Third-party Punishment is more effective on Women: Experimental Evidence

Pablo Brañas-Garza – GLOBE: University of Granada – Spain

Stefania Ottone – University of Milano-Bicocca – Italy

Abstract. Existing experimental studies mainly focus on motivations and choices of third-party punishers, but only few of them detect sanction efficacy contradictory results. Our paper wants to shed light on this point. In particular, we want to detect whether the threat of being punished for unfair actions is credible and affects subjects' choices thus, making it rational to behave fairly. To disentangle the effect of expected punishment on behaviour, we implement in the lab two experimental games - the standard Dictator Game, that is used as baseline, and the Third-Party Punishment Game that incorporates a third player who observes and may punish the Dictator. The idea is that, if the Dictator in treatment TPP believes punishment is a credible threat, s/he may decide to change her/his behaviour, that is, to behave generously in order to avoid sanctions. We find a clear gender bias: women reacted to the punishment threat by increasing their transfer to the Recipient, while men did exactly the opposite.

1.0 **Introduction**

The relevant and peculiar role played by social norms in human societies is well-known and generally accepted. Consequently, it is not surprising that several interdisciplinary studies have been devoted to social norms and to their endogenous enforcement mechanisms. Special attention has been given to punishment, given that norms compliance is mainly due to the expectations that violations will be sanctioned. In particular, third-party punishment[1] is considered the essence of social norms. Its evolutionary dominance with respect to second-party punishment – especially in large-scale societies – is to be ascribed to the fact that, as Fehr and Fischbacher (2004) correctly observe, "*if only second parties imposed sanctions, a very limited number of social norms could be enforced because norm violations often do not directly hurt other people*" (p. 64). Even if its costly nature makes third-party punishment a "*tendency that would be selected against*" (Kurzban et al. 2007 p.75), group selection as well as reputation models seem to justify its existence (see for example Barclay, 2006; Boyd et al. 2003; Gintis, 2000).

Though empirical evidence supports the existence of third-party punishment, experimental economics makes a meaningful contribution to understanding how social norms emerge and are endogenously enforced. The reason is that, unlike empirical evidence, appropriately designed experiments have been the key to rigorously disentangle the different forces that drive norm compliance. In particular, it turns out that:

a) third-party punishment exists and its strength is proportional to deviance to the norm (see for example Fehr and Fischbacher, 2004; Ottone, 2008);

b) in-group members are more severely punished than out-group subjects when deviating (see for example Shinada et al., 2004; Bernard et al., 2006);

[1] Two kind of punishment exists – second-party and third-party. Second-party punishment implies that the victim of an unfair action can react and punish the oppressor. Third-party punishment comes from a subject who is not directly hurt by the oppressor.

c) when third parties have the opportunity not only to punish a deviating subject but also to help the victims of injustice, they do that (Ottone, 2008).

However, what emerges is that existing experimental studies mainly focus on motivations and choices of third-party punishers, but only few of them detect sanction efficacy obtaining contradictory results (see Charness et al., 2007; Fehr and Fischbacher, 2004). Our paper wants to shed light on this point. In particular, we want to detect whether the threat of being punished for unfair actions is credible and affects subjects' choices thus, making it rational to behave fairly.

To disentangle the effect of expected punishment on behaviour, we implement in the lab two experimental games - the standard Dictator Game, that is used as baseline, and the Third-Party Punishment Game that incorporates a third player who observes and may punish the Dictator (see Fehr and Fischbacher, 2004). The idea is that, if the Dictator in the TPP treatment believes punishment is a credible threat, s/he may decide to change her/his behaviour, that is, to behave generously in order to avoid sanctions.[2]

We find a clear gender bias: women reacted to the punishment threat by increasing their transfer to the Recipient, while men did exactly the opposite.

This gender effect is curious but not surprising. Delinquency is more widespread among men and a large number of scholars have tried to explain this factor by exploring gender issues within deterrence. Emerging results are not univocal, with some of these studies emphasizing that women are more likely to perceive higher punishment threats than men (see Carmichael, 2004 for a brief survey). This result conforms also to the evidence that women are more risk averse (see e.g. Hartog et al., 2002; Agnew et al., 2008, Eckel and Grossman, 2008) and more sensible to threat of shame and embarrassment (Blackwell, 2000).

After the introduction, the paper is organized as follows. Section 2 describes the design and the procedures. Section 3 is devoted to results and the fourth concludes.

[2] In this experiment we refer to monetary sanctions only.

2.0 **Experimental design and procedures**

The experimental design consists of two treatments: the Dictator Game (DG) and the Third-Party Punishment Game (TPP).

In the DG, groups of 2 participants were formed (see figure 1) and subjects played a Dictator game: player A had the chance of passing money to player B who could not react to any decision of her/his partner. In the TPP, groups of 3 participants were formed (see figure 1) and the Third-Party Punishment Game was implemented. In the first stage player A and player B played a Dictator Game, while in the second stage player C entered the game and had to decide whether to bear a cost in order to sanction A or to keep the whole initial endowment.[3][4] The strategy method at the Observer's stage was implemented.[5]

In each treatment, subjects were asked to play the game once. In other words, it was a one-shot game. At the beginning of the experiment, participants were informed about the sequential nature of the game protocol. The instructions were read by participants on their computer screen while an experimenter read them loudly. After reading the instructions and before subjects were invited to take decisions, some control questions were asked in order to be sure that players understood the rules of the game. Then, in each treatment, each subject was randomly assigned a role – player A-Dictator or player B-Recipient in the DG and player A-Dictator, player B-Recipient or player C-Observer in the TPP. At the end of each session, subjects were asked to fill in a brief survey to collect their socio-demographic data. Each subject participated in one session only. In order to minimize unobservable variables affecting our results (beauty, gender, empathy, etc.) or post-session effects, the complete experiment

[3] Instructions do not contain terms such as punish or sanction. We use more neutral term such us deduction.
[4] In each treatment, A's and C's initial endowments are the same (20 tokens), while B's initial endowment is 10 tokens. Participant C has to pay 1 token to reduce participant A's payoff of 2 tokens. Each token value is 0.50 Euro.
[5] When the strategy method is used, subjects are asked their choices in each possible scenario. In this case, the Observers are asked to declare their level of punishment for each possible transfer from the Dictator to the Recipient. The final payoff is determined on the basis of the situation that actually occurs.

preserved anonymity among participants.[6] Subjects received their payments after the experiment and they were paid by an assistant. Each session lasted for about 20 minutes for the DG and 40 minutes for the TPP. Each subject earned on average 7.5 Euros.

The experiment was run in the Experimental Economics Laboratory (EELAB), at the University of Milano-Bicocca in Milan, Italy. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). Overall, 3 sessions for each treatment were run, with a total of 175 participants (70 participants in the DG and 105 in the TPP), recruited through a web-based recruitment system. In the DG, 66% of participants were women, while in the TPP the female representation was 46% of the sample.

3.0 **Results**

This paper explores whether the possibility of being punished by a third party is a credible and effective threat for our Dictators. Nash equilibrium predicts that player C will never punish player A and, consequently, player A does not anticipate any threat. Hence, player A will transfer nothing to player B.

However, the results showed that the Observers deviate from the Nash equilibrium by punishing selfish Dictators. Given that the actions by players A are observed by players C, the question is: may the Dictators anticipate sanctions? If it is the case, they may deviate from the predicted behavior and consequently behaving generously. Thus, we checked if there exists any difference in Dictators' behavior under the two treatments.

Table 1 (columns 1 and 2) and Figure 2 show the results of both treatments. At a first sight, there seems to be no significant differences. The average transfer from the Dictator to the Recipient is very similar (0.94 in the DG and 1.17 in the TPP), the number of subjects playing the Nash equilibrium is fairly the same (around 54% vs. 46%) and even more identical is the percentage of dictators using the equal split (2.9% in both cases). Both the

[6] In Dufwenberg and Muren (2006) dictators do contributions in front of the audience and subject contribute less.

Mann-Whitney (Z = -0.803; p = 0.422) and the Kolmogorov-Smirnov (Z = 0.0857; p = 0.979) tests do not reject the null hypothesis that the distributions of Dictators' giving in DG and TPP are drawn from the same population.

However, when we detect whether there exists any gender effect, we find out this is the case. We examine both within-treatment and between-treatment differences due to gender.

First of all, women-Dictators significantly increase their transfer to the Recipients when a third subject has the possibility to punish them (average transfer: 0.66 in the DG, 1.63 in the TPP; Mann-Whitney test, Z = -2.044, p = 0.04). On the other hand, men-Dictators' transfer in the TPP is slightly and not significantly lower than in the DG (average transfer: 1.09 in the DG, 0.62 in the TPP; Mann-Whitney test, Z = 1.236, p = 0.21). Then, when we compare each treatment, we find out that in the DG no difference between men and women emerges (Mann-Whitney test, Z = 0.915, p = 0.36), while in the TPP women transfer significantly more than men (Mann-Whitney test, Z = -2.439, p = 0.01). More precisely, we may study how the threat of third-party punishment affects deviations from the Nash equilibrium. For men, the threat is not very effective since 70% chose the Nash equilibrium. In sharp contrast, the finding is exactly the opposite for women: only 26.3% plays according to. See Table 1 and Figure 2 for a more detailed description.

4.0 **Discussion**

In this paper, we want to detect whether third-party punishment is effective. In particular, we want to check whether the possibility for a third party to sanction unfair choices in a Dictator Game leads the Dictator to transfer more to the Recipient in order to avoid punishment. What we found out is that such a deterrence effect emerges among women only.

Our result may have a twofold explanation – a methodological and a physiological one. On the one hand, it may happen that women are more responsive in the lab than in real-life situations. For instance, Levitt and List (2007) find out that women in the lab feel more

obligated to behave socially than males. Brañas et al. (2009) confirm their result and they add a relevant issue: the sample that is involved in lab experiments belongs to a reduced population with peculiar demographic characteristics. This may have some effects when analysing differences between genders. In order to test these lab effects, it would be interesting to run a field experiment with a wider and more heterogeneous sample, not aware of being observed as participants in a scientific study.

On the other hand, this difference may be intrinsic of human nature. We have a lot of evidence that women act differently from men in many situations and they are expected to do so (see Aguiar et al., 2009), due to their more fair-minded and risk-averse nature.

A possible explanation lies exactly in different attitudes when facing risky situations. If women are more risk averse, they are more likely to transfer money in order to avoid sanctions. Moreover, Eckel and Grossman (2006) report that women are more likely to change their behavior according to the situation. It may be the case that, they are more sensible to changes in scenarios, in order to avoid negative consequences of wrong behaviour.

Another possible reason is that implemented punishment is a hard enough threat to women, but not to men. This would explain why boys generally receive more severe punishments (Encyclopedia of women and gender, 2002). In line with this hypothesis, it would be interesting to check whether there is any kind of gender difference when the third party rewards instead of punishing.

Finally, a possibility is that women perceive the punishment as a norm-enforcement device, while men consider it as a mean to increase relative payoffs. In this case, there would be no reason for men to increase their transfer. This hypothesis is in line with a previous study (Hong and Bohnet, 2007) that reports men as status oriented.

References

- Agnew, J.R., Anderson, L.R., Gerlach, J.R., and Szykman, L.R. (2008). Who Chooses Annuities? An Experimental Investigation of the Role of Gender, Framing, and Defaults. *American Economic Review,* 98, 418–442.
- Aguiar F., Brañas, P., Cobo-Reyes, R., Jimenez, N., Miller, L.M. (2009). Are women expected to be more generous? *Experimental Economics.* 12, 93-98.
- Barclay, P. (2006). Reputational benefits for altruistic punishment. *Evolution and Human Behavior*, 26, 325-344.
- Bernhard, H., Fischbacher, U., Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442, 912-915.
- Blackwell, B.S. (2000). Perceived Sanction Threats, Sex, and Crime: A Test and Elaboration of Power-Control Theory. *Criminology* 38, 439-488.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. Proceedings of the National Academy of Sciences USA, 100, 3531–3535.
- Brañas P., Cardenas, J.C., Rossi, M. (2009). Gender, education and reciprocal generosity: Evidence from 1,500 experiment subjects. ECINEQ, working paper 128.
- Carmichael S. (2004). Gender differences and perceived sanction threats: the effect of arrest ratios Master Thesis, University of California.
- Charness, G., Cobo-Reyew, R., Jimenez, N. (2007). An investment game with third-party intervention. Department of Economic Theory and Economics History, University of Granada, ThE Papers, 06/13.
- Dufwenberg, M. and Muren, A. (2006). Generosity, Anonymity, Gender. *Journal of Economic Behavior & Organization.* 61, 42-49
- Eckel, C. C., and Grossman, P. J. (1996). Altruism in Anonymous Dictator Games. *Games and Economic Behavior.* 16, 181-191.
- Encyclopedia Of Women And Gender. Sex Similarities and differences and the impact of society on gender. 2002. Volume 1. Academic Press.
- Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior.* 25, 63-87.
- Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics.* 10, 171-178.

- Gintis H. (2000). Strong Reciprocity and Human Sociality. *Journal of Theoretical Biology*, 206, 169-79.
- Hartog, J., Ferrer-i-Carbonell, A., Jonker, N. (2002). Linking Measured Risk Aversion to Individual Characteristics. *Kyklos*, 55, 3-26.
- Hong, K., and Bohnet, I. (2007). Status and Distrust: The Relevance of Inequality and Betrayal Aversion. *Journal of Economic Psychology,* 28, 197-213.
- Kurzban, R., DeScioli, P., O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28, 75-84.
- Ottone, S. (2008). Are people Samaritans or avengers? *Economics Bulletin*, 3, 1-8.
- Shinada, M., Yamagishi, T., Ohmura, Y. (2004). False friends are worse than bitter enemies: "Altruistic" punishment of in-group member. *Evolution and Human Behavior*, 25, 379-393.

Tables and Figures

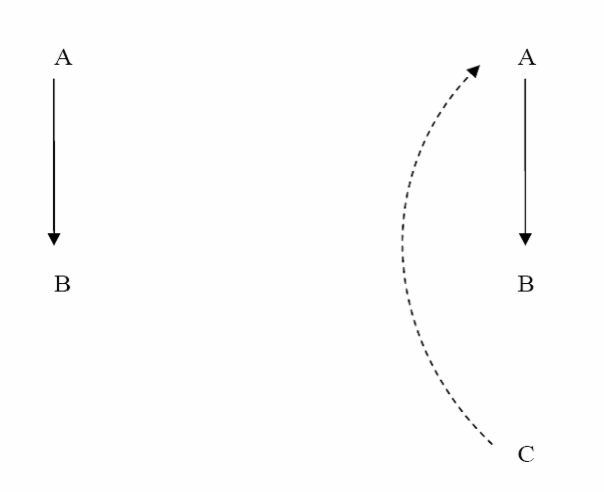
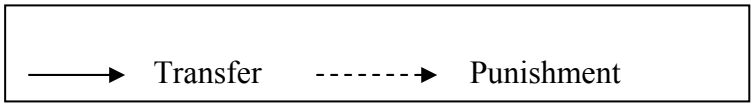


Figure 1: Experimental Games



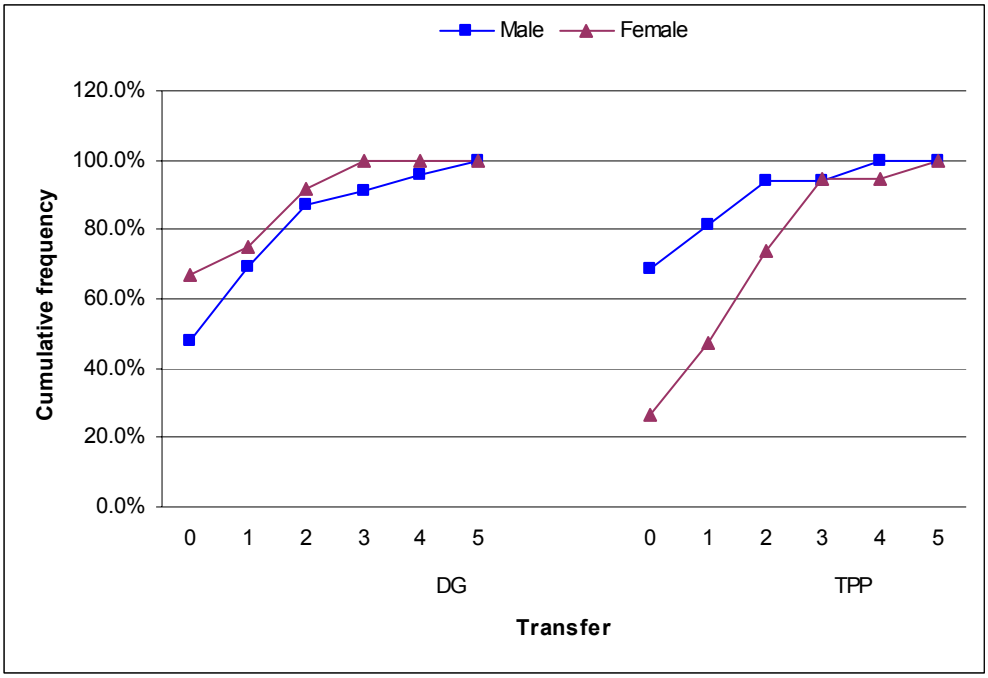


Figure 2: Dictators' transfer by treatment and gender

Table 1: Dictators' Behavior

Donations (in tokens)	DG	TPP	DG		TPP	
			M	F	M	F
0	54.3%	45.7%	47.8%	66.7%	68.8%	26.3%
1	17.1%	17.1%	21.7%	8.3%	12.5%	21.1%
2	17.1%	20.0%	17.4%	16.7%	12.5%	26.3%
3	5.7%	11.4%	4.3%	8.3%	0.0%	21.1%
4	2.9%	2.9%	4.3%	0.0%	6.3%	0.0%
5	2.9%	2.9%	4.3%	0.0%	0.0%	5.3%
N	35	35	23	12	16	19
Mean	0.94	1.17	1.09	0.66	0.62	1.63
Median	0	1	1	0	0	2
Mode	0	0	0	0	0	0;2

Average individual transfer from the Dictator to the Receiver			
	M	**F**	**Mann-Whitney test**
DG	1.09	0.66	$p = 0.36$
TPPG	0.62	1.63	$\mathbf{p = 0.01}$
Mann-Whitney test	$p = 0.21$	$\mathbf{p = 0.04}$	

Instructions

3.1. Instructions for the Dictator Game Treatment

1.A&B. Welcome to the experiment, and thank you for participating. Please follow the instructions that will appear on your screen. There is nothing complicated, nor tricky questions. Your answers will be absolutely anonymous. It will not be possible for the experimenters to match the answers with the person who provided them. For the success of the experiment, it is necessary that you do not communicate with each other.

2.A&B. The experiment involves two different kinds of participants – Participant A and Participant B. At the beginning of the experiment you will be randomly assigned a role (A or B) and you will be randomly paired with another participant. You will never know the identity of your partner.
Participant A has an initial endowment of 20 tokens. She has to decide the amount of tokens – between 0 and 5 – she wants to transfer to participant B. Participant B has an initial endowment of 10 tokens and she cannot take any decision.
Each token corresponds to 0.5 eurocents.

3.A&B. Payoffs are computed in the following way:
- payoff of Participant A is the initial endowment of 20 tokens minus the number of tokens she transfers to Participant B;
- payoff of Participant B is the initial endowment of 10 tokens plus the number of tokens she receives from Participant A.

4.A&B. In particular, it should be clear that each token Participant A transfers to Participant B increases the payoff of Participant B by 1 token – that is, 50 eurocents.

5.A&B. Control questions – see Appendix Figure A.1

6.A. Decision screen for Participant A – see Figure A.2
6.B. Waiting screen for Participant B – Participant A is making her choice. Please, wait.

7.A&B. Generally, how many tokens should Participant A send to Participant B?

8.A&B. Payoff screen

9.A&B. Now we ask you to fill in a brief questionnaire. After that the experiment will terminate and you will receive your payment.

3.2. Instructions for the Third-Party Punishment Game Treatment

1.A&B&C. Welcome to the experiment, and thank you for participating. Please follow the instructions that will appear on your screen. There is nothing complicated, nor tricky questions. Your answers will be absolutely anonymous. It will not be possible for the experimenters to match the answers with the person who provided them. For the success of the experiment, it is necessary that you do not communicate with each other.

2.A&B&C. The experiment involves three different kinds of participants – Participant A, Participant B and Participant C – and it has a two-stage structure. At the beginning of the first stage you will be randomly assigned a role (A, B or C) and you will be randomly paired with two participants. You will never know the identity of your partners.

In the FIRST STAGE Participant A has an initial endowment of 20 tokens. She has to decide the amount of tokens – between 0 and 5 – she wants to transfer to participant B. Participant B has an initial endowment of 10 tokens and she cannot take any decision.
In the SECOND STAGE Participant C has an initial endowment of 20 tokens. She can either reduce the payoff of Participant A at a cost – she has to spend 1 token to reduce the payoff of Participant A by 2 tokens – or keep the whole endowment.
To sum up, Participant C can allocate her endowment as she wants by taking into account that the sum of the tokens she spends to reduce the payoff of Participant A and she keeps for herself cannot be more than 20 tokens.
Participant C is asked to declare her choice for each possible transfer from Participant A to Participant B. The final payoffs are computed on the basis of the actual transfer from Participant A to Participant B. Each token corresponds to 0.5 eurocents.

3.A&B&C. Check whether it is clear that:
- if Participant A transfers 2 tokens to Participant B and Participant C spends 2 tokens to reduce the payoff of Participant A, the final payoff will be 14 tokens for Participant A, 12 tokens for Participant B and 18 tokens for Participant C;
- if Participant A transfers 2 tokens to Participant B and Participant C keeps her whole endowment, the final payoff will be 18 tokens for Participant A, 12 tokens for Participant B and 20 tokens for Participant C;

4.A&B&C. In particular, it should be clear that:
- each token Participant A transfers to Participant B increases the payoff of Participant B by 1 token – that is, 50 eurocents;
- each token – that is, for each 50 eurocents – Participant C spends reduces the payoff of Participant A by 2 tokens – that is, 1 euro.

5.A&B&C. Control questions – see Figure A.3

6.A. Decision screen for Participant A – see Figure A.4
6.B. Waiting screen for Participant B – Participant A is making her choice. Please, wait.
6.C. Decision screen for Participant C – see Figure A.5

7.A&B&C. Generally, how many tokens should Participant A send to Participant B?

8.A&B&C. Payoff screen

9.A&B&C. Now we ask you to fill in a brief questionnaire. After that the experiment will terminate and you will receive your payment.

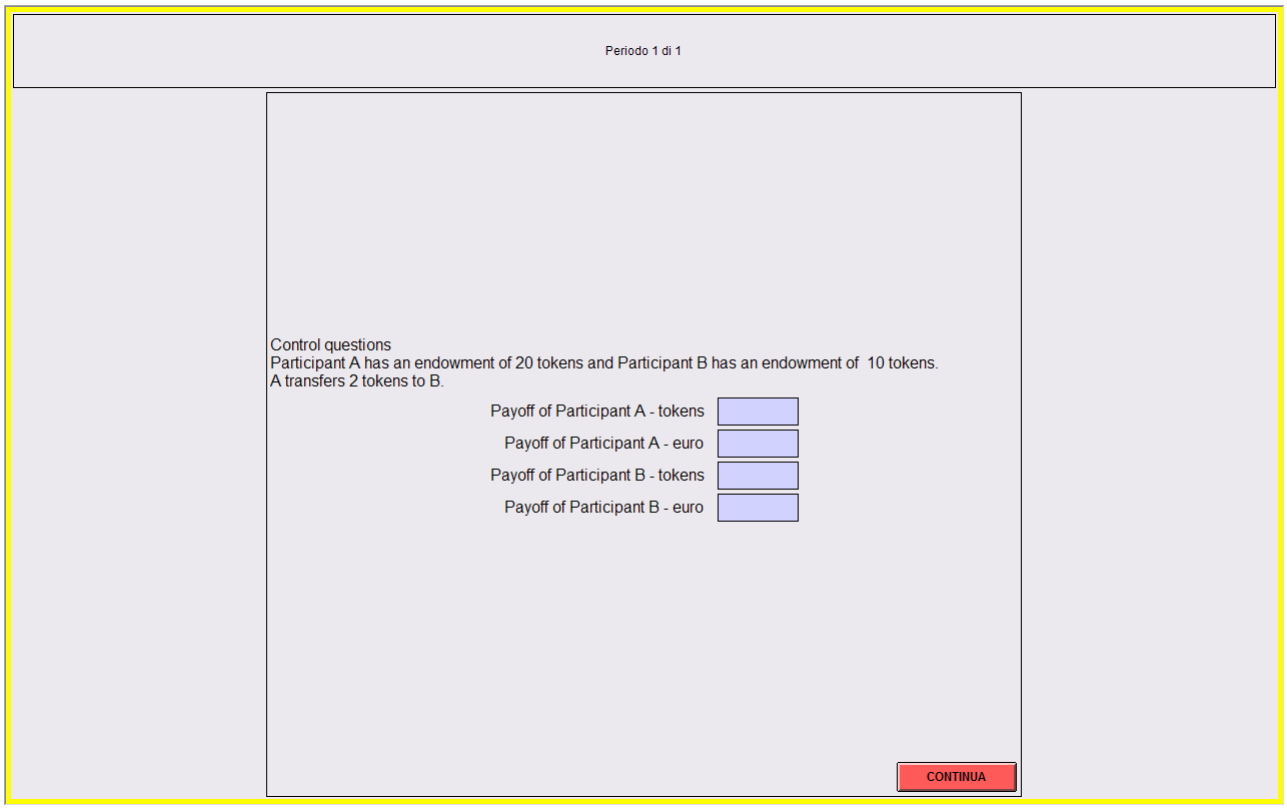


Figure A.1 – Control questions in the Dictator Game Treatment

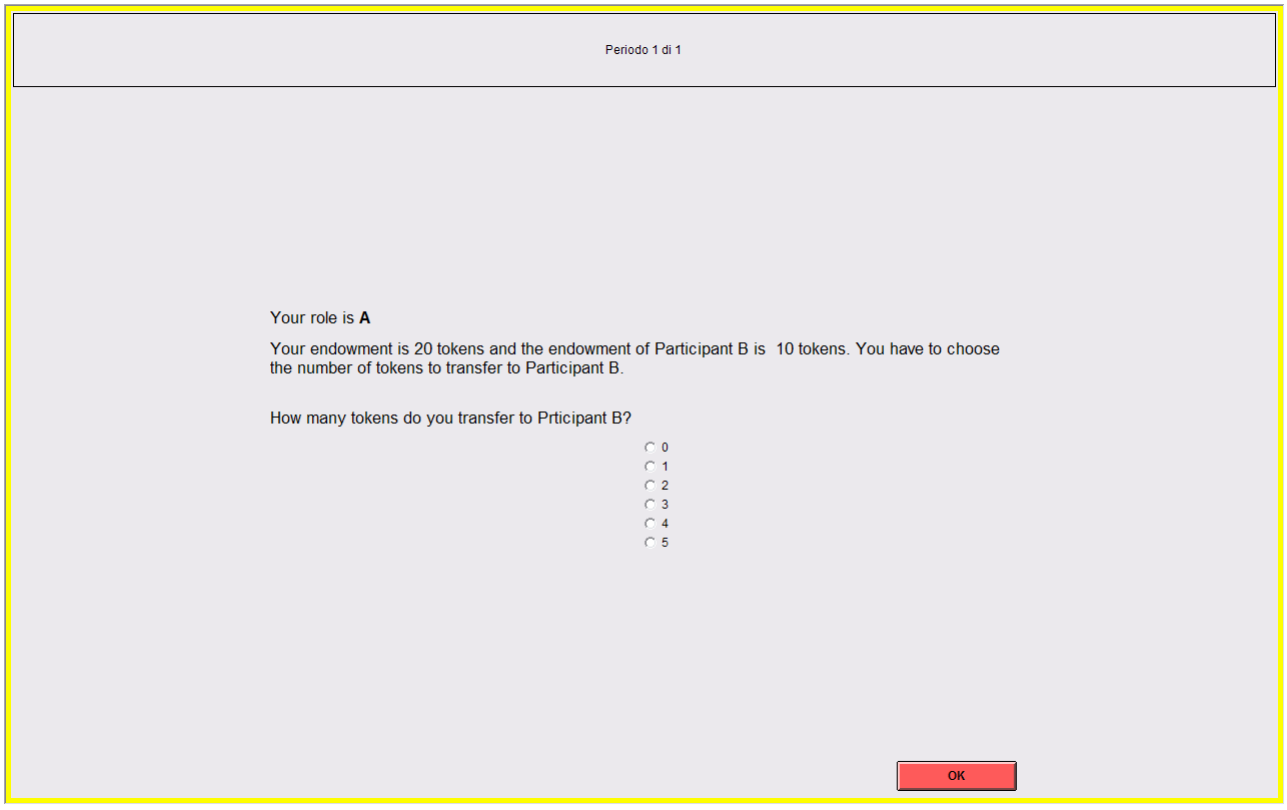


Figure A.2 – Decision screen of Participant A in the Dictator Game Treatment

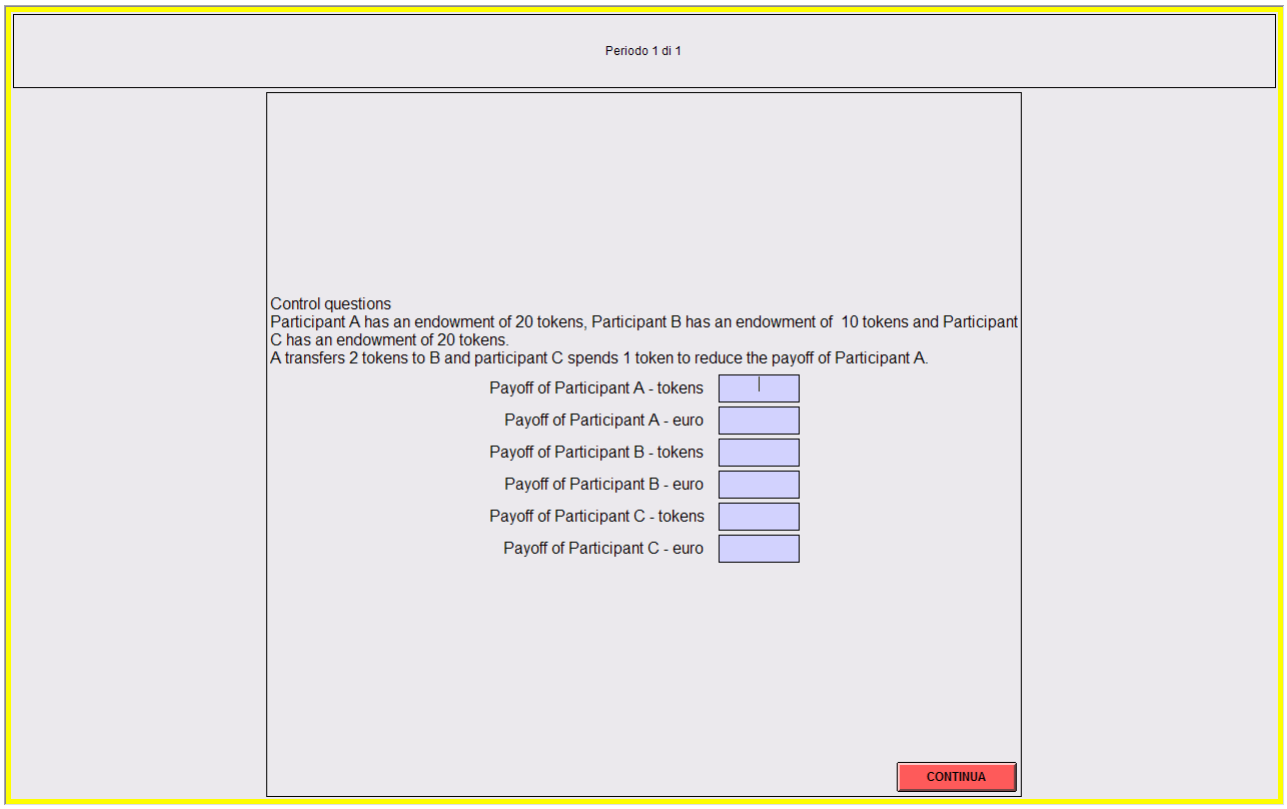


Figure A.3 – Control questions in the Third-Party Punishment Game Treatment

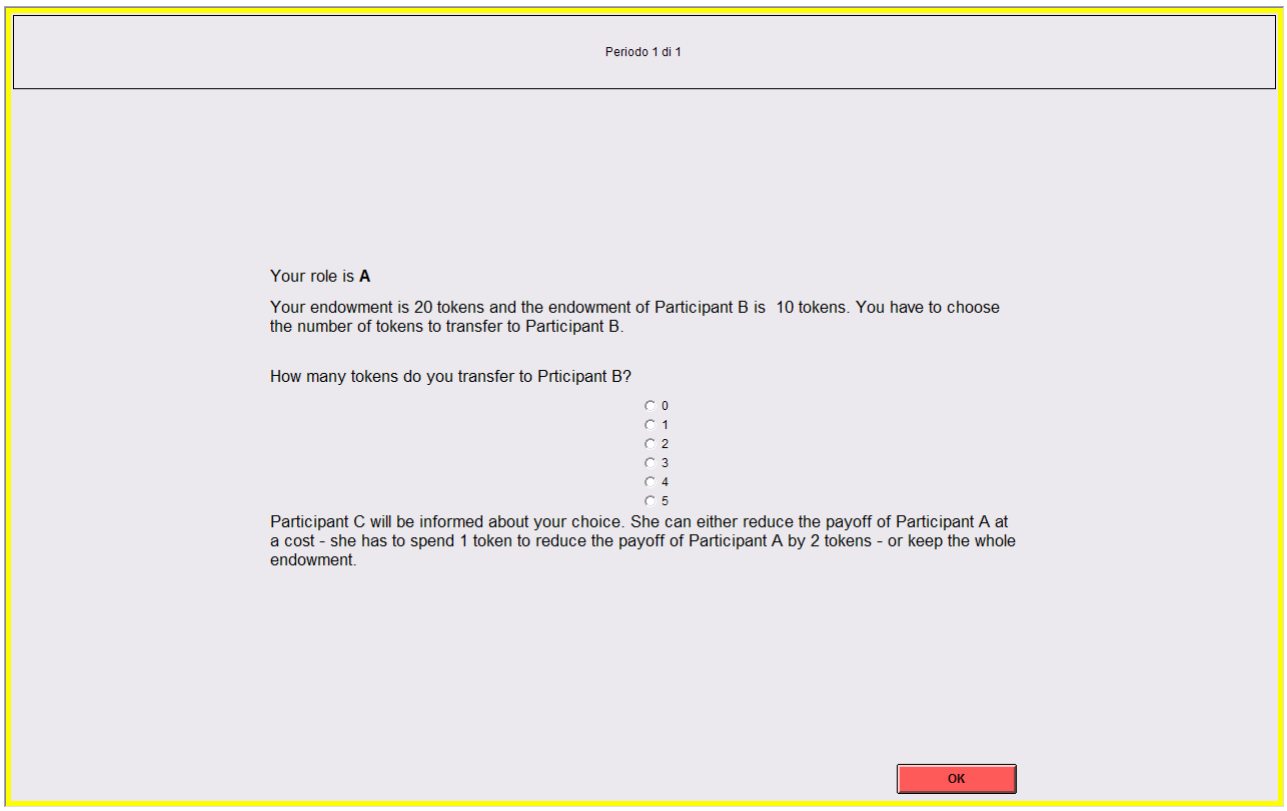


Figure A.4 – Decision screen of Participant A in the Third-Party Punishment Game Treatment

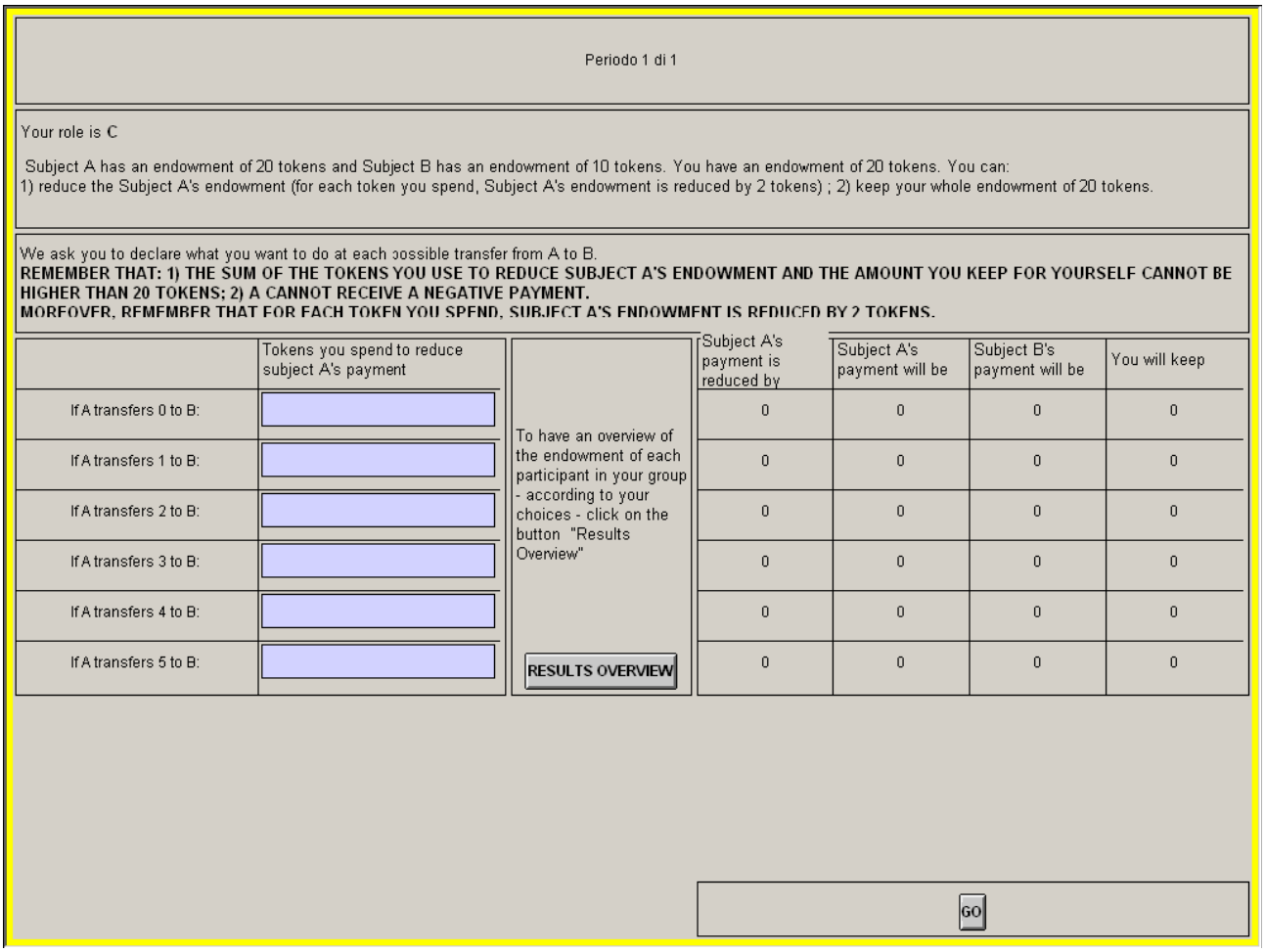


Figure A.5 – Decision screen of Participant C in the Third-Party Punishment Game Treatment