Proceedings of FIKUSZ '10 Symposium for Young Researchers, 2010, 111-138 © The Author(s). Conference Proceedings compilation © Obuda University Keleti Faculty of Business and Management 2010. Published by Óbuda University Keleti Károly Faculty of Business and Management, Tavaszmező u. 15-17. H-1084 Budapest, Hungary. http://kgk.uni-obuda.hu/fikusz

Do People Accurately Anticipate Sanctions?

Raúl López-Pérez and Hubert János Kiss[1]

WORK IN PROGRESS; DO NOT CIRCULATE

September 21, 2010

Abstract:

We consider an experiment with four different games. In each, a first mover can choose between two allocations of payoffs for her and a co-player, who has then the option to either punish or reward the first mover. Two treatments are examined: One with monetary sanctions and another one with non-monetary ones (i.e., approval and disapproval). In both, we elicit first movers' expectations to study whether they accurately predict the average sanction at each allocation. The results show that the average first mover anticipates correctly the average sanction at most allocations in both treatments. The only exceptions occur in allocations where first movers happen to be rewarded, as they tend to underestimate the average strength of the reward.

Keywords: *Approval; disapproval; expectations; monetary sanctions; non-monetary sanctions; punishment; rewards; social norms.*

JEL Classification: C70, C91, D63, D74, Z13.

1 López-Pérez (corresponding author): Department of Economic Analysis, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain. Fax: + 34 91 497 6930. E-mail address: raul.lopez@uam.es.

Kiss: Department of Economic Analysis, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain. Fax: + 34 91 497 6150. E-mail address: hubert.kiss@uam.es.

1. Introduction

Social researchers (Akerlof, 1980; Axelrod, 1986; Elster, 1989; Fehr and Fischbacher, 2004) often stress that punishment and reward are important to foster cooperation and enforce social norms.[2] This claim seems well vindicated by controlled experimental evidence. On the one hand, there is abundant evidence that the availability of *monetary* punishment and rewards promotes cooperation, generosity, and fairness (Güth et al. 1982; Roth et al, 1985; Camerer and Thaler, 1989; Ostrom et al., 1992; Andreoni et al, 1993; Fehr and Gächter, 2000; Falk et al., 2005; Sefton et al, 2007; Vyrastekova and van Soest; 2008). On the other hand, a growing experimental literature shows that even *non-monetary* punishment and rewards can foster pro-social behaviors (Masclet et al, 2003; Rege and Telle, 2004; Noussair and Tucker, 2005; Ellingsen and Johannesson, 2008; Xiao and Houser, 2009, López-Pérez and Vorsatz, 2010).[3]

For any type of sanction to be effective in promoting a certain behavior (like compliance with a norm or the law), however, it seems important that the individuals correctly anticipate the *expected* sanction. For instance, it has been suggested that disapproval from work colleagues may prevent highly productive employees from exceeding the formally agreed rate of output (Homans, 1961), or preclude workers from underbidding the prevailing wage in a community (Akerlof, 1980). Yet these sanctions would be probably useless if the agents were not able to foresee them. Of course, this seems a realistic assumption in repeated interactions, as in the examples cited. However, many of our daily exchanges occur with strangers, and it is less clear that the assumption of accurate expectations is correct in this kind of settings. If people are unable to anticipate the sanctions, this might affect cooperation and norm compliance.

This paper experimentally investigates whether people have accurate expectations regarding sanctions. Subjects in our experiment play four simple games with a two-stage structure. In the first stage, one A-player chooses between two allocations of pecuniary payoffs for her/him and a B-player. In the second stage, the B-player observes A's choice and can punish or reward the A-player. Important for our research question, we elicit the expectations of the A-subjects regarding B's behavior. Furthermore, our design has two treatments across subjects: One with monetary sanctions, the other with non-monetary sanctions. The punishment/reward technology in both treatments is similar and in both cases equally costly for the B-player. The key difference is that the payoff of the A-player can be affected in the monetary-treatment but not in the non-monetary treatment. Instead, in the non-monetary treatment B-players can send "disapproval" or "approval" points to the A-players. The four games are selected so as to induce different motivations for punishment and reward, and hence investigate whether A-players are able to anticipate these motivations and the differences across games. Moreover, our two treatments allow

2 As in the classical study by Coleman (1990), we will use the term sanctions as an umbrella term to refer both to punishment (negative sanction) and rewards (positive sanction).

3 Examples of non-monetary punishment include humiliation, insults, shaming and ostracism, while social approval, honors, and praise are examples of non-monetary rewards. Throughout the paper, we use the term disapproval (approval) as a synonym of non-monetary punishment (reward).

us to check whether people predict better some type of sanctions, or whether they expect differences across treatments.

The task of forecasting sanctions in our games is arguably a complex one. For this reason, the accuracy of the *average* subject in forecasting both monetary and non-monetary sanctions in our games is remarkable. To start, most subjects are able to anticipate correctly the sign (positive/negative) of the average sanction in most games in both treatments (focusing on those allocations where the average actual sanction is significantly different than zero and hence has a clear sign). Even more: The point prediction by the average subject is not significantly different from the actual average sanction in 6 out of the 8 allocations in the monetary treatment, and in 7 out of 8 the allocations in the non-monetary treatment. At the *individual* level, however, we observe that subjects commit errors but also that these errors are not systematically biased in most allocations. Interestingly, subjects only predict with a bias at allocations where rewards are relatively more frequent; in these cases they tend to underestimate the average sanction. Overall, the evidence hence suggests that subjects are much better at predicting the strength of the 'stick' than that of the 'carrot', a result which might have important implications regarding the efficiency of negative and positive sanctions. A possible reason for this is that A-subjects fail to anticipate that they will be rewarded if they choose an allocation where they have a lower payoff than the B-player. We explore this and other potential determinants of the individual prediction error with a regression analysis, finding suggestive evidence in this line. In a comparison across treatments, finally, we observe that error rates are significantly smaller in most allocations in the non-monetary treatment. Hence, it seems that people anticipate better this kind of sanctions.

In the rest of the paper, we first describe and motivate our experimental design and procedures in section 2. Section 3 presents and discusses our results in the monetary treatment, and section 4 our results in the non-monetary treatment. Section 5 concludes with a brief review of some related literature and some ideas for further research.

2. Experimental Design and Procedures

There are two treatments in our experimental design: The Monetary treatment (M), and the Non-Monetary treatment (NM). Each subject participated only in one treatment. We describe first the M-treatment. Participants play four games, all of them two-player games with a similar two-stage structure. In the first stage, one player (called A) chooses between two allocations of money for herself and another player (called B). Table 1 shows the two (A, B) payoff allocations available in each game (called left and right; in what follows, we denote by 1L the left-hand allocation in game 1, by 2R the right-hand allocation in game 2, and so on). We motivate our choice of the payoff constellation later in this section. Payoffs are presented in points, at the exchange rate 10 points = 1 Euro.

TABLE 1—THE ALLOCATIONS IN THE 4 GAMES		
Game	Allocation	
	Left	Right
1	(250, 100)	(200, 150)
2	(250, 100)	(250, 250)
3	(100, 200)	(150, 150)
4	(100, 200)	(100, 300)

In the second stage, B observes A's choice and can then *monetarily* punish or reward A, at a *fixed* cost of five points for B. More precisely, B can either increase or decrease A's payoff by up to 100 points, but for that she must previously pay five points from her allocation share. If B does not want to affect A's balance, no points are deducted from her allocation share. As an example, suppose that A chooses allocation (x_A, x_B) in a game. If B decides not to pay the five points, allocation (x_A, x_B) is implemented. If she pays the five points fee, however, she can choose a "point score" $s \in [-100, 100]$ so that A's payoff in the game is $x_A + s$, while B gets a payoff of x_B -5. For simplicity, s had to be a multiple of 10.

The NM-treatment is identical to the M-treatment, except for one important difference: In the second stage of each game, B cannot affect the monetary payoff of the co-player, but approve or disapprove her choice –i.e., she can reward or punish A in a *non-monetary* manner. As in the M-treatment, B must pay a fixed cost of five points for that. If B pays the five points, she can send an "evaluation score" $s \in [-100, 100]$ to A expressing either approval or disapproval of A's choice. In this respect, the instructions explicitly used the words "approval" for positive scores and "disapproval" for negative scores. Hence, an evaluation score of s = +100 means maximal approval, and s = -100 means maximal disapproval. The interpretation of score s is common knowledge. As an illustration of the payoff structure, suppose that A chooses allocation (x_A, x_B) in a game. Since B cannot affect A's balance in this treatment, A is then sure to get a payoff of x_A. With respect to B, she would get a payoff of x_B -5 if she sends an evaluation score, and a payoff of x_B otherwise.

The experimental procedures were identical in both treatments. We conducted seven sessions at the University of (location), and a total of 92 subjects participated in the M-treatment, and 84 in the NM-treatment. Subjects were students from different disciplines (9.6 percent came from the faculty of economics) and not students of the experimenters. Before the start of each session, we distributed instruction and decision sheets (dependent on role) in a class room, leaving enough space between seats to ensure anonymity. Then the subjects entered the room. The sheets were initially covered and the subjects could freely choose their seat; in that manner, we assigned them to be either an A- or a B-player. Subjects could read the instructions at their own pace and questions were answered in private. We avoided terms such as "punishment" or "reward" in the instructions. Before proceeding with their decisions, participants had to fill out control questions to make sure that they understood the rules.

In each treatment, subjects played the same four games in the same role and with the same anonymous co-player. We presented all four games on the same decision sheet; we believe that this feature makes our results more robust, as subjects take their decisions with the maximum information available. In particular, we prevent potential order effects that could appear if the games were presented one by one. Further, no subject was informed of her counterpart's actual choice in any game in order to avoid repeated game effects and changes of mood which would severely complicate the data analysis (e.g., the mood of a B-player could change depending on the A-player's choice in a preceding game). Therefore, we employed the *strategy method* to elicit the decisions of the B-players, i.e., they indicated in each allocation of each game whether they wanted to pay the five points fee, and if this was the case, we asked them which score $s \in [-100, 100]$ they wanted to assign to their co-player.[4]

In both treatments, we elicited the A-players' expectations with regard to the B-players' behavior. More precisely, we asked two questions for each allocation: (i) The percentage of B-players who would pay the 5 points fee at that allocation, and (ii) the average score s assigned by those B-players. Note that the product of these two numbers equals the expected sanction at that allocation. Subjects were not paid for answering these questions. While this might reduce accuracy in comparison with a situation in which beliefs are incentivized, we do not see this as a problem given our research question; we believe that the external validity of our study would be in doubt if we provided incentives and those incentives changed accuracy (people are not frequently paid for the accuracy of their beliefs about punishment and reward *out of the lab*).[5]

After subjects made their decisions in the four games, they answered a brief questionnaire. Then we collected their decision sheets and only thereafter selected one game randomly for payment in order to prevent income effects – of course, subjects were informed of this at the beginning of the experiment. Subjects were paid privately, and earned on average 18.3 Euros in the M-treatment and 20.6 Euros in the NM-treatment. All A-playerswere informed at the time of payment about the score sent by his/her co-player B (if any) at the payoff relevant allocation. Each session lasted approximately 60 minutes.[6]

4 A further advantage of the strategy method is that it maximizes the amount of statistical data gathered. In principle, this method might induce different behavior than the specific response method where participants know the choice made by the co-player. Yet Brandts and Charness (2009) review the experimental studies that use both methods and find no treatment differences in most of them. Moreover, they find that differences are particularly unlikely in experiments in which players make numerous choices (as in ours).

5 Admittedly, some subjects might answer in a random manner if they are not paid. However, provided that their choices are uniformly distributed, this should at least not affect the average expectation. In any case, we note that there is still an ongoing debate on whether incentivized belief elicitation increases accuracy. For example, Gächter and Renner (2010) elicit subjects' beliefs in a 10-times repeated public good game, and compare incentivized and non-incentivized elicitation. They find no differences across treatments in the absolute average error in period 1, whereas the error is significantly smaller in the incentivized treatment in the remaining periods. Hence, their data suggests that incentives affect the way in which people's beliefs change, but not the initial beliefs. Provided that this result can be generalized, it would be clearly relevant in our games, as they are one-shot.

6 The instructions for the A-players in the M-treatment and the NM-treatment are in appendix B and

We now motivate the payoff constellation in our games. For this, it is worthy to note first that the ability to accurately predict sanctions depends crucially on a good understanding of the other players' potential motives to sanction. In this respect, several experimental studies (Leibbrandt and López-Pérez, 2010a survey this literature) suggest that both inequity-aversion and reciprocity are two key motivations behind *monetary* sanctions. With this in mind, we selected our four games so that they include allocations in which (i) both motivations predict punishment/reward, (ii) only one of these motivations predicts punishment/reward, and (iii) none of these two motivations predicts sanctions. For clarification, table 2 summarizes the predictions of inequity-aversion and reciprocity in each allocation of the four games in the M-treatment. Since the B-players have arguably different motivations to sanction at the allocations, we expected them to sanction differently at them (as indeed happened). Our selection of games allows us to analyze whether the A-players anticipate the varying behavior of the B-players, and whether some situations are more prone to error (thus suggesting that the error is due to an inaccurate anticipation of B's motives).[7]

TABLE 2—PREDICTIONS OF MONETARY PUNISHMENT/REWARD

Game	A's and B's Allocation Share			Predictions Left		Predictions Right	
	Left		Right	Punishment	Reward	Punishment	Reward
1	(250, 100)	vs.	(200, 150)	IA, RP	----	IA	RP
2	(250, 100)	vs.	(250, 250)	IA, RP	----	----	RP
3	(100, 200)	vs.	(150, 150)	----	IA, RP	RP	----
4	(100, 200)	vs.	(100, 300)	RP	IA	----	IA, RP

The following notation is used: IA = Inequity aversion, RP = Reciprocity.

Our payoff constellation is also convenient for the NM-treatment as well, as we again expected some variability in the strength of the actual sanctions across allocations. Our design allows us to investigate if the A-subjects in the NM-treatment are able to foresee these variations. The reason why we expected sanctions to differ across allocations is that,

C, respectively. We also provide one example of a decision sheet for the M-treatment, to illustrate how beliefs were elicited.

7 To understand the predictions in table 2, note that models of inequity-aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) predict punishment of the A-player in the M-treatment if she has a larger payoff than the B-player, and reward if she has a smaller payoff; whereas reciprocity models like Dufwenberg and Kirchsteiger (2004) predict punishment if the B-player was harmed by the A-player's choice (i.e., in our games if A chooses the allocation where B's payoff is smallest), and reward if she was helped. These predictions hold for a very large range of the models' parameters. The reader may consult Leibbrandt and López-Pérez (2010a) for a more detailed discussion.

although the available evidence regarding the motivations behind non-monetary sanctions is rather scant, we hypothesize that motivations akin to inequity-aversion and reciprocity might explain the occurrence of approval and disapproval (as we report later, our evidence happened to be somehow consistent with this). A motivation akin to inequity-aversion means that individuals approve (disapprove) choices that leave them advantaged (disadvantaged), whereas a motivation akin to reciprocity means that individuals approve (disapprove) choices that help (harm) them -see Holländer (1990) and Kandel and Lazear (1992).[8] That is, these motivations respectively predict approval/disapproval in the same allocations where inequity-aversion and reciprocity predict monetary reward/punishment (see Table 2), thus generating differences in sanctions across allocations.

To finish, we make several remarks on the punishment/reward technology used in our games. This is an extremely simple technology, thus facilitating the already complex task of predicting sanctions. Since there is still not much evidence on expectations about sanctions, we believe that a simple technology is convenient for our research question. Although an even simpler technology would allow the B-players to sanction for free, we chose to introduce a (small) fee in order to prevent random choices from selfish players, which could occur if sanctioning was not costly and would complicate the interpretation of our results. We also note that our approval/disapproval technology is an adaptation of the technology used by Masclet et al. (2003) in their treatment with non-monetary punishment.

3. Results: Monetary Treatment

This section focuses on monetary sanctions and whether subjects anticipate them correctly. We first provide a succinct overview of *actual* sanctions by the B-players, underlining several key patterns. With this in mind, we then investigate whether the A-players are able to anticipate the average sanction in each allocation of the four games, and why they commit errors sometimes. The choices of the A-players in both treatments are shown in appendix A.

3.1 Monetary sanctions by the B-players: Overview

Table 3 shows for each allocation of each game in the M-treatment, the frequency of B-players who invested five points to punish, average punishment among those players who punished, and the same figures with respect to rewards. In game 1 (250/100 vs. 200/150), for instance, we observe that 45.7 percent of the B-players punish at the left-hand allocation (250/100), and that the average punishment is s = -99.5. The corresponding numbers for rewards in the same allocation are 13 percent and s = 83.3.

8 In what follows, we use the terms "reciprocity" and "inequity-aversion" in a flexible manner to refer not only to the theories that apply to monetary punishment/reward, but also to their analogues for disapproval/approval.

TABLE 3—Punishment and reward in the M-treatment: Frequency and strength

				monetary punishment				monetary reward			
Game	Allocation			%	average	%	average	%	average	%	average
	Left		*Right*	*Left*	*Left*	*Right*	*Right*	*Left*	*Left*	*Right*	*Right*
1	(250,100)	vs.	(200,150)	45.7	99.5	37	68.2	13	83.3	28.3	81.5
2	(250,100)	vs.	(250,250)	56.5	93.5	17.4	77.5	10.9	70	28.3	92.3
3	(100,200)	vs.	(150,150)	15.2	87.1	15.2	75.1	52.2	90	15.2	95.7
4	(100,200)	vs.	(100,300)	28.3	80	6.5	80	28.3	82.3	63	92.1

The first thing that catches the eye in table 3 is the large heterogeneity regarding punishment and reward. For instance, in each game and at each allocation there are always some B-subjects who punish but also some who reward and these figures are never lower than 10 percent . However, the data exhibits several patterns, which we briefly comment in what follows.[9] *First*, disadvantageous payoff differences and the existence of harm both affect the frequency of monetary punishment, suggesting a crucial role for both inequity-aversion and reciprocity. In effect, the frequency of this behavior reaches a highest level in allocations 1L and 2L, that is, the only allocations in which both forces predict punishment; an intermediate level in allocations 1R and 4L, where just one of these forces predicts punishment; and a lowest level in allocations 2R, 3L and 4R, where none of these forces predict punishment. *Second*, enjoying advantageous payoff differences and being helped by the A-player are both relevant for monetary rewards. We observe that rewards are more likely in allocations 3L and 4R, where both factors are at play; reach an intermediate level in those allocations in which one factor predicts rewards (R1, R2, L4), and a lower level in the remaining allocations (L1, L2, R3). This suggests that both inequity-aversion and reciprocity are crucial to account for monetary rewards. In line with the previous two points, *third*, the regression analysis reported in Leibbrandt and López-Pérez (2010a) shows that the *strength* of monetary punishment and reward positively depends on the size of the payoff difference and the size of harm/help (i.e., the difference in B's payoffs across the two allocations of a game). In summary, both inequity-aversion and reciprocity seem crucial to explain both the occurrence and strength of monetary sanctions.[10] In the next section we study if the A-types are able to anticipate average sanctions in this complex setting.

9 For an extensive analysis of the determinants of punishment and reward in these games, we refer the reader to Leibbrandt and López-Pérez (2010a).

10 Yet we note that other (minor) motives apart of inequity-aversion and reciprocity seem to be at work behind monetary sanctions. In effect, while the classification analysis in Leibbrandt and López-Pérez (2010b) finds two main groups of subjects - (i) subjects who sanction taking reciprocity and/or inequity into account, and (ii) subjects who never punish or reward, thus behaving in accordance with the standard selfish model-, the analysis also finds smaller fractions of subjects who punish in a spiteful or maybe competitive manner, and apparently altruistic subjects who tend to reward always.

3.2 Expectations on monetary sanctions of the A-players

Table 4 presents some aggregate data for each allocation of each game of the M-treatment. To start, we report the *average expected sanction* at each allocation, across all A-players – e.g., the average A-player expects a sanction of -26.47 in allocation 1L. Note in this respect that, for each subject and each allocation, we compute the sanction expected by her/him as the following product: Percentage of B-players who are expected to pay the 5 points fee multiplied by the average expected score from those players. For comparison, we report as well the *average actual sanction* at each allocation by the B-players (i.e., the average score s across all B-players; note that we set s = 0 when a B-player does not pay the 5 points fee). In allocation 1L, for instance, we observe that the average actual sanction is s = -34.56. We also report the results from two-sided Mann-Whitney tests comparing the average sanction and the average expectation.

TABLE 4—Expectations and sanctions at each allocation (M-treatment)

	Game 1		Game 2		Game 3		Game 4	
Allocation	(250,100)	(200,150)	(250,100)	(250,250)	(100,200)	(150,150)	(100,200)	(100,300)
Average expected sanction	-26.47	-4.3	-42.83	4.9	10.49	-4.06	-9.27	26.61
Average actual Sanction	-34.56	-2.17	-45.22	12.61	33.69	3.04	0.65	52.83
P-value (Mann-Whitney)	0.4674	0.7589	0.3008	0.9111	*0.0985*	0.2531	0.3959	*0.0067*
% expecting sanction = 0	8.7	13	13	13	6.5	13	10.9	10.9
% expecting sanction < 0	69.6	43.5	76.1	34.8	26.1	52.2	52.2	21.7
Average error (expectations)	35.6	30.49	38.55	32.16	46.01	19.58	33.43	46.42

Note: P-values come from two-sided Mann-Whitney tests. Differences significant at the 10 percent level are present in italics. N = 46.

Do the A-players accurately estimate the average sanction at each allocation? To answer this question, we first study whether they at least anticipate correctly the *sign* of the sanction, and later move a step further and analyze whether they accurately anticipate the *strength* of the sanction.

Regarding the sign, we focus on those allocations in which the average actual sanction is significantly different from zero at the 10 percent level, that is, allocations 1L, 2L, 3L and 4R.[11] Since the average sanction has a clear sign (either positive or negative) at these allocations, it makes sense to ask if subjects can anticipate it. To start, Table 4 shows that the average expected sanction has the same sign as the average actual sanction in these

11 At the remaining allocations 1R, 2R, 3R and 4L, the p-values of the two-sided Mann-Whitney test are 0.39, 0.22, 1 and 1, respectively.

allocations. Hence, the average A-player anticipates correctly the sign. On the individual level, we also observe that most subjects tend to predict correctly the sign of the average sanction. To clarify this, Table 4 reports for each allocation the percentage of A-players who expect a null sanction or a negative sanction. For instance, 8.7 and 69.6 percent of the A-players expect a null and a negative sanction, respectively, at allocation 1L. We observe that always more than two thirds of the subjects correctly anticipate the sign in the selected allocations –more precisely, the percentage of A-subjects who correctly anticipate the sign in allocations 1L, 2L, 3L and 4R are 69.6, 76.1, 67.4, and 67.4, respectively. We summarize our findings as follows:

Result 1: Most A-subjects accurately anticipate the sign of the average actual sanction when it is significantly different from zero.

Regarding the strength, Table 4 shows that the average A-player tends to estimate it rather accurately. In effect, the two-sided Mann-Whitney test reveals that the average expected sanction is not significantly different from the average actual sanction at most allocations. The only exceptions are 3L and 4R, and it is worthy to note that these are the two allocations where the average sanction reaches a highest positive level.[12] To phrase it like this, the average player anticipates better the stick than the carrot in our games, a point that we will address further later. In addition, we can also see that the average A-player roughly anticipates the ranking of the allocations depending on their corresponding average sanction. For this, Figure 1 presents the average sanction and expectation at each allocation, where the allocations have been ordered from lowest (i.e., more negative) to highest (i.e., more positive) average sanction. As we can see, the actual ordering of the allocations basically coincides with that expected by the average A-player (note: the average actual sanction is not significantly different from zero at the allocations 4L and 3R where it does not coincide).

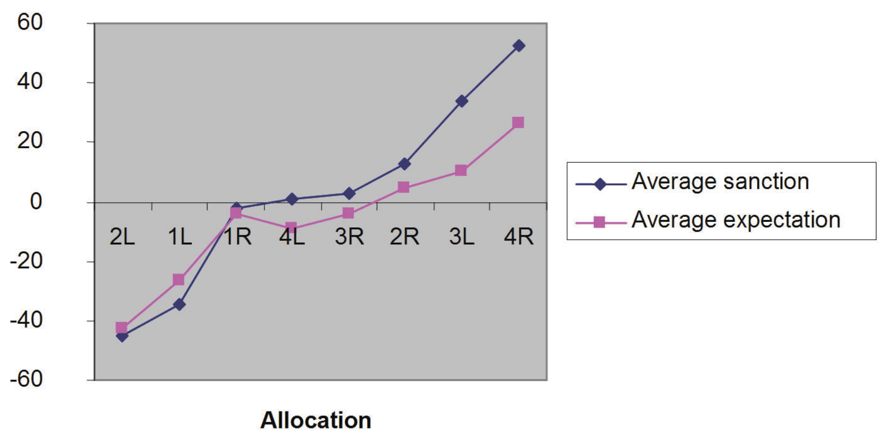


Figure 1: Average sanction and expectations at each allocation (M-treatment)

12 In fact, the average actual sanction is significantly higher than zero at the 10 percent level only at 3L and 4R, the p-values being less than 0.01 in both cases (one-sided Mann-Whitney test).

On the individual level, nevertheless, subjects frequently estimate the strength of the sanction with some error. We define a player's prediction error at one allocation as the *absolute* difference between the average sanction expected by that player at that allocation and the corresponding average actual sanction. Table 4 provides the average prediction error at each allocation; this is 35.6 at allocation 1L, for instance.[13] We can see that the average error is high in many allocations and in fact significantly higher than zero in all allocations (one-sided Mann-Whitney test, $p < 0.01$).

How is it possible then that the average A-player has often accurate expectations? The reason is that the errors caused by overestimation of the average sanction compensate the errors caused by underestimation at most allocations. In effect, if we compute the prediction error at each allocation as the difference between expectations and actual values (but not in absolute value), and consider the distribution of individual errors thus obtained at each allocation, we observe the following. First of all, if we test the hypothesis that the average of the distribution is zero (using the two-sided Mann-Whitney test) and the hypothesis that the median is zero (single-sample Wilcoxon signed-rank test), we find that *both* hypotheses are rejected at the 10 percent level only at allocations 3L, 3R, and 4R (p-value < 0.06 in all cases). This is an indication that subjects' expectations are not systematically biased in most allocations, except in the exceptions cited (recall in this respect that allocations 3L and 4R are the only ones in which the average expected sanction is significantly different that the average actual sanction). In fact, there seems to be a systematic downward error in allocations 3L, 3R, and 4R, and we note incidentally that the actual average sanction in these allocations is positive. Hence, we have further evidence that people foresee better the punishment than the reward. Finally, the third standardized moment of each distribution, which is a well-known coefficient of skewness is never larger than 0.6 (in absolute terms) at any allocation, suggesting that the errors are symmetrical.[14] We summarize our discussion as follows:

Result 2: The average A-subject accurately anticipates the strength of the average actual sanction at most allocations (in 6 out of 8 cases). Although subjects tend to commit errors of a significant size, they are not systematically biased, except in three allocations. Subjects tend to underestimate the average strength of the actual sanction in these three allocations, where rewards are relatively frequent.

Since subjects tend to predict the strength with a considerable error, we analyze in the next section some potential determinants of this prediction error.

3.3 Determinants of the prediction errors

We use here a regression approach in which the dependent variable is the (absolute) prediction error committed by each individual at each allocation. As independent variables,

13 To put this figure in context, consider a hypothetical individual who predicts the actual average sanction s by picking a number between -100 and 100 in a randomly uniform manner. One can show that the average absolute error would then be equal to 50 + s2/200. In allocation 1L, for instance, s = -26.47 so that the average error should therefore be 53.5. This figure is larger than the average prediction error in this allocation (35.6), and this happens to occur in all allocations.

14 Going from 1L to 4R, the coefficient equals -0.18; -0.55; 0.23; -0.45; -0.59; 0.52; -0.19 and -0.48.

we use individual socio-demographic variables collected from our questionnaire (gender, political ideology, and religiosity), and several key variables.

To motivate some of these variables, recall from section 3.1 that both payoff differences (inequity-aversion) and harm-help (reciprocity) seem important to explain the actual occurrence and strength of sanctions. With this in mind, we conjecture that one possible reason why individuals make errors is because they are unable to anticipate these patterns. For instance, they might fail to accurately forecast an increase in sanctions due to increasing payoff differences.[15] Since this prediction failure might depend on whether the other player is advantaged or disadvantaged (a potential explanation of the bias reported in Result 2 for three allocations), we distinguish two variables. The first one is called *paydiff-*, and is defined as the difference $x_B - x_A$ between B's payoff and A's payoff in the corresponding allocation, when B has a smaller payoff than A (otherwise, it takes value zero). For example, this variable equals 50 in allocation 1R, and 0 in allocation 3L. In turn, variable *paydiff+* is analogously defined for the case when B has a larger payoff than A –e.g., it takes value 0 and 100 in allocations 1R and 3L, respectively. In addition, we consider a dummy variable called *strict*, which takes value 1 if players' payoffs at the corresponding allocation are identical.

Players might also fail to anticipate that reciprocity is important to explain sanctions, and based on Result 2 we conjecture that this could be more pronounced with respect to rewards. For this reason, we introduce a dummy variable *Rec+*, taking value 1 when reciprocity (Dufwenberg and Kirchsteiger, 2004) predicts rewards, and value 0 otherwise (i.e., if reciprocity predicts punishment; see Table 2 for further clarification). For instance, this variable equals 0 in allocation 2L, and 1 in allocation 4R.

Finally, we consider three additional variables. One of them (*spiteful type*) is a dummy variable that takes value 1 if the corresponding A-player chose allocation 2L. Since this choice harms the B-player and provides no benefit to the A-player, it signals some negative predisposition towards the other player. We conjecture that this might have an effect on the prediction error: Since these types are least collaborative, they might be more likely to misreport their expectations or choose them randomly. Alternatively, if these types tend to believe that many others are like them (False consensus; consult Offerman, 2002, for a discussion and some references), this would lead to significant mistakes. Pursuing this kind of idea further, we also consider a dummy variable (IA type) that takes value 1 if the corresponding A-player chooses allocation 4L, a choice which could be a signal of inequity-aversion.[16] Another variable (*pessimism*) refers to the expectation pattern of each subject, as it equals the number of allocations at which that subject expects a negative sanction. For instance, if a subject expects to be sanctioned at 4 allocations, then the value of this variable is 4. Using this variable, we aim to check whether pessimistic people are more likely to commit larger errors than optimistic people (i.e., those who think that punishment is very unlikely). Given all this, Table 5 presents the results of two OLS regres-

15 Alternatively, subjects might be able to anticipate that payoff distances affect sanctions, but since the prediction problem probably becomes more complicated when the distance grows, they might make larger errors. For instance, forecasting seems easier when the allocations are strictly egalitarian (e.g., 3R) than otherwise. We try to distinguish between these two accounts later (see footnote 16).

16 We have considered additional variables that refer to the choices made by the A-players but they are not significant.

sions. Regression (1) considers all variables, while regression (2) does not consider the socio-demographical variables and other variables which do not happen to be significant at the 10 percent level. We focus in what follows on regression (2).

TABLE 5: RESULTS OF REGRESSION ANALYSIS
Dependent Variable: Prediction error

	Paydiff -	Paydiff +	Rec. +	Spiteful type	Pessimism	Female dummy	Political ideology	Religiosity	Strict	IA type	Constant
(1)	0.067 (0.052)	0.113*** (0.037)	5.621 (3.986)	32.685*** (10.678)	1.709 (1.131)	-7.451 (5.148)	0.474 (1.214)	-0.626 (0.717)	-3.457 (6.138)	-3.350 (4.856)	21.513** (8.753)
(2)	0.095*** (0.031)	0.097*** (0.022)	7.535*** (2.757)	36.204*** (10.296)	----	----	----	----	----	----	18.617*** (3.567)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N = 368. In parentheses, standard errors clustered on individual level. The R-squared of regressions (1) and (2) is 0.22 and 0.20, respectively.

We observe that the independent variables paydiff- and paydiff+ are highly significant. The estimated coefficients tell us that for each 100 points difference, the prediction error increases in 9.5 and 9.7 points, respectively. This shows that subjects' anticipation of punishment/reward tend to be less accurate when the payoff differences grow larger. Similarly, the variable Rec+ is also significant, so if reciprocity predicts reward/punishment, the prediction error increases/decreases in almost 8 points. This is another indication that subjects anticipate better punishment than rewards. We also observe that the players classified as spiteful make very significantly higher errors than the other players.[17]

Result 3: The magnitude of the prediction error that an A-subject makes grows significantly higher if (i) the payoff distance with the co-player increases, (ii) reciprocity predicts rewards at the corresponding allocation, and (iii) the subject is among those classified as spiteful.

17 The previous analysis uses the absolute value of the prediction error as the dependent variable. Hence, it focuses on the magnitude of the error, not its sign –i.e., whether errors are under- or over-estimations of the actual value. Yet one could conjecture that some of our variables are related to the sign of the error. For instance, A-players might underestimate the increase in rewards due to an increased payoff advantage by the B-player. To explore this kind of questions, we have conducted analogous regressions where the error by individual *i* is simply defined as the difference between the sanction expected by i and the actual average sanction. The results show that paydiff+, spiteful type, and pessimism are significant variables with negative coefficients, suggesting that these factors tend to induce under-estimations -the fact that paydiff+ is significant also suggests that subjects fail to understand that advantageous payoff differences affect rewards. The other variables are not significant. Hence, it seems that these other variables have (i) no significant effect on the error (as previous regression (1) indicated for the socio-demographic variables) or (ii) a significant effect which is not biased below or above. The results of these regressions are available from the authors.

4. Results: Non-Monetary Treatment

This section focuses on non-monetary sanctions and whether subjects anticipate them correctly. We first provide a succinct overview of *actual* sanctions, underlining several key patterns. With this in mind, we then investigate whether the A-players are able to anticipate the average sanction in each allocation of the four games, and why they commit errors sometimes.

4.1 Non-Monetary sanctions by the B-players: Overview

Table 6 presents for each allocation of each game, (a) the percentage of B-players who invested five points to punish in a non-monetary manner (i.e., disapprove), (b) average disapproval among those players who disapproved, and the same figures with respect to non-monetary reward (i.e., approval). In game 1, for instance, we observe that 38.1 percent of the B-players disapprove the choice of allocation 1L, and that the average disapproval is of s = -78.8. The corresponding numbers for approval in the same allocation are 4.8 percent and s = 75.

TABLE 6—Disapproval and approval in the NM-treatment: Frequency and strength

				Disapproval				Approval			
Game	Allocation			%	Average	%	Average	%	average	%	average
	Left		*Right*	*Left*	*Left*	*Right*	*Right*	*Left*	*Left*	*Right*	*Right*
1	(250,100)	vs.	(200,150)	38.1	78.8	14.3	58.3	4.8	75	28.6	73.3
2	(250,100)	vs.	(250,250)	52.4	82.7	2.4	60	4.8	75	47.6	90.5
3	(100,200)	vs.	(150,150)	16.7	65.7	2.4	100	38.1	80.6	42.9	84.4
4	(100,200)	vs.	(100,300)	35.7	83.3	7.1	80	20.1	78.8	52.4	86.8

We briefly comment three key patterns in the data -see again Leibbrandt and López-Pérez (2010a) for a more extensive discussion. *First*, it seems that both harm and *strict* payoff equality are crucial to explain the frequency of disapproval. Table 6 shows that the frequency of disapproval is above the average and even at an arguably high level (larger than 35 percent) *if and only if* the B-player is harmed and strict payoff equality is not achieved (i.e. in allocations 1L, 2L, and 4L). In contrast, disapproval is significantly smaller in the remaining allocations, where B is not harmed and/or strict equality is achieved. In a symmetric manner, *second*, it appears that both help and strict equality are important to explain approval: Table 6 shows that a choice is approved above the average (and by more than 35 percent of the subjects) *only* if (i) it helps the B-player and/or (ii) it achieves strict payoff equality (2R, 3L, 3R, 4R). In contrast, approval is very low (< 5%) only in some allocations like 1L and 2L, where conditions (i) and (ii) are not satisfied. The data also suggests that a force akin to inequity-aversion (I disapprove/approve if you have more/less than me) plays a less important role in explaining approval and disapproval. In an allocation like 4L where only inequity-aversion predicts approval, for instance, approval is below average and smaller than in 3L, where also reciprocity predicts approval (it must be incidentally noticed that both allocations are identical).

Third, the regression analysis reported in Leibbrandt and López-Pérez (2010a) shows that the *strength* of non-monetary punishment/reward positively depends on the size of harm/help (i.e., the difference in B's payoffs across the two allocations of a game). While there is also some dependence on the size of the payoff difference between the players, the coefficient for harm/help is significantly larger than that for the payoff difference. In comparison with the other treatment (see section 3.1), this suggests that non-monetary sanctions are *less* dependent on the existence of payoff differences between players than monetary sanctions (leaving aside the fact that strict equality fosters/diminishes approval/disapproval).

4.2 Expectations on non-monetary sanctions of the A-players

Table 7 is the analogue of Table 4, presenting aggregate data on expectations at each allocation of each game of the NM-treatment.

TABLE 7—Expectations and sanctions at each allocation (NM-treatment)

	Game 1		Game 2		Game 3		Game 4	
Allocation	(250,100)	(200,150)	(250,100)	(250,250)	(100,200)	(150,150)	(100,200)	(100,300)
Average expected sanction	-24.93	10.69	-42.49	17.62	18.46	0.86	-13.74	30.33
Average actual Sanction	-26.43	12.62	-39.76	41.67	19.76	33.81	-14.76	39.76
P-value (Mann-Whitney)	0.1170	0.2570	0.6261	0.4158	0.4445	*0.0024*	0.7322	0.6981
% expecting sanction = 0	11.9	26.2	16.7	38.1	31	33.3	21.4	21.4
% expecting sanction < 0	85.7	14.3	83.3	0	4.7	30.1	59.5	0
Average error (expectations)	20.19	13.5	31.95	32.53	21.74	35.92	21.23	26.93

Note: P-values come from two-sided Mann-Whitney tests. Differences significant at the 10 percent level are present in italics. N = 42.

As before, our main research question is whether the A-players' correctly anticipate the average sanction. We first investigate if A-players anticipate correctly the *sign* of the average sanction. For this, we focus on those allocations where the average actual sanction is significantly different from zero at the 10 percent level, so that the average sanction has a definite sign. In fact, this happens at all allocations except 1R (two-sided Mann-Whitney test, p-value = 0.12).[18] As Table 7 indicates, the majority of the subjects anticipate correctly the sign of the average sanction at almost every of the allocations considered. For instance, 85.7 percent of the subjects predict a negative sanction at 1L, and at least 59.5 percent anticipate the sign correctly at any other allocation, except 3R. There only 36.6 percent of the subjects properly forecast a positive sanction. To summarize:

18 The second highest p-value is that of allocation 4L (0.09).

Result 4: Most A-subjects accurately anticipate the sign of the average actual sanction in 6 out of the 7 allocations where it is significantly different from zero.

We now pass to study whether A-players anticipate correctly the *strength* of the average sanction, on the aggregate and individual level. We first note that the average A-player approximately anticipates the ordering of the allocations depending on their corresponding average sanction. To illustrate this, Figure 2 presents the average sanction and expectation at each allocation, where the allocations have been ordered from lowest (i.e., more negative) to highest (i.e., more positive) average sanction (note that this ranking is different than that in Figure 1). Note that the only failures occur in allocations 3R and 2R, where rewards are frequent, so that again we observe that the "stick" is better anticipated than the "carrot".

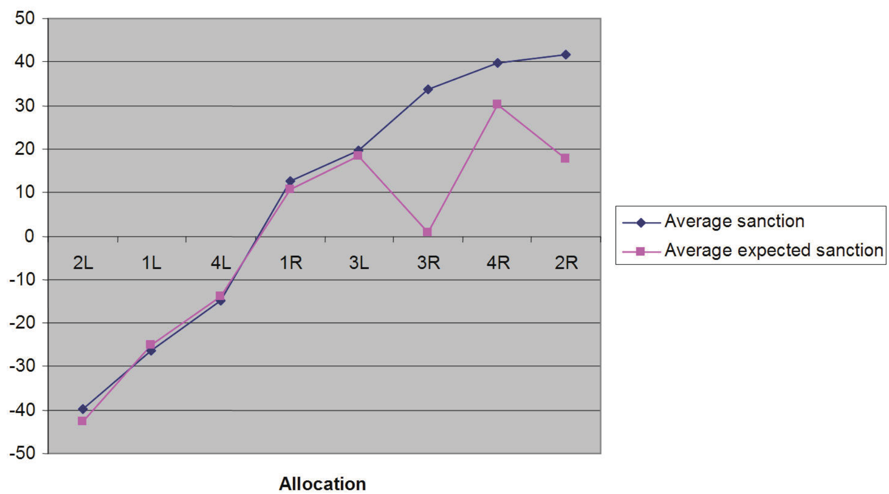


Figure 2: Average sanction and expectations at each allocation (NM-treatment)

On the other hand, the Mann-Whitney test reveals that the average expected sanction is different from the average actual sanction at the 10 percent level only in allocation 3R.[19] Hence, the average subject anticipates rather accurately the average sanction. As in the monetary treatment, however, the averages mask substantial heterogeneity between the subjects. As shown in Table 7, subjects generally estimate the strength of the sanction with some error. Using the same definition of error as in Table 4, the average error is again considerable in many allocations and in fact significantly higher than 0 in all allocations according to the one-sided Mann-Whitney test ($p < 0.01$).

Do subjects commit systematic errors at any allocation? Using the single-sample Wilcoxon signed-rank test we find that the median is not zero at 1R, 2R, 3R and 4R (p-value < 0.05). At the same allocations, we reject the hypothesis that the average of the distribution is zero (two-sided Mann-Whitney test, p-value < 0.05 in all cases). Note that

19 At visual inspection, the actual and expected average sanction differ considerably at 2R, suggesting that they come from different distributions. The non-parametric test we use finds them proceeding from the same distribution because the points coming from the distributions "mix" well when pooled.

these allocations present the highest positive sanctions (except 1R which is not significantly different from zero), evidence again that rewards are predicted with a systematic error. More precisely, there seems to be a systematic downward bias at these allocations. Finally, the skewness coefficients are higher in some allocations than they were in the monetary treatment, suggesting that the values come from more skewed distributions than in the monetary treatment.[20]

Result 5: In the NM-treatment, the average A-subject accurately anticipates the strength of the average actual sanction at most allocations (in 7 out of 8 cases). Although subjects tend to commit errors of a significant size, they are not systematically biased, except in 4 allocations. Subjects tend to underestimate the average strength of the actual sanction in these 4 allocations, where incidentally rewards are frequent.

We finish with a comparison across treatments. Do A-players commit larger mistakes in predicting a particular type of sanctions? We observe that the error is significantly smaller at any allocation of the NM-treatment than in the corresponding allocation of the M-treatment ($p < 0.03$ always; one-sided Mann-Whitney test), with just two exceptions. At allocation 2R and 3R, the average errors are smaller in the M-treatment (in both cases, $p < 0.02$). It seems, therefore, that the A-players have substantially more problems to foresee monetary sanctions.

4.3 Determinants of the prediction errors

To investigate the causes of prediction errors we use regressions as in section 3.3. The dependent variable is the prediction error, whereas socio-demographic variables and several other variables related to reciprocity and payoff differences are used as independent variables.

To motivate some of these variables, recall from section 4.1 that both harm-help (reciprocity) and strict equality seem crucial to explain the actual occurrence and strength of sanctions, whereas payoff differences play a relatively minor role. With this in mind, we conjecture that one possible reason why individuals make errors is because they are unable to anticipate these patterns. As in section 3.3, therefore, we again consider the same variables. Given all this, Table 8 presents the results of two OLS regressions. Regression (1) considers all variables, while regression (2) does not consider the variables which are never significant at the 10 percent. We focus in what follows on regression (2). In this respect, we observe that the payoff difference between the players is very significantly correlated with the prediction error, as happened in the M-treatment. In contrast to the M-treatment, however, people do not commit significantly larger mistakes when reciprocity predicts rewards, and they do when the allocation is strictly egalitarian. In fact, the absolute error increases in almost 23 points at those allocations where both players get identical payoffs. This suggests that the A-subjects do not anticipate that strict equality has a significant role on the occurrence of non-monetary sanctions. We also find marginally significant evidence that those A-subjects classified as inequity-averse (i.e., those

20 The coefficients from 1L to 4R are the following: -1.14; 1.70; -0.26; 1.93; 0.32; 0.49; -0.359 and 0.92.

choosing allocation 4L) make larger errors; intriguingly, we do not find the same result for the subjects classified as spiteful, contrary to what happened in the M-treatment.[21]

Result 6: The magnitude of the prediction error that an A-subject makes grows significantly higher if (i) the payoff distance with the co-player increases, and (ii) the allocation is strictly egalitarian, and (iii) the subject is among those classified as inequity-averse.

TABLE 8: RESULTS OF REGRESSION ANALYSIS

Dependent Variable: Prediction error

	Paydiff -	Paydiff +	Rec. +	Spiteful type	Pessimism	Female dummy	Political ideology	Religiosity	Strict	IA type	Constant
(1)	0.08^{**} (0.032)	0.087^{***} (0.020)	-2.555 (2.537)	-3.807 (4.999)	-0.644 (1.026)	-3.429 (2.249)	-1.181 (0.895)	0.858^{**} (0.359)	21.96^{***} (3.448)	9.176^{**} (4.530)	18.889^{***} (6.745)
(2)	0.095^{***} (0.027)	0.088^{***} (0.019)	-----	-----	----	----	----	0.650 (0.396)	22.962^{***} (3.170)	6.857^{*} (3.478)	7.769^{**} (3.233)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. N = 320 in regression (1) and 328 in regression (2).[1] In parentheses, standard errors clustered on individual level. The R-squared of regressions (1) and (2) is 0.186 and 0.165, respectively.

5. Conclusion

This paper analyzes whether people accurately anticipate sanctions, an important question in order to explore compliance with social norms. Our results show that the average subject predicts with remarkable accuracy not only the sign but also the strength of the average sanction. This occurs at most allocations, although subjects tend to fail in a systematic manner in those allocations where rewards are frequent. In general, our results both at the aggregate and the individual level indicate that the stick is better anticipated than the carrot. A possible reason, suggested by a regression analysis, is that many subjects fail to anticipate the dependence between rewards and the payoff difference between the players. Finally, we also find that subjects tend to forecast non-monetary sanctions with a smaller error than monetary ones. Since monetary sanctions are arguably not so frequently used out of the lab as non-monetary ones, subjects could be more familiar with the latter.

To our knowledge, the question of whether people anticipate sanctions has not received systematic attention in any previous study in Experimental Economics. Yet our study contributes to a literature on the broader field of beliefs (consult Palfrey and Wang (2009) and Blanco et al. (2010) for surveys of the Experimental Economics literature). A few papers consider games with punishment and/or rewards and elicit beliefs about sanctions. Thus, Costa-Gomes and Zauner (2001) elicit proposers' beliefs in the ultimatum game, and Offerman (2002) elicit beliefs about sanctions in a social dilemma game. They found that [accuracy; in progress].

21 This suggests that the reason why these subjects commit larger mistakes in the M-treatment is not because they choose in a random, careless manner. Otherwise, one would expect a similar result in the NM-treatment.

Our paper is also related to a literature with a long history in the social sciences and which stresses how groups of independently-deciding individuals come up with an accurate *average* prediction. For instance, Galton (1907) found that the crowd at a county fair accurately guessed the weight of an ox when their individual estimates were averaged. Recently, Griffiths and Tenenbaum (2006) show that people are able to predict remarkably well the duration or extent of everyday phenomena (e.g. life span, movie run time). Our paper relates to this literature in its interest (studying predictive abilities), but differs in its subjects since we study how people anticipate sanctions of others (and not objective statistics).

Our study is a first attempt to systematically study players' expectations about monetary and non-monetary sanctions and may encourage more research on this topic. To start, additional studies with different games are required to check that our results extend to other situations. As another example, although our evidence shows that the average subject predicts remarkably well the average sanction in most allocations, we also observe substantial heterogeneity in accuracy (for instance, the subjects classified as spiteful commit substantial errors in the M-treatment). Studying the sources of this heterogeneity seems a promising endeavour, with potential applications to issues like crime. In a study about shoplifting, for instance, Kraut (1976) shows that those who shoplifted generally underestimated both the risk of apprehension and the severity of the sanction. Finally, another possible research question appears when both types of sanctions are available: Do people correctly anticipate when each one is used by the co-players?

References

[1] Andreoni, James; Harbaugh, William and Vesterlund, Lise. "The Carrot or Stick: Reward, Punishment and Cooperation". American Economic Review, 2003, 93(3), pp. 893-902.

[2] Akerlof, George A. "A Theory of Social Custom, of which Unemployment may be One Consequence", Quarterly Journal of Economics, 1980, 94, pp. 749-775.

[3] Axelrod, Robert. "An Evolutionary Approach to Norms", American Political Science Review, 1986, 80, pp. 1095-1111.

[4] Blanco, Mariana; Engelmann, Dirk; Koch, Alexander K.; and Normann, Hans-Theo. "Belief Elicitation in Experiments: Is there a Hedging Problem?", Experimental Economics, forthcoming .

[5] Bolton, Gary E. and Ockenfels, Axel. "ERC: A Theory of Equity, Reciprocity, and Competition", American Economic Review, 2000, 90(1), pp. 166-93.

[6] Brandts, Jordi and Charness, Gary. "The Strategy versus the Direct-response Method: A Survey of Experimental Comparisons", mimeo, 2009.

[7] Camerer, Colin and Thaler, Richard. "Ultimatums, Dictators, and Manners", Journal of Economic Perspectives, 1995, 9, pp. 209-219.

[8] Coleman, James S. "Foundations of Social Theory", 1990, Cambridge, MA: Harvard University Press.

[9] Costa-Gomes, Miguel and Zauner, Klaus. "Ultimatum Bargaining Behavior in Israel, Japan, Slovenia, and the United States: A Social Utility Analysis", Games and Economic Behavior, 2001, 34, pp. 238-269.

[10] Dufwenberg, Martin and Kirchsteiger, Georg. “A Theory of Sequential Reciprocity”, Games and Economic Behavior, 2004, 47, pp. 268-98.

[11] Ellingsen, Tore and Magnus Johannesson. “Anticipated Verbal Feedback Induces Pro-social Behavior”, Evolution and Human Behavior, 2008, 29, pp. 100–105.

[12] Elster, Jon. “Social Norms and Economic Theory”, Journal of Economic Perspectives, 1989, 3(4), pp. 99-117.

[13] Falk, Armin; Fehr, Ernst and Fischbacher, Urs. “Driving Forces behind Informal Sanctions” Econometrica, 2005, 7(6), pp. 2017-30.

[14] Fehr, Ernst and Gächter, Simon. “Cooperation and Punishment in Public Goods Experiments”, American Economic Review, 2000, 90, pp. 980-994.

[15] Fehr, Ernst and Fischbacher, Urs. “Third Party Punishment and Social Norms”, Evolution and Human Behavior, 2004, 25, 63-87.

[16] Fehr, Ernst and Schmidt, Klaus. “A Theory of Fairness, Competition and Cooperation”, Quarterly Journal of Economics, 1999, 114(3), pp. 817-68.

[17] Galton, Francis. “Vox Populi”, Nature, 1907, 75, pp. 450-451.

[18] Gächter, Simon and Renner, Elke. “The Effects of (Incentivized) Belief Elicitation in Public Good Experiments”, Experimental Economics, 2010, 13, pp. 364–77.

[19] Griffiths, Thomas L. and Tenenbaum, Joshua B. “Optimal Predictions in Everyday Cognition”, Psychological Science, 2006, 17, pp. 767-773.

[20] Güth, Werner, Schmittberger, Rolf and Schwarze, Bernd. “An Experimental Analysis of Ultimatum Bargaining”, Journal of Economic Behavior and Organization, 1982, 3, pp.367-388.

[21] Holländer, Heinz. “A Social Exchange Approach to Voluntary Cooperation”, American Economic Review, 1990, 80, pp. 1157-1167.

[22] Homans, George. “Social Behavior”, 1961, New York: Harcourt, Brace & World.

[23] Kandel, Eugene and Lazear, Edward. “Peer Pressure and Partnerships”, Journal of Political Economy, 1992, 100, pp. 801–817.

[24] Kraut, Robert E. “Deterrent and Definitional Influences on Shoplifting”, Social Problems, 1976, 23(3), pp. 358-369.

[25] López-Pérez, Raúl and Vorsatz, Marc. “On Approval and Disapproval: Theory and Experiments”, Journal of Economic Psychology, 2010, 31, pp. 527-541

[26] Masclet, David; Noussair, Charles; Tucker, Steven and Villeval, Marie-Claire. “Monetary and Non-monetary Punishment in the Voluntary Contributions Mechanism”, American Economic Review, 2003, 93, pp. 366–380.

[27] Noussair, Charles and Tucker, Steven. “Combining Monetary and Social Sanctions to Promote Cooperation”, Economic Inquiry, 2005, 43, pp. 649–660.

[28] Offerman, Theo. “Hurting Hurts more than Helping Helps”, European Economic Review, 2002, 46, pp. 1423-1437.

[29] Ostrom, Elinor; Walker, James and Gardner, Roy. “Covenants with and without a Sword: Self-Governance is Possible”, American Political Science Review, 1992, 86(2), pp. 404-417.

[30] Rabin, Matthew. “Incorporating Fairness into Game Theory and Economics”, American Economic Review, December 1993, 83(5), pp. 1281-1302.

[31] Rege, Mari and Telle, Kjetil. “The Impact of Social Approval and Framing on Cooperation in Public Good Situations”, Journal of Public Economics, 2004, 88, pp. 1625–1644.

[32] Roth, Alvin E. "Bargaining Experiments", in J. Kagel and A. Roth (eds.): Handbook of Experimental Economics, 1995, Princeton, Princeton University Press.

[33] Sefton, Martin; Shupp, Robert and Walker, James. "The Effect of Rewards and Sanctions in Provision of Public Goods", Economic Inquiry, 2007, 45(4), pp. 671-690.

[34] Vyrastekova Jana and van Soest, Dan. "On the (in)effectiveness of Rewards in Sustaining Cooperation", Experimental Economics, 2008, 11, pp. 53–65.

[35] Xiao, Erte and Houser, Daniel. "Avoiding the Sharp Tongue: Anticipated Written Messages Promote Fair Economic Exchange", Journal of Economic Psychology, 2009, 30(3), pp. 393-404.

Appendix A: Choices of the A-players

TABLE 1A—A-players' Choices (M-treatment)

Game	Allocation			percentage	
	Left		Right	Left	Right
1	(250,100)	vs.	(200,150)	60.9%	39.1%
2	(250,100)	vs.	(250,250)	10.9%	89.1%
3	(100,200)	vs.	(150,150)	30.4%	69.6%
4	(100,200)	vs.	(100,300)	26.1%	73.9%

TABLE 2A—A-players' Choices (NM-treatment)

Game	Allocation			percentage	
	Left		Right	Left	Right
1	(250,100)	vs.	(200,150)	78.6%	21.4%
2	(250,100)	vs.	(250,250)	9.5%	90.5%
3	(100,200)	vs.	(150,150)	4.8%	95.2%
4	(100,200)	vs.	(100,300)	23.8%	76.2%

Appendix B: Instructions for Participant A (Monetary Treatment)

Welcome to this experiment on decision making. At the end of the experiment, you will be paid some money; the precise amount will depend on your decisions and the decisions of another participant. During the experiment we always speak of points; note that

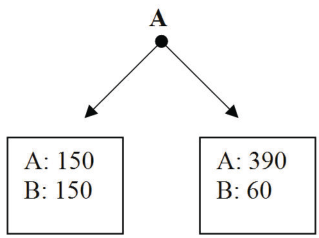
10 points = 1 Euro

Please, do not talk to any other participant during the experiment. If you do not follow this rule we will have to exclude you from the experiment and you will not earn any money. If you have questions, please raise your hand and we will attend you.

There are two types of participants in this experiment: A and B. There is the same number of participants of each type. Previously, the instructor has distributed in a random manner the same number of instructions for each type across the room. Given your seat choice, **you are a type A participant. Further, you will be anonymously matched with a type B participant** (in what follows, we call him/her B)**.** You will never know the type of any other participant, nor will any other participant get to know your type. The decisions in this experiment are anonymous. This means no participant will ever know which participant made which choice.

Description of the Experiment

You, as player A, and B will take decisions in four scenarios, all of them with a two-stage structure. In the first stage of each scenario, you have to decide between two allocations of points for you and B. In the hypothetical example of the figure, the left-hand allocation gives 150 points to you and 150 points to B. The right-hand allocation gives 390 points to you and 60 points to B.



Remember: 10 points = 1 Euro.

In the second stage of each scenario, B can affect your balance. For this, B must pay previously 5 points. If B pays the 5 points, B can then assign to you any amount of points between -100 and +100. This amount will decrease or increase your balance by the same amount. If B chooses not to pay the 5 points, she cannot assign any points to you so that the allocation chosen by you is implemented.

Example 1: Suppose that you choose the left-hand allocation in the previously illustrated scenario and that B then decides to spend the 5 points and assigns to you +60 points. Then you would have a balance of 150 + 60 = 210, and B would get 150 – 5 = 145 points.

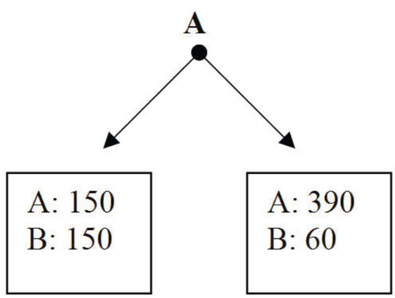
Example 2: Suppose that you choose the right-hand allocation in the previously illustrated scenario and that B then decides to spend the 5 points and assigns to you -30 points. Then you would have a balance of 390-30 = 360, and B would have 60-5 = 55 points.

Important: When deciding, B will not know the allocation actually chosen by you in any scenario. For this reason, B will indicate her decision for any possible choice by you at any scenario. Following with the example of the figure, B should answer four questions: (1) Would you pay the 5 points if A had chosen (150, 150)?, (2) in case you pay the 5 points, what amount of points (between -100 and +100) would you assign then to A?, (3) and (4) the same questions if A had chosen (390, 60).

After all participants have taken their decisions in the four scenarios and answered a brief questionnaire, the instructor will collect your form. Afterwards, one scenario will be chosen randomly (with the roll of a die). This is important because any participant will be paid only for her/his final point score in that scenario (the instructor will divide that score by 10). To finish, note that you will be paid in private and that we will inform you in that moment about B's choice in the payment-relevant game (without, of course, revealing B's identity).

Before we proceed with the experiment, please answer the following control questions. Raise your hand after that so that we can verify that the answers are correct.

In the hypothetical example of the figure, assume the following: (a) B decides to pay the 5 points if A had chosen allocation (A: 150, B: 150), and assigns then +100 points to A, (b) B decides not to pay the 5 points if A had chosen allocation (A: 390, B: 60).



Taking into account all this, answer the following questions,

- What would be the final point score of A if she/he chooses (A: 150, B: 150)?

  ____

- What would be the final point score of B if A chooses (A: 150, B: 150)? ___________
- What would be the final point score of A if she/he chooses (A: 390, B: 60)? ________
- What would be the final point score of B if A chooses (A: 390, B: 60)? ____________

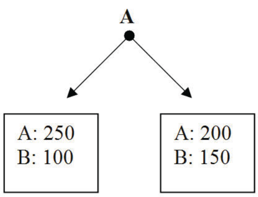
In addition:

- Will you know any of the decisions taken by B before you have made your decision in all four scenarios? Yes No
- Will B know any of your effective decisions before B has made her/his decision in all four scenarios? Yes No
- How many scenarios has this experiment? __________
- How many scenarios will be relevant for your payment? __________
- Can B ever affect your balance without spending 5 points? Yes No

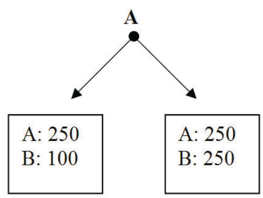
Decisions of a type-A participant

The 4 scenarios

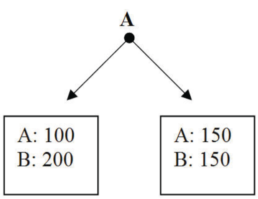
For your information, we present here the point allocations available in each of the 4 scenarios. In the next sheets, you can take your decisions in each scenario.



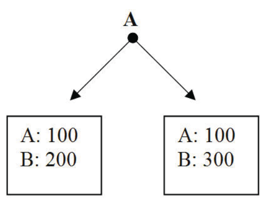
Escenario 1



Scenario 2



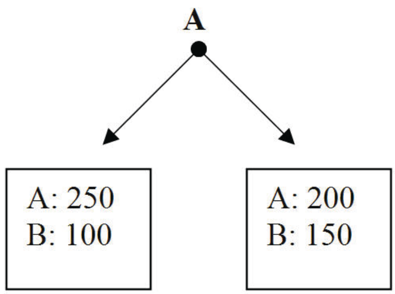
Scenario 3



Scenario 4

Note: In the next sheets, you can take your decisions in any order as you wish (that is, you do not need to start deciding in scenario 1). Until we collect your decision form, moreover, you can always change your decision in any scenario if you decide so (to facilitate this, you can initially use a pencil; write down your final decision with a pen, though).

Scenario 1



Recall: 10 points = 1 Euro

The point allocation that I choose in this scenario is (select it with a circle):

- A: 250, B: 100
- A: 200, B: 150
-

Independently of your previous choice, we kindly ask you to make a series of estimations (your answers here will not affect your final payoff):

- What is the percentage of participants B that will pay the 5 points if A chooses (250, 100)? ________ (this must be a number between 0 and 100, both included)
- In the previous case, how many points (in average) will these B-participants assign to the A-participant? ________ (this must be a number between -100 and 100, both included)
- What is the percentage of participants B that will pay the 5 points if A chooses (200, 150)? ________ (this must be a number between 0 and 100, both included)
- In the previous case, how many points (in average) will these B-participants assign to the A-participant? ________ (this must be a number between -100 and 100, both included)

Appendix C: Instructions for Participant A (Non-Monetary Treatment)

Welcome to this experiment on decision making. At the end of the experiment, you will be paid some money; the precise amount will depend on your decisions and the decisions of another participant. During the experiment we always speak of points; note that

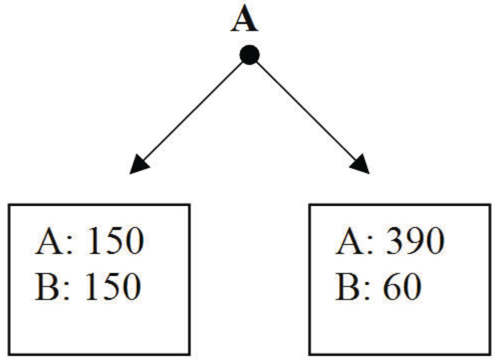
10 points = 1 Euro

Please, do not talk to any other participant during the experiment. If you do not follow this rule we will have to exclude you from the experiment and you will not earn any money. If you have questions, please raise your hand and we will attend you.

There are two types of participants in this experiment: A and B. There is the same number of participants of each type. Previously, the instructor has distributed in a random manner the same number of instructions for each type across the room. Given your seat choice, **you are a type A participant. Further, you will be anonymously matched with a type B participant** (in what follows, we call him/her B)**.** You will never know the type of any other participant, nor will any other participant get to know your type. The decisions in this experiment are anonymous, that is, no participant will ever know which participant made which choice.

Description of the Experiment

You, as player A, and B will take decisions in four scenarios, all of them with a two-stage structure. In the first stage of each scenario, you have to decide between two allocations of points for you and B. In the hypothetical example of the figure, the left-hand allocation gives 150 points to you and 150 points to B. The right-hand allocation gives 390 points to you and 60 points to B.



Remember: 10 points = 1 Euro.

In the second stage of each scenario, B cannot affect your balance, but can approve or disapprove your prior choice. For this, B must pay 5 points. If B pays the 5 points, B can then assign an evaluation score between -100 and +100 to you. A negative score indicates that B disapproves your choice (-100 is maximum disapproval), while a positive score indicates that B approves your choice (+100 is maximum approval). We note again that, whatever its sign, this score will not affect your balance. If B chooses not to pay the 5 points, B cannot assign a score to you

Example 1: Suppose that you choose the left-hand allocation in the previously illustrated scenario and that B then decides to spend the 5 points and assign a score of +60 to you. That means that B approve your choice with intensity equal to 60 out of 100. Note also that your balance is unchanged (you get 150 points), whereas B would get 150 – 5 = 145 points.

Example 2: Suppose that you choose the right-hand allocation in the previously illustrated scenario and that B then decides to spend the 5 points and assign a score of -30 to you. That means that B disapproves your choice with intensity equal to 30 out of 100. Note also that your balance is unchanged (you get 390 points), whereas B would get 60 – 5 = 55 points.

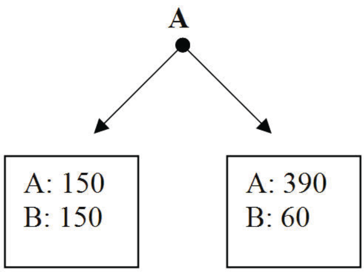
Important: When deciding, B will not know the allocation actually chosen by you in any scenario. For this reason, B will indicate her decision for any possible choice by you at any scenario. Following with the example of the figure, B should answer four questions: (1) Would you pay the 5 points if A had chosen (150, 150)?, (2) in case you pay the 5 points, what score (between -100 and +100) would you assign then to A?, (3) and (4) the same questions if A had chosen (390, 60).

After all participants have taken their decisions in the four scenarios and answered a brief questionnaire, the instructor will collect your form. Afterwards, one scenario will be chosen randomly (with the roll of a die). This is important because any participant will be paid only for her/his final point score in that scenario (the instructor will divide that score by 10). To finish, note that everyone will be paid in private and that we will inform you in that moment about the evaluation score that B assigned to you in the payment-relevant game (without, of course, revealing B's identity).

Before we proceed with the experiment, please answer the following control questions. Raise your hand after that so that we can verify that the answers are correct.

In the hypothetical example of the figure, suppose that A chooses allocation (A: 150, B: 150) and that B decides to pay the 5 points, and assigns then a score +100 to A. In this case:

- What would be A's final balance? _______
- Does B approve or disapprove A's choice? ___________



Suppose now that A chooses allocation (A: 390, B: 60) and that B decides not to pay the 5 points.

- What would be A's final balance then? ________
- What would be B's final balance then? ____________

In addition:

- Will you know any of the decisions taken by B before you have made your decision in all four scenarios? Yes No
- Will B know any of your decisions before she/he has made her/his decision in all four scenarios? Yes No
- How many scenarios has this experiment? __________ How many scenarios will be relevant for your payment? __________
- Can B ever affect your balance? Yes No

(Footnotes)

1 There were two subjects who did not answer the question about their political ideology. One of the two did not reveal the degree of religiosity. Therefore, in regression (1) instead of the possible 336 observations we have only 320, while in regression (2) we have 328.