WORKING PAPERS

N° 14-08

Are we more fearful than greedy? Outbounding the incentives to defect in cooperation dilemmas

Cesar Mantilla

IAST
INSTITUTE for ADVANCED STUDY in TOULOUSE

Are we more fearful than greedy? Outbounding the incentives to defect in cooperation dilemmas

César Mantilla *

Institute for Advanced Study in Toulouse (IAST)

Abstract

Previous studies analyzing the impact of payoffs' cardinality in cooperation dilemmas have concluded that the additional benefits of defecting against a cooperator, the *greed* dimension of the game, are more salient than the additional costs of cooperating against a defector, the *fear* dimension of the game. We conduct an experiment to show that when the costs of cooperation exceed its gains, this pattern is reversed. The larger effect of *fear* with respect to *greed* on the likelihood to defect is robust to random rematching and to repeated matching, and is mostly driven by a relative rather than an absolute perception of the incentives to cooperate across different dilemmas.

JEL Classification: C91, D03

Keywords: experiment; motivational pressures to defect; prisoner's dilemma

*__Contact info:__ Toulouse School of Economics, 21 allée de Brienne F-31015 Toulouse CEDEX 6. *E-mail address:* cesar.mantilla@iast.fr. *Telephone:* +33 0561122512

1 Introduction

Social dilemmas describe situations characterized by a tradeoff between individual and collective incentives. If someone is willing to incur in an individual cost to provide a benefit to the group he belongs to we say he is willing to cooperate. This situation is a "dilemma" if individuals are better off when the other group members cooperate, but for each one of them is individually rational to not incur in the costs of cooperation.

The two "motivational pressures" to defect, i.e. to not cooperate, were proposed by Rapoport (1967) in its definition of the *Cooperation Index* for the most popular of the cooperation dilemmas, the Prisoner's Dilemma (henceforth PD)[1]. These incentives to defect are not simultaneously activated, but rather they depend on the expected action of the opponent. *Fear* is associated to the extra costs from cooperation if opponent's defection is expected. *Greed*, on the other hand, is associated to the extra benefits from defection if opponent's cooperation is expected.

The notions of *fear* and *greed* were extended to n-players cooperation dilemmas by Schelling (1973) in his study of binary choices. Although he did not make the reference to these labels, Schelling drew the attention to the separate analysis of individual incentives to stay out of a coalition when the majority already joined it, i.e. the benefits of defection under the majority's cooperation, and the lack of incentives to join a coalition when almost nobody have joined it, i.e. the costs of cooperation under the majority's defection.

The acknowledgement of these two separate forces was followed by empirical testings of their differences. Komorita et al. (1980) conducted a three-player PD, finding that *greed* has a larger negative impact on the cooperation rate than *fear*. More recently, Ahn et al. (2001) proposed a set of four PD games to analyze the incentives to defect in a between-subjects protocol involving fixed (or Partners) matching and random (or Strangers) matching. They found that *fear* and *greed* affect negatively the cooperation rates with

[1]For a meta-analysis of experimental evidence for Prisoner's Dilemmas see Sally (1995).

Strangers matching but not with Partners matching, suggesting that history of play is more important than the motivational pressures to defect. When these dimensions have an effect, as in the case of Strangers matching, *greed* was more relevant than *fear* in the reduction of cooperation.

The manipulation of the *fear* and *greed* dimensions has been also helpful to analyze gender differences in the likelihood to cooperate (Simpson, 2003; Kuwabara, 2005) and to test the neurobiological foundations of in-group and out-group perceptions towards cooperation (De Dreu et al., 2010, 2012).

Empirical evidence suggests that the *greed* dimension is more salient than the *fear* dimension in the individual decisions to engage in cooperative behavior. It implies, in terms of contributions to public goods, that individuals are more aware of the increasing benefits of not contributing, conditional on group members' cooperation, than to the increasing costs of contributing, conditional on group members' defection. Nevertheless, the relative effects of *fear* and *greed* were previously tested in situations in which the gains from cooperation were at least as large as the gains from defection.

Our conjecture is that *fear* was less likely to induce defection in Komorita et al. (1980) and Ahn et al. (2001) because the gains from cooperation were more "attractive" than the costs of engaging in this behavior. If cooperation is not very costly, the likelihood to assume mutual cooperation as default outcome increases. These conditions suppress the *fear* of being exploited by a defector, while at the same time they make more salient the effects of *greed* in the cooperation decision.

We propose an alternative experimental setting in which the costs of cooperation, *fear* and *greed*, exceed the cooperator's gain. We hypothesize that the dominant effect *greed* over *fear* will be reversed if there is more uncertainty regarding the expected play of the opponent.

Blonski et al. (2011) state that cooperation gets more risky in a PD if the "suckers's payoff", the payoff received by a cooperator when he faces a defector, gets smaller. This is equivalent to say that cooperation becomes a riskier strategy when the *fear* dimension increases. They show that, for

experimental data with an infinitely repeated PD, a discount factor depending on the value of the "sucker's payoff" provides a better prediction of the cooperation rates.

In our experimental design we manipulate the *fear* dimension by reducing the sucker's payoff, as in Blonski et al. (2011); but also the *greed* dimension by increasing the payoff received by a defector when faces a cooperator, the parameter known as the "temptation" payoff. We increase these dimensions long enough to alter the outcomes maximizing and minimizing the joint payoffs, leaving the PD conditions. A sufficiently large increase in the *greed* dimension leads to a point in which the aggregate benefits from the asymmetric outcome exceed those under mutual cooperation, even if the additional earnings are captured by the defector. Correspondingly, an increase in the *fear* dimension may enlarge the costs inflicted to the cooperator in the asymmetric outcome up to a point that mutual defection will be socially most preferred than the asymmetric outcome.

We find that, when the benefits from defection are larger than the benefits from cooperation, *fear* has a larger impact than *greed* in the likelihood to cooperate. The effect is observed not only in the Strangers treatment, but also in the Partners treatment. For the latter matching type, the effects are identifiable after disentangling the complementarities between the history of the game and the incentives to defect, in particular the *fear* dimension. We also show evidence that the effects of fear and greed are driven by a relative rather than an absolute perception of the incentives to cooperate across the different dilemmas. This is particularly clear within the partners condition.

The remainder of this paper is organized as follows. In Section 2 are defined the conditions for the three cooperation dilemmas used in the experimental setting, which is fully described in Section 3. The results are presented in Section 4 and discussed in Section 5. Section 6 concludes.

2 Defining *fear* and *greed* in cooperation dilemmas

A cooperation dilemma is characterized by a unique Pareto inefficient Nash equilibrium. *Defection* strictly dominates *Cooperation*, leading self-regarding players to an undesirable outcome. Following Rapoport (1967), we define the four payoffs of this game as the temptation (T) to defect when the other player cooperates, the reward (R) earned when both players cooperate, the punishment (P) received when both players defect; and the "sucker's payoff" (S), received by cooperating against a defector. The game is presented in its strategic form in Table 1.

Table 1: Strategic form of a cooperation dilemma

Self \ Opp.	**(C)ooperate**	**(D)efect**
(C)ooperate	R, R	S, T
(D)efect	T, S	P, P

The condition (1) is shared by all cooperation dilemmas and is known as the *suboptimality condition*. It is because $T > R$ and $P > S$ that cooperation is strictly dominated in this game. In the particular case of the PD, it exists a second constraint defined as the *externality larger than internality* condition (Frohlich and Oppenheimer, 1996). As shown in expression (2), it guarantees that the externality cost imposed on the other player ($R - S$) will be larger than internality derived from the benefits of defecting against the cooperator ($T - R$).

$$T > R > P > S \tag{1}$$

$$2R > T + S \tag{2}$$

Now we can attach our definitions of *fear* and *greed* to the payoff parameters in cooperation dilemmas. Going back to Table 1, the *greed* dimension

will be the difference between row payoffs in the first column $T - R$, i.e., when opponent's cooperation is expected. Similarly, the *fear* dimension is given by the difference between row payoffs in the second column $P - S$, i.e., when opponent's defection is expected. The remaining difference between two consecutive payoffs according to condition (1), $R - P$, is known as the *cooperator's gain*. It reflects the additional benefits received by each player when they move from mutual defection to mutual cooperation.

Holding the cooperator's gain constant, the only way to enlarge the *greed* dimension is by increasing the temptation payoff. If $T-R$ is increased enough the *externality larger than internality* condition is no longer held. The sum of payoffs will be maximized under the asymmetric outcome and mutual cooperation will turn into a second best from an utilitarian perspective. If the internality would exceed the externality, the defector could compensate the coooperator with an *ex-post* transfer δ under the threat that $S + \delta > P$, and then he could keep for himself the excedent from mutual cooperation $T - (R + \delta)$. In a repeated game with a sufficiently long horizon the players may establish an alternating strategy, allowing them to maximize mutual payoffs while they minimize the difference between their intertemporal earnings.

Another *externality larger than internality* condition, often neglected in the definition of a PD, is displayed in expression (3). It guarantees that the costs of leaving mutual defection $P-S$ are smaller than the opponent's benefits of keep choosing defection $T-P$. If the condition holds, mutual defection minimizes the players' aggregate payoffs. We could think of condition (3) as a reflection of condition (2) in the domain of losses within the asymmetric outcomes. Both are guaranteeing that the PD outcomes lie in a convex hull. But unlike (2), in which the loss of convexity leads to a set of alternating strategies mutually beneficial, if (3) is violated the set of feasible payoffs under mixed strategies lead to worst expected payoffs than mutual defection. Holding the cooperator's gain constant, the violation of condition (3) occurs by a substantial reduction in the "sucker's payoff", or equivalently, an increase in the *fear* dimension.

$$T + S > 2P \tag{3}$$

If conditions (2) or (3) are not satisfied we no longer consider the game a PD. Nevertheless, as long as the *suboptimality condition* holds these games are still considered cooperation dilemmas. We will define as a *Greedy Dilemma (GD)* a game in which expression (2) is no longer valid, which may appear due to an outbounded *greed* level. Likewise, we will define as a *Fearful Dilemma (FD)* a game that violates condition (3).

3 Experimental Design

The effects of *fear* and *greed* were tested on a within-subjects basis using three different cooperation dilemmas: a *Fearful Dilemma (FD)*, a *Prisoner's Dilemma (PD)* and a *Greedy Dilemma (GD)*. Participants interacted for thirty rounds, divided into three stages of ten rounds, one for each cooperation dilemma. The matching protocols, Strangers and Partners, were introduced on a between-subjects basis, as well as the order in which the three cooperation dilemmas were presented to players. The full description of the six treatment cells is presented in Table 2.

Table 2: Experimental design: Treatment cells

Cell	**Matching**	**Games' sequence**
1	Strangers	FD - PD - GD
2	Partners	FD - PD - GD
3	Strangers	PD - GD - FD
4	Partners	PD - GD - FD
5	Strangers	GD - FD - PD
6	Partners	GD - FD - PD

The payoff matrices for the three dilemmas are displayed in Table 3. Each unit earned in these games was paid at \$100 Colombian pesos (cop)[2]. We take the *PD* as the baseline in our experiment. In this game, the *fear* and

[2]At October 2011 the exchange rate for \$1 usd was \$1,903 cop. This means that each point in Table 3 was paid at 5.25 dollar cents.

greed dimensions have a value of 1 unit. The *cooperator's gain*, which will be held constant across games, has a value of 2 units.

To convert the *PD* into the *FD* we reduce the "sucker's payoff" from 5 to 1 units as is shown in Table 3. This is equivalent to an increase of four times the *fear* dimension. Following a similar procedure, the *GD* game is obtained from the standard PD by increasing the temptation payoff from 9 to 13 units, also an increase of four times the *greed* dimension.

Table 3: Payoff matrices for the three cooperation dilemmas

Fearful Dilemma

	C	D
C	8; 8	1; 9
D	9; 1	6; 6

Prisoner's Dilemma

	C	D
C	8; 8	5; 9
D	9; 5	6; 6

Greedy Dilemma

	C	D
C	8; 8	5; 13
D	13; 5	6; 6

We define the *cooperator's cost* as the maximum level between the *fear* and *greed* dimensions. If we compute the ratio between the *cooperator's cost* and the *cooperator's gain* for each one of the three dilemmas, we have a value of 0.50 in the *PD*, and a value of 2.50 in the *FD* and *GD* games. In Ahn et al. (2001), the ratio is 0.25 in the baseline *PD* game and increases to 1.00 in the other *PD* games used to test the effects of *fear* and *greed*.

The ratio between costs and benefits increases four times in Ahn et al.'s design and five times in ours, which are comparable magnitudes. However, it is only in our design in which the absolute value of this ratio surpasses the unity. This is a key feature of our design, since we claim that the reversal in the relative weights of *fear* and *greed* in the cooperation decision only occurs when the *cooperator's cost* exceed the *cooperator's gain* and therefore the default outcome is more likely to be mutual defection than mutual cooperation.

The experiment was conducted at Universidad de los Andes, in Bogotá (Colombia), between September and October 2011. The 156 participants were selected from a total of 349 applicants who were initially contacted by email or flyers delivered in the university. The participants were undergraduate and graduate students from different majors, mainly economics and

engineering. Descriptive statistics of the participants are shown in Table 4.

A total of thirteen sessions, with twelve players each, were conducted. There were two sessions per treatment cell, except for cell number 5, for which we conducted three sessions. They lasted from sixty to eighty minutes. There was no show-up fee and the participants were initially informed that they will receive the payoffs' sum of their decisions from all the thirty rounds of play at the end of the game. Participants earned on average 20,500 cop, corresponding to 10.7 usd[3] at the time of the experiment.

Table 4: Descriptive statistics from participants

	Mean	Standard deviation
Male	0.576	0.495
Age	20.603	2.747
Academic terms	6.013	3.088
Economics	0.359	0.481
Engineering	0.250	0.434
Art / Design	0.179	0.384
Social Sciences	0.128	0.335
Business	0.122	0.328
Natural Sciences	0.122	0.328
Earnings (cop)	20,506	2,227

The experiment was conducted using z-Tree (Fischbacher, 2007). Once participants fully understood the instructions and signed the informed consentment, they started with a practice round (a 2×2 coordination game) to get familiarized with the decision screen. After the practice round, participants interacted for thirty rounds divided into three sets of ten rounds each. Each one of the games shown in Table 3 was played in one of the sets. Once rounds 10 and 20 were finished, the experimenter publicly announced that the game they were facing was about to change in the following round, and that it will remain the same for ten rounds. After the thirty rounds participants completed a post-experimental survey and receive their earnings. Instructions for one of the sessions are presented in Appendix A.

[3]Daily minimum wage in Colombia for 2011 was approximately 9.3 usd.

4 Results

4.1 The asymmetric effects of *fear* and *greed*

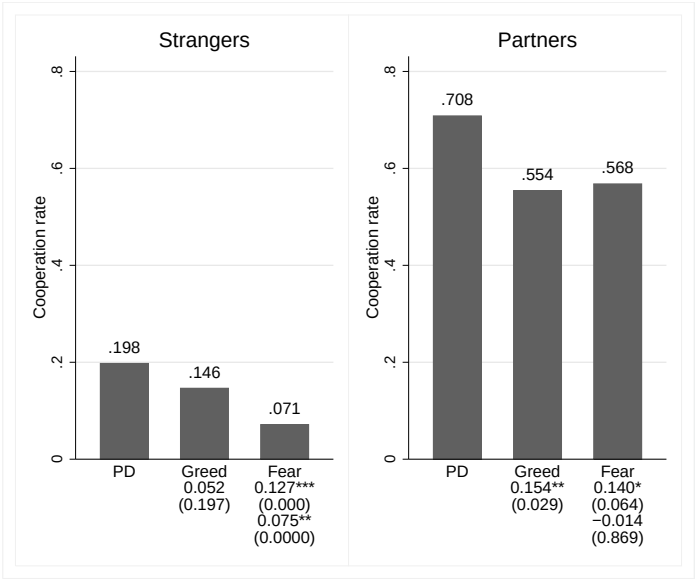
The cooperation rates for the three dilemmas are shown in Figure 1 for both matching protocols. For the Strangers matching we observe that the cooperation rate is higher in the *PD* (0.198), followed by the *GD* (0.146) and the *FD* (0.071). The differences in the cooperation rate between these games, as well as the p-values of the statistical tests, are shown at the bottom of the figure[4]. An adjusted Chi-squared test is used to compare cooperation rates across games. Observations are clusterized at the subject level for its computation. Under Strangers matching the *PD* and the *GD* are statistically different from the *FD*, but the difference between these two games is not statistically significant. For Partners matching, the cooperation rate in the PD (0.708) is higher than in the *FD* (0.568) and the *GD* (0.554). Although the differences with respect to the *PD* are statistically significant, cooperation rates between the *FD* and the *GD* are not statistically different.

Results from the Partners matching suggest that, in presence of reputational mechanisms, the *fear* and *greed* dimensions have a similar effect on the decision to cooperate. Nevertheless, data aggregation at the *game* × *matching* level in Figure 1 makes challenging the detection of between-subjects differences across dilemmas. Figure 2 provides an alternative representation of the results, displaying the cumulative frequency of the fraction of cooperative moves per participant. The more frequent the cooperative behavior, the steeper the cumulative frequency when the fraction of cooperative moves is close to unity. Following this description, one might interpret the area above the curve as a proxy for the cooperation rate in each dilemma.

For the Strangers matching the differences between the three dilemmas are consistent with the results in Figure 1. The largest percentage of subjects that never cooperated belongs to the *FD* (64.3%), followed by the *GD*

[4]In the first row we present the differences for the *GD* and the *FD* with respect to the *PD*. In the second row we presente the difference for the *FD* with respect to the *GD*.

Figure 1: Cooperation rate by game and matching protocol



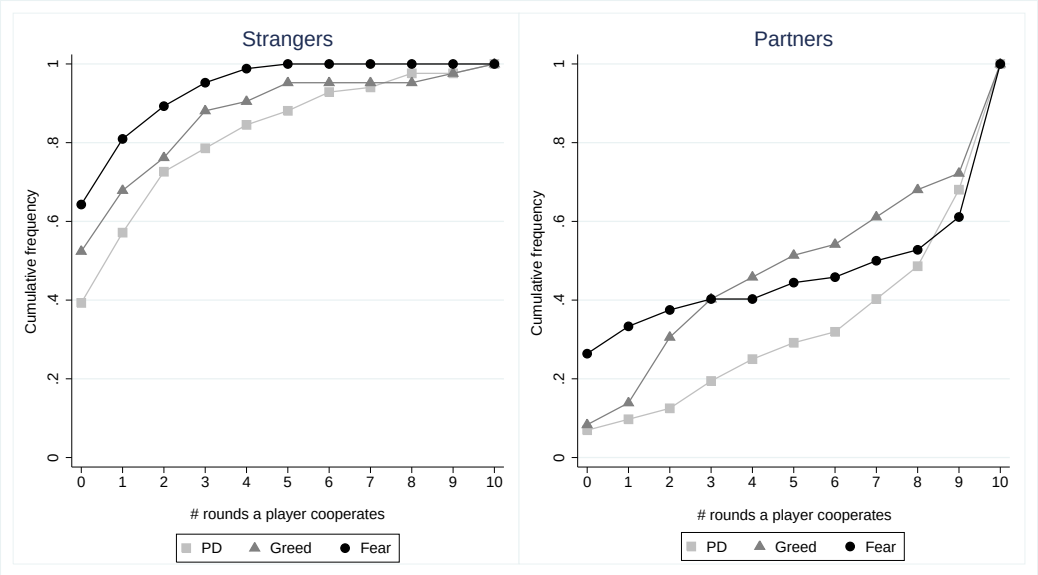
(52.4%) and the *PD* (39.3%). The subjects that cooperated more frequently in the *FD* did it for four rounds, while for the *GD* and the *PD* we found that at least one player cooperated for the whole ten rounds. The three curves do not cross each other[5], remarking a similar behavioral pattern in all three dilemmas. They differ only in the fraction of participants that never cooperated and the speed of convergence to full defection.

Under Partners matching, although the average cooperation rate does not differ statistically from one setting to another, we observe very different cumulative distributions for the *FD* and the *GD* games. The percentage of subjects that did not cooperate in any round in the *GD* (8.3%) is almost as lower as in the *PD* (6.9%). In the *FD* a larger proportion of subjects, about one of every four (26.4%), did not cooperate in any round. On the other hand, the largest proportion of subjects that cooperated in all ten rounds appear in the *FD* (38.9%), followed by the *PD* (31.9%) and the *GD* (27.8%).

In the *FD* is highly dependent on the trajectory. We find for this game the

[5]The only exception occurs between the *GD* and the PD for 8 rounds of cooperation. However, this effect is driven by the single player that cooperated for more than five rounds in the *GD* dilemma.

Figure 2: Cumulative frequency of cooperative moves (per player) by game and matching protocol



largest fraction of subjects that never cooperated but also the largest fraction of subjects that always cooperated. The differences between *fear* and *greed* also exist under the Partners matching protocol, but they strongly depend on the history of the game. A past cooperative choice is more rewarded in the *FD* than in the *GD* because, by definition, the pressure from the *fear* dimension is overcame once mutual cooperation is reached.

We run a set of logistic regressions to test if the differences between *fear* and *greed* are *(i)* directly observable in the Strangers treatment; and *(ii)* observable in the Partners treatment after controlling for the game history in each dilemma separately. To disentangle the complementarities between reciprocity and the motivational pressures to defect we introduce an interaction term between the opponent's lagged choice and the *fear* and *greed* dimensions.

The marginal effects for the probability of cooperation are reported in Table 5. Models (1) and (2) correspond to Strangers and models (3) and (4) to Partners matching. Columns (1) and (3) do not include the interaction

terms mentioned above, while columns (2) and (4) do. All four models include individual and round fixed effects, as well as other individual characteristics such as gender, age, career and academic terms in the university. The χ^2 statistics resultant from the comparison between the coefficients from *fear* and *greed* for each regression, as well as their correspondent p-value, are shown at the bottom of the table.

Columns (1) and (3) replicate the results we initially observed in Figure 1. For Strangers, facing the *FD* reduces the probability of cooperation in 18.7 percentage points (pp) with respect to the *PD*; while facing the *GD* reduces the probability of cooperation in 6.9 pp with respect to the *PD*. The difference between these coefficients is statistically significant. For Partners, the reduction in the probability of cooperation of 9.7 pp when facing the *FD* is not statistically different from the reduction of 10.1 pp in the probability of cooperation in the *GD* (p-value 0.856), both compared to the *PD*.

Table 5: Logit estimates: Probability of cooperation (marginal effects)

Probability of cooperation	Strangers		Partners	
(marginal effects)	(1)	(2)	(3)	(4)
Fear (FD)	-0.187***	-0.188***	-0.097***	-0.175***
	(0.022)	(0.025)	(0.021)	(0.036)
Greed (GD)	-0.069***	-0.071***	-0.101***	-0.069**
	(0.018)	(0.021)	(0.021)	(0.031)
Round	-0.0079***	-0.0079***	0.0016	0.0014
	(0.0009)	(0.0010)	(0.0010)	(0.0010)
Opponent's choice$_{t-1}$	0.138***	0.133***	0.298***	0.276***
	(0.018)	(0.026)	(0.013)	(0.031)
Opponent's choice$_{t-1}$ × Fear		0.006		0.141***
		(0.054)		(0.049)
Opponent's choice$_{t-1}$ × Greed		0.011		-0.057
		(0.041)		(0.043)
Fear = Greed (χ^2 statistic)	25.92***	20.73***	0.03	10.56***
	(0.000)	(0.000)	(0.856)	(0.001)
Observations	1,728	1,728	1,701	1,701

Individual fixed effects included in all regressions. Controls included: gender, age, terms in university, and career dummies for economics and business. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

With the interaction terms included in model (4), the differences between *fear* and *greed* under Partners matching become evident. The reduction in

the probability of cooperation in the *FD*, with respect to the *PD*, is of 17.5 pp. This effect is, in magnitude, larger than the reduction of 6.9 pp in the *GD*. Unlike model (3), for model (4) the difference between the marginal effects of *fear* and *greed* is statistically significant, even at the one percent level (p-value 0.001).

Note that in model (4) the only interaction term that is statistically significant is *Opponent's choice*$_{t-1}$ $\times$ *Fear*. Its effect can be interpreted as an additional increase of 14.1 pp in the probability of cooperation in the *FD* (for a total of 41.7 pp with respect to the 27.6 pp in the *PD*) if the opponent cooperated in the previous interaction. The term *Opponent's choice*$_{t-1}$ $\times$ *Greed*, which implies a reduction of 5.7 pp in the likelihood to cooperate in a *GD* if the opponent cooperated in the previous round, is not statistically significant.

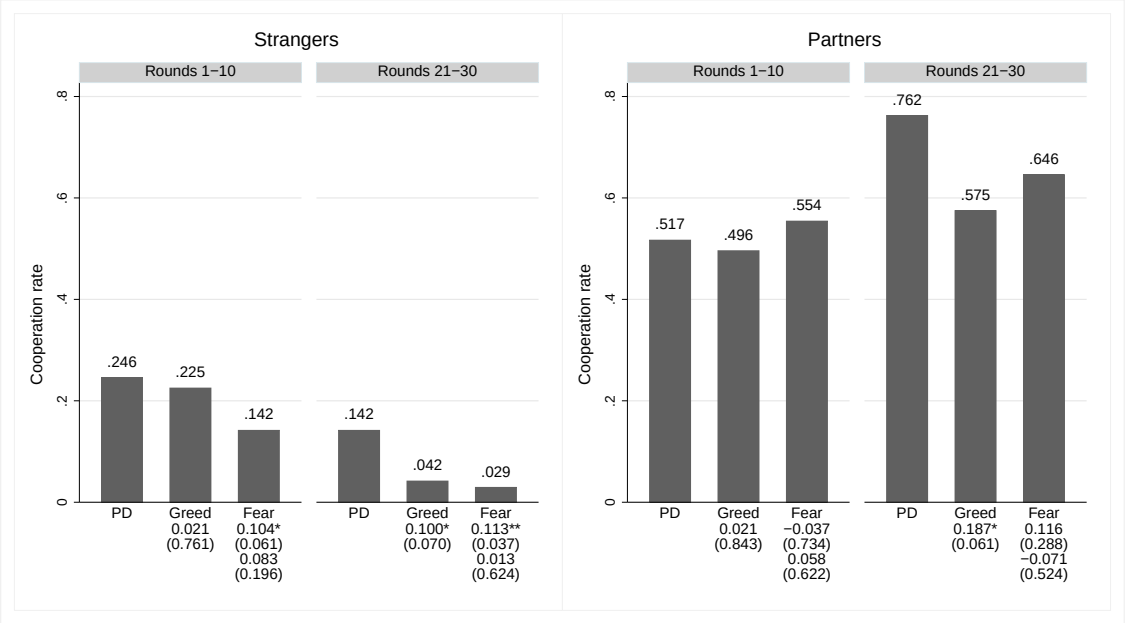
4.2 The relative effects of *fear* and *greed*

Within-subjects variation in the reward structures of a game usually holds on the *sequential independence assumption* (Vlaev and Chater, 2006). It means that choices in a game are assumed to be based only on the attributes of the current game, and that they are independent from the attributes of previous games. In terms of our experimental design, the validity of the *sequential independence assumption* would require that the differences in cooperation rates between dilemmas would be orthogonal to the order in which these games were faced. Following this argument, the differences between stages for a given game should be explained uniquely by the gradual decay of cooperation for Strangers matching, and its gradual increase for Partners matching.

The cooperation rates aggregated at the *game*$\times$*stage*$\times$*matching* level are reported in Figure 3. For each matching protocol, the average cooperation rate is shown separately for the cases when the dilemma was played in the first stage, from Round 1 to 10; and when it was played in the third and last stage, from Round 21 to 30. Differences in cooperation rates are reported,

for each stage, at the bottom of the figure. We also report the p-value of the adjusted Chi squared test with data clusterized at the individual level.

Figure 3: Cooperation rates for each dilemma in the first stage (rounds 1 to 10) and the last stage (rounds 21 to 30)



The only difference that is statistically significant in the the first stage for Strangers matching is the comparison between the *PD* and the *FD*. For the last stage, if the *sequential independence assumption* is held, this should be the only difference between games remaining statistically significant. Nevertheless, what we observe is that when the most unequal games are compared to the *PD*, both differences turn out to be statistically different from zero. With respect to the *FD*, not only the magnitude increases but also its significance level.

For Partners matching the between-stages comparison is more striking. In the first stage, the differences between cooperation rates for the *PD*, the *GD* and the *FD* are not statistically significant. Therefore, the differences between games reported in Figure 2 cannot be reconciled with the *sequential independence assumption*. The higher and statistically significant coopera-

tion rate in the *PD* obtained with aggregate data must therefore come from advanced stages in the experiment. For the last stage, Rounds 21 to 30, we find that the difference between the *PD* and the *GD* is statistically significant, but not the difference between the *PD* and the *FD*. Our explanation for the lack of significance despite having larger differences than in the Strangers treatment is that the high inter-cluster correlations reduce notably the power in our sample.

5 General discussion

5.1 Why *fear* is more salient than *greed*?

When payoff inequality within the asymmetric outcomes of a cooperation dilemma is large enough, the *fear* dimension has a larger impact in the likelihood to cooperate than the *greed* dimension. At first sight this result seems to contradict the previous findings from Komorita et al. (1980) and Ahn et al. (2001), but we argue that the relative effects of *fear* and *greed* depend on the cost-benefit analysis of cooperation *versus* defection.

We claim that an increase in the *fear* and *greed* dimensions alter the expectations regarding the other's decision. Engel and Zhurakhovska (2012) show that subjects beliefs' on their counterpart's likelihood to cooperate is positively correlated with an increase in the gains from cooperation relative to its costs. If this is the case, the operating "motivational pressure" is *greed* and *fear* is left with a secondary role in the decision. Assuming we can extend this argument in the opposite direction, i.e. that the counterpart's expected likelihood to cooperate is reduced by an increase in the cost of cooperation, we can rationalize why the salience of the *fear* dimension increases for the *FD* and *GD* games. The larger the increase in the *fear* or *greed* dimension, the higher the probability that subjects take mutual defection as the "default outcome" of the game.

An alternative explanation could be loss aversion, which means that under uncertainty expected losses are more weighted than expected earnings

(Kahneman and Tversky, 1979). Under this hypothesis, *fear* would be more salient because the reduction in utility of being exploited in the *FD* is larger than the increase in utility of exploiting a cooperator in the *GD*, even if the payoff differences are the same. Nevertheless, loss aversion cannot explain why the relative weights of *fear* and *greed* are the opposite when the costs of cooperation are at most equal to its benefits, as in Ahn et al. (2001).

Our claim that *fear* becomes more salient than *greed* when the costs of cooperation are high enough to switch the "default outcome" is more evident in the absence of additional mechanisms that may alter the costs of cooperation. This is the case of the Strangers environment, characterized by random interactions, in which the differential effects of the "motivational pressures" were directly observed. On the contrary, in the Partners matching protocol, in which the costs of cooperation are reduced by the presence of reputation and retaliation mechanisms, the effects of *fear* and *greed* seemed symmetric at first sight.

We unveil the asymmetric effects of *fear* and *greed* by considering the game history as complementary rather than substitute of the pressures to defect. We disentangle the asymmetries between these two effects by introducing interaction terms between the opponent's lagged choice and the *FD* and *GD* categorical variables. Of these interactions, only the *opponent's* $choice_{t-1}\times$ *fear*'s coefficient was statistically significant. Its positive effect indicates that an opponent's cooperative move is more rewarded under the *FD* than in the *PD*, suggesting that the pressure from the *fear* dimension ceases when counterpart's cooperation is expected. In addition, if cooperative moves are highly rewarded in the *FD* due to reputational effects, then we should not find these differences for Strangers matching. Indeed, results in Table 5 reveal that none of the interaction terms are statistically significant in model (2).

The natural inquiry after these arguments is to analyze to which extent the mechanisms under Strangers and Partners are consistent with each other. Our answer is that even if for both matching protocols the explanation could

be reduced to "how the default outcome switches as a function of the relative cost of cooperation", the updating process of the counterpart's expected behavior is very different. For Strangers, the updating relies on what a subject can infer from his last opponent with respect to the behavior of the entire population, which is not too much. For Partners, this inference is much more stronger because the subject is updating his opponent's expected behavior based on the actual history of their pairwise interactions. This is the reason why we observe path dependency for Partners matching. A signal of cooperative behavior is strong enough to induce cooperation with a fixed partner, but it does not provide enough information for a subject randomly rematched after each interaction.

Going back to the differences between *fear* and *greed* under fixed matching, one might think on an alternative explanation for the intermediate cooperation rate in the *GD*. Given that $T + S > 2R$, the intertemporal payoff maximization will be achieved by a *coordinate alternation* between cooperation and defection reflected in a cooperation rate close to one half. If subjects were pursuing this strategy, the effect of *greed* on the likelihood to cooperate will not be driven by the relative perceptions of the costs of cooperation but rather by a notable degree of rationality and a capacity to establish tacit coordination agreements.

Nevertheless, we do not find evidence supporting the *coordinate alternation* strategy. To inspect the experimental data, we define it as an attempt to alternate when a subject switched from cooperation to defection or viceversa during at least four consecutive rounds. For the Partners treatment, 16.6% attempted to alternate in the *GD*. However, using the same definition of alternation, 9.7% of subjects also tried to alternate under the *PD*. After performing a Wilcoxon rank-sum test[6] we find that the difference between games is not statistically significant. Therefore, we do not find evidence in favor of the alternation hypothesis given that this strategy does not maximize the expected joint payoffs in the *PD*.

[6]We compare the distributions of the number of attempts to alternate between games.

5.2 When do we observe the effects of *fear* and *greed*?

Our findings suggest that the effects of *fear* and *greed* are mainly driven by a relative rather than an absolute perception of the asymmetric payoffs. Consider the results of the first stage of the game. At that point, each participant has faced only one of the three cooperation dilemmas for ten rounds. The differences in cooperation rates across treatment cells is what a between-subjects design would have shown us. For Strangers, we would have only find an effect of the *fear* dimension. For Partners, we would not have found any effect of the motivational pressures in the likelihood to cooperate. Therefore, it is necessary to explore how the effects depend on the transitions from one payoff matrix to another.

We can think on the differences in cooperation rates across games and between the first and the third stage as a differences-in-differences estimator at the population level. Controlling for the gradual decrease of cooperation for Strangers and its gradual increase for Partners, any remaining difference across cooperation dilemmas must be explained by the relative effects of *fear* and *greed* dimensions. In other words, the absence of order effects should be validated by identical drops in cooperation from the first to the last stage across the three games.

For Strangers matching, the slopes describing these drops are -0.183 for the *GD*, -0.113 for the *FD* and -0.104 for the *PD*; and they correspond to reductions of 81.4%, 79.6% and 42.3% in the cooperation rate with respect to the first stage, respectively. However, the drops in the *GD* and *FD* might be bounded given the low cooperation rates for these games since the first stage of the experiment. Therefore, we might have a lower-bound estimates for these slopes. For Partners matching case we have at our disposition a cleaner test for the order effects. Cooperation rates do not differ statistically across games for the first stage, so differences would be then attributed to changes in the perception of the relative costs and benefits of cooperation. The slopes for the *GD*, the *FD* and the *PD* are +0.079, +0.092 and +0.245 respectively, corresponding to increases of 15.9%, 16.6% and 47.4% in the

cooperation rates with respect to the first stage.

Although the increase of the cooperation rate between-stages is about three times larger in the *PD* than in the *GD* and *FD* games, we acknowledge that our tests fell short in terms of statistical significance. We argue that, after correcting for the non-independence of observations by clustering them at the individual level, the power in our test was dramatically reduced because of the high inter-clustering correlations. In fact, a naïve Wilcoxon rank-sum test considering all observations as independent would say that the comparisons of the *GD* and the *FD* with respect to the *PD* were statistically significant at the one percent level.

Summing up, our interpretation of the order effects is the following: subjects that played the *PD* after facing more unequal games were more likely to cooperate than those that started playing the *PD*, and it is because they did not have as reference point a more unequal dilemma. For Strangers, the evidence is the lowest drop of the cooperation rate in the *PD*, about only half of the drop in the *GD* and the *FD*. For Partners the evidence goes in the opposite direction: the cooperation rate in the *PD* increases between two-and-a-half and three times more than in the *GD* and the *FD*. Nevertheless, our loss of statistical power due to the inter-clustering correlations limit our results to suggestive evidence.

6 Concluding remarks

We explored experimentally what Rapoport (1967) defined as the two "motivational pressures" inducing defection in cooperation dilemmas. Unlike previous studies (Komorita et al., 1980; Ahn et al., 2001), we found that the fear of being exploited reduces more the likelihood to cooperate than the additional benefits of exploiting the counterpart. This reversal in the relative effects of *fear* and *greed* is the consequence of larger payoff differences within the asymmetric outcomes in 2×2 cooperation dilemmas.

We conducted a laboratory experiment where subjects faced three differ-

ent cooperation dilemmas for blocks of ten rounds each. When subjects were randomly rematched after each round, the larger effect of the *fear* dimension was immediately evident. When subjects interacted with the same participant repeatedly, the additional effect of *fear* needed to be disentangled from the history of the game. As the opponent's past cooperative decisions were largely rewarded in the game with a large degree of *fear*, the complementarity between reputational effects and the *fear* dimension suggests that path dependency becomes particularly relevant if the costs of being exploited, but not the benefits for the exploiter, are high.

The effects of *fear* and *greed* are reference dependent, as they only appear after a between-game comparison of the relative incentives to cooperate. Further research should be undertaken to validate this result in a broader set of social dilemmas such as public goods and common-pool resource games. A reduction in the perception of the costs of cooperation, even if it remains as a dominated strategy, could be a cost-effective strategy to increase cooperative behavior.

Acknowledgements

I want to thank to Samuel Bowles and Juan Camilo Cárdenas for their advise. I also acknowledge the contributions from Pablo Brañas, Marc Hofstetter and participants in the seminars at the Santa Fe Institute and Universidad de los Andes. The experiment was conducted with a grant provided by CeiBA (Centro de estudios Interdisciplinarios Básicos y Aplicados en Complejidad). Support through ANR - Labex IAST is gratefully acknowledged. Financial support was provided by Fundación Bolívar Davivienda.

References

Ahn, T., Ostrom, E., Schmidt, D., Shupp, R., Walker, J. 2001. Cooperation in PD games: Fear, greed and history of play. Public Choice, 106, 137–155.

Blonski, M., Ockenfels, P., Spagnolo, G. 2011. Equilibrium selection in the repeated prisoner's dilemma: Axiomatic approach and experimental evidence. American Economic Journal: Microeconomics 3, 164–192.

De Dreu, C., Greer, L., Handgraaf, M., Shalvi, S., Van Kleef, G., Baas, M., Ten Velden, F., Van Dijk, E., Feith, S. 2010. The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. Science, 328, 1408–1411.

De Dreu, C., Shalvi, S., Greer, L., Van Kleef, G., Handgraaf, M. 2012. Oxytocin motivates non-cooperation in intergroup conflict to protect vulnerable in-group members. PLOS ONE, 7 (11), 1–7.

Engel, C., Zhurakhovska, L. 2012. When the risk of cooperation is worth taking? the prisoner's dilemma as a game of multiple motives. Preprints of the Max Planck Institute for Research on Collective Goods, 1–34.

Fischbacher, U. 2007. z-tree: Zurich toolbox for ready-made economic experiments. Experimental Economics, 171–178.

Frohlich, N., Oppenheimer, J. 1996. When is universal contribution best for the group? characterizing optimality in the prisoners' dilemma. The Journal of Conflict Resolution, Vol. 40, No. 3, 502–516.

Kahneman, D., Tversky, A. 1979. Prospect theory: An analysis of decision under risk. Econometrica, 47 (2), 263–291.

Komorita, S., Sweeney, J., Kravitz, D. A. 1980. Cooperative choice in the n-person dilemma situation. Journal of Personality and Social Psychology, 504–516.

Kuwabara, K. 2005. Nothing to fear but fear itself: Fear of fear, fear of greed and gender effects in two-person asymmetric social dilemmas. Social Forces, 84 (2), 1257–1272.

Rapoport, A. 1967. A note on the "index of cooperation" for prisoner's dilemma. Journal of Conflict Resolution, 11 (1), 100–103.

Sally, D. 1995. Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. RATIONALITY AND SOCIETY, Vol. 7, No.1, 85–92.

Schelling, T. C. 1973. Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities. The Journal of Conflict Resolution, 17 (3), 381–428.

Simpson, B. 2003. Sex, fear and greed: A social dilemma analysis of gender and cooperation. Social Forces, 82 (1), 35–52.

Vlaev, I., Chater, N. 2006. Game relativity: How context influences strategic decision making. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32 (1), 131–149.

Appendix

A Experimental Instructions (translated from Spanish)

The following figures correspond to screenshots of the experimental environment using the z-Tree software. Figure A.1 displays the initial interface where the main instructions of the game can be read while the experimenter read them out loud. Figure A.2 shows the instructions to take the decision in the practice round. Figure A.3 displays the interface in which participants take their decisions for each one of the thirty rounds of play. Figure A.4 presents the feedback after each interaction: the subjects' own and their opponent's actions and earnings.

Figure A.1: Experimental instructions

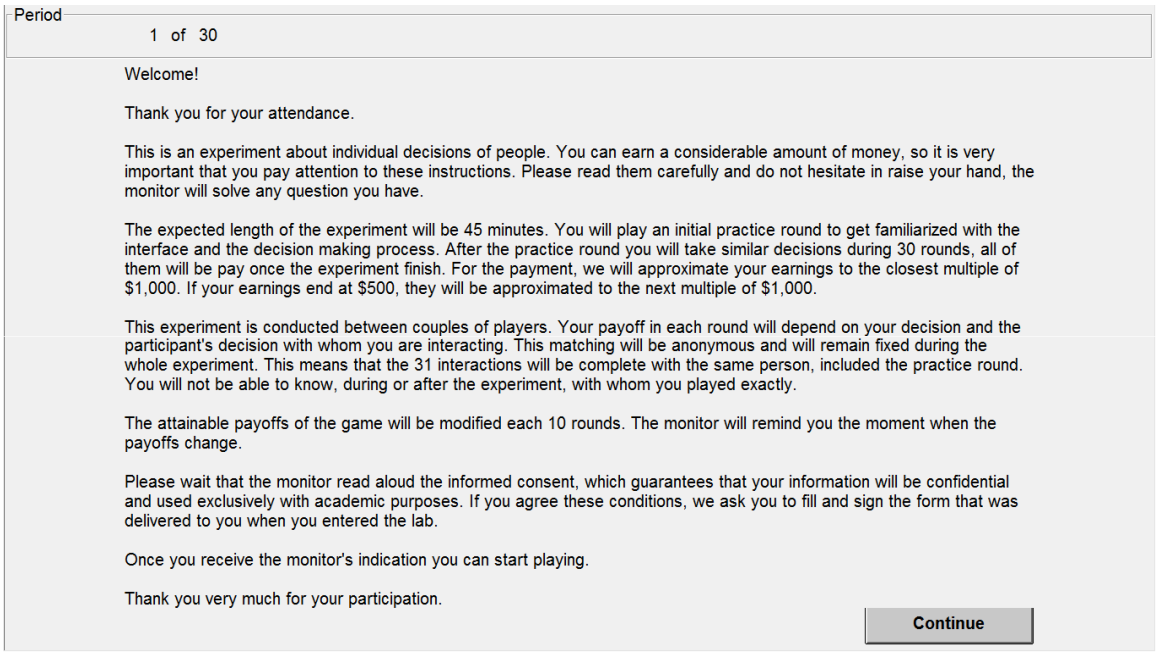


Figure A.2: Experimental instructions for the practice round

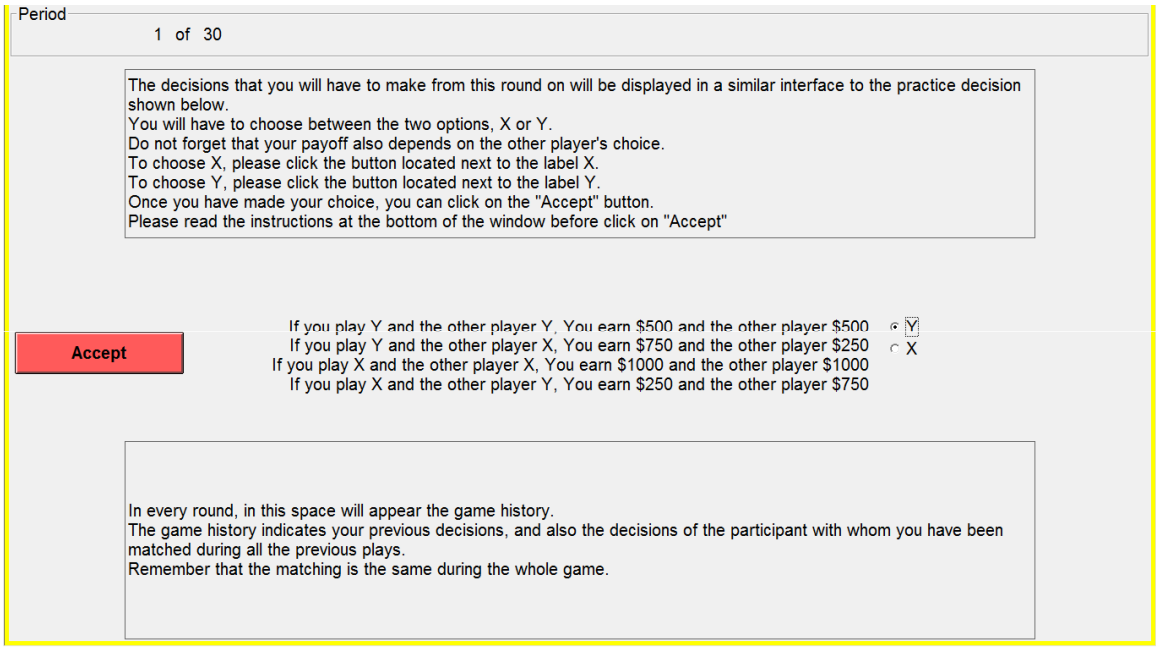


Figure A.3: Decision environment for a given round

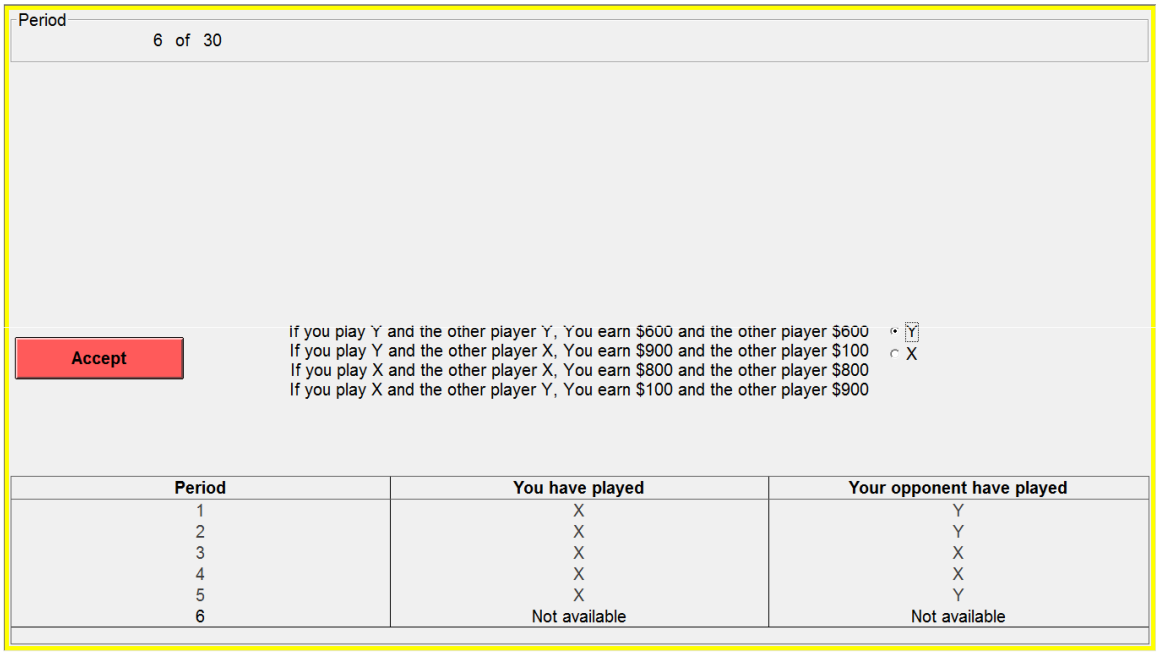


Figure A.4: Feedback after each round