PR_{C3}	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Intercept	1.814	0.556	1.960	0.580	-1.450	1.295	8.524	2.920	-0.128	0.546	-3.047	0.766
Round			-0.005	0.005					-0.001	0.007		
SVO_o					6.574	2.445					6.716	1.475
df	2		3		3		2		3		3	
AIC	930		931		924		565		567		546	
BIC	941		947		940		574		580		560	
Log-likelihood	-463		-463		-459		-280		-280		-270	
χ^2			0.862		7.993				0.021		20.29	
df χ^2			1		1				1		1	
p			0.353		0.005				0.884		<0.0001	

Table 6. Mixed-effect logistic regression models fitted to predict negative reciprocity choices in Games I and III in terms of round and own SVO (SVO_o) as fixed effects (subject is the random effect) with likelihood ratio tests comparing models (NR_{Bx} with $NR_{Ax,}$ NR_{Cx} with NR_{Bx}).

	Models											
	Game I						Game III					
	NR_{A1}		NR_{B1}		NR_{C1}		NR_{A3}		NR_{B3}		NR_{C3}	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Intercept	1.179	0.277	0.668	0.305	1.401	0.591	2.003	0.328	1.568	0.358	2.128	0.558
Round			0.017	0.005	0.017	0.005			0.014	0.005	0.014	0.005
SVO_o					-1.594	1.115					-1.407	1.075
df	2		3		4		2		3		4	
AIC	1201		1188		1188		1141		1134		1135	
BIC	1211		1204		1209		1152		1150		1156	
Log-likelihood	-598		-591		-590		-569		-564		-563	
χ^2			14.57		1.936				9.303		1.680	
df χ^2			1		1				1		1	
p			0.0001		0.164				0.002		0.195	

Discussion

In this study, we show how a combination of the exogenous incentive structure of PD games and the intrinsic social preferences of decision makers can facilitate or hinder the emergence of cooperation. Specifically, our results reveal the existence of three distinct phenomena: (1) some incentive structures can prevent prosocial individuals from promoting

cooperation-- when R+P is sufficiently large; some structures can lead prosocial individuals to be exploited by others-- when R is sufficiently small and R+P ≅ T+S; and some structures can instead promote cooperation which rewards prosocial players-- when R+P << T+S. These findings complement previous studies that have provided evidence that some specific personality traits (e.g., honesty, humility, agreeableness) can account for prosocial behavior as it is observed in some economic contexts such as dictator games, ultimatum games (Hilbig & Zettler, 2009, Hilbig et al., 2013), and public goods games (Hilbig et al., 2012). More specifically, such personality features appear to play an important role in explaining people's general tendency for acting fairly and/or forgiving (see, e.g., Hilbig et al., 2016) transgressions, both of which are crucial ingredients for the emergence and sustaining of cooperation. The current study highlights how endogenous factors like SVO can be contingent on different payoff structures that may accentuate or mitigate their effects and in some cases lead to joint cooperation.

Up to this point in the existing literature, the general relationship between prosocial preferences and cooperation has produced mixed results (e.g., Balliet et al 2009; Kümmerli et al., 2010), perhaps due to the overconcentration of research attention to particular regions of possible PD games. For example, Axelrod's PD game (R' = 0.6 and P' = 0.2), one of the most commonly used PD settings, has a relatively low R' (R' has a lower bound of 0.5 assuming that 2R > (T+S)) and is closest to our PD game I. As our results show, the unique characteristics of this incentive structure create a context where prosocials thrive, whereas other regimes of the PD space do not have these particular properties and do not afford prosocials fertile ground for engendering cooperation. In addition, our results demonstrate a positive relationship between people's own social preferences and their choice to cooperate only in games with a low R'. In the high R' region of PD games, cooperation is *not* influenced by social preferences but can homogeneously

emerge in games with low P'. In the low R' region, we show that games with high P' (resulting in a low K-index) can result in exploitation of prosocial individuals. Such exploitation can be accounted for by the irrelevance of beliefs whenever R+P is similar to T+S: in this case, individualists (prosocials) unconditionally prefer to defect (cooperate). On the other hand, games with low R' and P' are a fruitful domain for decision makers with high social preferences to establish outbreaks of cooperation. It is in this environment where nice decision makers can influence the self-interested decision makers to cooperate, allowing cooperative behavior to thrive. This conclusion is consistent with the fact that beliefs become more relevant to determine behavior whenever R+P is lower than T+S: an individualist may actually have highly contingent social preferences and thus only be nice and choose to cooperate if he strongly believes that the other player is likely to cooperate as well.

This research also offers important theoretical implications for the study of SVO and preferences in dynamic interactions. It is often assumed that prosociality can be captured by simply incorporating the welfare of others into one's own utility function (Equation 1). However, our results suggest more complex relationships that depend on the incentive structure of the PD and on the dynamics of the interaction between two players. For example, Game III offers some sharp predictions about what people would do depending on their social preferences (either always play C or always play D, regardless of their beliefs about the other player's behavior). But our results suggest that people with higher SVOs in Game III do not cooperate unconditionally and we observe significant variability in individual behavior (Figure 3). While prosocials positively reciprocate more than individualists do, we infer that both types of players have a similar attitude towards negative reciprocity, regardless of their social preferences. In other words, nice people are willing to elicit cooperative behavior, but not do so naively. We find

that prosocials do not tolerate being exploited more than other types of players. Note however, that such a self-protection mechanism is not sufficient to prevent nice people from being exploited, although it can reduce the negative consequences that interacting with self-regarding individuals can have (see Figure 5).

Our results also suggest that beliefs about the other player's actions are likely to play a major role in the emergence of cooperation, as it is predicted by a recent model (Murphy & Ackermann, 2015). Moreover, the belief dynamics may depend on the type of social preferences (e.g., prosocials may have more variability in their beliefs about the other's behavior than individualists) and the model from Murphy and Ackermann is static. In future studies, it could be worthwhile to investigate how the dynamics of the interaction gives rise to beliefs and how those beliefs are updated over sequential iterations of the game (this would also shed light on reciprocity and its emerges). For example, future studies may intermittently assess the beliefs a player has in anticipation of the other player's behavior and updated social preferences, contingent on their changing beliefs. Recent research showed that a cognitive model that dynamically adjusts the SVO index α in a joint utility function was able to account for the observed behavior in a PD game quite accurately (Gonzalez, et al., 2015).

In a world with heterogeneous decision makers with varied social preferences, researchers need to pay attention to the incentive structures that may promote selfish behavior or cooperative behavior in social dilemmas as neither intrinsic preferences nor exogenous incentives tell the whole story. When interacting with others repeatedly, such as in our workplace, it is important to facilitate the structural conditions that would promote civility and foster signals of unselfishness, thus allowing nice people to coordinate and reap the efficiency gains from sustained cooperation. Increasing the consequences of unilaterally deviating from

mutual cooperation is a necessary but not sufficient condition to these ends. Genuinely nice people may be exploited in such environments where mutual cooperation is not appropriately rewarding. However, it becomes possible for them to "contaminate" those less inclined to cooperate when the cost of mutual defection is significant. It is heartening to see evidence that prosocial decision makers can ascend and do well even in an austere social context where the temptation to defect is persistent.

References

Ackermann, K., Fleiß, J., & Murphy, R. O. (2016). Reciprocity as an individual difference. *Journal of Conflict Resolution, 60*, 2, 340 - 367.

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 1236-1239.

Axelrod, R. (1967). Conflict of interest: An axiomatic approach. *Journal of Conflict Resolution*, 87-99.

Axelrod, R. (1984), The Evolution of Cooperation, Basic Books, New York, NY.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*(4489), 1390-1396. http://dx. Doi.org/10.1126/science.7466396

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, *12*(4), 533-547. http://dx.doi.org/10.1177/1368430209105040

Becker, G. S. (1976). Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature, 14*(3), 817-826.

Burton-Chellew, M. N., Nax, H. H., & West, S. A. (2015). Payoff-based learning explains the decline in cooperation in public goods games. *Proceedings of the Royal Society of London B: Biological Sciences*, *282*(1801), 20142678. http://dx.doi.org/ 10.1098/rspb.2014.2678

Burton-Chellew, M. N., & West, S. A. (2013). Prosocial preferences do not explain human cooperation in public-goods games. *Proceedings of the National Academy of Sciences*, *110*(1), 216-221. http://dx.doi.org/10.1073/pnas.1210960110

Camerer, C. F., & Fehr, E. (2006). When does "economic man" dominate social behavior? *Science*, *311*(5757), 47-52. http://dx.doi.org/10.1126/science.1110600

Cooper, D., & Kagel, J. H. (2009). Other regarding preferences: a selective survey of experimental results. *Handbook of Experimental Economics*, Volume 2.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*(1), 169-193.

Dawes, R. M., & Thaler, R. H. (1988). Anomalies: Cooperation. *The Journal of Economic Perspectives*, *2*(3), 187-197.

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the Economics of Giving, Altruism and Reciprocity, 1*, 615-691. http://dx.doi.org/10.1016/S1574-0714(06)01008-6

Gonzalez, C., Ben-Asher, N., Martin, J. & Dutt, V. (2015). A Cognitive Model of Dynamic Cooperation with Varied Interdependency Information. *Cognitive Science, 39*(3), 457-495. http://dx.doi.org/10.1111/cogs.12170

Hamilton, W. D. (1964). The genetical evolution of social behaviour I and II. *Journal of Theoretical Biology*, *7*(1), 1-52. http://dx.doi.org/10.1016/0022-5193(64)90039-6

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859), 1243-1248.

Hilbig, B. E., Thielmann, I., Klein, S. A., & Henninger, F. (2016). The two faces of cooperation: On the unique role of HEXACO Agreeableness for forgiveness versus retaliation. *Journal of Research in Personality*, *64*, 69-78.

Hilbig, B. E., & Zettler, I. (2009). Pillars of cooperation: Honesty–Humility, social value orientations, and economic behavior. *Journal of Research in Personality*, *43*(3), 516-519.

Hilbig, B. E., Zettler, I., & Heydasch, T. (2012). Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional honesty–humility. *European Journal of Personality*, *26*(3), 245-254.

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*(5), 598-603.

Kümmerli, R., Burton-Chellew, M. N., Ross-Gillespie, A., & West, S. A. (2010). Resistance to extreme strategies, rather than prosocial preferences, can explain human cooperation in public goods games. *Proceedings of the National Academy of Sciences*, *107*(22), 10125-10130. http://dx.doi.org/10.1073/pnas.1000829107

Kreps, D. M., Milgrom, P., & Wilson, J. R. R. (1982). Rational cooperation in the repeated prisoner's dilemma. *Journal of Economic Theory*, *27(2), 245-252.*

Liebrand, W. (1984). The effect of social motives, communication and group-size on behavior in an n-person multi-stage mixed-motive game. *European Journal of Social Psychology*, *14*, 239–264.

Martin, J. M., Gonzalez, C., Juvina, I., & Lebiere, C. (2014). A Description–Experience Gap in Social Interactions: Information about Interdependence and Its Effects on Cooperation. *Journal of Behavioral Decision Making*, *27*(4), 349-362. http://dx.doi.org/10.1002/bdm.1810

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, *4*(1), 1-25. http://dx.doi.org/10.1016/0022-1031(68)90046-2

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision Making*, 6(8), 771-781. http://dx.doi.org/10.2139/ssrn.1804189

Murphy, R. O., & Ackermann, K. A. (2013). Social value orientation theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, http://dx.doi.org/10.1177/1088868313501745

Murphy, R. O., & Ackermann, K. A. (2015). Social Preferences, Positive Expectations, and Trust Based Cooperation. *Journal of Mathematical Psychology*, 1(67), 45-50. http://dx.doi.org/10.1016/j.jmp.2015.06.001

Ostrom, E. (2014). Collective action and the evolution of social norms. *Journal of Natural Resources Policy Research*, *6*(4), 235-252.

Ostrom, E., Burger, J., Field, C. B., Norgaard, R. B., & Policansky, D. (1999). Revisiting the commons: local lessons, global challenges. *Science, 284*(5412), 278-282. http://dx.doi.org/ 10.1126/science.284.5412.278

Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D., & Nowak, M. A. (2009). Positive interactions promote public cooperation. *Science*, *325*(5945), 1272-1275. http://dx.doi.org/ 10.1126/science.1177418

Rapoport, A. (1967). A note on the "index of cooperation" for Prisoner's Dilemma1. *The Journal of Conflict Resolution*, *11*(1), 100.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation* (Vol. 165). University of Michigan Press.

Roth, A. E., & Murnighan, J. K. (1978). Equilibrium behavior and repeated play of the prisoner's dilemma. *Journal of Mathematical Psychology*, *17*(2), 189-198. http://dx.doi.org/10.1016/0022-2496(78)90030-5

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences*, *106*(35), 15073-15078. http://dx.doi.org/10.1073/pnas.0904312106

Simon, H. A. (1993). Altruism and economics. *The American Economic Review*, *83*(2), 156-161.

Stivers, A., Murphy, R. O., Kuhlman, D. M. (2017). Indexing Prisoner's Dilemma games: Quantifying psychological factors underling cooperative behavior. Working paper.

Van Lange, P. A. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, *77*(2), 337. http://dx.doi.org/10.1037/0022-3514.77.2.337

Van Lange, P. A. M., De Cremer, D., & Van Dijk, E., & Van Vugt, M. (2007). Self-interest and beyond: Basic principles of social interaction. In A. W. Kruglanski & E. T. Higgings (Eds), Social Psychology: Handbook of Basic Principles (2nd Edition, pp. 540-561). New York: Guilford.

Van Lange, P. A., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes*, *120*(2), 125-141. http://dx.doi.org/10.1016/j.obhdp.2012.11.003

Appendix

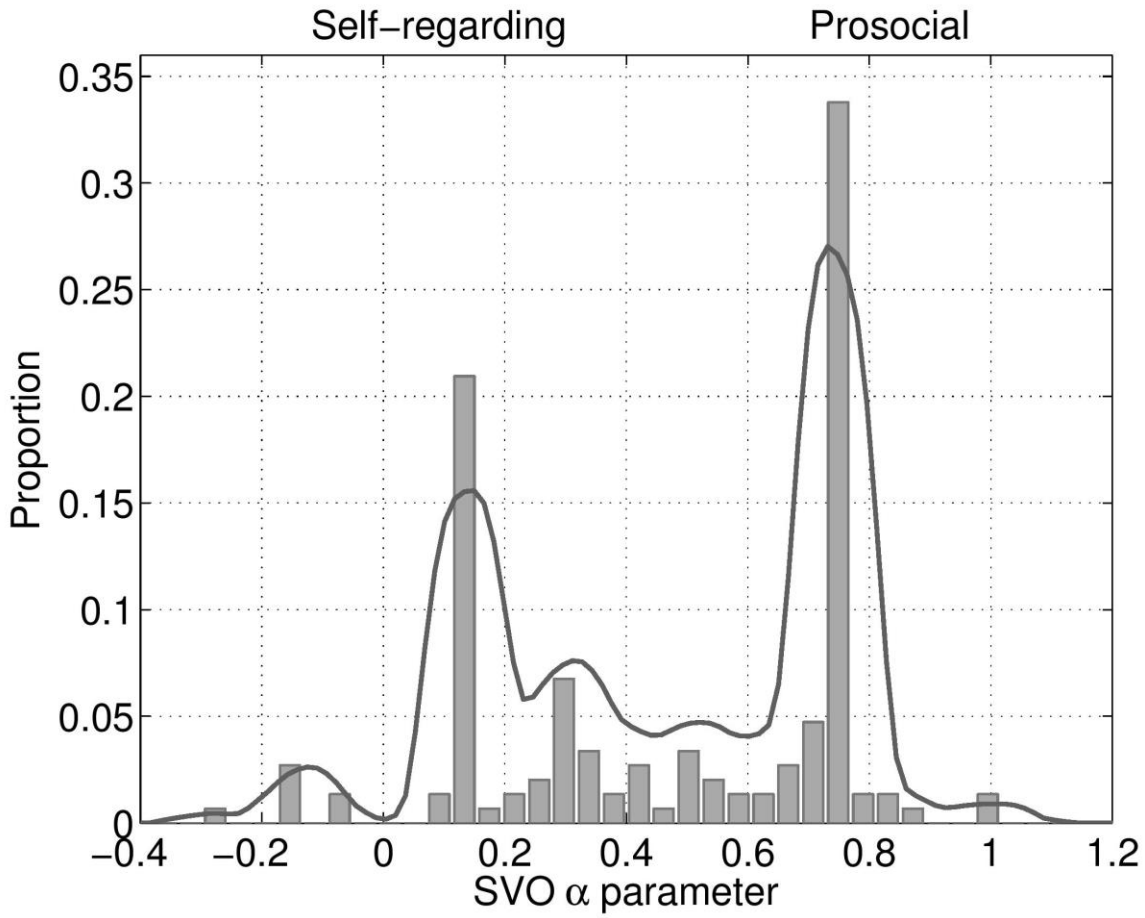

Figure 6. Distribution of social value orientation index in the experiment. This bimodal pattern is commonly found and what is shown here is very similar to other empirical distributions reported in the literature (see, e.g., Murphy et al., 2011 or the baseline condition from Ackermann et al., 2016).

Table 8. Linear regressions predicting individual cooperation rate (over 60 rounds per participant) in terms of one's own SVO (SVO_o) across all games (** $p<0.01$)

	Game I		Game II		Game III		Game IV	
Variable	SE_β	β	SE_β	β	SE_β	β	SE_β	β
SVO_o	0.129	0.414^{**}	0.179	-0.074	0.141	0.45^{**}	0.169	0.024
R^2	0.171		0.006		0.202		<0.001	
F	10.3^{**}		0.172		10.1^{**}		0.020	

Table 9. Linear regressions predicting mean individual payoff (over 60 rounds per participant) in terms of one's own SVO (SVO_o) in Games I and III (** $p<0.01$)

	Game I		Game III	
Variable	SE_β	β	SE_β	β
SVO_o	0.131	0.374**	0.157	0.115
R^2	0.14		0.013	
F	8.14**		0.539	

Table 10. Linear regressions predicting mean individual payoff (over 60 rounds per participant) in terms of the difference between one's own SVO (SVO_o) and the partner's SVO (SVO_p) in Games I and III ($p<0.05$(*))

	Game I		Game III	
Variable	SE_β	β	SE_β	β
SVO_o	0.144	-0.011	0.152	-0.356*
R^2	<0.001		0.127	
F	0.006		5.51*	

Table 11. Mixed-effect logistic regression models fitted to predict binary cooperation choices in Game I in terms of round, own SVO (SVO_o), and the partner's SVO (SVO_p) as fixed effects (subject is the random effect) with likelihood ratio tests comparing models (M_{B1} with M_{A1}, M_{C1} with M_{B1}, M_{D1} with M_{C1}).

	Models							
	M_{A1}		M_{B1}		M_{C1}		M_{D1}	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Intercept	1.016	0.469	1.076	0.480	-1.386	0.955	-4.366	1.215
Round			-0.008	0.003	-0.008	0.003	-0.008	0.003
SVO_o					5.286	1.729	5.594	1.599
SVO_p							5.627	1.599
df	2		3		4		5	
AIC	2385		2379		2373		2363	
BIC	2397		2397		2397		2393	
Log-likelihood	-1190		-1187		-1182		-1177	
χ^2			7.593		8.533		11.718	
df χ^2			1		1		1	
p			0.006		0.003		0.0006	

Table 12. Mixed-effect logistic regression models fitted to predict binary cooperation choices in Game III in terms of round, own SVO (SVO_o), and the partner's SVO (SVO_p) as fixed effects (subject is the random effect) with likelihood ratio tests comparing models (M_{B3} with M_{A3}, M_{C3} with M_{B3}, M_{D3} with M_{C3}).

	Models							
	M_{A3}		M_{B3}		M_{C3}		M_{D3}	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Intercept	-1.047	0.454	-0.605	0.469	-2.467	0.715	-3.283	0.858
Round			-0.015	0.003	-0.015	0.003	-0.015	0.003
SVO_o					4.343	1.362	4.062	1.326
SVO_p							2.184	1.325
df	2		3		4		5	
AIC	1869		1853		1846		1845	
BIC	1881		1871		1869		1874	
Log-likelihood	-932		-924		-919		-918	
χ^2			18.063		9.341		2.637	
df χ^2			1		1		1	
p			<0.0001		0.002		0.104	