SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Scienze Umanistiche “Riccardo Massa”
Dottorato di Ricerca in / PhD program Scienze dell’Educazione e della Comunicazione
Ciclo / Cycle XXVIII

Essays on Cooperation: scales of interactions, competition, punishment

Cognome / Surname Batistoni Nome / Name Tommaso

Matricola / Registration number 774820

Tutore / Tutor: Prof. Francesco Paoletti

Cotutore / Co-tutor:
(se presente / if there is one)

Supervisor:
(se presente / if there is one)

Coordinatore / Coordinator: Prof.ssa Laura Formenti

ANNO ACCADEMICO / ACADEMIC YEAR 2017 – 2018

to r,

whom copyrights belong to

Introduction

The work presented in Chapter 1 is a behavioural experiment studying the effect of partner-choice on punishing and helping decisions, under the assumption that they can serve as signal of trustworthiness. The work was conducted jointly with Nichola Raihani and Pat Barclay. Raihani, Barclay and I planned the study; I coded and performed the experiment, as well as collected and analysed the data; I also wrote the version of the paper presented in this dissertation.

The work presented in Chapter 2 is the design of a study aiming to investigate how interactions at different scales can undermine the effectiveness of a well-known mechanism to sustain cooperation in large communities, namely indirect reciprocity. It was conducted jointly with Nichola Raihani and Michael Muthukrishna. Raihani, Muthukrishna and I elaborated the study design; I wrote the version of the paper presented in this dissertation and I coded the experimental setup as reported here.

CHAPTER 1

The Reputation of Punishers: Does Partner Choice Escalate Third-Party Punishment?

Abstract

We investigate the hypothesis that partner choice, by creating competition for being chosen by the best partners, can amplify the function of third-party punishment as signal of trustworthiness. We also consider the case of signalling via generous acts in order to provide a direct test of the relative strength of the two types of signals. Our data show that both punishment and help are reliable signal of trustworthiness and are perceived as such by non-involved bystanders – although the signalling value of punishment is much more uncertain and weaker. We did not find, however, clear support for our main hypotheses: investments in neither of those types of signal escalate in response to partner choice. We discuss the discordance between previous and our results considering some key differences between the respective experimental designs.

Introduction

Peer punishment has been identified as a key factor for maintaining cooperation among non-relatives. Punishment refers to the act of paying a cost to inflict a fee on a social partner and it has, therefore, the immediate consequence to reduce the overall productivity of all the individuals involved (e.g. Dreber et al., 2008; Ohtsuki et al., 2009). Notwithstanding, peer punishment is widely observed in laboratory studies (e.g. Fehr and Gachter, 2002; Fehr and Fischbacher, 2004), it can increase cooperative levels within a group and can eventually result in higher overall payoffs (e.g. Gaechter et al, 2009). Although peer punishment can be group beneficial, however, it always implies an individual cost for the punisher, which ends up to be disadvantaged compared to non-punishers. For peer punishment to come under positive selection, therefore, a mechanism is needed for the individual cost of punishing to be ultimately recouped by the punisher. It has been argued that one option for punitive strategies to become adaptive at the individual-level is by carrying out reputation consequences that increase the punisher's likelihood to have profitable social interactions. Building a reputation as a punisher might result advantageous via two mechanisms: (i) by implementing a threat of punishment, which can deter current social partners or bystanders to adopt exploitative strategies towards punishers (e.g. dos Santos et al., 2011, 2013; Hilbe and Traulsen 2012); (ii) by signalling that punishers adheres to a norm of cooperation, which can give them access to rewards from and profitable interactions with new social partners (e.g. Gintis et al., 2001; Barclay, 2006; Raihani and Bshary, 2015). Here we focus on the latter possibility, which has been elaborated within the framework of costly signalling theory (CST; Zahavi, 1995; Gintis et al., 2001; Lotem et al., 2003).

From the perspective of CST, punitive acts can be conceptualized as type-separating signals that allow the punisher to convey an otherwise unobservable cooperative intent (Przepiorka and Liebe, 2016). Observers can then act contingently to the informative value of the signals and select punishers (over non-punishers) as partners to establish mutually beneficial interactions. Theoretical models (e.g. Gintis et al., 2001; Jordan et al, 2016) and laboratory studies (e.g. Barclay, 2006; Kurzban et al., 2007; Jordan et al., 2016) have mainly supported this account, although not univocally (Horita, 2010; Rockenbach & Milinski, 201; Przepiorka and Liebe, 2016).

The availability of type-separating signals becomes particularly salient when individuals are embedded in fluid social networks, i.e. when they are able (at least partially) to break undesired social ties and establish more profitable ones. The faculty to exert partner choice, indeed, can introduce a market-like logic in the dynamic of social interactions, resulting in an increased level of competition among individuals to be chosen by the best partners (Noe and Hammerstein 1994, 1995; Roberts, 1998). In this scenario, being able to signal desirable qualities becomes a crucial strategic advantage as it allows to attract high-quality individuals and, conversely, reduces the risk of interact with exploitative partners. Confirming this perspective, previous works have shown indeed that in this kind of *biological market* (Noe and Hammerstein 1994) levels of altruism are higher than those observed when individuals are only concerned with their social image, i.e. when they do not face any threat of social exclusion (e.g. Barclay and Willer, 2007; for recent reviews, see Barclay, 2016 and Hammerstein and Noe, 2016).

Here we investigate how punishers perform in such a biological market, under the hypothesis that punitive acts can function as costly signals of the punisher's trustworthiness. Specifically, we ask whether punishers escalate their punishing investments to compete for being selected as social partners by a bystander and, subsequently, whether they are more likely than non-punishers to actually be chosen. Coherently with a signalling account of peer-punishment, we also aim to test the *actual* and *perceived* reliability of punitive acts as type-separating signals of trustworthiness (Przepiorka and Liebe, 2016; Jordan et al., 2016). We then ask whether investments in punishment are predictive of *(i)* the punisher's trustworthiness in a subsequent interaction with a bystander and *(ii)* the level of trust exhibited by a bystander when interacting with the punisher.

As it has been previously argued (Raihani and Bhsary, 2015), the context in which punishment occurs is likely to play a major role in determining the information it conveys and its reputation consequences. Contexts where the punisher was harmed directly by the wrongdoer (i.e. in the case of second-party punishment) are more likely to be motivated by vengeful sentiments and then less likely to be interpreted as signals of the punisher's cooperativeness. On the other side, if the punisher was not impacted by the cheater (i.e. in the case of third-party punishment), the punitive act is more likely to convey an altruistic intent. In line with this argument, here we only focus on the signalling value of third-party punishment to give our hypotheses the best chance of being supported.

To summarize, in the current study we aim to extend the current signalling account of peer-punishment by testing whether:

- H1: reputation-based partner choice escalates investments in punishing as third-parties;
- H2: individuals who invest more in punishing as third-parties are more likely to be chosen as social partners by a bystander;
- H3: individuals who invest more in punishing as third-parties are trusted more by a bystander;
- H4: third-party punishment is a reliable signal of the punisher's trustworthiness.

Eventually, we aim to draw a full characterization of punishment as costly signal of cooperative intents in a context where partner choice is actually implemented.

While punishment can reduce inequalities and be group beneficial, it can also serve a more competitive function. Any act of punishment, by definition, imposes a cost on its target. This implies that – when the inflicted cost is higher than the cost incurred by the punisher – a punitive act can be implemented to alter the relative difference in payoffs between the punishing and the punished agents. This aspect of punishment opens a spectrum of possible applications that exceeds the mere enforcement of cooperation and includes inefficient behaviours such as anti-social forms of punishment (e.g. Hermann et al., 2008; Nikiforakis et al., 2012). The competitive side of punishment may imply a reduction of its informative value as signal of trustworthiness. Previous research, indeed, has shown that when allowed both a punishing and a helping option, individuals tend to prefer the latter as a mean to enhance their reputation (Raihani and Bshary, 2015; Jordan et al. 2016). Given the potential ambiguity of punishment as signal of trustworthiness, it is worth to investigate its informative value when compared to a less ambiguous signal such as helping. We then double our treatments to include a set of conditions where help is the only option available to third-parties. This allows us to perform an independent comparison of the extent that partner choice has different effects on punishment and help as signals of trustworthiness[1].

The hypothesises presented above can be translated to cover the case of helping behaviour by asking whether:

- H5: reputation-based partner choice escalates investments in helping by third-parties;
- H6: individuals who invest more to help as third-parties are more likely to be chosen as social partners by a bystander;
- H7: individuals who invest more to help as third-parties are trusted more by a bystander;
- H8: third-party help is a reliable signal of the helper’s trustworthiness.

Methods

Particpants. A total of 2253 participants were recruited through the online labour market Amazon Mechanical Turk (AMT). Each participant was allocated to one of three roles: Dictator (*n* = 902), Third-Party (*n* = 902) and Bystander (*n* = 449). Throughout the study roles where labelled using neutral terms. Participants in each role were recruited individually and then matched with participants in the other roles. Third-Parties were recruited first, followed by Bystanders and Dictators (details on the procedure and the matching protocol are reported in

[1] We note that our design does not include any treatment where both options (punishment and help) are available to the same subject.

the *Supplementary Information* (*SI*)). All participants received a basic fee contingent on the role assigned ($0.20 for Dictator, $0.50 for both Third-Party and Bystander) and were given the chance to earn a bonus based on their decisions in the experiment. Total average earnings for each role were $0.52 (Dictators), $1.12 (Third-Parties) and $0.80 (Bystanders). All data were collected anonymously and no deception was used. The study was approved by the University College of London Ethics Board (project 3720/001).

Experimental design. Third-Parties and Bystanders were assigned to one of six treatments (as described next). After reading the instructions, participants were administered a comprehension check and then transferred to the decision pages. Data were collected for all participants and all data collected are included in the main analysis to avoid selection bias, although we highlight cases where the results are significantly different when restricting the analyses to participants who correctly answered all comprehension questions (for all the analyses, including those restricted to participants with full comprehension, see the *SI*).

The experimental setting consisted of three stages. In Stage A, Dictators and Third-Parties played a variant of the Dictator Game (Kahneman et al. 1986). Each Dictator was endowed with $0.50 and faced a dichotomous decision between a fair ("Keep $.25 and give $0.25") and an unfair ("Keep $0.45 and give $0.05") share of the endowment with a passive receiver. Each Third-Party had to choose how much, if any, to invest to punish an unfair Dictator or, according to the experimental condition, to help a receiver who was given an unfair share. Third-Parties were endowed with $050 and could invest any amount between $0.00 and $0.45. In Stage B, Third-Parties were paired and one out of each pair was selected (either randomly or by a Bystander, according to the experimental condition) to take part in the last stage. In Stage C, the selected Third Party and a Bystander played a Trust Game (Berg et al, 1995) as the trustee and the trustor, respectively. The Bystander was given $0.30 and had to choose how much, if any, to send to the Third-Party. The amount sent was tripled and the Third-Party had to choose which percentage to keep for herself and which percentage to return to the Bystander.

Across treatments, we varied: *(i)* whether Third-Parties could punish an unfair Dictator or help the corresponding receiver in Stage A; *(ii)* whether Third-Parties were randomly selected or chosen by a Bystander in Stage B and *(iii)* whether the Bystander was informed or not of how the Third-Parties had behaved in Stage A. We implemented a 2 (Punish versus Help) X 2 (Random vs Choice) X 2 (Anonymous versus Knowledge) between-subjects fractional factorial designs, resulting in six experimental treatments: Punish/Random/Anonymous, Punish/Random/Knowledge, Punish/Choice/Knowledge, Help/Random/Anonymous, Help/Random/Knowledge, Help/Choice/Knowledge. No treatment pairing a “Choice” and an “Anonymous” condition was included in the design as not relevant for the study. Allocations to all treatments occurred randomly within each session and participants made their decision in isolation. We initially collected the decisions of Third-Parties in Stage A and Stage B, then we paired and matched them in chronologically order to Dictators and Bystanders (e.g. within a same condition, the first two Third-Parties we recruited were paired and then matched with the first Dictator and the first Bystander). In the Punish/Choice/Knowledge, Help/Random/Anonymous, Help/Random/Knowledge, Help/Choice/Knowledge conditions one Third-Party (i.e. 4 overall) could not be paired and, therefore, was not matched with a Bystander (their decisions, however, are included in the analysis). Full version of the instructions, comprehension questions, statistical analysis as well as the screenshots of the experiment as administered to the participants are provided in the *SI*.

Results

1. Did partner choice escalate investments in punishing/helping? (Hypotheses 1 and 5)

Descriptively, we observe overall law rate of positive investments in punishment (Median = \$0 in all the three conditions) and, when punishment does occur, investments are small ($Q_3 = \$0.10, \$0.20$ and $\$0.20$ in the Random/Anonymous, Random/Knowledge and Choice/Knowledge condition, respectively). Mean investments in punishment differed significantly between the three conditions (Kruskal-Wallis: $X^2 = 7.00$, *p* = 0.03), although this difference is no more significant when excluding participants who did not pass the comprehension check with a full score (Kruskal-Wallis: $X^2 = 3.61$, *p* = 0.16). When including all participants in the analysis, we observe that Third-Parties invested more when Bystanders could choose their partner than when their decisions were anonymous (Punish/Choice/Knowledge versus Punish/Random/Anonymous: Wilcoxon *W* = 9242, *p* = 0.025). We did not find any significant difference between investments in all the other pairwise comparisons (Punish/Choice/Knowledge versus Punish/Random/Knowledge: Wilcoxon *W* = 10901, *p* = 0.425; Punish/Random/Knowledge versus Punish/Random/Anonymous: Wilcoxon *W* = 9144, *p* = 0.118). These results suggest that, in our setting, only the opportunity to be chosen as partner based on their punishing decisions induced Third-Parties to escalate their investments. Our first hypothesis then seems to be partially confirmed, as only in the partner-choice protocol investments in punishing are significantly higher than when they are anonymised. However, we cannot draw any clear conclusion based on our data, as the observed difference is no more significant when restricting the analysis to participants with full comprehension and we did not find a significant difference between the Punish/Choice/Knowledge and the Punish/Random/Knowledge conditions.

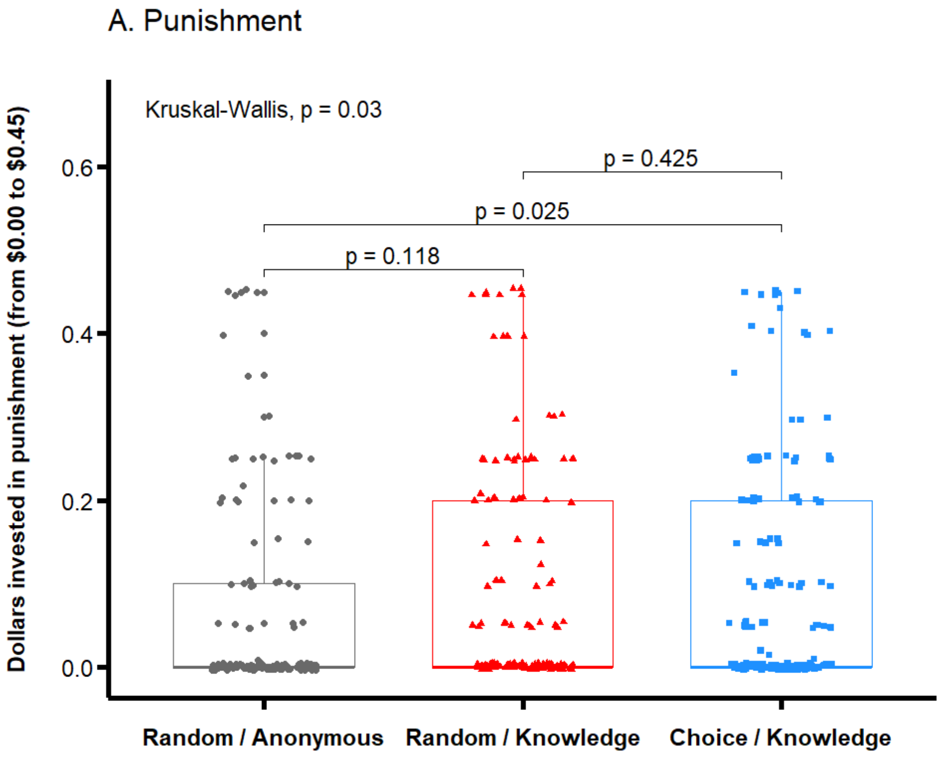


Figure 1. Dollars invested in punishment by Third-Parties according to treatment.
Number of dollars invested in third-party punishment in each treatment before the stage of partner assignment/choice. When including all participants, investments were significantly different across treatments ($X^2 = 7.00$, p = 0.03). A pairwise Wilcoxon test leads to identical results (Random/Anonymous versus Random/Knowledge: W = 9144, p = 0.118; Random/Anonymous versus Choice/Knowledge: W = 9242, p = 0.025; Random/Knowledge versus Choice/Knowledge: W = 10901, p = 0.425).

Investments in help differed significantly between the three conditions (Kruskal-Wallis: $X^2 = 26.43$, $p < 0.001$), even when excluding participants who did not pass the comprehension check with a full score (Kruskal-Wallis: $X^2 = 12.93$, $p = 0.001$). When including all participants in the analysis, we observe that Third-Parties invested more when their decisions were observed than when they were anonymised (Help/Random/Knowledge versus Help/Random/Anonymous: Wilcoxon $W = 9091$, $p < 0.001$). Similarly, investments were higher when Bystanders could choose their partner than when Third-Parties' decisions were anonymous (Help/Choice/Knowledge versus Help/Random/Anonymous: Wilcoxon $W = 8425$, $p < 0.001$). However, we did not find any evidence that investments escalated under the partner choice protocol compared to when Third-Parties' decisions were only observable (Help/Choice/Knowledge versus Help/Random/Knowledge: Wilcoxon $W = 9953$, $p = 0.17$). An identical pattern is observed when excluding participants without full comprehension, although all the effects become smaller (see details in section 3 of the *SI*). Thus, we reject our hypothesis that partner-choice escalates investments in helping to signal trustworthiness: according to our data, mere reputation incentives and partner-choice succeed both in increasing helping behaviour, but with no difference between them.

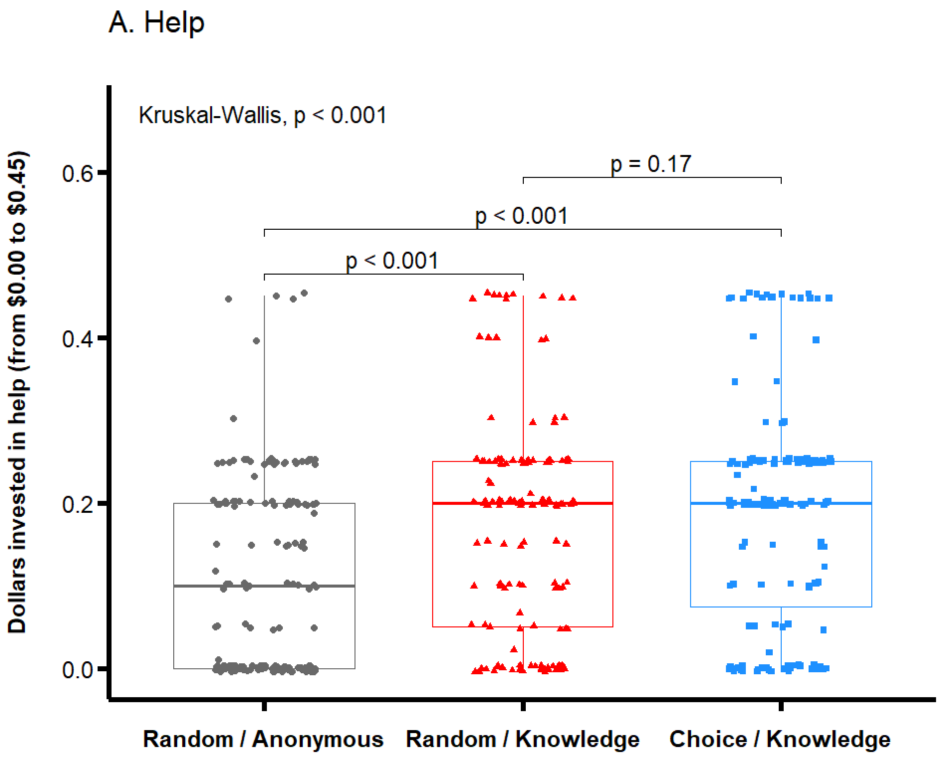


Figure 2. Dollars invested in help by Third-Parties according to treatment.

Number of dollars invested in third-party help in each treatment before the stage of partner assignment/choice. When including all participants, investments were significantly different across treatments ($X^2 = 26.43$, *p* < 0.001). A pairwise Wilcoxon test specify the differences between groups (Random/Anonymous versus Random/Knowledge: *W* = 9091, *p* < 0.001; Random/Anonymous versus Choice/Knowledge: *W* = 8425, *p* < 0.001; Random/Knowledge versus Choice/Knowledge: *W* = 9953, *p* = 0.17).

2. *Were investments in punishing/helping used to choose their partners by Bystanders? (Hypotheses 2 and 6)*

Third-Parties who invested the most in punishment were chosen on 36/56 occasions, indicating that Bystanders were keen to use this information to identify the most desirable partner (Binomial, *p* = 0.02). We note, however, that when considering only participants with full comprehension, the highest investors were not more likely to be chosen as partners (highest chosen on 15/27 occasions; Binomial, *p* = 0.35). Therefore, we cannot draw clear conclusions on whether investing in punishing created a competitive advantage to be selected as social partners.

Conversely, helping behaviour was a much clearer determinant of Bystanders' choices, with highest investors in help being chosen on 64/65 occasions (Binomial, *p* < 0.001). The same result is observed when considering the subset of full comprehenders (highest chosen on 32/33 occasions; Binomial, *p* < 0.001). Helping at higher rates, therefore, clearly gave a competitive advantage to be selected as social partners.

3. *Were higher investors trusted more by Bystanders? (Hypotheses 3 and 7)*

Across treatments we did not find significant differences in sent amount to Third-Parties who had the option to punish a stingy Dictator (Kruskal-Wallis: $X^2 = 1.507$, *p* = 0.47; see Figure S5 in the *SI*). A more thorough look via regression analysis, however, shows a significative positive effect of investment on the amount sent by Bystanders in the Random/Knowledge and Choice/Knowledge conditions: for each percentage point of the endowment invested by the selected Third-Party, Bystanders sent 0.84 ($p < 0.001$) and 0.45 ($p < 0.01$) percentage point more, respectively (this effect becomes statistically non-significant in the Choice/Knowledge condition when considering only full comprehenders, see Column 4 of Table S1 in the *SI*). Pairing this result with our previous findings, we conclude that Bystanders used investments in punishment as a signal of the Third-Party's trustworthiness – and particularly so when Bystanders could not choose their partner.

Across treatments where Third-Parties had the option to help, we did find significant differences in the amount sent by Bystanders (Kruskal-Wallis: $X^2 = 7.978$, p = 0.018), driven by the higher amount sent in the Choice/Knowledge condition compared to the Random/Anonymous condition (W = 2285, p = 0.014; see Figure S6 in the *SI*). Furthermore, a more thorough look via regression analysis shows that the amount invested was determinant of the amount sent by Bystanders only in the Random/Knowledge condition: here, for each percentage point of the endowment invested by the selected Third-Party, Bystanders sent 0.49 ($p < 0.05$) percentage point more (although this effect is not more significant when considering only full comprehenders, see Table S2 in the *SI*). Coherently with the findings reported above, therefore, we observe that Bystanders interpreted investments in punishing as a signal of a Third-Party's trustworthiness. Interestingly, the amount invested by Third-Parties influenced the amount sent by Bystanders only when the latter could not choose their partners. When partner-choice was an option, Bystanders tended to choose the most helpful partners and then send their entire endowment to them (Help/Choice/Knowledge condition, Median = $0.30, i.e. 100% of Bystanders' endowment). This result suggests that the perceived reliability of the signal was not reduced when its net benefit increased (by increasing the likelihood to take part in the TG), further confirming that Bystanders perceived signalling via helping as reliable.

4. *Were investments in punishing/helping a reliable signal of the Third-Party's trustworthiness? (Hypotheses 4 and 8)*

We took the amount returned in the TG by Third-Parties as proxy of their trustworthiness. Across treatments we did not find significant differences in returned amount by Third-Parties who had the option to punish a stingy Dictator (Kruskal-Wallis: $X^2 = 4.05$, p = 0.132; see Figure S5 in the *SI*). A more thorough look via regression analysis, however, shows a significative positive main effect of investment on the amount returned by Third-Parties: for each percentage point of the endowment invested, Third-Parties returned 0.25 ($p < 0.001$) percentage point more (see Column 1 of Table S3 in the *SI*). Interestingly, this effect is reversed in the Choice/Knowledge condition: here, investments in signalling have a significant negative effect on the amount returned (coeff = -169, p < 0.05). This result suggests that when the net benefit of signal production increases, the temptation to cheat reduces the reliability of the signal itself.

Across treatments we did not find significant differences in returned amount by Third-Parties who had the option to help in the first stage (Kruskal-Wallis: $X^2 = 2.44$, *p* = 0.29; see Figure S6 in the *SI*). A more thorough look via regression analysis, however, shows a significative positive main effect of investment on the amount returned by Third-Parties: for each percentage point of the endowment invested, Third-Parties returned 0.4 (p < 0.001) percentage point more (see Column 1 of Table S4 in the *SI*). No significant interaction with experimental conditions is found, showing that the reliability of help-based signals is more resistant to the temptation to cheat.

Discussion

This study aimed at filling a gap in the current literature by investigating the link between partner choice and peer-punishment. In particular, it asked whether investments in punishing behaviour escalate as a result of the competition created by partner-choice. We found only partial support for this hypothesis: only when punishing decisions were determinant to take part in a subsequent (and potentially profitable) social interaction, investments in punishment differed significantly from a context were punishment did not receive any strategic incentive. However, punishing decisions did not escalate compared to a context where punishment was strategically incentivized only by the possibility of acquiring a better social image. Furthermore, our results on the competitive advantage of punishing investments in a partner-choice market are ambiguous: the positive effect observed when considering all participants disappear when taking into account comprehension levels of the participants (as defined by the correct answers to the comprehension check included in the instructions). Notwithstanding, we found a clear association between punishing as third-party and trustworthiness/trust, coherently with a signalling account of punishment. Overall, our results also suggest that punishing behaviour tends to be perceived by the signaller as a weaker signal of trustworthiness compared to helping (even when it is not a direct alternative to punishment, as in Raihani and Bshary, 2015 and Jordan et al., 2016). Indeed, when the mere reputation gains of costly signalling are diluted (as in our Random/Knowledge condition), punishment doesn't seem to be considered an investment worth its cost.

We also note some important differences between some previous results and ours. Some key aspects of our design might explain this discordance. The arguments that we propose in what follow, however, are purely hypothetical and should benefit from further empirical testing.

Jordan et al. (2016) found that TTP was used to signal trustworthiness to a future partner in a TG. Conversely, we did not find evidences for signalling via TPP when partner-choice was not implemented. The reason for this discrepancy might be related to an important difference between the designs of the two studies. In Jordan et al. (2016) participating to the TG of the last stage was automatic, while in our study Third-Parties had only the 50% of chance to take part in the last stage.

Barclay and Willer (2007) claimed to find that the competition introduced by partner choice caused an escalation of investments in signalling behaviour via generosity. In our study, we did not find support for a similar conclusion, as investments in the Help/Random/Knowledge condition did not differ significantly from investments in the Help/Choice/Knowledge condition. We propose two arguments to explain this discrepancy. First, we note that in Barclay and Willer (2007)'s design the net cost of signalling might have been reduced in the Choice/Knowledge condition compared to the Random/Knowledge condition. In the first stage of their experiment, indeed, participants played a continuous PD, where the relative impact of the signalling cost on the initial welfare of a player can be considered as a negative function of the partner's cooperation (e.g. signalling by giving \$10 when receiving \$0 reduces the initial endowment of \$10; signalling by giving \$10 when receiving \$10 reduces the initial endowment of \$0). As treatment conditions were common knowledge in the study, participants might have expected a higher level of cooperation from the partner in the PD played in the Choice/Knowledge condition. Under this belief, the expected impact of the altruistic signal on the initial welfare of a player might have been reduced, potentially confounding the interpretation of the results in terms of competitive altruism. Another key difference between the previous study and the current one relates to the type of game that was played in the final stage. In Barclay and Willer (2007), participants played a PD and received a *new* endowment at the beginning of the stage. Taking part in the second PD, therefore, gave *direct* access to new resources. Conversely, in our design taking part to the second game did not give access *per se* to any new resource. In our experiment, therefore, recouping the cost of signalling in the second stage was (somehow more realistically) much more uncertain.

To conclude, in our study we found that the empirical evidences in favour of a signalling account of third-party punishment are overall inconclusive. First, consistently with some of previous studies (e.g. Jordan et al., 2016), but in contrast with another one (Przepiorka and Liebe, 2916), we found that third-party punishment is a reliable signal of trustworthiness and interpreted as such by non-involved observers. Importantly, however, Bystanders only displayed a weak inclination to actually choose higher investors in punishment as social partners.

Supplementary Information

S1 – Recruitment and Procedure

Participants were recruited on Amazon Mechanical Turk (AMT), an online labour market where workers can be hired to perform so called Human Intelligence Tasks (HITs), consisting of usually short tasks completed in exchange of a commensurate small pay. Workers recruited via AMT have proved to be a valid alternative to more conventional subject pools (e.g. undergraduate students) for running experiments in the social sciences. The recruitment process is faster and cheaper, allowing for the collection of larger sample sizes. Furthermore, workers are likely to be a demographically more varied and representative sample of the population compared to undergraduate students (Horton et al., 2011). Considering that workers perform their tasks in remote (ether via the AMT online platform or on a website they are redirected to), concerns have been raised on the reliability of the results obtained on AMT. Compared to traditional experiments run in physical laboratories, indeed, AMT implies a reduced control on subjects' behaviour and on the way the experimental task is carried out. These and other potential drawbacks, however, have been directly addressed by some previous studies (e.g. Bernsky et al., 2012; Amir et al., 2012), which have overall validated the results of experiments conducted on AMT.

The experiment was run using oTree, a free open-source software for running economic experiments in physical labs, online and in the field (Chen et al., 2016).

In what follows we describe the procedure implemented for the participants in each role.

Procedure for Third-Parties Participants assigned to the role of Third-Party were recruited first. This procedure allowed us to use Third-Parties' actual decisions in the part of the experiment administered to Bystanders. After reading the instructions, participants went through a comprehension check aiming to assess their understanding of (i) the structure of the payoff for each role and (ii) the strategic nature of the interactions in each stage of the task. Participants were then asked to make a unique decision about how much to invest for punishing (helping) an unfair Dictator (a worker who received an unfair share). Thereafter, they were asked to make their decision as trustee in a TG. They were fully informed that their decisions will have been implemented only *if* they were matched with an unfair Dictator and/or selected (chosen by a Bystander) to take part in the last stage. They were also told that their final payoff will have been calculated once the decisions from all participants in the study had been collected.

Procedure for Bystanders Participants assigned to the role of Bystander were recruited after the recruitment of Third-Parties had been completed. After reading the instructions, participants went through a comprehension check, which had a comparable structure to the one used for Third-Parties. Participants were then presented with a page displaying a pair of Third-Parties (stage B of the experiment). Contingent to the treatment they had been assigned, on this page they could: see the decisions of the Third-Parties in the previous stage and choose whom to interact with in the TG (Choice/Knowledge); see the decisions of the Third-Parties in the previous stage but not choose whom to interact with in the TG (Random/Knowledge); not see the decisions of the Third-Parties in the previous stage and not choose whom to interact with in the TG (Random/Anonymous). Thereafter, they were asked to make their decision as trustor in the TG. to make a unique decision about how much to invest for punishing (helping) an unfair Dictator (a worker who received an unfair share). They received immediate feedback about their final payoff.

Procedure for Dictators Participants assigned to the role of Dictator were recruited after the recruitment of Third-Parties had been completed. After reading the instructions, participants went through a comprehension check, which had a comparable structure to the one used for the other roles, but only concerned the first stage of the experiment. Participants then participated in the dichotomous DG and, thereafter, they received immediate feedback about the decision of the Third-Party and their final payoff. Recipients of the DG were randomly selected from the pool of Mturk workers who had previously participated in studies conducted on AMT by the Raihani Lab.

We note that in the experimental settings for each role we framed the options that Third-Parties could choose from as “punish” (“help”) or not an “unfair Player 1” (a “Player 2 who received an unfair share”). We used this kind of explicit wording for the sake of comparability with Jordan et al. (2016).

S2 – Statistical Analysis

Overview All analyses are performed in R 3.0.3, using Kruskal tests, Wilcox tests, Binomial tests and regression analyses as appropriate. All statistical tests are two-sided. To account for multiple comparisons, threshold levels for significance of Wilcox tests are adjusted using the Benjamini–Hochberg method (Benjamini and Hochberg 1995). To facilitate the interpretation of the results, in all regressions we take as variables the percentage of the endowment invested in punishing/helping or sent in the TG. We use robust standard errors in all regression analyses. We do not report any analysis relative to Dictators, as they are not relevant in this study.

Comprehension Questions Third-Parties were presented with 8 comprehension questions and 51.7% of participants answered all questions correctly. Bystanders were presented with 8 comprehension questions and 51.4% of participants answered all questions correctly.

Regression Tables

Table S1: Linear regression model estimating the effect of investment in punishment (as percentage of the endowment) by the selected Third-Party on amount sent (as percentage of the endowment) by the Bystander in the Trust Game. Treatment conditions are dummy coded and included as covariates in Columns 3 and 4. Column 1 and 3 report results from all participants; Column 2 and 4 only report results from participants with full comprehension. Robust standard errors in parentheses.

	Dependent variable:			
	Sent			
	All Subjects	Full Comp	All Subjects	Full Comp
	(1)	(2)	(3)	(4)
Punish by Selected	0.278^{**}	0.319^{*}	-0.143	-0.268
	(0.089)	(0.127)	(0.185)	(0.295)
Random/Know			-22.113^{**}	-33.746^{**}
			(7.851)	(11.116)
Choice/Know			-17.506^{*}	-17.079
			(8.040)	(12.069)
Chosen Punish x Random/Know			0.840^{***}	1.094^{**}
			(0.213)	(0.334)
Chosen Punish x Choice/Know			0.451^{*}	0.609
			(0.229)	(0.357)
Constant	50.161^{***}	52.402^{***}	63.169^{***}	68.467^{***}
	(3.363)	(4.981)	(5.349)	(7.434)
Observations	221	115	221	115
R^2	0.041	0.049	0.104	0.155
Adjusted R^2	0.037	0.041	0.083	0.116
Note:				$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$ Robust standard errors in parenthesis

Table S2: Linear regression model estimating the effect of investment in help (as percentage of the endowment) by the selected Third-Party on amount sent (as percentage of the endowment) by the Bystander in the TG. Treatment conditions are dummy coded and included as covariates in Columns 3 and 4. Column 1 and 3 report results from all participants; Column 2 and 4 only report results from participants with full comprehension. Robust standard errors in parentheses.

	Dependent variable:			
	Sent			
	All Subjects	Full Comp	All Subjects	Full Comp
	(1)	(2)	(3)	(4)
Help by Selected	0.253^{**}	0.432^{***}	-0.101	0.056
	(0.083)	(0.126)	(0.173)	(0.242)
Random/Know			-10.849	-16.824
			(10.779)	(15.108)
Choice/Know			4.078	8.604
			(13.003)	(18.820)
Chosen Help x Random/Know			0.495^{*}	0.543
			(0.233)	(0.310)
Chosen Help x Choice/Know			0.293	0.241
			(0.239)	(0.352)
Constant	52.801^{***}	45.899^{***}	57.728^{***}	52.796^{***}
	(4.737)	(6.921)	(6.677)	(10.575)
Observations	228	116	228	116
R^2	0.038	0.093	0.073	0.138
Adjusted R^2	0.033	0.085	0.052	0.099
Note:				$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$ Robust standard errors in parenthesis

Table S3: Linear regression model estimating the effect of investment in punishment (as percentage of the endowment) on amount returned (as percentage of the sent amount) by the by Third-Parties in the TG. Treatment conditions are dummy coded and included as covariates in Columns 3 and 4. Column 1 and 3 report results from all participants; Column 2 and 4 only report results from participants with full comprehension. Robust standard errors in parentheses.

	Dependent variable:			
	Returned			
	All Subjects	Full Comp	All Subjects	Full Comp
	(1)	(2)	(3)	(4)
Punish	0.259***	0.228***	0.333***	0.256***
	(0.037)	(0.045)	(0.050)	(0.061)
Random/Know			2.896	0.461
			(3.392)	(4.734)
Choice/Know			6.584	0.870
			(3.467)	(4.558)
Punish x Random/Know			-0.044	0.023
			(0.088)	(0.111)
Punish x Choice/Know			-0.169*	-0.072
			(0.078)	(0.095)
Constant	26.717***	27.506***	23.615***	27.008***
	(1.405)	(1.844)	(2.420)	(3.455)
Observations	443	245	443	245
R^2	0.098	0.081	0.109	0.085
Adjusted R^2	0.096	0.077	0.099	0.065
Note:				*p<0.05; **p<0.01; ***p<0.001 Robust standard errors in parenthesis

Table S4: Linear regression model estimating the effect of investment in help (as percentage of the endowment) on amount returned (as percentage of the sent amount) by the by Third-Parties in the TG. Treatment conditions are dummy coded and included as covariates in Columns 3 and 4. Column 1 and 3 report results from all participants; Column 2 and 4 report only results from participants with full comprehension. Robust standard errors in parentheses.

	Dependent variable:			
	Returned			
	All Subjects	Full Comp	All Subjects	Full Comp
	(1)	(2)	(3)	(4)
Help	0.401^{***}	0.371^{***}	0.497^{***}	0.558^{***}
	(0.036)	(0.050)	(0.057)	(0.081)
Random/Know			2.983	3.915
			(3.597)	(5.344)
Choice/Know			-0.358	0.617
			(3.741)	(5.083)
Help x Random/Know			-0.138	-0.214
			(0.087)	(0.128)
Help x Choice/Know			-0.107	-0.235^{*}
			(0.086)	(0.112)
Constant	17.855^{***}	18.136^{***}	16.788^{***}	16.240^{***}
	(1.497)	(2.130)	(2.120)	(3.177)
Observations	459	221	459	221
R^2	0.255	0.227	0.263	0.256
Adjusted R^2	0.253	0.224	0.255	0.239

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$
Robust standard errors in parenthesis

Figures

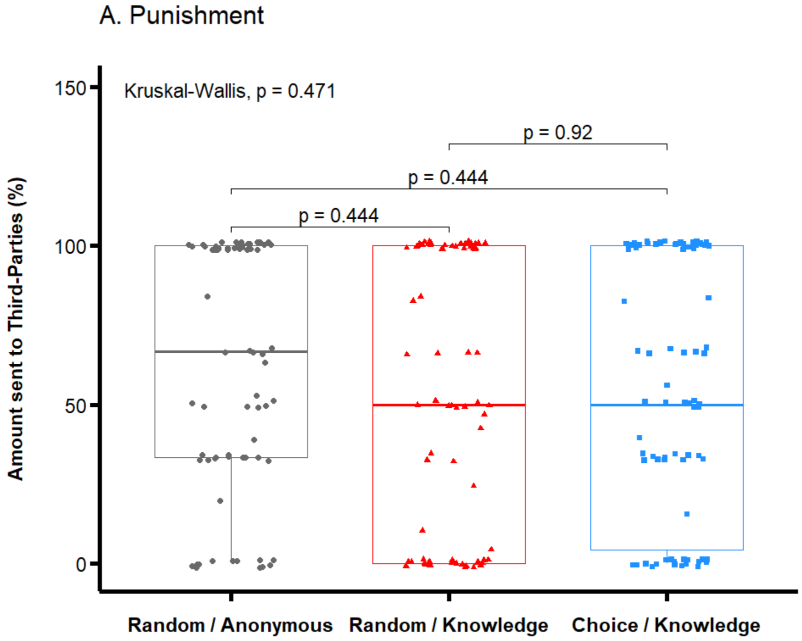


Figure S5. Amount sent to Third-Party Punishers according to treatment.
Amount sent (as percentage of the endowment) by Bystanders in the TG to Third-Parties who had to decide how much to invest in punishment. The amount sent was not significantly different across treatments ($X^2 = 1.507$, $p = 0.471$). A pairwise Wilcoxon test leads to identical results (Random/Anonymous versus Random/Knowledge: $W = 2813$, $p = 0.44$; Random/Anonymous versus Choice/Knowledge: $W = 2992$, $p = 0.44$; Random/Knowledge versus Choice/Knowledge: $W = 2820$, $p = 0.92$).

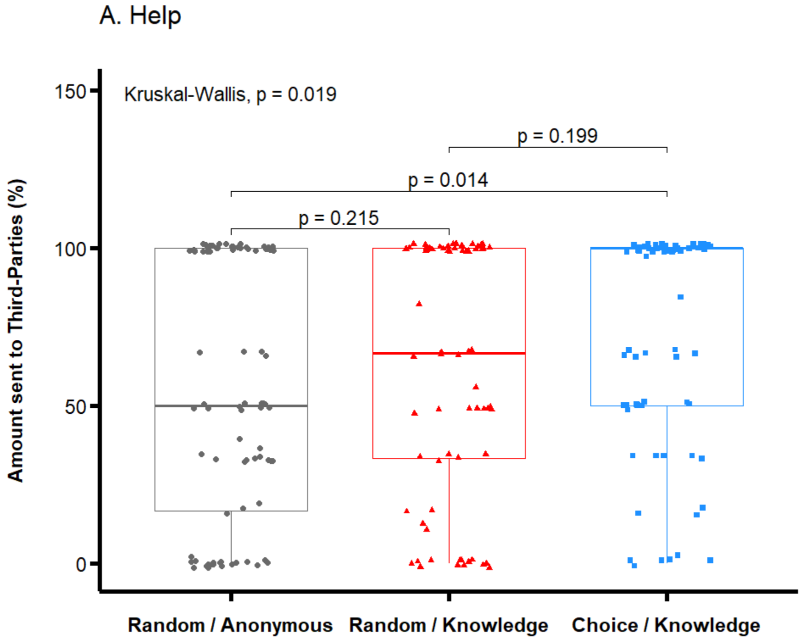


Figure S6. Amount sent to Third-Party Helpers according to treatment.
Amount sent (as percentage of the endowment) by Bystanders in the TG to Third-Parties who had to decide how much to invest in help. Amount sent was significantly different across treatments ($X^2 = 7.97$, $p = 0.019$). A pairwise Wilcoxon test specify the differences between groups (Random/Anonymous versus Random/Knowledge: *W* = 2592, *p* = 0.21; Random/Anonymous versus Choice/Knowledge: *W* = 2285, *p* = 0.014; Random/Knowledge versus Choice/Knowledge: *W* = 2340, *p* = 0.19).

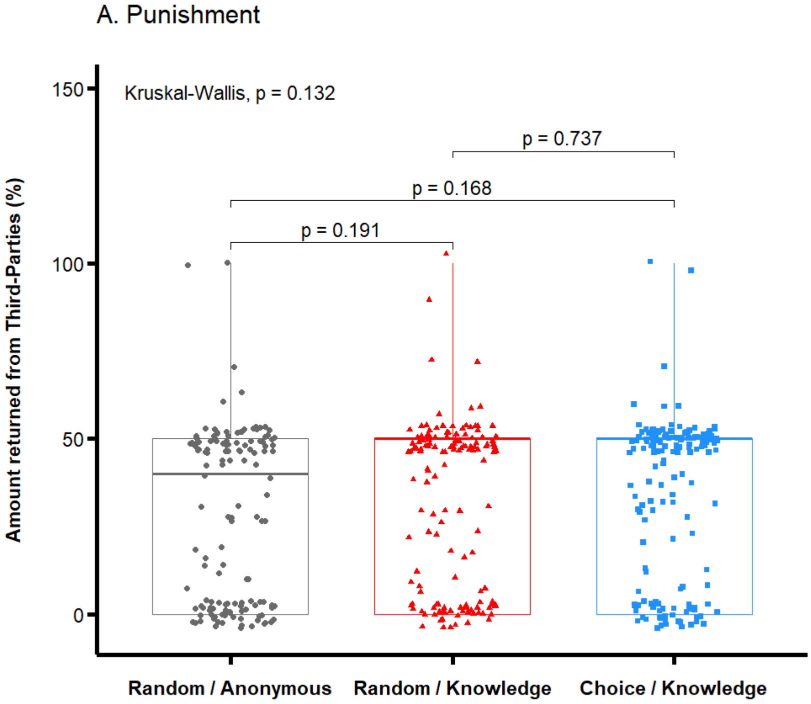


Figure S6. Amount sent to Third-Party Helpers according to treatment.
Amount sent (as percentage of the endowment) by Bystanders in the TG to Third-Parties who decided how much to invest in help. Amount sent was significantly different across treatments ($X^2 = 4.051$, $p = 0.132$). A pairwise Wilcoxon test leads to identical results (Random/Anonymous versus Random/Knowledge: *W* = 9223, *p* = 0.19; Random/Anonymous versus Choice/Knowledge: *W* = 9672, *p* = 0.16; Random/Knowledge versus Choice/Knowledge: *W* = 11224, *p* = 0.73).

A. Help

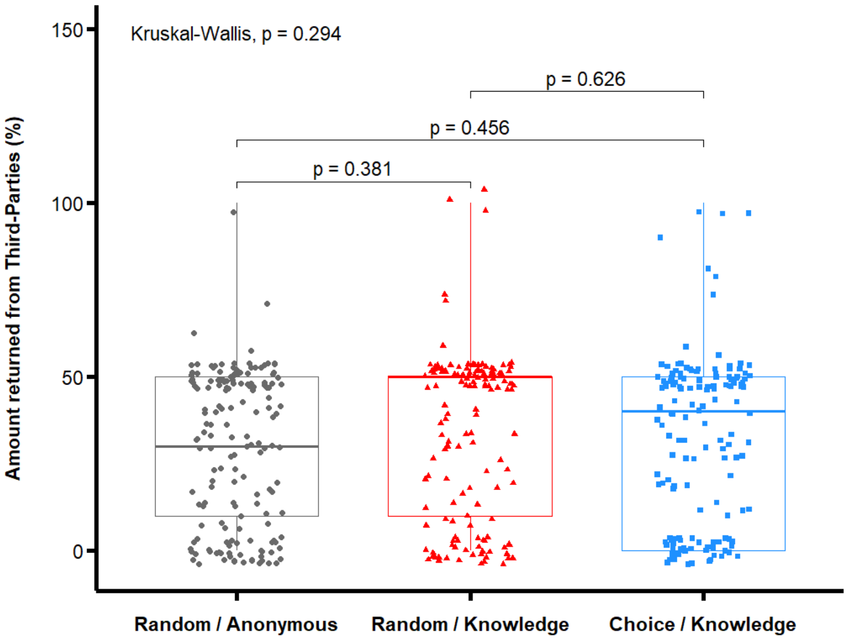


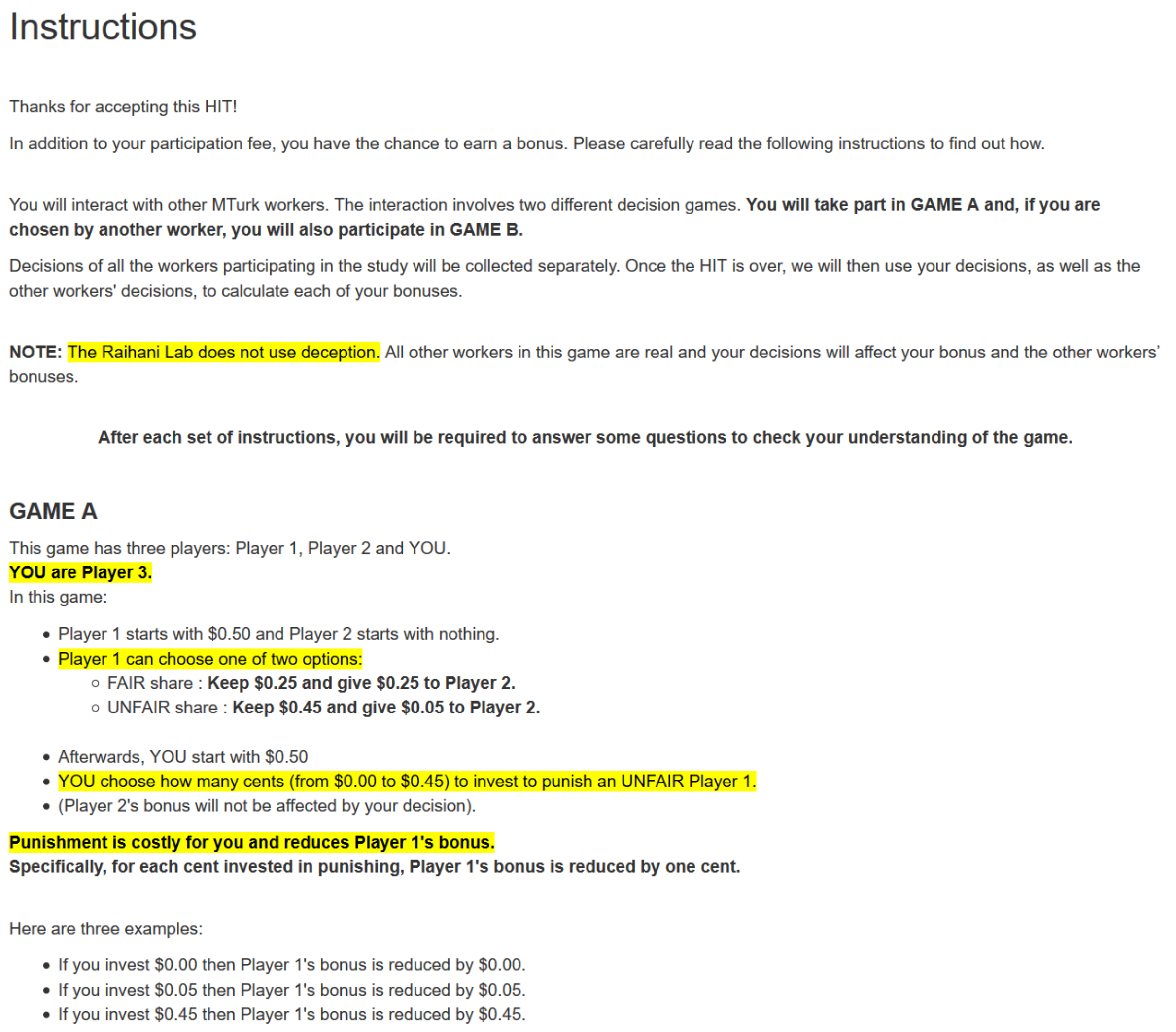
Figure S6. Amount sent to Third-Party Helpers according to treatment.

Amount sent (as percentage of the endowment) by Bystanders in the TG to Third-Parties who decided how much to invest in help. Amount sent was significantly different across treatments ($X^2 = 2.447$, *p* = 0.294). A pairwise Wilcoxon test leads to identical results (Random/Anonymous versus Random/Knowledge: *W* = 10680, *p* = 0.38; Random/Anonymous versus Choice/Knowledge: *W* = 11511, *p* = 0.45; Random/Knowledge versus Choice/Knowledge: *W* = 11288, *p* = 0.62).

S3 – Experimental Instructions

Here we present the experimental instructions, comprehension questions and decision pages used for the two games implemented in the study. The material is presented separately for each role (Dictators and Receivers are omitted as not relevant). Shown below is the version presented in the Punishment + Knowledge/Choice treatment. After each part of the content, we highlight the differences with the other treatments. Across all treatments, each role was neutrally labelled as follow: Dictators were "Player 1", Receivers were "Player 2", Third-Parties (in either the Help or the Punish treatment) were "Player 3", Bystanders were "Player 4".

Third-Parties Shown below are the screenshots of the instructions and the comprehension questions presented to Third-Parties. Instructions and comprehension questions were presented on the same page.



Instructions

Thanks for accepting this HIT!

In addition to your participation fee, you have the chance to earn a bonus. Please carefully read the following instructions to find out how.

You will interact with other MTurk workers. The interaction involves two different decision games. **You will take part in GAME A and, if you are chosen by another worker, you will also participate in GAME B.**

Decisions of all the workers participating in the study will be collected separately. Once the HIT is over, we will then use your decisions, as well as the other workers' decisions, to calculate each of your bonuses.

NOTE: The Raihani Lab does not use deception. All other workers in this game are real and your decisions will affect your bonus and the other workers' bonuses.

After each set of instructions, you will be required to answer some questions to check your understanding of the game.

GAME A

This game has three players: Player 1, Player 2 and YOU.
YOU are Player 3.
In this game:

- Player 1 starts with $0.50 and Player 2 starts with nothing.
- Player 1 can choose one of two options:
  - FAIR share : **Keep $0.25 and give $0.25 to Player 2.**
  - UNFAIR share : **Keep $0.45 and give $0.05 to Player 2.**

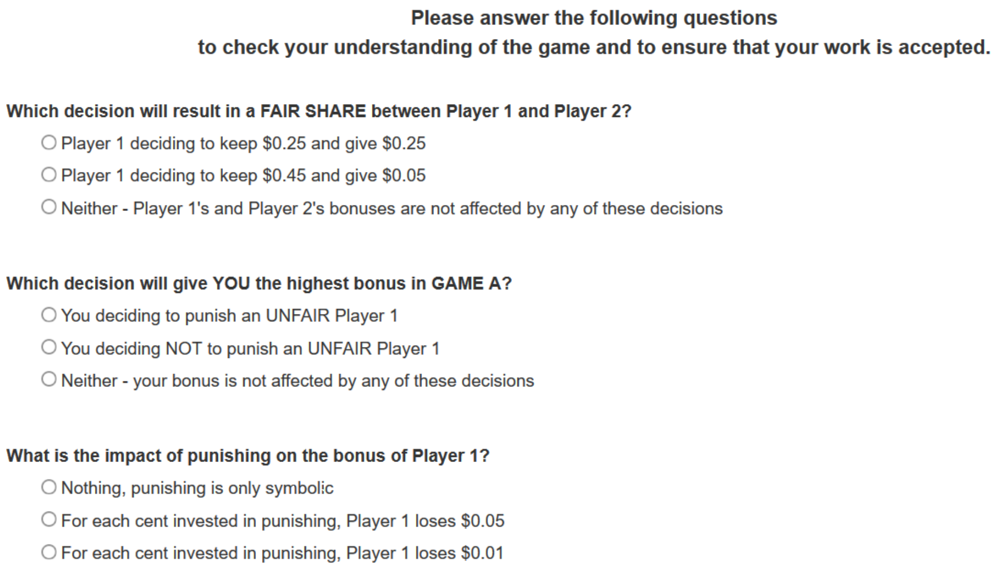
- Afterwards, YOU start with $0.50
- YOU choose how many cents (from $0.00 to $0.45) to invest to punish an UNFAIR Player 1.
- (Player 2's bonus will not be affected by your decision).

Punishment is costly for you and reduces Player 1's bonus.
Specifically, for each cent invested in punishing, Player 1's bonus is reduced by one cent.

Here are three examples:

- If you invest $0.00 then Player 1's bonus is reduced by $0.00.
- If you invest $0.05 then Player 1's bonus is reduced by $0.05.
- If you invest $0.45 then Player 1's bonus is reduced by $0.45.

NOTE: In the *Help condition*, the decision available to Player 3 was framed as "help a Player 2 who received an UNFAIR share." The same reference to the cost associated with the decision was made here.



Please answer the following questions
to check your understanding of the game and to ensure that your work is accepted.

Which decision will result in a FAIR SHARE between Player 1 and Player 2?

○ Player 1 deciding to keep $0.25 and give $0.25
○ Player 1 deciding to keep $0.45 and give $0.05
○ Neither - Player 1's and Player 2's bonuses are not affected by any of these decisions

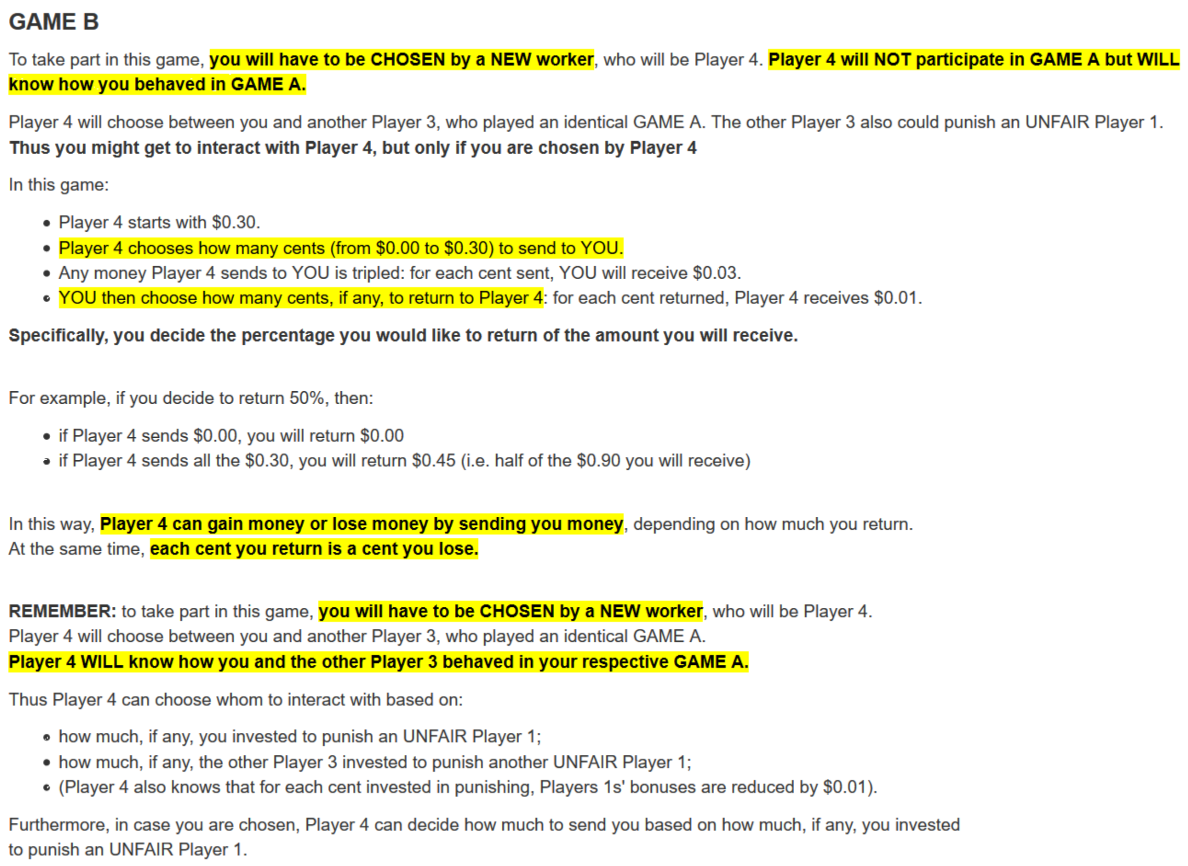
Which decision will give YOU the highest bonus in GAME A?

○ You deciding to punish an UNFAIR Player 1
○ You deciding NOT to punish an UNFAIR Player 1
○ Neither - your bonus is not affected by any of these decisions

What is the impact of punishing on the bonus of Player 1?

○ Nothing, punishing is only symbolic
○ For each cent invested in punishing, Player 1 loses $0.05
○ For each cent invested in punishing, Player 1 loses $0.01

Correct answers: 1) "Player 1 deciding to keep $0.25 and give $0.25; 2) "You deciding NOT to punish an UNFAIR Player 1"; 3) "For each cent invested in punishing, Player 1 loses $0.01".



GAME B

To take part in this game, **you will have to be CHOSEN by a NEW worker**, who will be Player 4. **Player 4 will NOT participate in GAME A but WILL know how you behaved in GAME A.**

Player 4 will choose between you and another Player 3, who played an identical GAME A. The other Player 3 also could punish an UNFAIR Player 1.
Thus you might get to interact with Player 4, but only if you are chosen by Player 4

In this game:

- Player 4 starts with $0.30.
- Player 4 chooses how many cents (from $0.00 to $0.30) to send to YOU.
- Any money Player 4 sends to YOU is tripled: for each cent sent, YOU will receive $0.03.
- YOU then choose how many cents, if any, to return to Player 4: for each cent returned, Player 4 receives $0.01.

Specifically, you decide the percentage you would like to return of the amount you will receive.

For example, if you decide to return 50%, then:

- if Player 4 sends $0.00, you will return $0.00
- if Player 4 sends all the $0.30, you will return $0.45 (i.e. half of the $0.90 you will receive)

In this way, **Player 4 can gain money or lose money by sending you money**, depending on how much you return.
At the same time, **each cent you return is a cent you lose.**

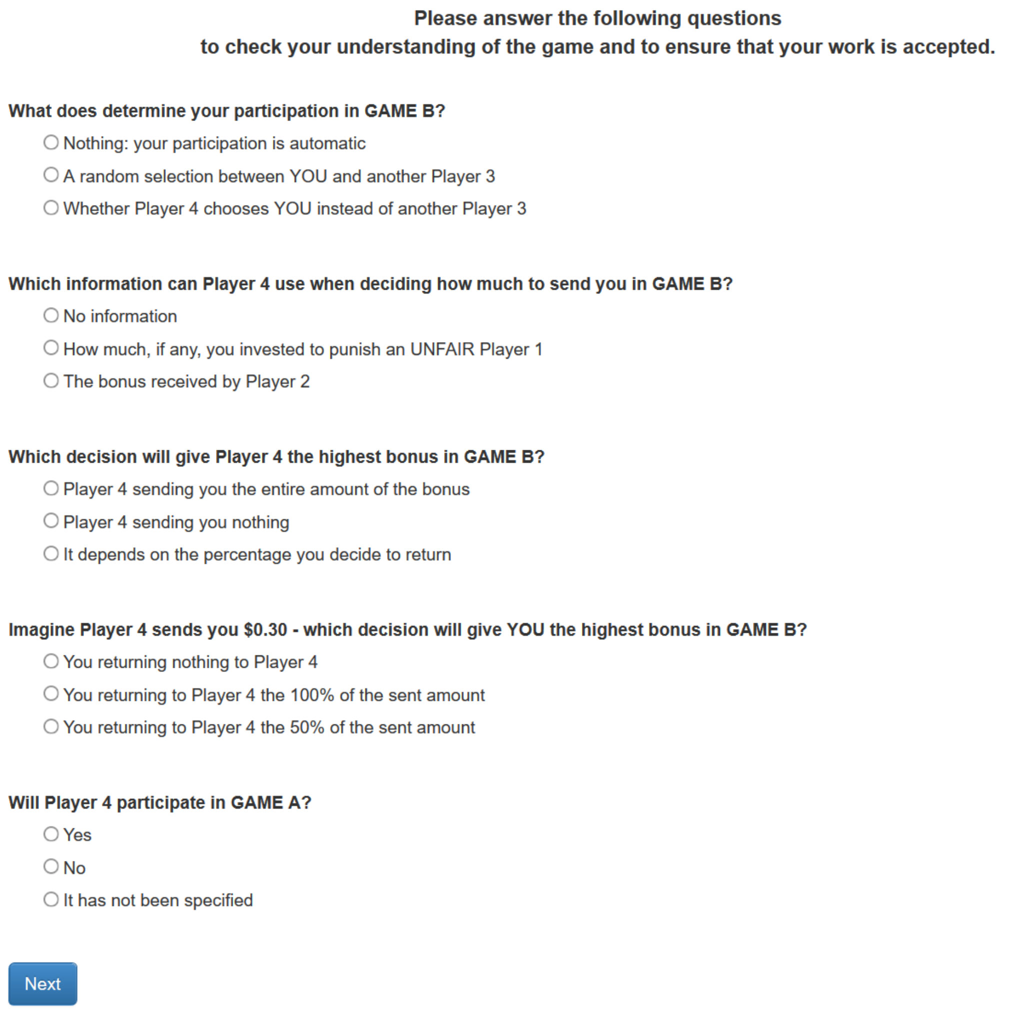
REMEMBER: to take part in this game, **you will have to be CHOSEN by a NEW worker**, who will be Player 4.
Player 4 will choose between you and another Player 3, who played an identical GAME A.
Player 4 WILL know how you and the other Player 3 behaved in your respective GAME A.

Thus Player 4 can choose whom to interact with based on:

- how much, if any, you invested to punish an UNFAIR Player 1;
- how much, if any, the other Player 3 invested to punish another UNFAIR Player 1;
- (Player 4 also knows that for each cent invested in punishing, Players 1s' bonuses are reduced by $0.01).

Furthermore, in case you are chosen, Player 4 can decide how much to send you based on how much, if any, you invested to punish an UNFAIR Player 1.

NOTE: In the *Help condition*, the decision available to Player 3 was framed as "help a Player 2 who RECEIVED an UNFAIR SHARE." The same reference to the cost associated with the decision was made here. In the *Random conditions*, the partner choice protocol was substituted by a random process as follow: "you will have to be RANDOMLY selected to interact with a NEW worker" (the rest of the instructions was modified accordingly and using the same wording). In the *Anonymous condition*, it was specified that "Player 4 will NOT participate in GAME A and will NOT know how you behaved in GAME A."



Please answer the following questions
to check your understanding of the game and to ensure that your work is accepted.

What does determine your participation in GAME B?

○ Nothing: your participation is automatic
○ A random selection between YOU and another Player 3
○ Whether Player 4 chooses YOU instead of another Player 3

Which information can Player 4 use when deciding how much to send you in GAME B?

○ No information
○ How much, if any, you invested to punish an UNFAIR Player 1
○ The bonus received by Player 2

Which decision will give Player 4 the highest bonus in GAME B?

○ Player 4 sending you the entire amount of the bonus
○ Player 4 sending you nothing
○ It depends on the percentage you decide to return

Imagine Player 4 sends you $0.30 - which decision will give YOU the highest bonus in GAME B?

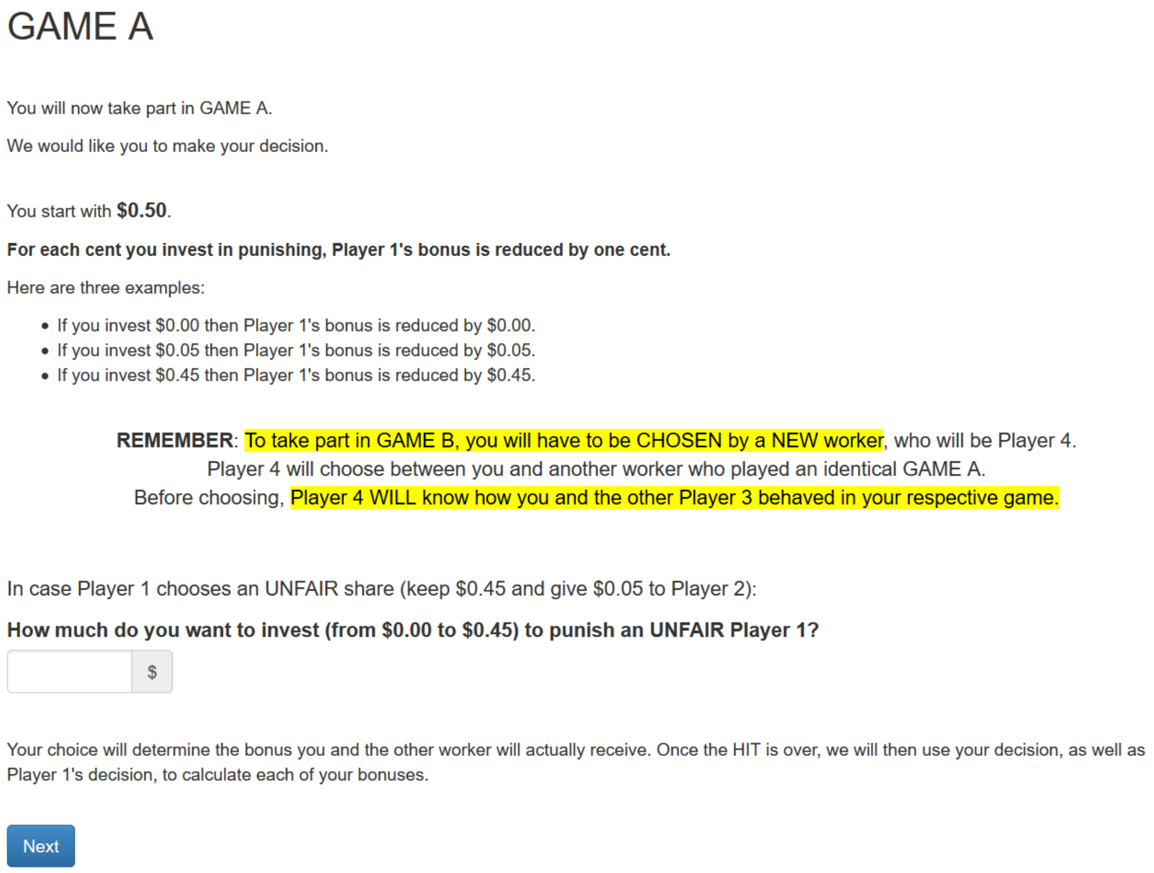
○ You returning nothing to Player 4
○ You returning to Player 4 the 100% of the sent amount
○ You returning to Player 4 the 50% of the sent amount

Will Player 4 participate in GAME A?

○ Yes
○ No
○ It has not been specified

Correct answers: 1) (*Choice condition*) "Whether Player 4 chooses YOU instead of another Player 3", (*Random condition*) "A random selection between YOU and another Player 3"; 2) (*Knowledge condition*) "How much, if any, you invested to punish an UNFAIR Player 1", (*Anonymous condition*) "No information"; 3) "It depends on the percentage you decide to return"; 4) "You returning nothing to Player 4"; 5) "No".

Third-Parties then made their decision in GAME A on the following page:



GAME A

You will now take part in GAME A.

We would like you to make your decision.

You start with **$0.50**.

For each cent you invest in punishing, Player 1's bonus is reduced by one cent.

Here are three examples:

- If you invest $0.00 then Player 1's bonus is reduced by $0.00.
- If you invest $0.05 then Player 1's bonus is reduced by $0.05.
- If you invest $0.45 then Player 1's bonus is reduced by $0.45.

REMEMBER: To take part in GAME B, you will have to be CHOSEN by a NEW worker, who will be Player 4.
Player 4 will choose between you and another worker who played an identical GAME A.
Before choosing, Player 4 WILL know how you and the other Player 3 behaved in your respective game.

In case Player 1 chooses an UNFAIR share (keep $0.45 and give $0.05 to Player 2):

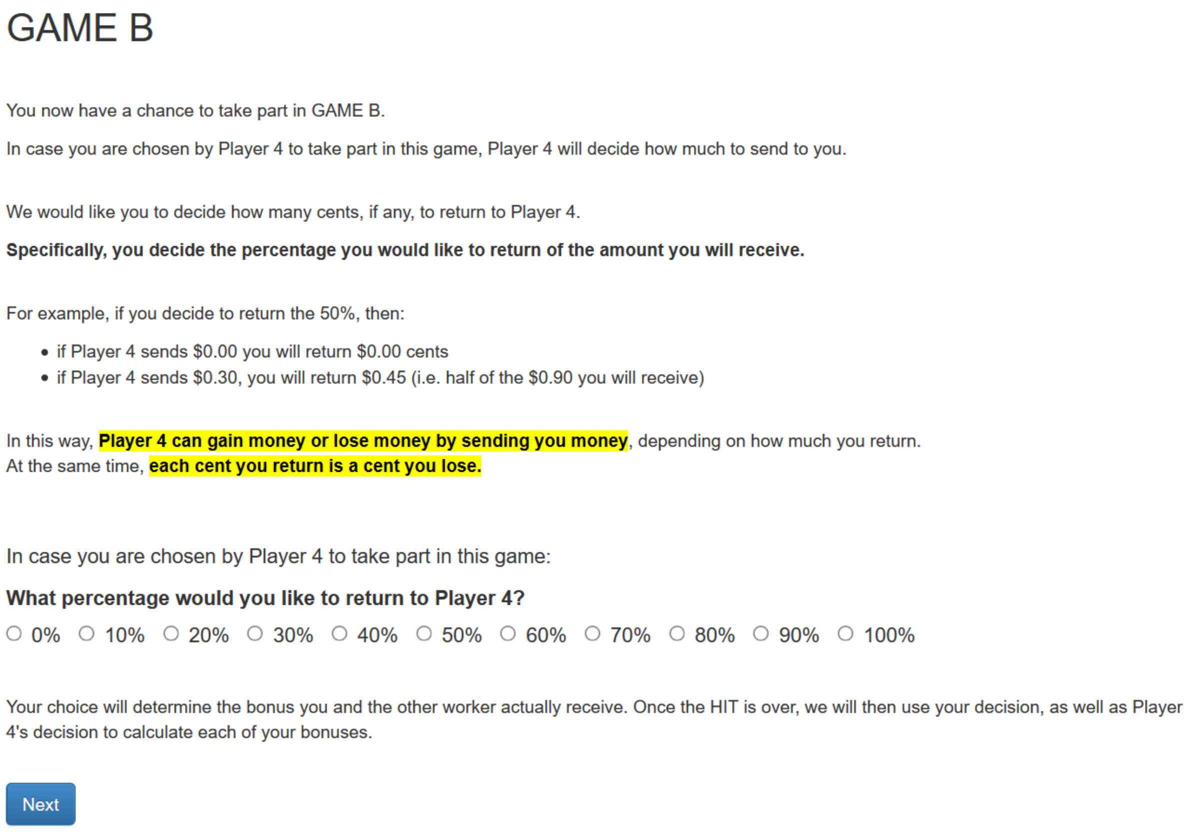
How much do you want to invest (from $0.00 to $0.45) to punish an UNFAIR Player 1?

Your choice will determine the bonus you and the other worker will actually receive. Once the HIT is over, we will then use your decision, as well as Player 1's decision, to calculate each of your bonuses.

NOTE: In the *Help condition*, the cost associated with the decision available to Player 3 was framed as “For each cent you invest in helping, Player 2’s bonus is increased by one cent” (the rest of the instructions was modified accordingly and using the same wording). In the *Random conditions*, the partner choice protocol was substituted by a random process as follow: “To take part in GAME B, you will have to be RANDOMLY selected to interact with a NEW worker”. In the *Anonymous condition*, it was specified that “In GAME B, Player 4 will NOT know how you behaved in this game”.

Next, Third-Parties decided how much to return (in percentage) to Player 4 in GAME B. They were also reminded of the ex-post matching protocol used to calculate this part of their bonus.

The decision was made on the following page:



GAME B

You now have a chance to take part in GAME B.

In case you are chosen by Player 4 to take part in this game, Player 4 will decide how much to send to you.

We would like you to decide how many cents, if any, to return to Player 4.

Specifically, you decide the percentage you would like to return of the amount you will receive.

For example, if you decide to return the 50%, then:

- if Player 4 sends $0.00 you will return $0.00 cents
- if Player 4 sends $0.30, you will return $0.45 (i.e. half of the $0.90 you will receive)

In this way, **Player 4 can gain money or lose money by sending you money**, depending on how much you return.
At the same time, **each cent you return is a cent you lose.**

In case you are chosen by Player 4 to take part in this game:

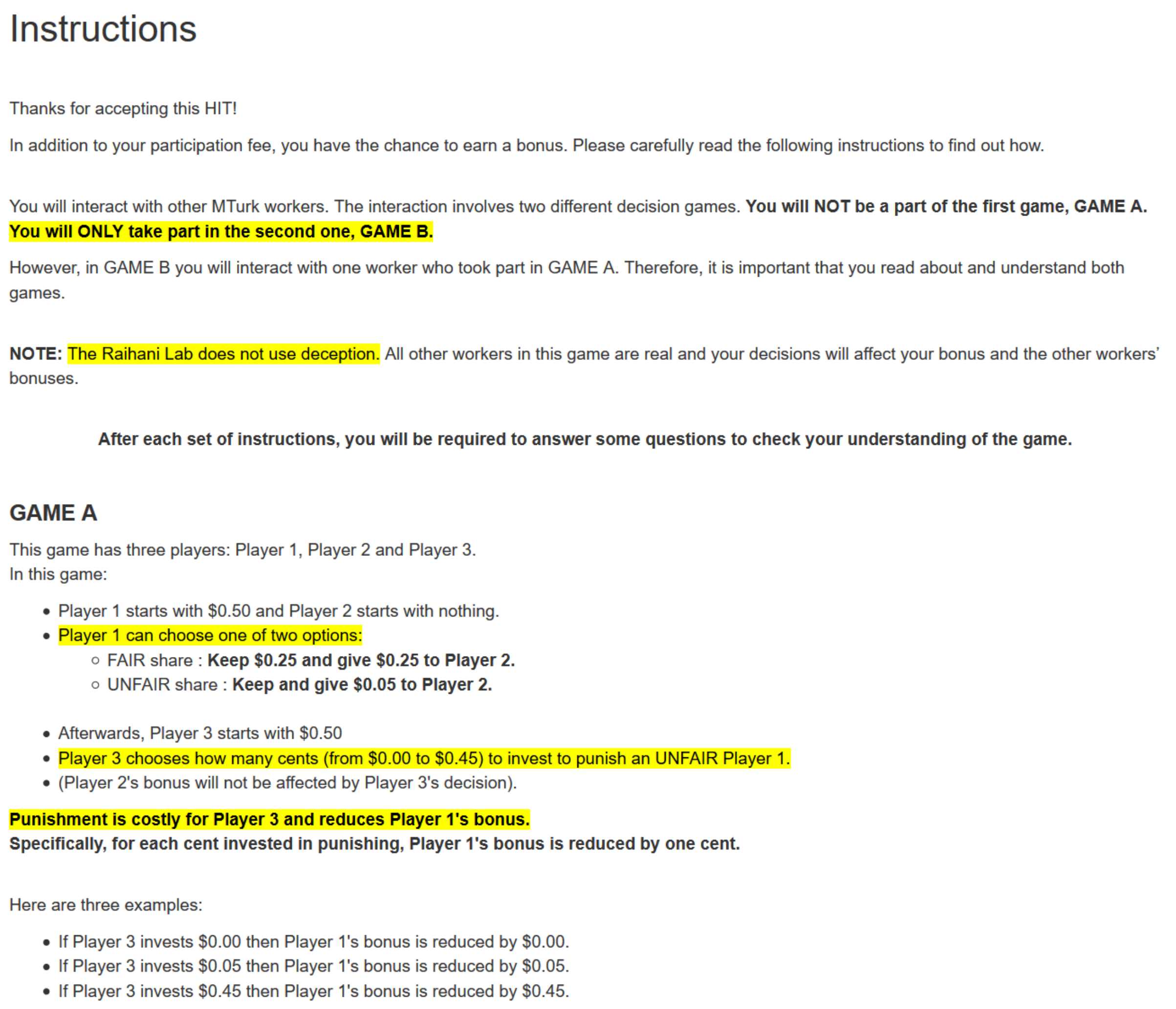
What percentage would you like to return to Player 4?

○ 0% ○ 10% ○ 20% ○ 30% ○ 40% ○ 50% ○ 60% ○ 70% ○ 80% ○ 90% ○ 100%

Your choice will determine the bonus you and the other worker actually receive. Once the HIT is over, we will then use your decision, as well as Player 4's decision to calculate each of your bonuses.

NOTE: In the *Random condition*, the partner choice protocol was substituted by a random process as follow: “In case you are selected to take part in this game”.

Bystanders Shown below are the screenshots of the instructions and the comprehension questions presented to Third-Parties. Instructions and comprehension questions were presented on the same page.



Instructions

Thanks for accepting this HIT!

In addition to your participation fee, you have the chance to earn a bonus. Please carefully read the following instructions to find out how.

You will interact with other MTurk workers. The interaction involves two different decision games. **You will NOT be a part of the first game, GAME A. You will ONLY take part in the second one, GAME B.**

However, in GAME B you will interact with one worker who took part in GAME A. Therefore, it is important that you read about and understand both games.

NOTE: The Raihani Lab does not use deception. All other workers in this game are real and your decisions will affect your bonus and the other workers' bonuses.

After each set of instructions, you will be required to answer some questions to check your understanding of the game.

GAME A

This game has three players: Player 1, Player 2 and Player 3.
In this game:

- Player 1 starts with $0.50 and Player 2 starts with nothing.
- Player 1 can choose one of two options:
    - FAIR share : **Keep $0.25 and give $0.25 to Player 2.**
    - UNFAIR share : **Keep and give $0.05 to Player 2.**

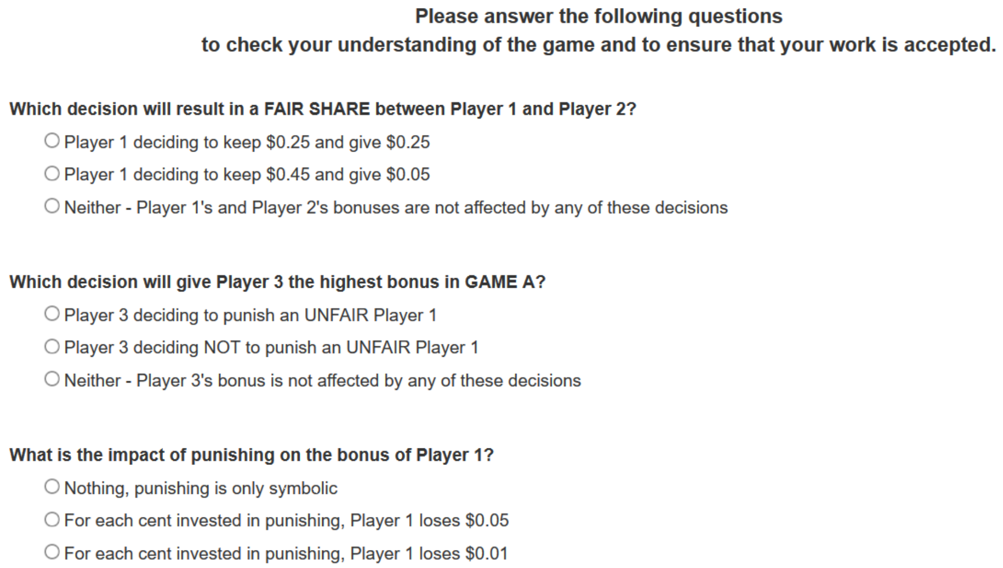
- Afterwards, Player 3 starts with $0.50
- Player 3 chooses how many cents (from $0.00 to $0.45) to invest to punish an UNFAIR Player 1.
- (Player 2's bonus will not be affected by Player 3's decision).

Punishment is costly for Player 3 and reduces Player 1's bonus.
Specifically, for each cent invested in punishing, Player 1's bonus is reduced by one cent.

Here are three examples:

- If Player 3 invests $0.00 then Player 1's bonus is reduced by $0.00.
- If Player 3 invests $0.05 then Player 1's bonus is reduced by $0.05.
- If Player 3 invests $0.45 then Player 1's bonus is reduced by $0.45.

NOTE: In the *Help condition*, the decision available to Player 3 was framed as "help a Player 2 who received an UNFAIR share." The same reference to the cost associated with the decision was made here.



Please answer the following questions
to check your understanding of the game and to ensure that your work is accepted.

Which decision will result in a FAIR SHARE between Player 1 and Player 2?

○ Player 1 deciding to keep $0.25 and give $0.25
○ Player 1 deciding to keep $0.45 and give $0.05
○ Neither - Player 1's and Player 2's bonuses are not affected by any of these decisions

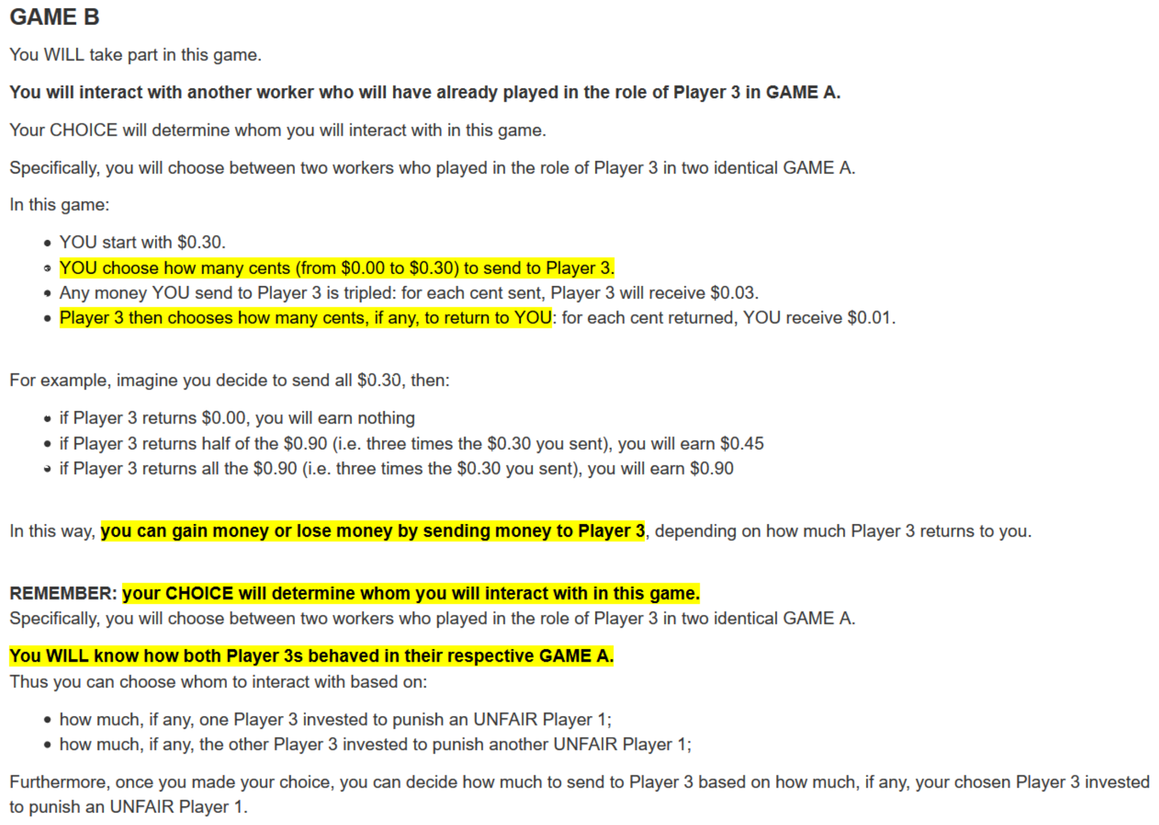
Which decision will give Player 3 the highest bonus in GAME A?

○ Player 3 deciding to punish an UNFAIR Player 1
○ Player 3 deciding NOT to punish an UNFAIR Player 1
○ Neither - Player 3's bonus is not affected by any of these decisions

What is the impact of punishing on the bonus of Player 1?

○ Nothing, punishing is only symbolic
○ For each cent invested in punishing, Player 1 loses $0.05
○ For each cent invested in punishing, Player 1 loses $0.01

Correct answers: 1) “Player 1 deciding to keep $0.25 and give $0.25; 2) “Player 3 deciding NOT to punish an UNFAIR Player 1”; 3) “For each cent invested in punishing, Player 1 loses $0.01”.



GAME B

You WILL take part in this game.

You will interact with another worker who will have already played in the role of Player 3 in GAME A.

Your CHOICE will determine whom you will interact with in this game.

Specifically, you will choose between two workers who played in the role of Player 3 in two identical GAME A.

In this game:

- YOU start with $0.30.
- YOU choose how many cents (from $0.00 to $0.30) to send to Player 3.
- Any money YOU send to Player 3 is tripled: for each cent sent, Player 3 will receive $0.03.
- Player 3 then chooses how many cents, if any, to return to YOU: for each cent returned, YOU receive $0.01.

For example, imagine you decide to send all $0.30, then:

- if Player 3 returns $0.00, you will earn nothing
- if Player 3 returns half of the $0.90 (i.e. three times the $0.30 you sent), you will earn $0.45
- if Player 3 returns all the $0.90 (i.e. three times the $0.30 you sent), you will earn $0.90

In this way, **you can gain money or lose money by sending money to Player 3**, depending on how much Player 3 returns to you.

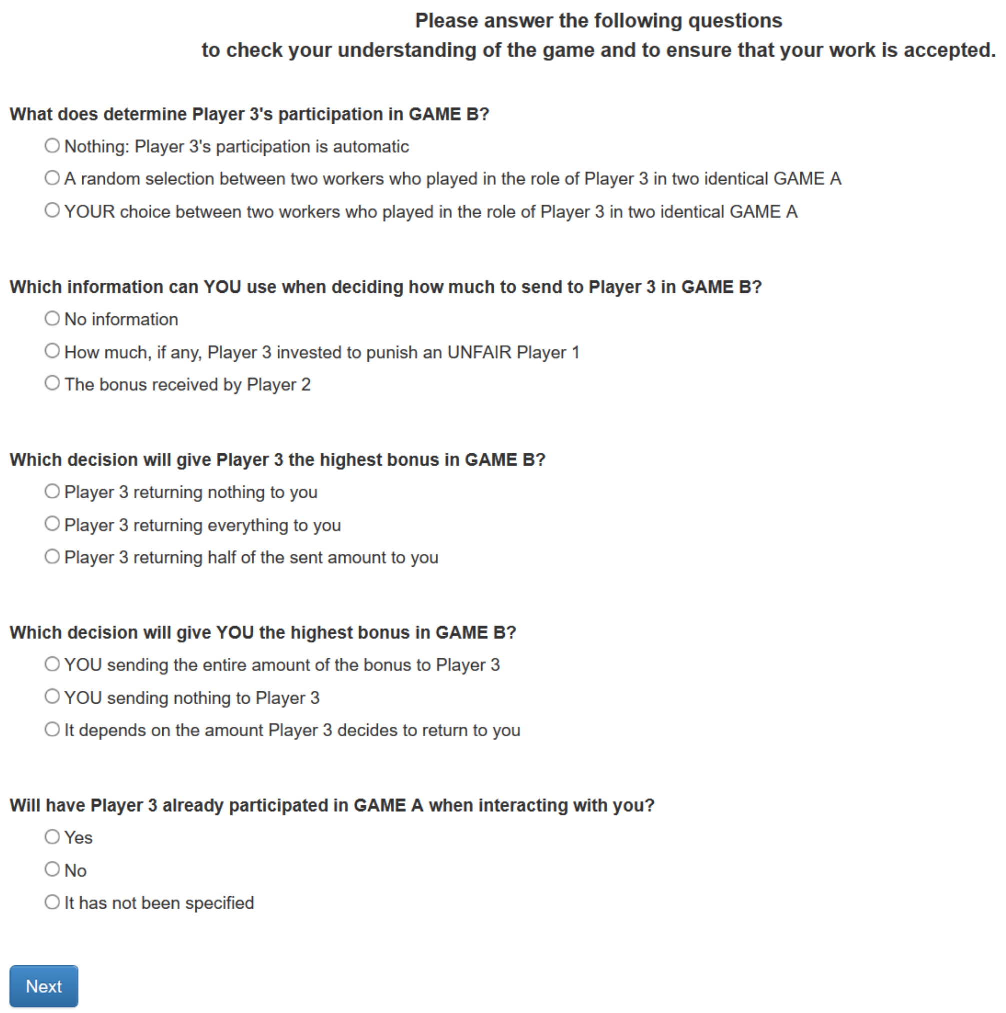
REMEMBER: your CHOICE will determine whom you will interact with in this game.
Specifically, you will choose between two workers who played in the role of Player 3 in two identical GAME A.

You WILL know how both Player 3s behaved in their respective GAME A.
Thus you can choose whom to interact with based on:

- how much, if any, one Player 3 invested to punish an UNFAIR Player 1;
- how much, if any, the other Player 3 invested to punish another UNFAIR Player 1;

Furthermore, once you made your choice, you can decide how much to send to Player 3 based on how much, if any, your chosen Player 3 invested to punish an UNFAIR Player 1.

NOTE: In the *Help condition*, the decision available to Player 3 was framed as “help a Player 2 who RECEIVED an UNFAIR SHARE.” In the *Random conditions*, the partner choice protocol was substituted by a random process as follow: “A RANDOM selection will determine whom you will interact with in this game. Specifically, the random selection will be between two workers who played in the role of Player 3 in two identical GAME A." (the rest of the instructions was modified accordingly and using the same wording). In the *Anonymous condition*, it was specified that "When making your decision, you will NOT know how your assigned Player 3 behaved in GAME A".



Please answer the following questions
to check your understanding of the game and to ensure that your work is accepted.

What does determine Player 3's participation in GAME B?

○ Nothing: Player 3's participation is automatic

○ A random selection between two workers who played in the role of Player 3 in two identical GAME A

○ YOUR choice between two workers who played in the role of Player 3 in two identical GAME A

Which information can YOU use when deciding how much to send to Player 3 in GAME B?

○ No information

○ How much, if any, Player 3 invested to punish an UNFAIR Player 1

○ The bonus received by Player 2

Which decision will give Player 3 the highest bonus in GAME B?

○ Player 3 returning nothing to you

○ Player 3 returning everything to you

○ Player 3 returning half of the sent amount to you

Which decision will give YOU the highest bonus in GAME B?

○ YOU sending the entire amount of the bonus to Player 3

○ YOU sending nothing to Player 3

○ It depends on the amount Player 3 decides to return to you

Will have Player 3 already participated in GAME A when interacting with you?

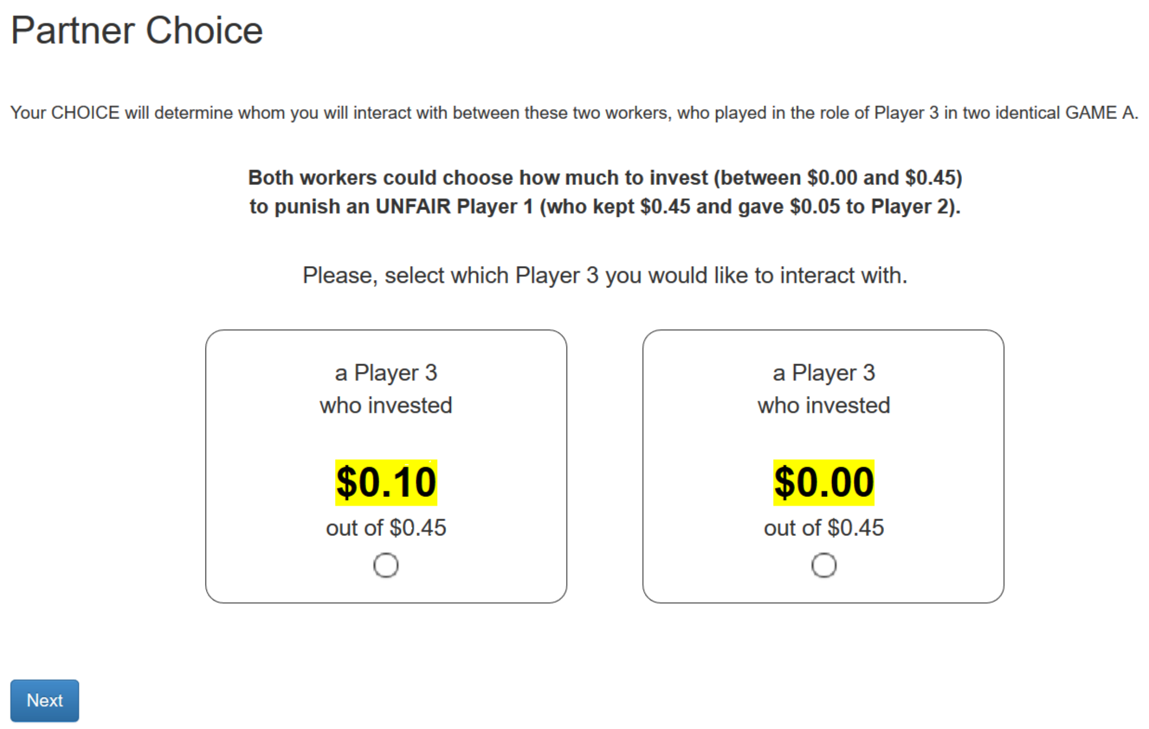
○ Yes

○ No

○ It has not been specified

Correct answers: 1) (*Choice condition*) "YOUR choice between two workers who played in the role of Player 3 in two identical GAME A", (*Random condition*) "A random selection between two workers who played in the role of Player 3 in two identical GAME A"; 2) (*Knowledge condition*) "How much, if any, Player 3 invested to punish an UNFAIR Player 1", (*Anonymous condition*) "No information"; 3) "Player 3 returning nothing to you"; 4) "It depends on the amount Player 3 decides to return to you"; 5) "Yes".

In the *Choice/Knowledge condition*, Bystanders then chose their partner for GAME B on the following page:



Partner Choice

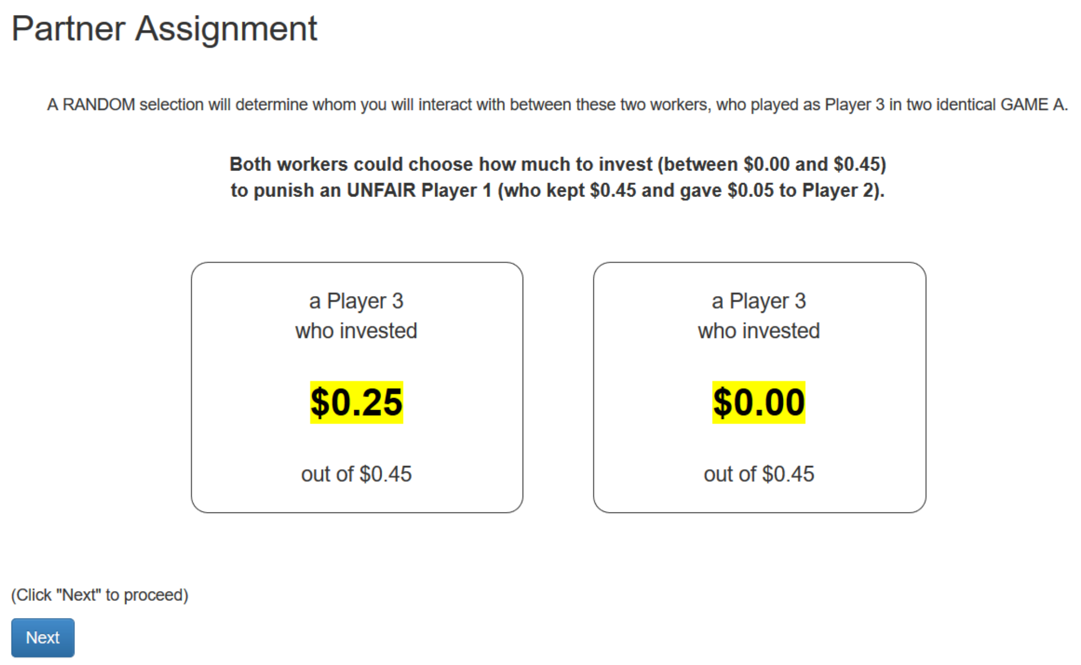
Your CHOICE will determine whom you will interact with between these two workers, who played in the role of Player 3 in two identical GAME A.

Both workers could choose how much to invest (between $0.00 and $0.45) to punish an UNFAIR Player 1 (who kept $0.45 and gave $0.05 to Player 2).

Please, select which Player 3 you would like to interact with.

a Player 3 who invested **$0.10** out of $0.45 ○	a Player 3 who invested **$0.00** out of $0.45 ○

In the *Random/Knowledge condition*, Bystanders then were shown their potential partners for GAME B on the following page:



Partner Assignment

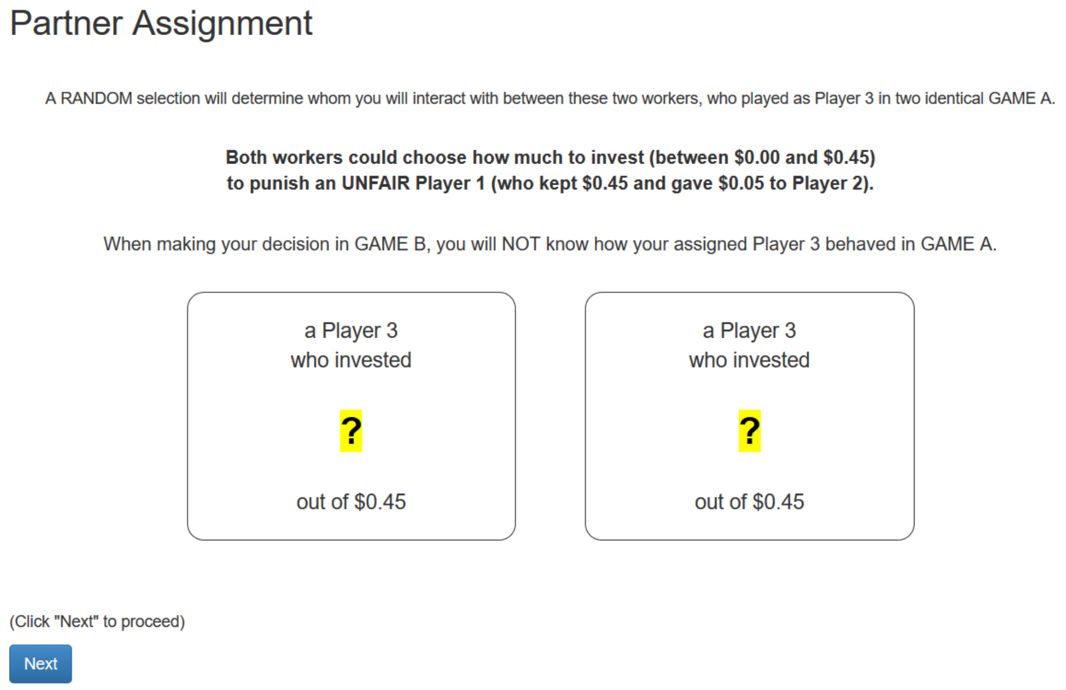
A RANDOM selection will determine whom you will interact with between these two workers, who played as Player 3 in two identical GAME A.

Both workers could choose how much to invest (between $0.00 and $0.45) to punish an UNFAIR Player 1 (who kept $0.45 and gave $0.05 to Player 2).

a Player 3 who invested **$0.25** out of $0.45	a Player 3 who invested **$0.00** out of $0.45

(Click "Next" to proceed)

In the *Random/Anonymous condition*, Bystanders then were shown their potential partners for GAME B on the following page:



Partner Assignment

A RANDOM selection will determine whom you will interact with between these two workers, who played as Player 3 in two identical GAME A.

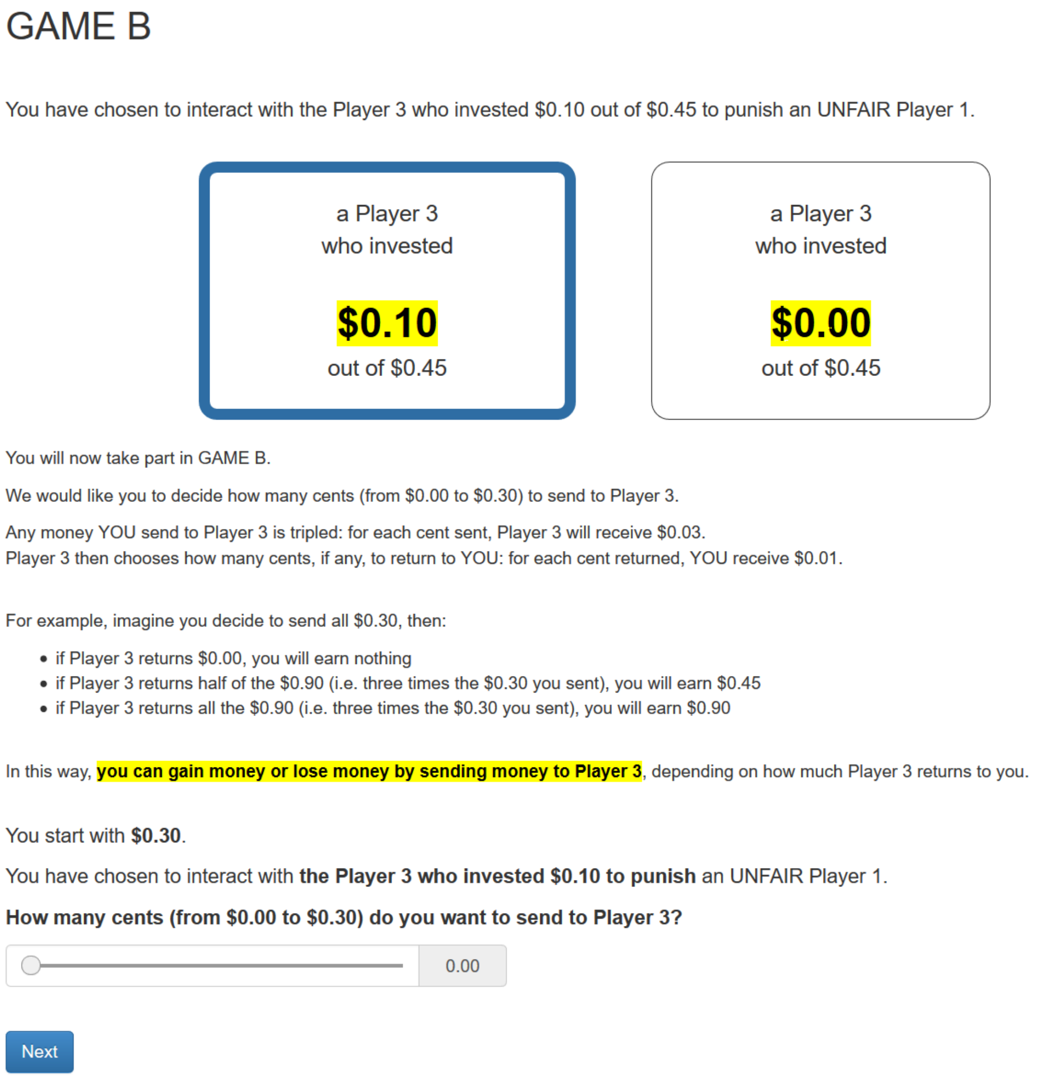
Both workers could choose how much to invest (between $0.00 and $0.45) to punish an UNFAIR Player 1 (who kept $0.45 and gave $0.05 to Player 2).

When making your decision in GAME B, you will NOT know how your assigned Player 3 behaved in GAME A.

a Player 3 who invested ? out of $0.45	a Player 3 who invested ? out of $0.45

(Click "Next" to proceed)

Next, Bystanders were presented on the same page with the result of the partner choice (partner assignment) stage and decided how many cents to send to Player 3 in GAME B. The decision was made on the following page:



GAME B

You have chosen to interact with the Player 3 who invested $0.10 out of $0.45 to punish an UNFAIR Player 1.

a Player 3 who invested **$0.10** out of $0.45

a Player 3 who invested **$0.00** out of $0.45

You will now take part in GAME B.

We would like you to decide how many cents (from $0.00 to $0.30) to send to Player 3.

Any money YOU send to Player 3 is tripled: for each cent sent, Player 3 will receive $0.03.
Player 3 then chooses how many cents, if any, to return to YOU: for each cent returned, YOU receive $0.01.

For example, imagine you decide to send all $0.30, then:

- if Player 3 returns $0.00, you will earn nothing
- if Player 3 returns half of the $0.90 (i.e. three times the $0.30 you sent), you will earn $0.45
- if Player 3 returns all the $0.90 (i.e. three times the $0.30 you sent), you will earn $0.90

In this way, **you can gain money or lose money by sending money to Player 3**, depending on how much Player 3 returns to you.

You start with **$0.30**.

You have chosen to interact with **the Player 3 who invested $0.10 to punish** an UNFAIR Player 1.

How many cents (from $0.00 to $0.30) do you want to send to Player 3?

NOTE: In the *Help condition*, the decision available to Player 3 was framed as “help a Player 2 who RECEIVED an UNFAIR SHARE.” In the *Random/Knowledge condition*, the partner choice protocol was substituted by a random process as follow: “You have been assigned to interact with the Player 3 who invested $0.10 out of $0.45”. In the *Random/Anonymous condition*, the partner choice protocol was substituted by a random process as follow: “You have been assigned to interact with one of two workers who played in the role of Player 3 in two identical GAME A”.

CHAPTER 2

Cooperation and Scale of Interactions: Testing Indirect Reciprocity in a Nested Public Goods Game

Abstract

We present a study design aiming to test the hypothesis that large scale cooperation can be undermined by the presence of cooperative equilibria at smaller scales. When interests across scale of interactions are not aligned, we predict that social agents tend to disregard large-scale cooperation so as to invest their resources in cooperative interactions at smaller scales, as the latter are usually less risky and/or more profitable. We applied this argument to a scenario where cooperation is promoted via indirect reciprocity, which has been often proposed as a key mechanism to explain the evolution and the maintenance of large scale cooperation.

Introduction

By definition, cooperation is individually costly and socially optimal. Explaining how it can be selected by evolution and how it can be encouraged in modern human societies is a major challenge across the biological and social sciences. Mathematical evolutionary models and experimental investigations have identified multiple mechanisms that can, under specific conditions, sustain cooperation (Nowak, 2006), such as direct (Bó, 2005; Trivers, 1971) and indirect reciprocity (Milinski et al., 2002a; Ohtsuki and Iwasa, 2006; Wedekind and Milinski, 2000), multilevel selection (Erev et al., 1993; Puurtinen and Mappes, 2009; Traulsen and Nowak, 2006), spatial selection (Apicella et al., 2012; Rand et al., 2011; Wang et al., 2012) and kin selection (Hamilton, 1964; for a general review, mirroring theoretical and empirical findings, see Rand and Nowak (2013).

Here we focus on indirect reciprocity, which occurs when an individual pays a cost to give a benefit to a partner who had previously done the same toward a third party. Social agents are then incentivized to build a positive social image so as to be rewarded in future interactions. When a social norm is in place that links a positive social image to contributions that benefit a whole group and when the reputation gains are large enough, indirect reciprocity becomes a solution to potentially any collective action problem (Panchanathan and Boyd, 2004). Since its initial formulation (Alexander, 1987), indeed, the efficacy of indirect reciprocity to promote cooperative acts has been confirmed across a wide range of experimental settings (Cuesta et al., 2015; Engelmann and Fischbacher, 2009; Feinberg et al., 2014; Rockenbach and Milinski, 2006; dos Santos et al., 2015; Stanca et al., 2011) and domains (Bateson et al., 2006; DellaVigna et al., 2012; Ernest-Jones et al., 2011; Kandori and Obayashi, 2014; Milinski et al., 2002b). These results have led to define reputation (which indirect reciprocity is based on) as a "universal currency for social interactions" (Milinski, 2016).

The vast majority of the experimental literature on this topic has focused on small groups, where peer monitoring can be assumed to be more feasible and effective. As a group becomes larger, however, the probability for individuals to engage in pairwise interactions with each other is reduced and, most importantly, monitoring is likely to become ineffective or error prone, thus eroding the beneficial impact of indirect reciprocity on cooperation (e.g. Carpenter, 2007; Suzuki and Akiyama, 2005, 2007). Recent studies (e.g. Milinski et al., 2006; Hauser et al., 2016) have addressed these potential shortcomings by investigating indirect reciprocity in extremely large group (e.g. in the context of a global social dilemma, such as climate mitigation) and have found results consistent with the literature on small groups. Local (pairwise) interactions seem then to be instrumental in fostering contributions to collective actions that involve also social actors located outside of an individual's social circle and for whom social feedback is scarce or absent.

Departing from previous approaches, here we hypothesize that neither the size of a group nor the lack of social feedback *per se* play a major role in potentially undermining cooperative equilibria in large-scale interactions. We note that as groups become larger, the possibility of forming coalition and subgroups based on shared interests increases. Plausibly, therefore, when smaller levels of interactions present conditions that are more favorable to the emergence of a cooperative equilibrium, cooperation at the larger scale may collapse. From this perspective, it becomes crucial to investigate whether contributions to large scale collective actions are undermined by the opportunity to take part to collective actions at a smaller scale (assuming misalignment of interests across scales).

In the case of indirect reciprocity, virtually all previous studies have focused on settings wherein only a single opportunity for reputation building was considered (e.g. charity donation, climate protection, a public good shared within a small or a large group). In more realistic settings, however, people are presented with different opportunities of manipulating their social image and social agents can, in principle, assign different strategic values to different opportunities, selecting the most advantageous in terms of personal gain. Exploring the effect of multiple reputation opportunities seems particularly relevant when considering large-scale collective action dilemmas. In this context, a large-scale public good can be juxtaposed with a public good that admits partial exclusion based on location and is not-excludable only at the local level (Blackwell and McKee, 2003; Fellner and Lünser, 2014). Evidence from the field, where multiple scales of interaction and opportunities for reputation building naturally occur, seems anyway to confirm the effectiveness of reputation incentives in fostering contributions to large-scale public goods (e.g. Yoeli et al., 2013; Rogers et al., 2016). We note, however, that those field experiments did not directly tested whether contributors were actually rewarded by peers; furthermore, they did not measure participants' behavior in repeated interactions, where strategies (at both the individual and aggregate level) are more likely to respond to the structural incentive of the interaction setting.

Methods

The experiment is designed to test whether smaller scales of (group) interactions can undermine cooperative equilibria at larger scale of interactions in a setting where cooperation is promoted via indirect reciprocity. We consider a basic scenario represented by a Nested Public Goods Game, where interactions occur at the level of a smaller group of 3 players and at the level of a larger group of 9 players. The game is nested in the sense that the larger group comprises the smaller groups within itself. This implies that each participant (hereafter, players) is member of the large group and of one of the smaller group *at the same time*. We then test under which conditions of social efficiency and peer monitoring indirect reciprocity is more likely to sustain large scale cooperation (i.e. contributions to the large group PG). Crucially, in our setting, indirect reciprocity is bounded at the smaller scale of interaction.

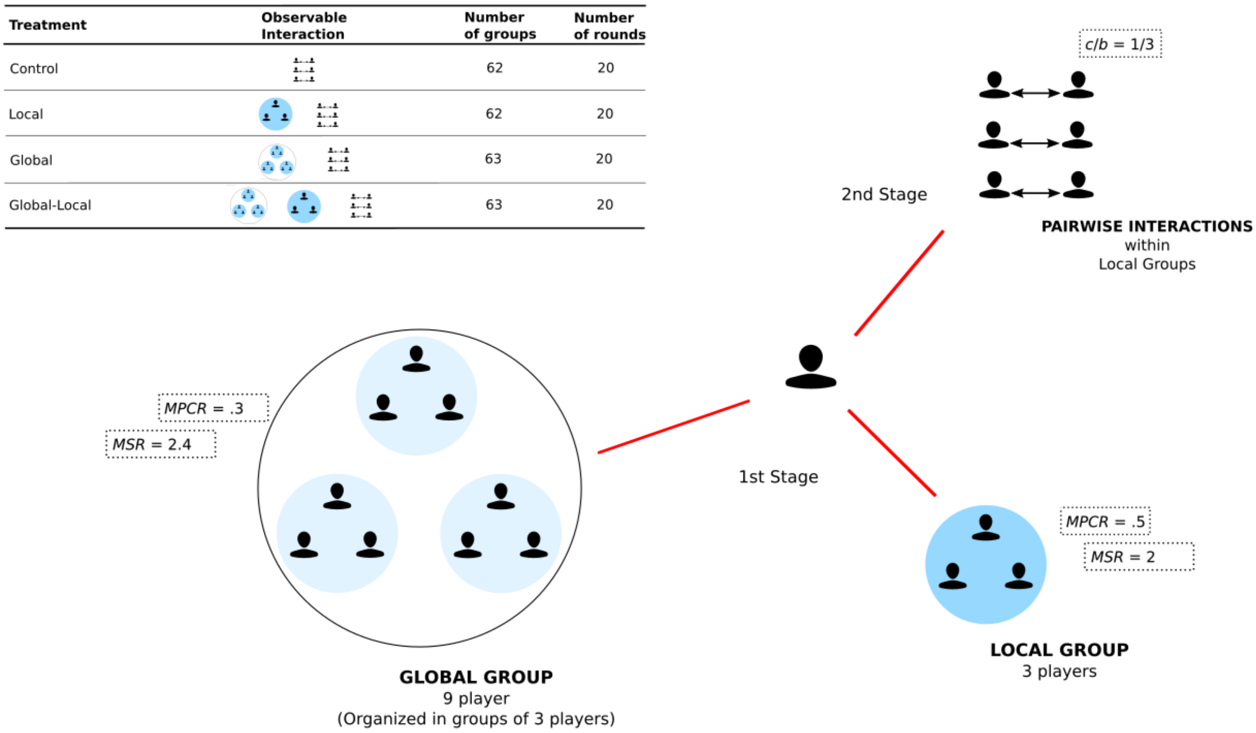
General overview

Overall, the experiment comprises 20 rounds, each including two decision tasks (stages), implemented in the following order:

- Stage A, consisting of the nested Public Goods Game;
- Stage B, consisting of the set of pairwise interactions between the members affiliated to the same small group, modeled as a dichotomous Prisoner Dilemma.

The two versions of the PGG nested in Stage A differ with respect to three components:

- group size;
- Marginal Social Return (MSR), which is always higher in the large PG
- Degree of peer monitoring (i.e. social feedback), which is always bounded within the smaller group.



Treatment	Observable Interaction	Number of groups	Number of rounds
Control		62	20
Local		62	20
Global		63	20
Global-Local		63	20

Figure 1. Illustration of the general experimental setup. In each round, participant first take part in a nested PGG, where they have to decide how to split their endowment between a private account, the large group PG and the small group PG. Thereafter, each participant plays a PD with each member of the same small group. Pairwise interactions and social feedback are always bounded at the small group level. Degree of observability and MPCR of the Large PG are manipulated as treatments variables (in the figure, only the condition with lower MPCR is shown).

An important consequence of the last point is that players are provided with complete information about the *structure* of the game, but limited information about its *dynamic*:

- they know all the rules and parameters of the games;
- they can never observe the behavior of players that are member of the same large group but members of different small subgroups.

The overall structure of the game, therefore, simulates imperfect peer monitoring and limited range of direct social interactions, as it is reasonable to assume they should be in a large scale social dilemma.

In Stage A, players receive an endowment of 20 Experimental Unit (EU) and take part in the Nested Public Goods Game. The game is played simultaneously by all players belonging to the same large group. Each player decides how much keep in a personal account, how much to contribute to the small group PG and how much to the large PG. Next, after receiving a private feedback on their earning from the two PGs, all players in the same large group take part in Stage B. Here they engage in a set of pairwise interactions with all the other players in the same small group (direct interactions with payers in other small groups are precluded). The pairwise interactions are modelled as dichotomous Prisoner Dilemmas, where defecting is always the strictly dominant strategy. Across treatments, we vary: *(i)* what social feedback (None versus Local versus Global versus Local-Global) that players receive in Stage B; *(ii)* the Marginal Social Return (MSR) of the large group PG (Low versus Medium versus High). We then implement a 4 X 3 between-subjects full factorial design, resulting in 12 experimental cells. Next, we describe the parameters of the games and the payoff specification of the Nested PGG.

Large Public Good Game The group comprises 9 players, MSR varies according to the experimental conditions between 2.4, 2.8 and 3.2 (resulting in a MPCR of 0.3, 0.4 and 0.5, respectively).

Small Public Good Game The group comprises 3 players, MSR is 1.5 (resulting in a MPCR of 0.5) and is fixed across all experimental conditions.

The individual payoff of each player in Stage A is specified as follow:

$$\pi_i = d\, -\, l_i - g_i + \alpha \sum_{j=1}^{n} l_j + \beta \sum_{j=1}^{N} g_j$$

where:

- n = number of local players
- N = number of global players
- d = endowment
- l = contribution to the Local PG
- g = contribution to the Global PG
- α = MPCR of the small PG
- β = MPCR of the large PG
- $\boldsymbol{\alpha n} < \boldsymbol{\beta N}$ (i.e. the social efficiency of the Large PG is always greater than the social efficiency of the Small PG).

Dichotomous Prisoner Dilemma Set of pairwise interactions among members of each local group. Each player received an endowment of 6 EU and is asked to make a binary decision toward each of the other 2 members of the same local group:

- pay 3 ECU for the other participant to get 9 EU
- pay 0 ECU for the other participant to get 0 EU

When players make their decisions, the social information available to them depends on the treatment (see Figure 1) but is always bounded to the small group the focus player belongs to.

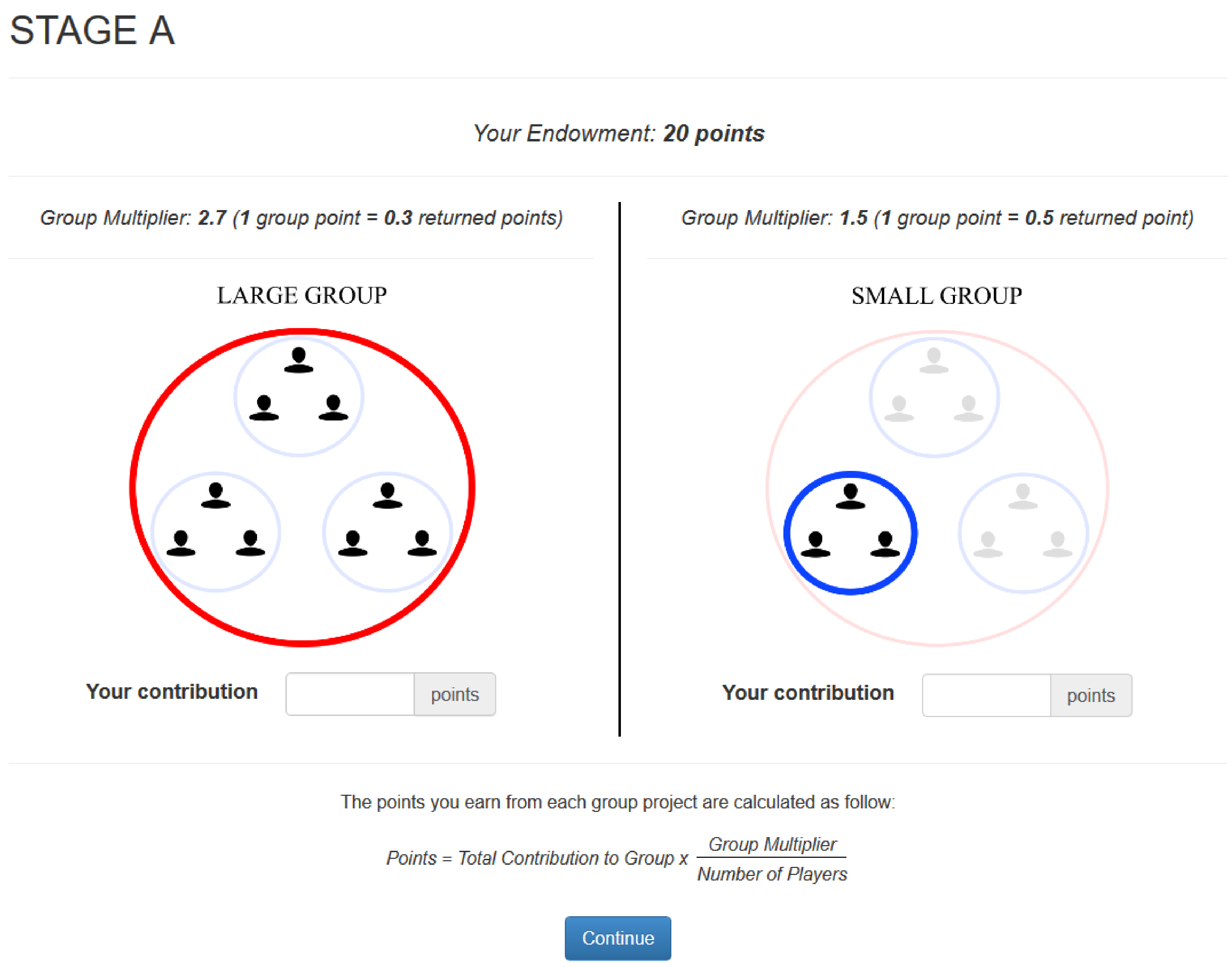
In light of the arguments reported in the introduction, we formulate the following predictions:

- *H1*: When social feedback about contributions to both the large-scale and a small-scale PGs is absent, contributions to both PGs decline over time;
- *H2*: When social feedback is limited to contributions to the small PG, contributions to the small PG are sustained over time, while contributions to the large PG decline;
- *H3*: When social feedback is limited to contributions to the large PG, contributions to the large PG are sustained over time, while contributions to the small PG decline;
- *H4*: When social feedback covers contributions to both PGs, contributions to the small PG are sustained over time, while contributions to the large PG decline;
- *H5*: Increasing the social efficiency of the large PG slows down the collapse of the large PG predicted in H1, H2 and H4.

Instructions and Experimental Setup

In what follows, we present the instructions and (within them) the screenshots of the experiment as it has to be presented to the participants. Specifically, we report the material relative to the Local-Global/Low-MSR condition.

The experimental setup is coded using oTree, a free open-source software for running economic experiments in physical labs, online and in the field (Chen et al., 2016).



STAGE A

Your Endowment: ***20 points***

Group Multiplier: ***2.7*** *(**1** group point =* ***0.3*** *returned points)*

LARGE GROUP

Your contribution ____ points

Group Multiplier: ***1.5*** *(**1** group point =* ***0.5*** *returned point)*

SMALL GROUP

Your contribution ____ points

The points you earn from each group project are calculated as follow:

$$Points = Total\ Contribution\ to\ Group\ x\ \frac{Group\ Multiplier}{Number\ of\ Players}$$

STAGE A

This screen is the normal screen for making your decisions in STAGE A.

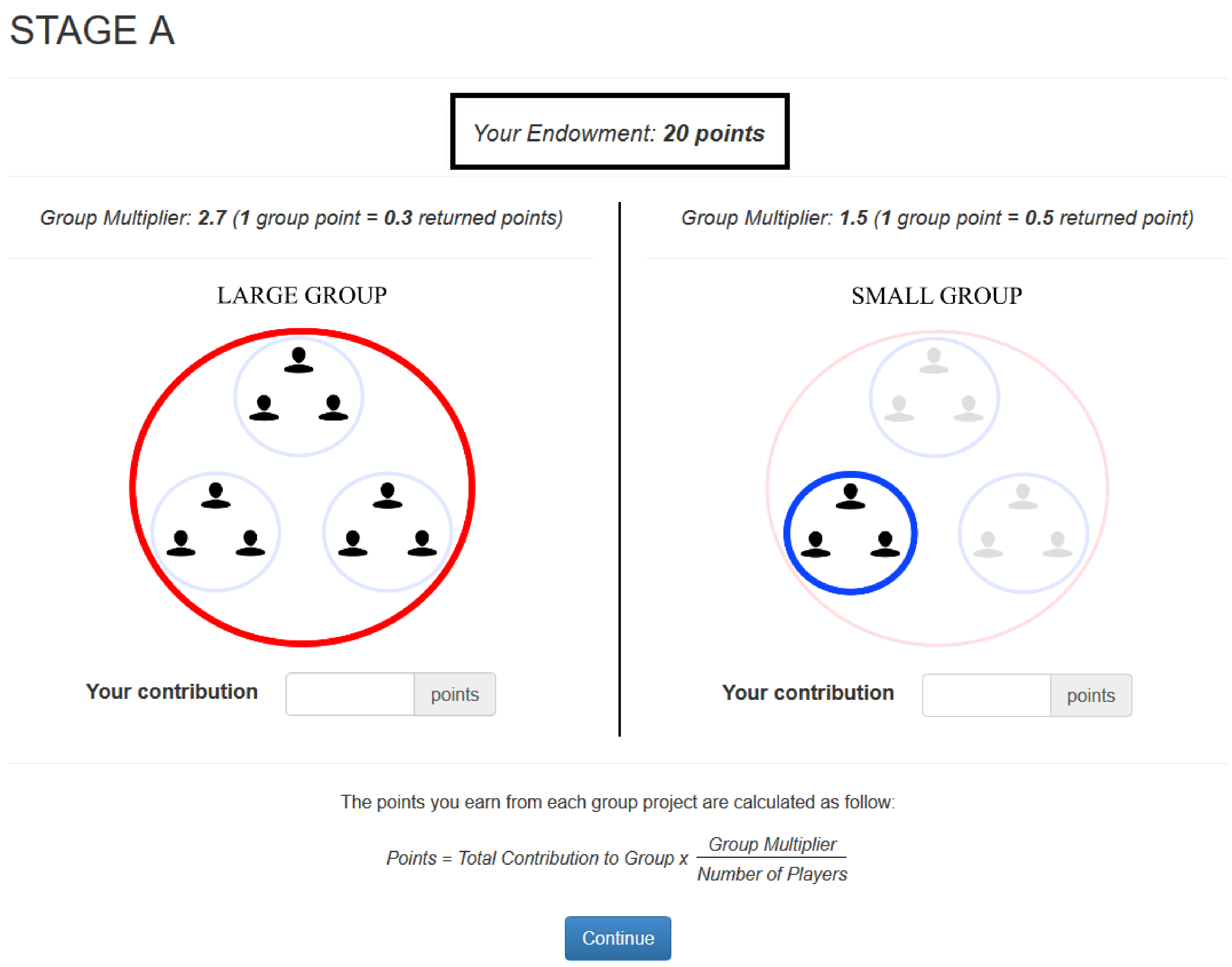
As highlighted in the image, throughout the game you are

member of a SMALL GROUP

of 3 participants, which is

part of a LARGE GROUP

of 9 participants.



STAGE A

*Your Endowment: **20 points***

*Group Multiplier: **2.7** (**1** group point = **0.3** returned points)*

LARGE GROUP

Your contribution [ ] points

*Group Multiplier: **1.5** (**1** group point = **0.5** returned point)*

SMALL GROUP

Your contribution [ ] points

The points you earn from each group project are calculated as follow:

$$\textit{Points} = \textit{Total Contribution to Group} \times \frac{\textit{Group Multiplier}}{\textit{Number of Players}}$$

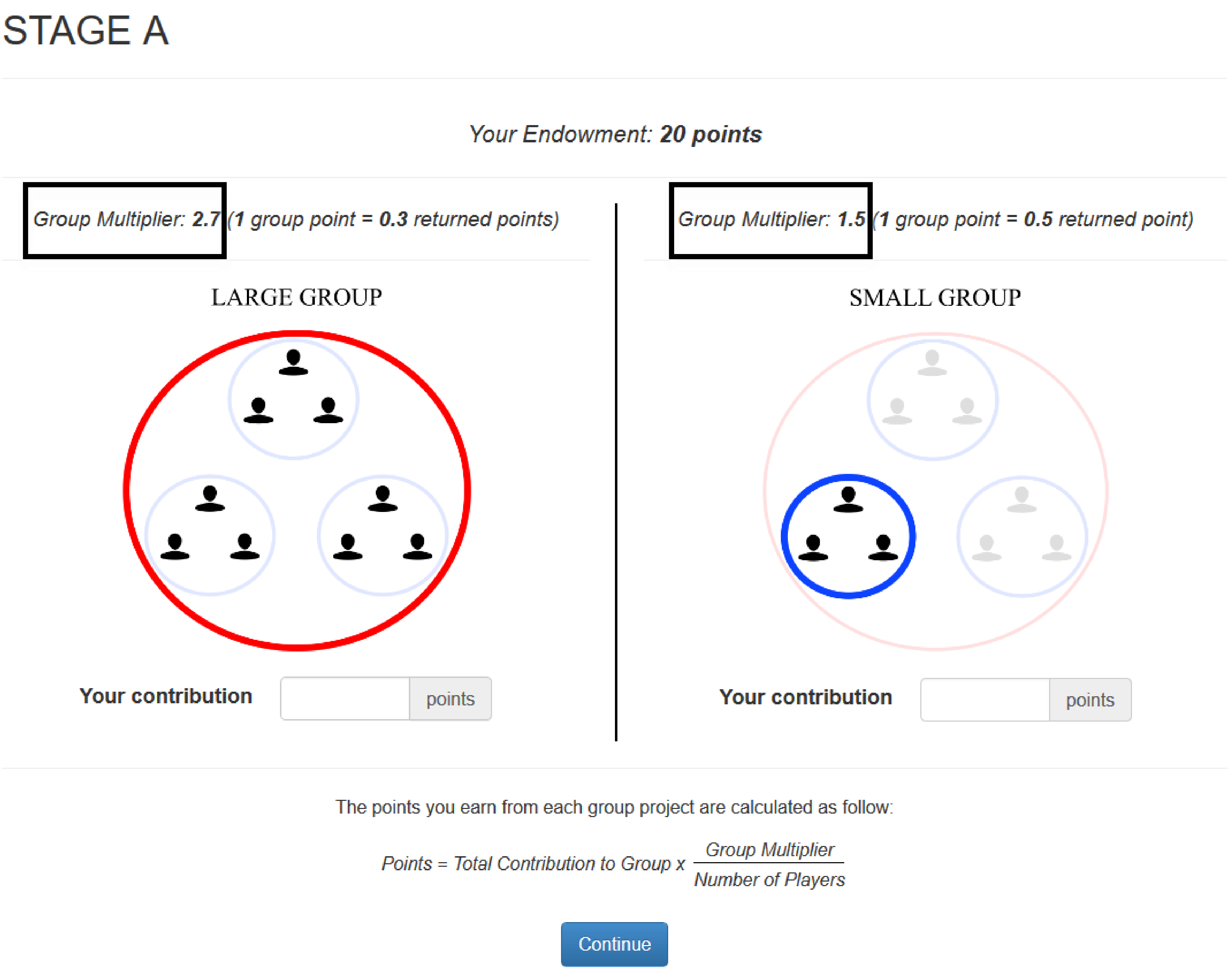
STAGE A

At the beginning of the stage each participant receives an endowment of **20 points**.

YOUR TASK is to decide how to use your endowment. You have to decide how many of the points you want to:

- contribute to the SMALL GROUP project
- contribute to the LARGE GROUP project
- keep for yourself

<u>You **must** make your decision **within 25 seconds**</u>.



STAGE A

Your Endowment: ***20 points***

Group Multiplier: ***2.7*** *(1 group point =* ***0.3*** *returned points)*

LARGE GROUP

Your contribution points

Group Multiplier: ***1.5*** *(1 group point =* ***0.5*** *returned point)*

SMALL GROUP

Your contribution points

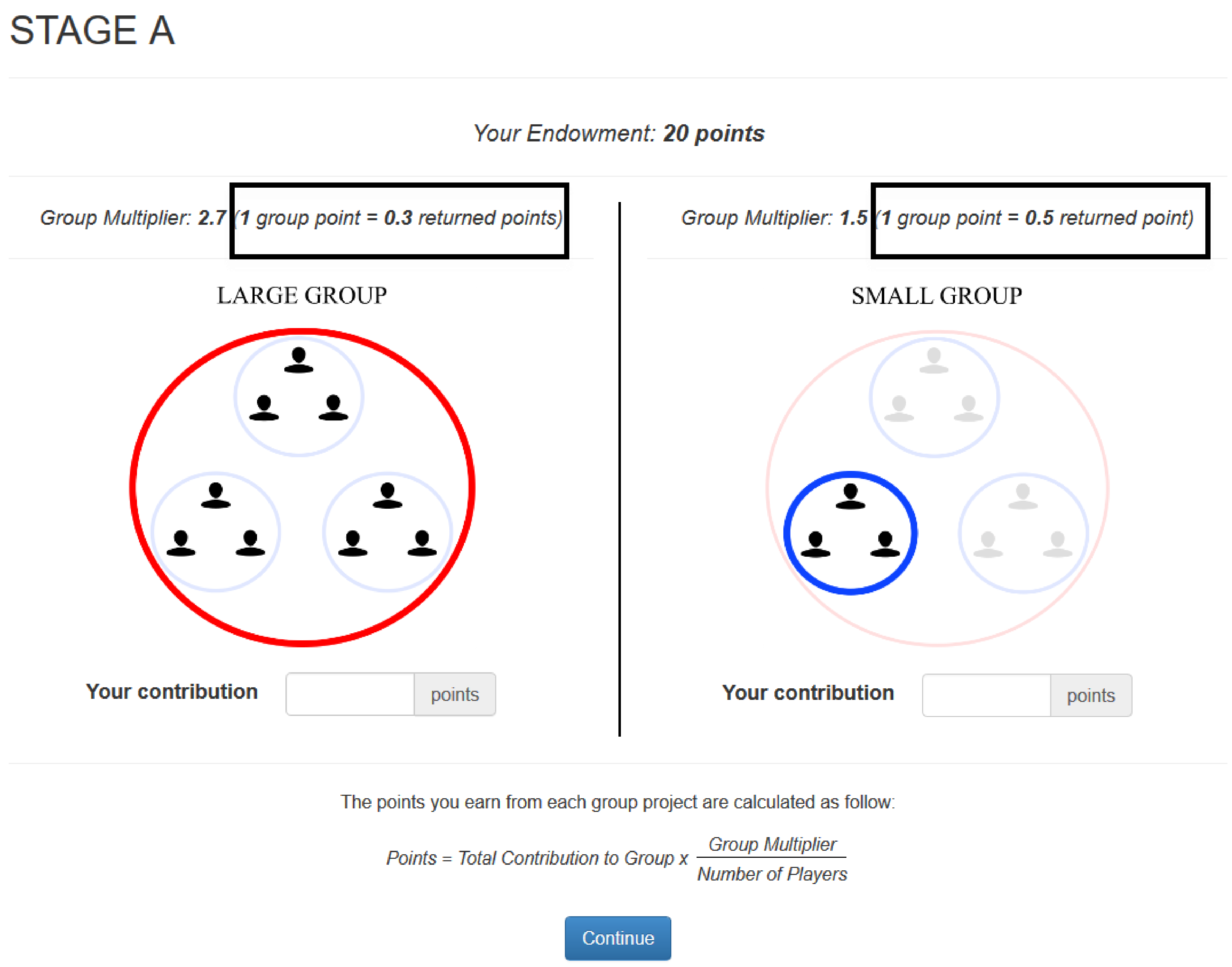
The points you earn from each group project are calculated as follow:

$$\textit{Points} = \textit{Total Contribution to Group} \times \frac{\textit{Group Multiplier}}{\textit{Number of Players}}$$

GROUP MULTIPLIER

When group members allocate points to a group project, the total allocated by the group is multiplied by this number.

This number represents the **gain from the joint project for the group as a whole.**



STAGE A

Your Endowment: ***20 points***

Group Multiplier: ***2.7*** *(****1*** *group point =* ***0.3*** *returned points)*

LARGE GROUP

Your contribution ______ points

Group Multiplier: ***1.5*** *(****1*** *group point =* ***0.5*** *returned point)*

SMALL GROUP

Your contribution ______ points

The points you earn from each group project are calculated as follow:

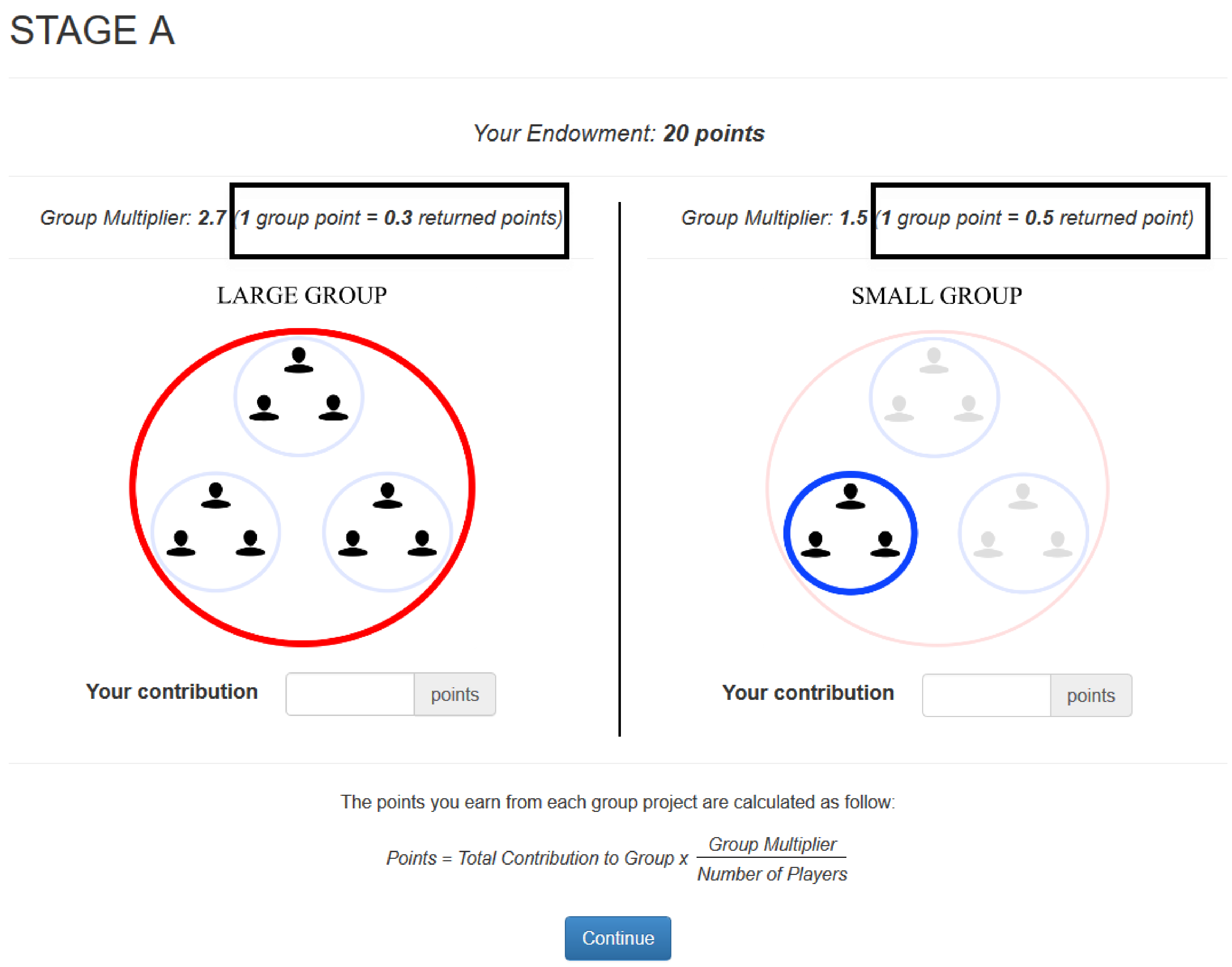
$$\textit{Points} = \textit{Total Contribution to Group} \times \frac{\textit{Group Multiplier}}{\textit{Number of Players}}$$

RETURNED POINTS

The returned points shown in the parenthesis represent how much each point allocated to the joint project is worth to each player.

They equal the total allocated to the joint project, multiplied by the group multiplier **and divided by the number of members in a group**.

This number represents the **personal gain from the joint project for each individual in the group**.



STAGE A

Your Endowment: ***20 points***

Group Multiplier: ***2.7*** *(****1*** *group point =* ***0.3*** *returned points)*

Group Multiplier: ***1.5*** *(****1*** *group point =* ***0.5*** *returned point)*

Your contribution points

Your contribution points

The points you earn from each group project are calculated as follow:

$$\textit{Points = Total Contribution to Group x } \frac{\textit{Group Multiplier}}{\textit{Number of Players}}$$

RETURNED POINTS

In both group projects, the return from each point contribute is smaller than 1.

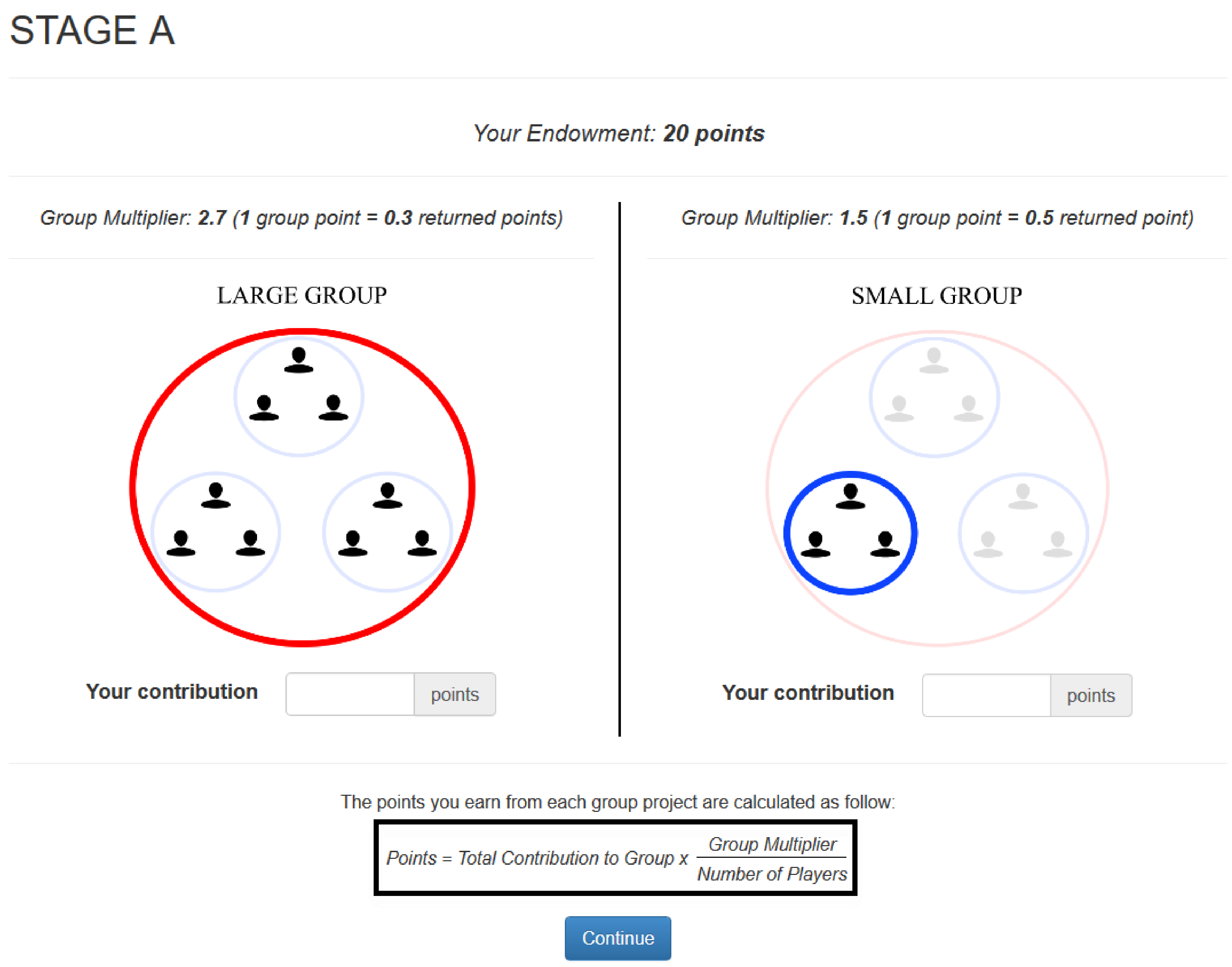
Therefore, contributions are

beneficial for the group

but

individually costly.

Each point you keep for yourself remains as 1 point.



STAGE A

Your Endowment: ***20 points***

Group Multiplier: ***2.7*** *(****1*** *group point =* ***0.3*** *returned points)*

LARGE GROUP

Your contribution points

Group Multiplier: ***1.5*** *(****1*** *group point =* ***0.5*** *returned point)*

SMALL GROUP

Your contribution points

The points you earn from each group project are calculated as follow:

$$\boxed{\textit{Points} = \textit{Total Contribution to Group x} \frac{\textit{Group Multiplier}}{\textit{Number of Players}}}$$

PROFIT IN STAGE A

At the bottom of the page, you can see a reminder of how your profit from each group project is calculated:

the total allocated to the joint project is

multiplied by the group multiplier and

evenly divided by the number of players in a group,

Each player gets the same share of the project, irrespective of her contribution.

PROFIT IN STAGE A

Here you have an EXAMPLE.

Imagine that, for both projects, the sum of the contributions of all group members is 40points.

In this case, **each member of the group**, irrespective of her contribution, gets a profit from each joint project calculated as follow:

SMALL GROUP

40points times **1.5** (Group Multiplier) divided by **3** (number of group members) = **20points**

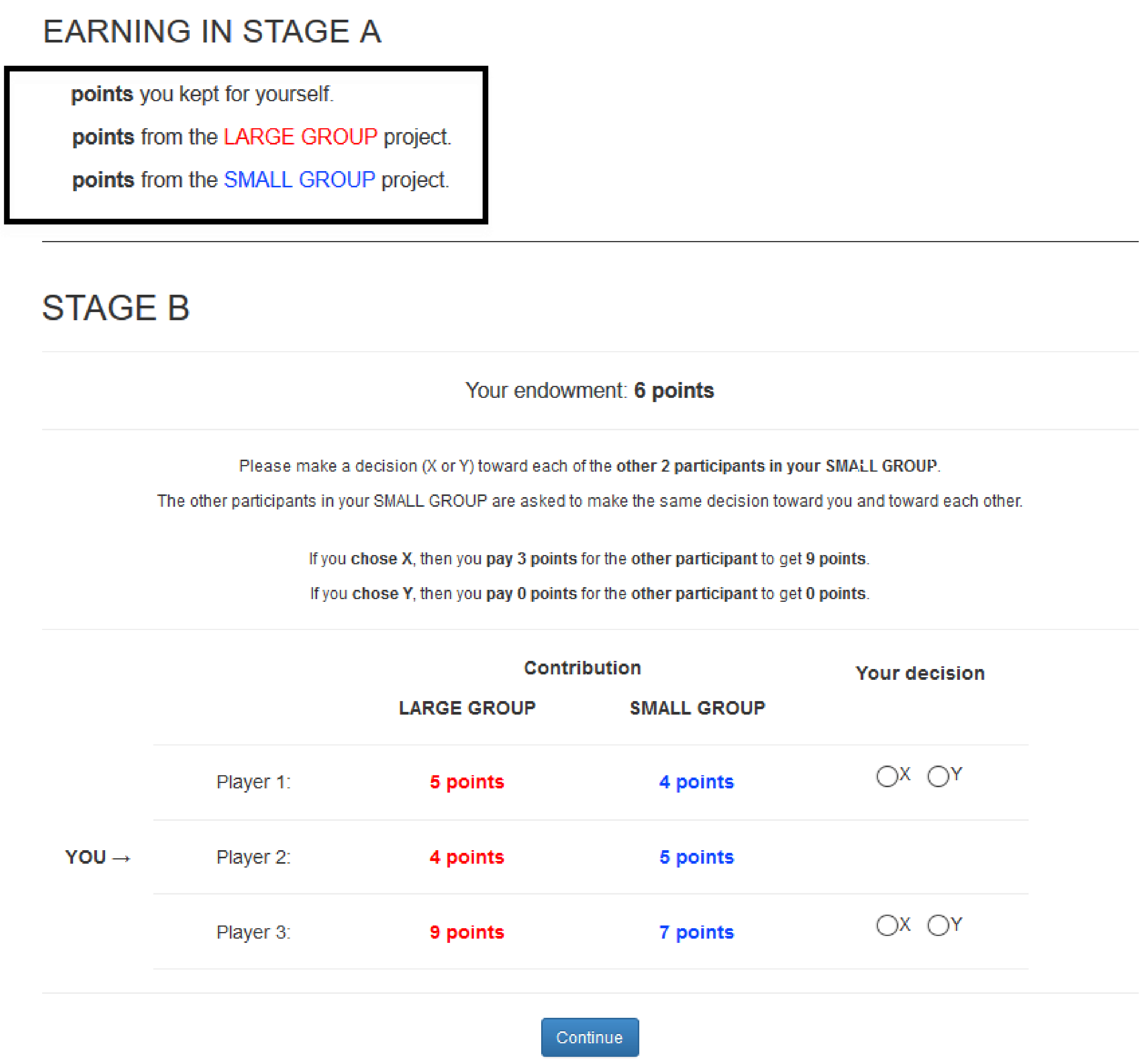
LARGE GROUP

40points times **2.7** (Group Multiplier) divided by **9** (number of group members) = **12points**

TOTAL EARNING IN STAGE A

To summarize, your **total earning** in STAGE A is equal to

Your Endowment − Your Contributions + Your Profit from SMALL GROUP + Your Profit from LARGE GROUP



EARNING IN STAGE A

points you kept for yourself.

points from the LARGE GROUP project.

points from the SMALL GROUP project.

STAGE B

Your endowment: **6 points**

Please make a decision (X or Y) toward each of the **other 2 participants in your SMALL GROUP**.

The other participants in your SMALL GROUP are asked to make the same decision toward you and toward each other.

If you **chose X**, then you **pay 3 points** for the **other participant** to get **9 points**.

If you **chose Y**, then you **pay 0 points** for the **other participant** to get **0 points**.

		Contribution		Your decision
		LARGE GROUP	SMALL GROUP	
	Player 1:	5 points	4 points	○X ○Y
YOU →	Player 2:	4 points	5 points	
	Player 3:	9 points	7 points	○X ○Y

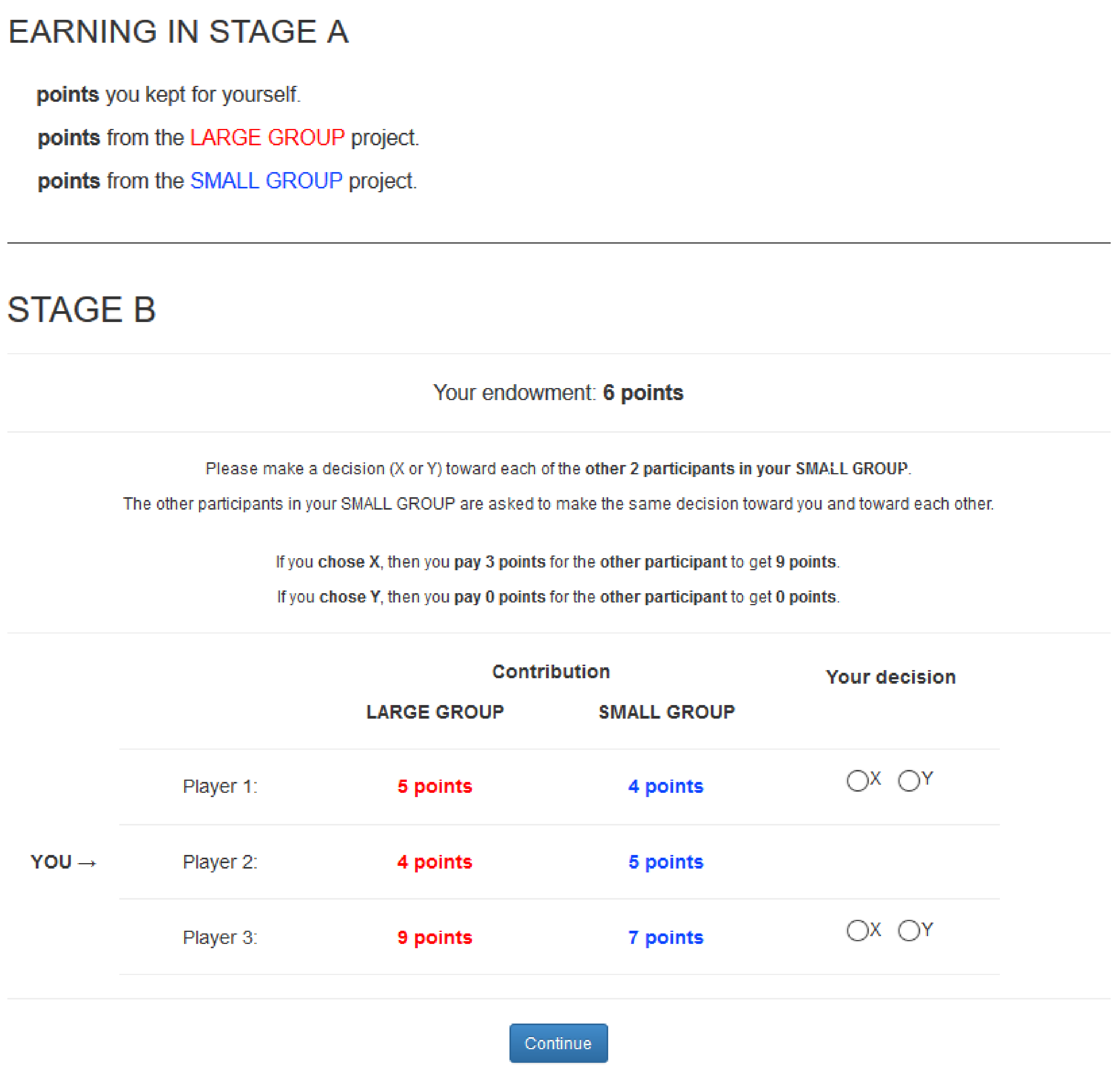
EARNING IN STAGE A

Once all participants have made their decisions, you will be shown details of your earning in STAGE A in the current round.

NOTE

Once you have made your decision in STAGE A, **you might need to wait few seconds** for the other participants to make their decisions as well.

It is very important that you **don’t abandon the page** at any time and stay ready to continue the game.



EARNING IN STAGE A

points you kept for yourself.

points from the LARGE GROUP project.

points from the SMALL GROUP project.

STAGE B

Your endowment: **6 points**

Please make a decision (X or Y) toward each of the **other 2 participants in your SMALL GROUP**.

The other participants in your SMALL GROUP are asked to make the same decision toward you and toward each other.

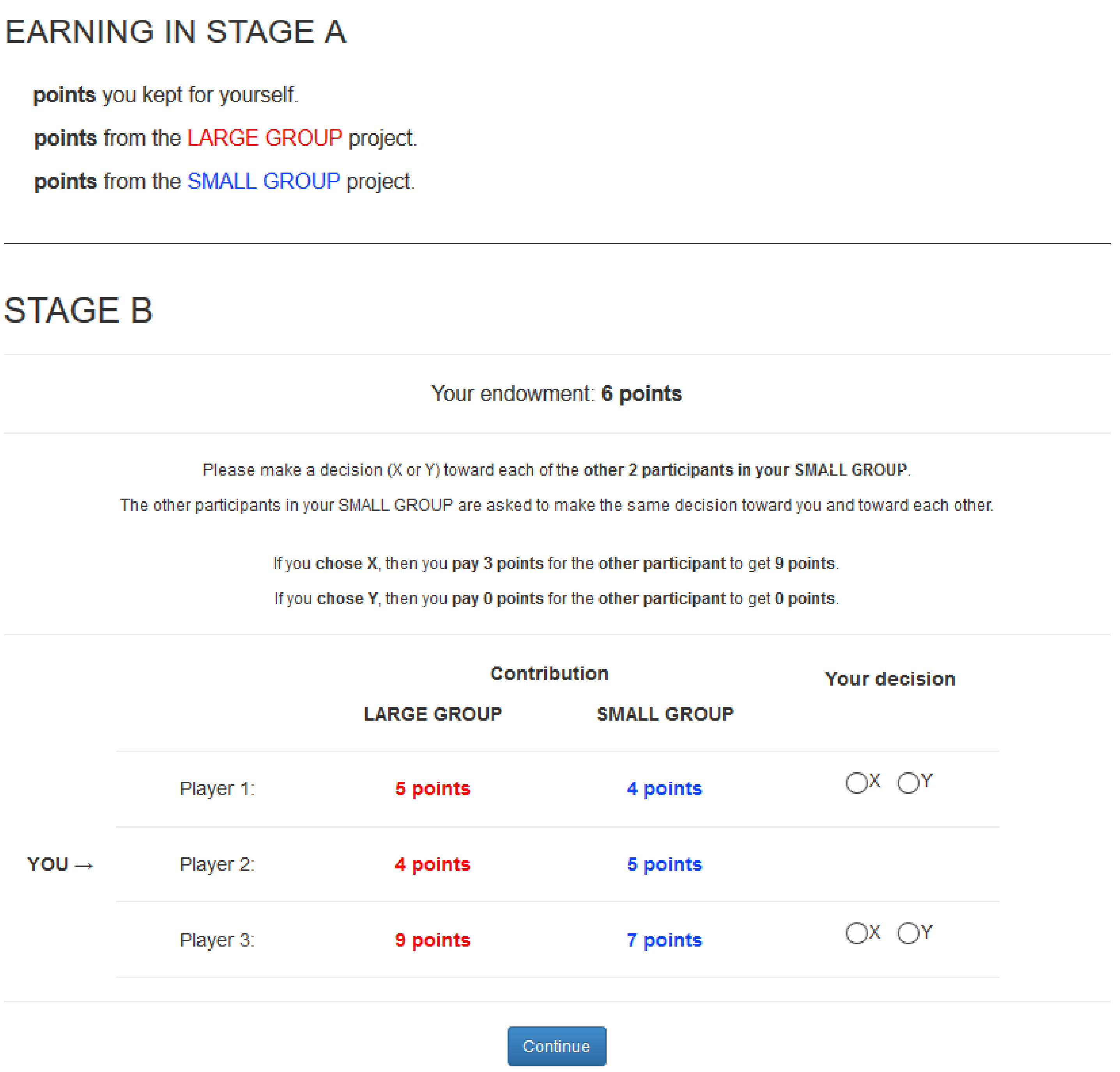
If you **chose X**, then you **pay 3 points** for the **other participant** to get **9 points**.

If you **chose Y**, then you **pay 0 points** for the **other participant** to get **0 points**.

		Contribution		Your decision
		LARGE GROUP	SMALL GROUP	
	Player 1:	5 points	4 points	○X ○Y
YOU →	Player 2:	4 points	5 points	
	Player 3:	9 points	7 points	○X ○Y

STAGE B

This screen is the normal screen for making your decisions in STAGE B.



EARNING IN STAGE A

points you kept for yourself.

points from the LARGE GROUP project.

points from the SMALL GROUP project.

STAGE B

Your endowment: **6 points**

Please make a decision (X or Y) toward each of the **other 2 participants in your SMALL GROUP**.

The other participants in your SMALL GROUP are asked to make the same decision toward you and toward each other.

If you **chose X**, then you **pay 3 points** for the **other participant** to get **9 points**.

If you **chose Y**, then you **pay 0 points** for the **other participant** to get **0 points**.

		Contribution		**Your decision**
		LARGE GROUP	**SMALL GROUP**	
	Player 1:	**5 points**	**4 points**	○X ○Y
YOU →	Player 2:	**4 points**	**5 points**	
	Player 3:	**9 points**	**7 points**	○X ○Y

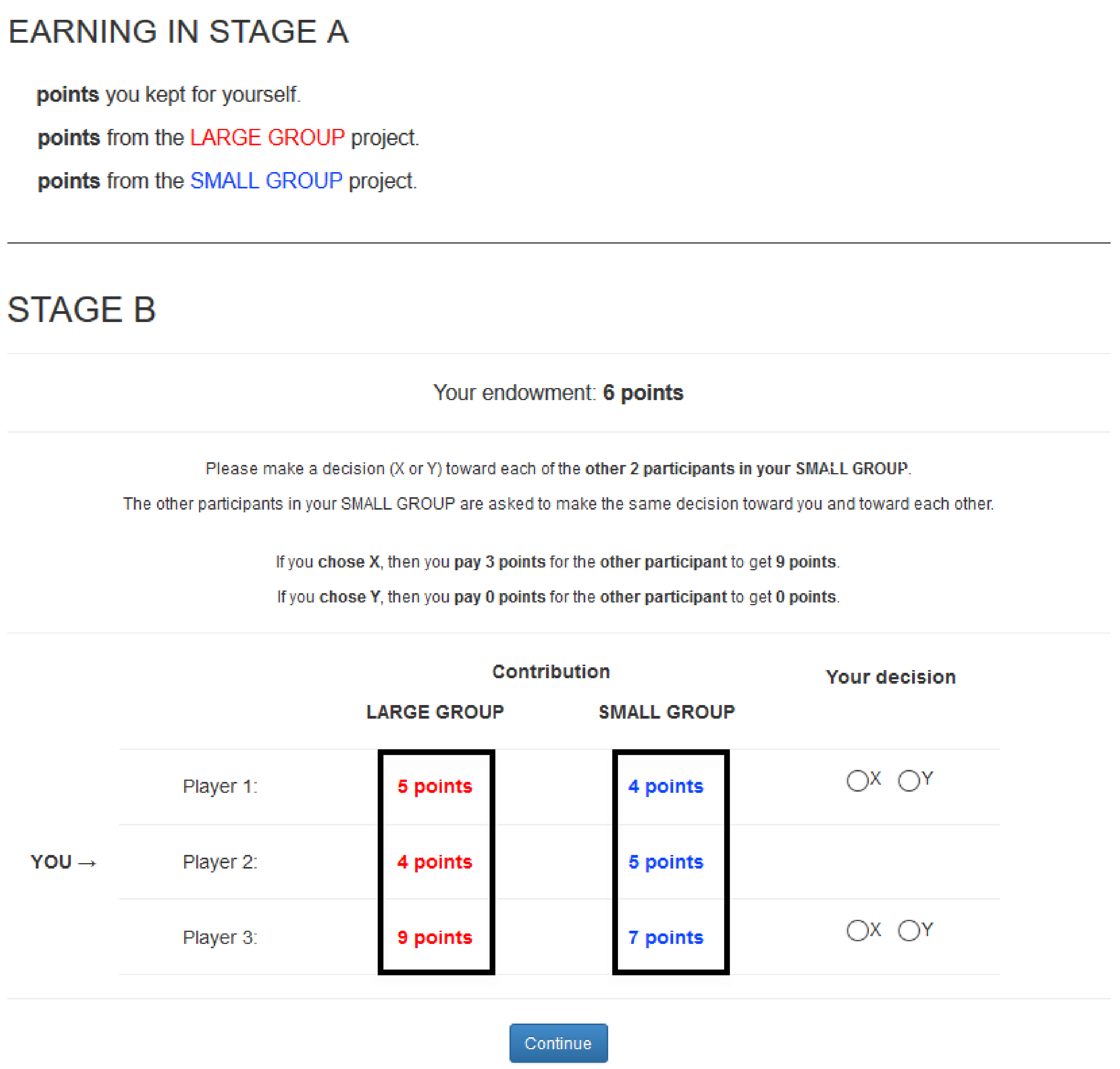
STAGE B

Your task is to make a decision toward each of the **other 2 participants in your SMALL GROUP**.

The other participants in your SMALL GROUP are asked to make the same decision toward you and toward each other.

You can choose whether you want:

- **Pay 3 points** for the **other participant** to get **9 points**
- **Pay 0 points** for the **other participant** to get **0 points**



EARNING IN STAGE A

points you kept for yourself.

points from the LARGE GROUP project.

points from the SMALL GROUP project.

STAGE B

Your endowment: **6 points**

Please make a decision (X or Y) toward each of the **other 2 participants in your SMALL GROUP**.

The other participants in your SMALL GROUP are asked to make the same decision toward you and toward each other.

If you **chose X**, then you **pay 3 points** for the **other participant** to get **9 points**.

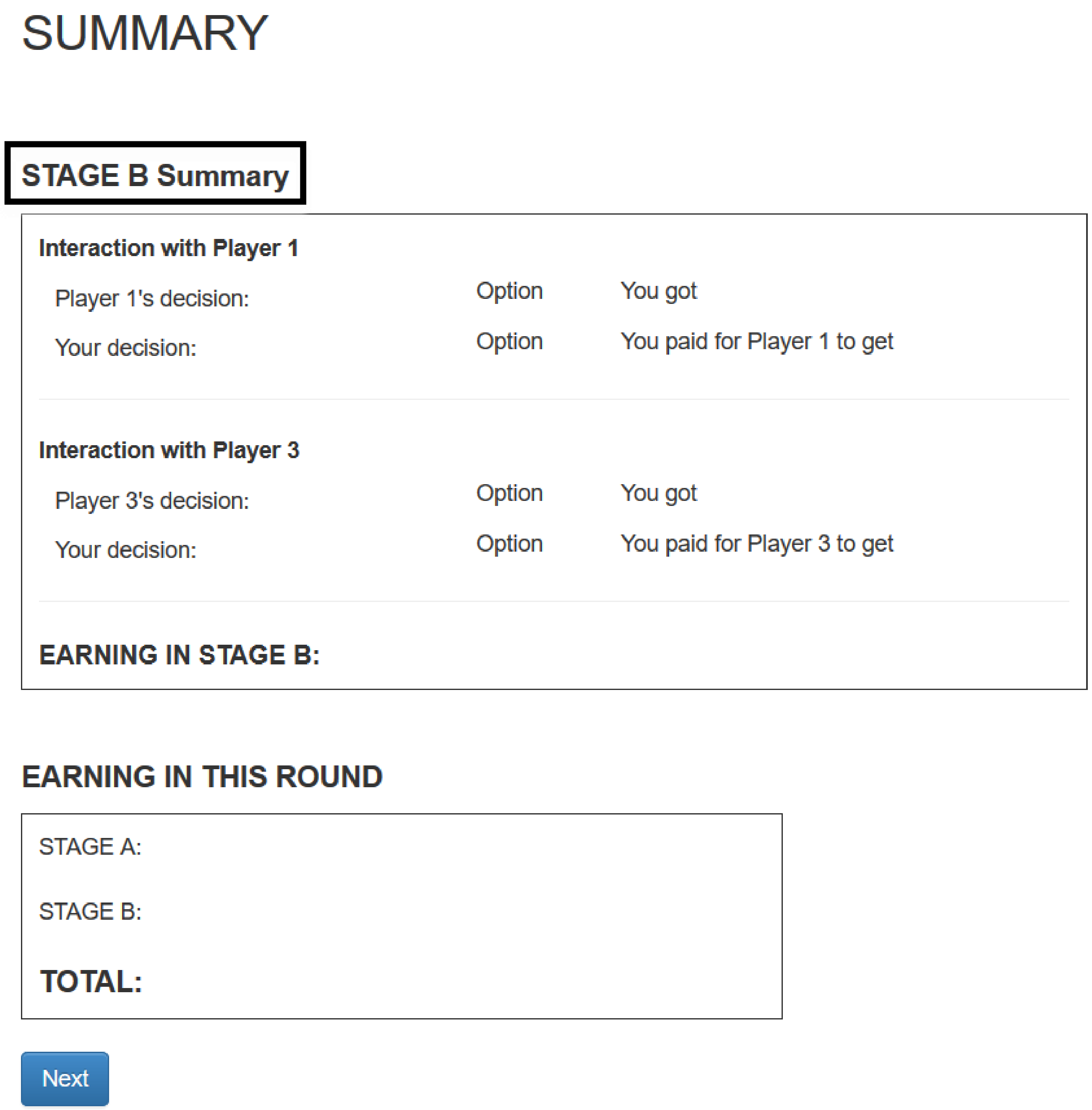
If you **chose Y**, then you **pay 0 points** for the **other participant** to get **0 points**.

		Contribution		Your decision
		LARGE GROUP	SMALL GROUP	
	Player 1:	5 points	4 points	○X ○Y
YOU →	Player 2:	4 points	5 points	
	Player 3:	9 points	7 points	○X ○Y

STAGE B

When making your decisions, **you will see the contributions** of the other participants in your SMALL GROUP to both the SMALL and the LARGE GROUP projects.

Once you have made your decisions, you will **move to the next round**.
Every round consists of the same two stages. You always interact with the same participants, who **keep their identification numbers during the game**.



SUMMARY

STAGE B Summary

Interaction with Player 1

Player 1's decision:	Option	You got
Your decision:	Option	You paid for Player 1 to get

Interaction with Player 3

Player 3's decision:	Option	You got
Your decision:	Option	You paid for Player 3 to get

EARNING IN STAGE B:

EARNING IN THIS ROUND

STAGE A:

STAGE B:

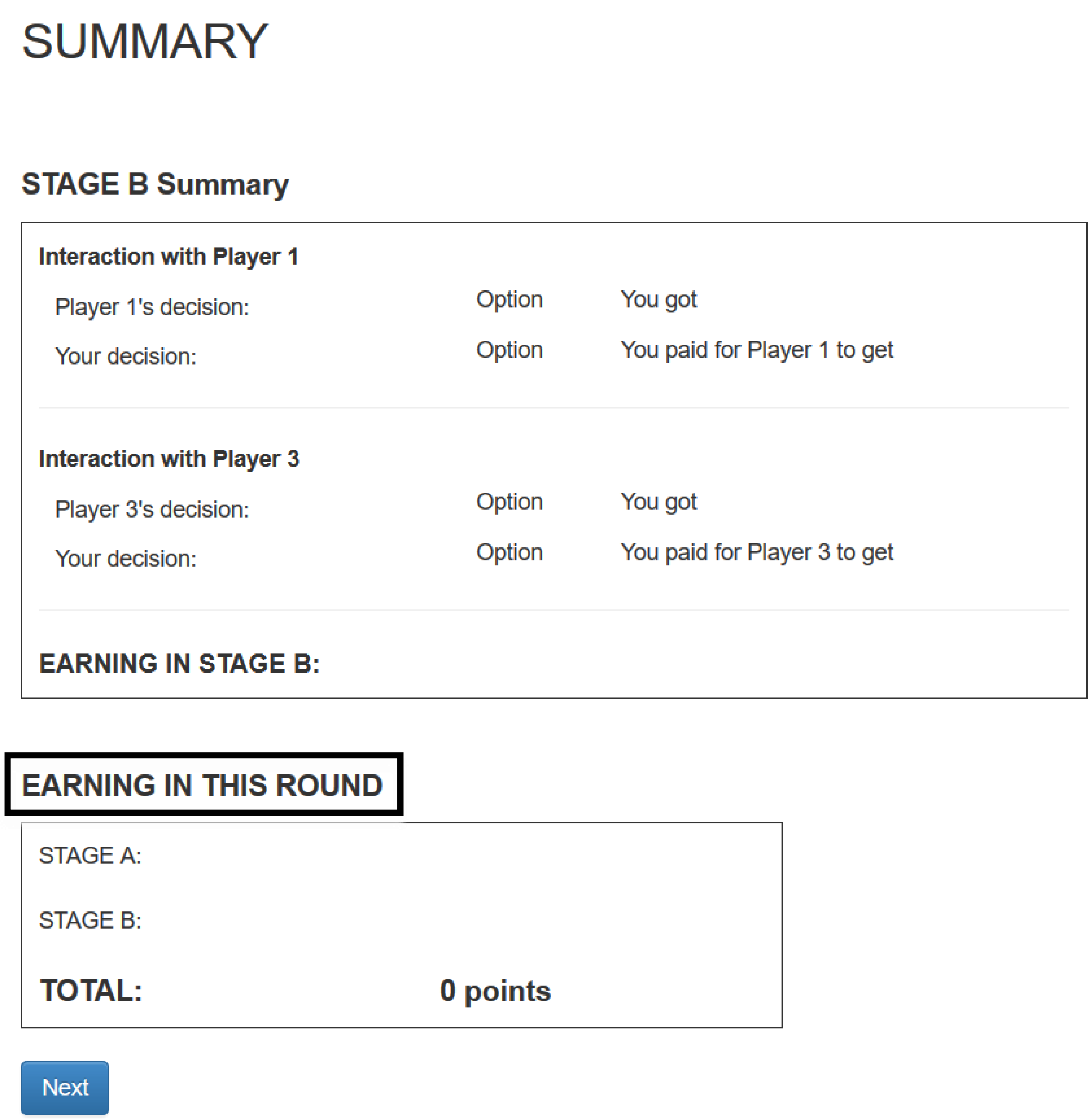
TOTAL:

SUMMARY

Once all participants have made their decisions, you will be shown a summary of STAGE B.

NOTE
Once you have made your decision in STAGE B, **you might need to wait few seconds** for the other participants to make their decisions as well.

It is very important that you **don’t abandon the page** at any time and stay ready to continue the game.



SUMMARY

STAGE B Summary

Interaction with Player 1

Player 1's decision:	Option	You got
Your decision:	Option	You paid for Player 1 to get

Interaction with Player 3

Player 3's decision:	Option	You got
Your decision:	Option	You paid for Player 3 to get

EARNING IN STAGE B:

EARNING IN THIS ROUND

STAGE A:	
STAGE B:	
TOTAL:	**0 points**

SUMMARY

You will also see a summary of your combined payoff in the current round.

Then you will move to the **next round**.

REREFERENCES

Alexander, Richard D. *The Biology of Moral Systems*. Hawthorne, N.Y: A. de Gruyter, 1987.

Amir, Ofra, David G. Rand, and Ya'akov Kobi Gal. "Economic Games on the Internet: The Effect of $1 Stakes." Edited by Matjaz Perc. *PLoS ONE* 7, no. 2 (February 21, 2012): e31461. https://doi.org/10.1371/journal.pone.0031461.

Apicella, Coren L., Frank W. Marlowe, James H. Fowler, and Nicholas A. Christakis. "Social Networks and Cooperation in Hunter-Gatherers." *Nature* 481, no. 7382 (January 25, 2012): 497–501. https://doi.org/10.1038/nature10736.

Barclay, Pat. "Biological Markets and the Effects of Partner Choice on Cooperation and Friendship." *Current Opinion in Psychology*, Evolutionary psychology, 7 (February 2016): 33–38. https://doi.org/10.1016/j.copsyc.2015.07.012.

———. "Reputational Benefits for Altruistic Punishment." *Evolution and Human Behavior* 27, no. 5 (September 2006): 325–44. https://doi.org/10.1016/j.evolhumbehav.2006.01.003.

Barclay, Pat, and Robb Willer. "Partner Choice Creates Competitive Altruism in Humans." *Proceedings of the Royal Society B: Biological Sciences* 274, no. 1610 (March 7, 2007): 749–53. https://doi.org/10.1098/rspb.2006.0209.

Bateson, M., D. Nettle, and G. Roberts. "Cues of Being Watched Enhance Cooperation in a Real-World Setting." *Biology Letters* 2, no. 3 (September 22, 2006): 412–14. https://doi.org/10.1098/rsbl.2006.0509.

Benjamini, Yoav, and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, 289–300.

Berg, Joyce, John Dickhaut, and Kevin McCabe. "Trust, Reciprocity, and Social History." *Games and Economic Behavior* 10, no. 1 (July 1, 1995): 122–42. https://doi.org/10.1006/game.1995.1027.

Berinsky, A. J., G. A. Huber, and G. S. Lenz. "Evaluating Online Labor Markets for Experimental Research: Amazon.com's Mechanical Turk." *Political Analysis* 20, no. 3 (July 1, 2012): 351–68. https://doi.org/10.1093/pan/mpr057.

Blackwell, Calvin, and Michael McKee. "Only for My Own Neighborhood?: Preferences and Voluntary Provision of Local and Global Public Goods." *Journal of Economic Behavior & Organization* 52, no. 1 (September 2003): 115–31. https://doi.org/10.1016/S0167-2681(02)00178-6.

Bó, Pedro Dal. “Cooperation under the Shadow of the Future: Experimental Evidence from Infinitely Repeated Games.” *American Economic Review*, 2005, 1591–1604.

Brandt, H., C. Hauert, and K. Sigmund. “Punishment and Reputation in Spatial Public Goods Games.” *Proceedings of the Royal Society of London B: Biological Sciences* 270, no. 1519 (May 22, 2003): 1099–1104. https://doi.org/10.1098/rspb.2003.2336.

Carpenter, Jeffrey P. “Punishing Free-Riders: How Group Size Affects Mutual Monitoring and the Provision of Public Goods.” *Games and Economic Behavior* 60, no. 1 (Luglio 2007): 31–51. https://doi.org/10.1016/j.geb.2006.08.011.

Chen, Daniel L., Martin Schonger, and Chris Wickens. “oTree—An Open-Source Platform for Laboratory, Online, and Field Experiments.” *Journal of Behavioral and Experimental Finance* 9 (March 2016): 88–97. https://doi.org/10.1016/j.jbef.2015.12.001.

Cuesta, Jose A., Carlos Gracia-Lázaro, Alfredo Ferrer, Yamir Moreno, and Angel Sánchez. “Reputation Drives Cooperative Behaviour and Network Formation in Human Groups.” *Scientific Reports* 5 (January 19, 2015): 7843. https://doi.org/10.1038/srep07843.

DellaVigna, Stefano, John A. List, and Ulrike Malmendier. “Testing for Altruism and Social Pressure in Charitable Giving.” *The Quarterly Journal of Economics* 127, no. 1 (February 1, 2012): 1–56. https://doi.org/10.1093/qje/qjr050.

Engelmann, Dirk, and Urs Fischbacher. “Indirect Reciprocity and Strategic Reputation Building in an Experimental Helping Game.” *Games and Economic Behavior* 67, no. 2 (November 2009): 399–407. https://doi.org/10.1016/j.geb.2008.12.006.

Erev, Ido, Gary Bornstein, and Rachely Galili. “Constructive Intergroup Competition as a Solution to the Free Rider Problem: A Field Experiment.” *Journal of Experimental Social Psychology* 29, no. 6 (November 1993): 463–78. https://doi.org/10.1006/jesp.1993.1021.

Ernest-Jones, Max, Daniel Nettle, and Melissa Bateson. “Effects of Eye Images on Everyday Cooperative Behavior: A Field Experiment.” *Evolution and Human Behavior* 32, no. 3 (May 2011): 172–78. https://doi.org/10.1016/j.evolhumbehav.2010.10.006.

Feinberg, Matthew, Robb Willer, and Michael Schultz. “Gossip and Ostracism Promote Cooperation in Groups.” *Psychological Science*, 2014, 0956797613510184.

Fellner, Gerlinde, and Gabriele K. Lünser. “Cooperation in Local and Global Groups.” *Journal of Economic Behavior & Organization* 108 (Dicembre 2014): 364–73. https://doi.org/10.1016/j.jebo.2014.02.007.

Gachter, S., E. Renner, and M. Sefton. “The Long-Run Benefits of Punishment.” *Science* 322, no. 5907 (December 5, 2008): 1510–1510. https://doi.org/10.1126/science.1164744.

Hamilton, W. D. “The Genetical Evolution of Social Behaviour. I.” *Journal of Theoretical Biology* 7, no. 1 (Luglio 1964): 1–16. https://doi.org/10.1016/0022-5193(64)90038-4.

Hauser, Oliver P., Achim Hendriks, David G. Rand, and Martin A. Nowak. “Think Global, Act Local: Preserving the Global Commons.” *Scientific Reports* 6 (November 3, 2016): 36079. https://doi.org/10.1038/srep36079.

Herrmann, B., C. Thoni, and S. Gachter. “Antisocial Punishment Across Societies.” *Science* 319, no. 5868 (March 7, 2008): 1362–67. https://doi.org/10.1126/science.1153808.

Horton, John J., David G. Rand, and Richard J. Zeckhauser. “The Online Laboratory: Conducting Experiments in a Real Labor Market.” *Experimental Economics* 14, no. 3 (September 1, 2011): 399–425. https://doi.org/10.1007/s10683-011-9273-9.

Jann, Ben, and Wojtek Przepiorka. *Social Dilemmas, Institutions, and the Evolution of Cooperation*. Walter de Gruyter GmbH & Co KG, 2017.

Jordan, Jillian J., Moshe Hoffman, Paul Bloom, and David G. Rand. “Third-Party Punishment as a Costly Signal of Trustworthiness.” *Nature* 530, no. 7591 (February 25, 2016): 473–76. https://doi.org/10.1038/nature16981.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. “Fairness as a Constraint on Profit Seeking: Entitlements in the Market.” *American Economic Review* 76, no. 4 (1986): 728–41.

Kandori, Michihiro, and Shinya Obayashi. “Labor Union Members Play an OLG Repeated Game.” *Proceedings of the National Academy of Sciences* 111, no. Supplement 3 (2014): 10802–9.

Lotem, Arnon, Michael A. Fishman, and Lewi Stone. “From Reciprocity to Unconditional Altruism through Signalling Benefits.” *Proceedings of the Royal Society of London B: Biological Sciences* 270, no. 1511 (January 22, 2003): 199–205. https://doi.org/10.1098/rspb.2002.2225.

Lotem, Arnon, Michael A Fishman, and Lewi Stone. “From Reciprocity to Unconditional Altruism through Signalling Benefits.” *Proceedings of the Royal Society B: Biological Sciences* 270, no. 1511 (January 22, 2003): 199–205. https://doi.org/10.1098/rspb.2002.2225.

Milinski, Manfred. “Reputation, a Universal Currency for Human Social Interactions.” *Phil. Trans. R. Soc. B* 371, no. 1687 (February 5, 2016): 20150100. https://doi.org/10.1098/rstb.2015.0100.

Milinski, Manfred, Dirk Semmann, and Hans-Jurgen Krambeck. “Donors to Charity Gain in Both Indirect Reciprocity and Political Reputation.” *Proceedings of the Royal Society B: Biological Sciences* 269, no. 1494 (May 7, 2002): 881–83. https://doi.org/10.1098/rspb.2002.1964.

Milinski, Manfred, Dirk Semmann, and Hans-Jürgen Krambeck. "Reputation Helps Solve the 'tragedy of the Commons.'" *Nature* 415, no. 6870 (January 24, 2002): 424–26. https://doi.org/10.1038/415424a.

Milinski, Manfred, Dirk Semmann, Hans-Jürgen Krambeck, and Jochem Marotzke. "Stabilizing the Earth's Climate Is Not a Losing Game: Supporting Evidence from Public Goods Experiments." *Proceedings of the National Academy of Sciences of the United States of America* 103, no. 11 (March 14, 2006): 3994–98. https://doi.org/10.1073/pnas.0504902103.

Nelissen, Rob M. A. "The Price You Pay: Cost-Dependent Reputation Effects of Altruistic Punishment." *Evolution and Human Behavior* 29, no. 4 (July 1, 2008): 242–48. https://doi.org/10.1016/j.evolhumbehav.2008.01.001.

Nikiforakis, Nikos, Charles N. Noussair, and Tom Wilkening. "Normative Conflict and Feuds: The Limits of Self-Enforcement." *Journal of Public Economics* 96, no. 9 (October 1, 2012): 797–807. https://doi.org/10.1016/j.jpubeco.2012.05.014.

Noë, Ronald, and Peter Hammerstein. "Biological Markets: Supply and Demand Determine the Effect of Partner Choice in Cooperation, Mutualism and Mating." *Behavioral Ecology and Sociobiology* 35, no. 1 (July 1, 1994): 1–11. https://doi.org/10.1007/BF00167053.

Nowak, Martin A. "Five Rules for the Evolution of Cooperation." *Science* 314, no. 5805 (2006): 1560–63.

Ohtsuki, Hisashi, and Yoh Iwasa. "The Leading Eight: Social Norms That Can Maintain Cooperation by Indirect Reciprocity." *Journal of Theoretical Biology* 239, no. 4 (April 2006): 435–44. https://doi.org/10.1016/j.jtbi.2005.08.008.

Przepiorka, Wojtek, and Ulf Liebe. "Generosity Is a Sign of Trustworthiness—the Punishment of Selfishness Is Not." *Evolution and Human Behavior* 37, no. 4 (July 1, 2016): 255–62. https://doi.org/10.1016/j.evolhumbehav.2015.12.003.

Puurtinen, Mikael, and Tapio Mappes. "Between-Group Competition and Human Cooperation." *Proceedings of the Royal Society of London B: Biological Sciences* 276, no. 1655 (January 22, 2009): 355–60. https://doi.org/10.1098/rspb.2008.1060.

Raihani, Nichola J., and Redouan Bshary. "The Reputation of Punishers." *Trends in Ecology & Evolution* 30, no. 2 (February 2015): 98–103. https://doi.org/10.1016/j.tree.2014.12.003.

———. "Third-Party Punishers Are Rewarded, but Third-Party Helpers Even More so." *Evolution; International Journal of Organic Evolution* 69, no. 4 (April 2015): 993–1003. https://doi.org/10.1111/evo.12637.

Rand, David G., Samuel Arbesman, and Nicholas A. Christakis. “Dynamic Social Networks Promote Cooperation in Experiments with Humans.” *Proceedings of the National Academy of Sciences* 108, no. 48 (2011): 19193–98.

Rand, David G., and Martin A. Nowak. “Human Cooperation.” *Trends in Cognitive Sciences* 17, no. 8 (August 2013): 413–25. https://doi.org/10.1016/j.tics.2013.06.003.

Rockenbach, Bettina, and Manfred Milinski. “The Efficient Interaction of Indirect Reciprocity and Costly Punishment.” *Nature* 444, no. 7120 (December 7, 2006): 718–23. https://doi.org/10.1038/nature05229.

Rogers, Todd, John Ternovski, and Erez Yoeli. “Potential Follow-up Increases Private Contributions to Public Goods.” *Proceedings of the National Academy of Sciences*, April 25, 2016, 201524899. https://doi.org/10.1073/pnas.1524899113.

Santos, Miguel dos, Sarah Placì, and Claus Wedekind. “Stochasticity in Economic Losses Increases the Value of Reputation in Indirect Reciprocity.” *Scientific Reports* 5 (December 14, 2015): 18182. https://doi.org/10.1038/srep18182.

Santos, Miguel dos, Daniel J. Rankin, and Claus Wedekind. “The Evolution of Punishment through Reputation.” *Proceedings of the Royal Society of London B: Biological Sciences* 278, no. 1704 (February 7, 2011): 371–77. https://doi.org/10.1098/rspb.2010.1275.

Stanca, Luca, Luigino Bruni, and Marco Mantovani. “The Effect of Motivations on Social Indirect Reciprocity: An Experimental Analysis.” *Applied Economics Letters* 18, no. 17 (November 2011): 1709–11. https://doi.org/10.1080/13504851.2011.560105.

Suzuki, S., and E. Akiyama. “Reputation and the Evolution of Cooperation in Sizable Groups.” *Proceedings of the Royal Society B: Biological Sciences* 272, no. 1570 (July 7, 2005): 1373–77. https://doi.org/10.1098/rspb.2005.3072.

Suzuki, Shinsuke, and Eizo Akiyama. “Evolution of Indirect Reciprocity in Groups of Various Sizes and Comparison with Direct Reciprocity.” *Journal of Theoretical Biology* 245, no. 3 (April 7, 2007): 539–52. https://doi.org/10.1016/j.jtbi.2006.11.002.

Traulsen, Arne, and Martin A. Nowak. “Evolution of Cooperation by Multilevel Selection.” *Proceedings of the National Academy of Sciences* 103, no. 29 (2006): 10952–55.

Trivers, Robert L. “The Evolution of Reciprocal Altruism.” *Quarterly Review of Biology*, 1971, 35–57.

Wang, J., S. Suri, and D. J. Watts. “Cooperation and Assortativity with Dynamic Partner Updating.” *Proceedings of the National Academy of Sciences* 109, no. 36 (September 4, 2012): 14363–68. https://doi.org/10.1073/pnas.1120867109.

Wedekind, Claus, and Manfred Milinski. “Cooperation Through Image Scoring in Humans.” *Science* 288, no. 5467 (May 5, 2000): 850–52. https://doi.org/10.1126/science.288.5467.850.

Yoeli, E., M. Hoffman, D. G. Rand, and M. A. Nowak. “Powering up with Indirect Reciprocity in a Large-Scale Field Experiment.” *Proceedings of the National Academy of Sciences* 110, no. Supplement_2 (June 10, 2013): 10424–29. https://doi.org/10.1073/pnas.1301210110.

Zahavi, Amotz. “Altruism as a Handicap: The Limitations of Kin Selection and Reciprocity.” *Journal of Avian Biology* 26, no. 1 (March 1995): 1. https://doi.org/10.2307/3677205.