Essays on Social Reference Points and the Compromise Effect in Bargaining

Patrick Maus

Thesis submitted to the University of Nottingham for the Degree of Doctor of Philosophy

School of Economics
University of Nottingham

May 2024

Abstract

This thesis contributes to the investigation of reference points, the behavioral effects of social comparisons and context effects in bargaining games. The work consists of three experimental studies which will be summarized below.

In the first study, we report a laboratory experiment testing whether social reference points impact effort provision. Subjects are randomly assigned the role of worker or peer and the worker observes the peer's earnings before participating in a real-effort task. Between treatments, we exogenously manipulate peer earnings. We find that the workers recall the earnings of their peer and are less satisfied with their own earnings when their peer earns more. Despite this, we do not observe a treatment effect in effort choices. Thus, although our subjects appear to care about income differentials, this does not translate to a change in behavior in our incentivized environment. We relate our results to recent studies of inequality and effort provision.

In the second study, we use a very similar design to test whether social reference points impact donation decisions. Subjects are randomly assigned the role of decision-maker or peer and the decision-maker observes the peer's earnings before making a real donation decision. Between treatments, we exogenously manipulate peer earnings. We derive our predictions using a model where the optimal donation decision is influenced by social comparison concerns. We find that subjects donate substantially more when peer earnings are relatively low. Using additional treatments, we show that our treatment effects can not be explained by confounding factors such as feelings of disappointment, anchoring, and experimenter demand effects.

In the third study, we experimentally investigate compromise effects, namely a tendency for bargainers to agree to an intermediate outcome, by comparing Ultimatum Games with varying sets of feasible offers. We find that when an extremely unfair offer is added to the set of feasible offers i) Responders are more likely to accept a somewhat unfair offer and ii) Proposers tend to switch from a fair offer to a somewhat unfair offer but only if the Fair contract is not exactly equal. To test for the role of reciprocal concerns in our findings, we additionally conducted a set of Dictator and Random Proposal Games. Together, they suggest that the change in responder behavior is not purely driven by reciprocity, and the change in proposer behavior is not driven by a change in proposer preferences over outcomes.

Acknowledgements

First, I would like to thank my supervisors Prof. Maria Montero and Prof. Martin Sefton for their dedicated, energetic and intellectually stimulating mentoring over the past years. I thank you for your trust, support and the patience you have always shown me along the way. It was a privilege to learn from your guidance and criticism. I could not have wished for better supervisors and nicer people to work with.

I would also like to thank other professors, colleagues and staff at the School of Economics, whose doors were always open for me. In particular, I would like to thank Diego Marino Fages who not only became my co-author but also one of my best friends and Jose Guinot Saporta who can only be described as an angel for all students who conducted experiments in Nottingham or will do so in the future. I am also grateful for the financial support from the School of Economics that enabled me to pursue my PhD at the University of Nottingham.

I am deeply grateful to all my friends, but especially Christian Bäurer and Valerius Disch, who always have an open ear for me and whose support I can always rely on. I thank my girlfriend Laura Isabel Reyes Navarro for an abundance of love, support and laughter whenever I need it. You let the sun shine even on rainy days.

Last but not least, I would like to thank my mum and my sister for their love and endless support. I couldn't ask for a better support system in my life. Thank you for always being there for me, for your understanding and for encouraging me to pursue my dreams. You have helped make one of them come true today.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In standard economic models, individuals are typically assumed to make their decisions in isolation from their social surroundings. However, human decisions are seldom made in isolation and observing others can be an important source of reference for individuals. For instance, the income of others might serve as a social reference point and affect decisions. Consequently, higher social reference points could motivate individuals to change their behavior in order to avoid unfavorable social comparisons.

Imagine for instance two individuals - both are equally wealthy but the first one is the richest in his social environment whereas the second one is the poorest. Their income in absolute terms is the same for both individuals but presumably, the second individual is less happy with the status quo and might have stronger incentives to change it. Do higher social reference points indeed change behavior because people want to avoid unfavorable comparisons? I investigate this research question in two different domains - labor supply and charitable giving.

In Chapter 2, we test experimentally whether social reference points impact effort provision. In our experiment, only two subjects participate in each session. Each subject is assigned one of the two roles, worker or peer, by a publicly observed coin toss. The peer receives a fixed payment, whereas the worker participates in a tedious real-effort task (counting lines). The fixed payment to the peer is varied across two treatments. Consequently, any difference in the average level of effort provision by the workers between treatments can be traced back to social reference points. In our experiment, we find that workers remember their peers' earnings and appear to care about them. The workers are significantly less happy with their earnings when peer earnings are relatively high compared to when they are relatively low. While workers are less happy when their respective peer earns more, this does not translate to a change in working behavior in the real-effort task. The difference in effort provision between treatments is very small and not statistically significant. Together, this suggests that people care about income differences but this does not necessarily translate to a change in behavior in incentivized environments. This chapter is joint work with my supervisors Maria Montero and Martin Sefton.

In Chapter 3, we use a very similar design to test whether social reference points impact donation decisions. Similarly to Chapter 2, in the social treatments, two subjects are randomly assigned the role of a decision-maker or peer and the decision-maker observes the peer's earnings before making a real donation decision. In a between-subject design, the peer earnings were exogenously varied between £5 and £2 and each decision-maker was endowed with £6. As the donation task, we gave the subjects the possibility to donate money from their initial endowment to WWF's Climate Relief Fund. The two social treatments are accompanied by two non-social treatments. The non-social treatments replicate the social treatments with a crucial change: only one subject participated in each session. A coin toss assigned each single subject per session to make a donation decision or leave with pre-determined earnings. The donation decision was exactly the same as in the social treatments. Before making their decision, decision-makers were exposed to the amount of money they would have earned if their coin had landed on the other side. Consequently, the donation decision differs between the social and non-social treatments only in the fact that exposure to counterfactual earnings referred to peer earnings in the social treatments and not obtained earnings in the non-social treatments. In the social treatments, decision-makers donated significantly more when the peer earnings were relatively low compared to when they were relatively high. As expected, decision-makers did not donate more in the non-social treatments when their reference earnings were relatively low. Consequently, our treatment effect in the social treatments can be fully attributed to social comparison concerns.

In Chapter 4, we examine compromise effects and the mechanism behind them by varying the set of feasible contracts in simple structured bargaining games. Specifically, we use Ultimatum Games as they have a very straightforward structure, where a proposer makes a take-it-or-leave-it proposal about how to split a pie and a receiver can either accept or reject the proposal. Using Dictator Games, where the receiver cannot reject a proposal, we can observe whether proposer preferences are affected by adding a menu option. Finally, we use Random Proposal Games, in which the receivers indicate the acceptability of proposals when the proposal is chosen by Nature. This allows us to observe whether the acceptability of contracts is subject to 'pure' menu effects, independent of the intentions of proposers. In each of these games we vary the set of feasible contracts and observe how this changes behavior and outcomes.

In our base scenarios, there is a Greedy contract (80-20) alongside a Fair contract. We examine two base scenarios where the Fair contract is either exactly equal (50-50) or nearly equal (60-40). In our other scenarios, we add a third Decoy contract. In some of these the Decoy is (90-10), making the Greedy contract a compromise, and in others it is (20-80), making the Fair contract a compromise.

We find that the addition of a third contract in the Ultimatum Games can matter for both responders and proposers but only when the Fair contract is unequal. Responders are significantly more likely to accept the Greedy contract once an even greedier Decoy contract is added to the set of feasible contracts. Similarly, proposers are more likely to switch to offering a more unfair contract once an even greedier contract is added to the set of feasible contracts.

In the Dictator Games, we do not observe that proposers are more likely to switch from the Fair to the Greedy contract once an even greedier contract is added to the set of feasible contracts. This suggests that proposers in the Ultimatum Game anticipate changes in the responders' rejection rates and that their change in behavior cannot be explained by pure menu effects.

In the Random Proposal Games, we see a similar picture as in the Ultimatum games. Responders are significantly more likely to accept the Greedy contract once an even greedier Decoy is added to the set of feasible contracts for the base Scenario if the Fair contract is unequal. This suggests that the effects observed in the Ultimatum Games survive even when possible reasons for reciprocity are removed. This chapter is joint work with my supervisors Maria Montero and Martin Sefton.

Chapter 2

Social reference points and real-effort provision

2.1 Introduction

In standard economic models, workers care only about wage levels and their cost of effort when making decisions about their labor supply. However, decisions about labor supply are seldom made in isolation and observing others can be an important source of reference for individuals. For example, the income of others could serve as a social reference point and affect labor supply decisions. Consequently, high social reference points could motivate individuals to change their behavior in order to avoid unfavorable social comparisons. For instance, imagine two identical individuals. Both individuals are equally wealthy but the first one is the richest in his social environment whereas the second one is the poorest. Their income in absolute terms is the same for both individuals but presumably, the second individual is less happy with the status quo and might have stronger incentives to change it.

In particular, social psychology offers a large body of evidence that shows that social comparisons affect the way we feel and behave (Festinger, 1954). People care about their relative status and their decisions are often influenced by what they can observe from others (Bandura, 1977). Social comparisons have been found to be important determinants for our subjective well-being (Ferreri-Carbonell, 2005; Perez-Truglia, 2020), fairness perceptions (Akerlof and Yellen, 1990; Fehr and Gächter, 2000), happiness (Clark and Oswald, 1996; Luttmer, 2005) and health

(Marmot, 2005).

Akerlof and Yellen (1990) coined the concept "fair-wage hypothesis". The core assumption of this concept is that people compare their wages to what they perceive as a fair wage and adjust their level of effort accordingly. An important question that remains in this context is what factors determine whether a wage level is perceived as fair or unfair. Austin et al. (1980) show that the comparison of the own income with a peer's income (as a social reference point) has a stronger impact on pay satisfaction than the comparison with an individual (private) reference point like previous earnings. These findings then also raise the question of how individuals would react to advantageous or disadvantageous income comparisons with relevant peers if they have the chance to change their position in the income distribution by modifying their behavior. Genicot and Ray (2017) point out that different exposure to peer outcomes within societies may lead to heterogeneous social reference points which in turn may set higher incentives to work hard for peers from a well-off strata as they are confronted with higher social reference points while peers from a more unfavorable background have fewer incentives to exert effort. Long-term this could foster equality within groups and inequality between groups.

The main goal of this study is to explore whether and how effort provision is affected by social reference points. Do higher social reference points cause people to supply more effort because they want to avoid unfavorable comparisons? The main challenge to this identification is that social reference points relevant to individuals are usually hard to observe and exogenous variation over social reference points is hard to obtain outside of the lab. We address this research question utilizing an experiment that allows us to exogeneously manipulate peer earnings between treatments.

In our experiment, subjects engage in a piece rate real-effort task given the earnings of a peer and they cannot affect their peer's earnings. This is arguably a feature of many work environments where a worker can reduce inequality relative to a social reference point provided by other workers' earnings by working harder, but cannot reduce it by reducing the other workers' earnings. It also provides a clean test of how workers respond to exogenously set social reference points.

In our experiment, only two subjects participate in each session. Each subject is assigned one of the two roles, worker or peer, by a publicly observed coin toss. The peer receives a fixed payment, whereas the worker participates in a tedious real-effort task (counting lines). The fixed payment to the peer is varied across two treatments. Consequently, any difference in the average level of effort provision by the workers between treatments can be traced back to social reference points. In our experiment, no worker can observe or be observed by other participants. This minimizes other peer effects, such as peer pressure, imitation, or learning (Sacerdote, 2001; Falk and Ichino, 2006; Mas and Moretti, 2009).

To derive the main hypothesis that our experiment is designed to test, we formalize a simple Fehr-Schmidt model of inequality aversion around the peer's earnings. The manipulation of the peer's earnings changes the incentives for the workers who want to avoid unfavorable social comparisons. Consequently, our main hypothesis is that workers provide more effort on average when the earnings of their peer are relatively high (*High treatment*) compared to when they are relatively low (*Low treatment*).

In our experiment, we find that workers remember their peers' earnings and appear to care about them. The workers are significantly less happy with their earnings when their peer earns £7.10 (*High treatment*) compared to when it earns £2.90 (*Low treatment*). While workers are less happy when their respective peer earns more, this does not translate to a change in working behavior in the real-effort task. The difference in effort provision between treatments is very small and not statistically significant. Together, this suggests that people care about income differences but this does not necessarily translate to a change in behavior in incentivized environments.

This experiment was pre-registered, with sample sizes determined through a power analysis used to ensure adequate statistical power. The power analysis was based on assumptions about effort provision using data from a previous experiment (Gagnon et al., 2020). A surprising result from our experiment was that efforts, in both treatments, substantially exceeded our expectations and were inconsistent with our assumptions. Indeed, many workers provided maximal effort, even though our experiment was designed so that workers would choose when to stop working and the maximal effort constraint would not bind. A consequence is that similar efforts across treatments may be due to a "ceiling effect" whereby real-effort costs were so low as to make effort insensitive to incentives.

Because the effort provided by our subjects in our pre-registered experiment was substantially greater than anticipated and many people solved the maximum amount of lines, we conducted a follow-up experiment with a modified and more difficult working task. We do this by raising workers' effort costs to ensure that in our experiment enough participants could be influenced by concerns of behindness aversion. Our re-calibration successfully reduced effort levels such that no subjects exerted maximal effort. The results of our re-calibrated experiment are qualitatively very similar to the original experiment and again we find no indication that social reference points affect real-effort provision.

The paper is organized as follows. Section 2 discusses the relevant literature. Section 3 explains the experimental design in detail. Section 4 discusses the behavioral predictions and derives our main hypothesis. In Section 5 we present the result of our original experiment. Section 6 discusses and presents the results of our re-calibrated experiment. Section 7 concludes.

2.2 Related Literature

This study aims to contribute to the strands of literature on reference dependence, peer effects, and wage inequalities. Firstly, the study is linked with the literature on reference dependence. So far, a lot of this literature has focused on individual decision making where reference points are based on expectations or the status quo (see e.g. Kőszegi and Rabin (2006)). For instance, Gill and Prowse (2012) investigate whether individuals are disappointment averse when they compete in a real-effort sequential-move tournament. They find that the second mover shies away from working hard when she observes that the first mover has worked hard, and tends to work relatively hard when she observes that her competitor has put in a low effort which is consistent with an expectation-based model of disappointment aversion. Gächter et al. (2018) compare "social" and "asocial" versions of the Gill and Prowse experiment, where the scope for social comparison is removed in the latter. They find behavior in social and asocial treatments to be similar, suggesting that social comparisons have little impact in this setting.

Abeler et al. (2011) also provide evidence that expectations-based reference points matter in the context of effort provision. In their experiment, subjects work on a tedious and repetitive task. After each repetition, they decide whether to continue or to stop working. They get a piece rate, but receive their accumulated piece-rate earnings only with a 50 percent probability, whereas with a 50 percent probability they receive a fixed, known payment instead. Which payment subjects receive is determined only after they have made their choice about when to stop working. The only treatment manipulation is a variation in the amount of the fixed payment. In their experiment workers worked substantially longer and earned more money if their expectations were high compared to low expectations. While there is some evidence that expectations-based reference points matter when it comes to effort provision, evidence is scarce when it comes to different sources of reference points. We complement the literature of reference points and effort provision by using social outcomes as reference points while our design allows us to keep expectations constant between treatments.

More recently, Schwerter (2023) investigates experimentally whether social reference points impact individual risk-taking. Decision-makers in his experiment observe the earnings of a peer subject before making a risky choice. This allows him to manipulate the peer earnings across two treatments exogenously. He finds a significant treatment effect on risk-taking, i.e. decision-makers make more risk-seeking choices if the earnings of their peer are relatively large. The experimental design of Schwerter (2023) is very similar to ours where people make effort choices but prior observe peer earnings.

Secondly, our study contributes to the literature on peer effects. The most closely related studies to ours investigate the impact of peers on performance. Card et al. (2012) find that workers with salaries below the median for their pay unit and occupation report lower pay and job satisfaction, while those earning above the median report no higher satisfaction. Likewise, below-median earners are significantly more likely to look for a new job, while above-median earners are unaffected. Their findings suggest that job satisfaction depends directly on relative pay comparisons.

Falk and Ichino (2006) recruited subjects to fill letters into envelopes with a remuneration independent of their performance. In the first treatment, subjects worked alone in a room, and consequently peer effects are ruled out. In the second treatment, two subjects work in one room at the same point in time which makes peer effects possible. They find that peer effects raise productivity and that low-productivity workers are the most sensitive to the behavior of peers. Mas and Moretti (2009) find strong evidence of positive productivity spillovers from the introduction of highly productive personnel into a shift using high-frequency data on worker productivity from a supermarket chain. The effort of the workers is positively correlated to the productivity of workers who see them but not workers who do not see them. In all these studies, individuals usually observe the results and behavior of others and expect others to observe their behavior and results. As a consequence, this does not allow them to distinguish between relative concerns and additional peer effects, such as social pressure, imitation, and learning. Our study aims to test a specific channel through which peer effects might work, namely that people compare their income to that of a peer and adjust their behavior accordingly.

Thirdly, our study is also related to the experimental literature that investigates how wage inequalities influence effort provision. Gächter and Thöni (2010) find that disadvantageous wage discrimination leads to lower levels of effort while advantageous wage discrimination does not increase the levels of effort on average using a three-person gift-exchange game. In a similar vein, Cohn et al. (2014) conducted a field experiment to examine how social comparison affects workers' effort provision if their wage and/or the wage of their co-worker is cut. Workers were assigned to groups of two, performed the same task, and received the same performance-independent hourly wage. They find that workers decrease their level of effort less in consequence of a pay cut if not only their wage but also the wage of their co-workers is cut. However, in the experimental setting of Gächter and Thöni (2010) and Cohn et al. (2014) the wage of the employees does not depend on performance and their behavior could also be influenced by reciprocal motives. In the environment of labor supply that we study here the performance of a subject is directly linked to its earnings and there are no reciprocal relationships.

More closely related to our study, Bracha et al. (2015) test the hypothesis that low wage levels compared to others decrease labor supply. To test this, they offer participants to choose how long they want to work on a real-effort task for a piece-rate pay level that is either high or low. In one treatment, the subjects are only aware of one pay rate, whereas in the second treatment, they have a natural reference point - the other piece-rate pay level offered. They find that relative pay comparisons affect labor supply, such that lower-paid individuals supplied significantly less work time relative to higher-paid individuals and significantly less time than when they were unaware of the higher piece rates.

Note that, Bracha et al. (2015) do not implement a social reference point that has predetermined reference earnings associated with it. The subjects are aware of the different piece-rate payments of their peers (and consequently might form beliefs about the performance of their counterparts) but they do not know the absolute income of their peers. The absence of a deterministic social reference point might modify working behavior. For instance, the earnings of the individual's peer are not salient and could be strongly correlated with the beliefs about its peer's working motivation. Moreover, the absence of a deterministic social reference point makes it impossible to know whether a decision-maker could avoid a disadvantageous social comparison by providing additional effort.

Gagnon et al. (2020) investigate the effect of neutral and gender-discriminatory unfair chances on real-effort provision. In their experiment, workers engage in a real-effort task for a piece-rate wage on an online labor platform. They randomize workers into treatments where they control relative pay and chances to receive a low or a high wage. They find that unequal pay affects the labor supply of discriminated workers but does not change the effort-provision of the high-wage workers significantly, irrespective of whether the low wage is the result of fair or unfair chances. However, similar to Bracha et al. (2015) they do not introduce a deterministic social reference point. In contrast to the experiments of Bracha et al. (2015) and Gagnon et al. (2020), we do not manipulate the piece-rate incentives between two workers performing an identical task. In our setting, one individual is endowed with a fixed payment without performing a task and consequently, this payment might serve as a social reference point for a worker who is rewarded according to the same piece-rate wage between treatments.

2.3 Experimental Design

The experiment is designed to create an environment that allows for a precise measurement of effort provision and in which we can exogenously influence a reference point. Between the two treatments, we manipulate predetermined peer earnings exogenously. This allows us to test the influence of a social reference point on effort provision by comparing the behavior of the subjects between treatments. Our study is preregistered at the Open Science Framework (Doi: 10.17605/OSF.IO/QVCE9).

The experiment was conducted in virtual rooms in Microsoft Teams with students of the University of Nottingham in June 2021. Each experimental session consisted of two subjects only. The subjects were recruited via Orsee (Greiner, 2015). The experimental materials are reproduced in Appendix 2B to Appendix 2E.

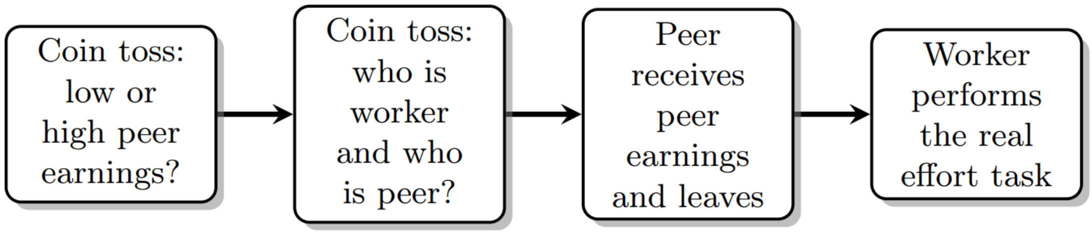


Figure 2.1: Procedural overview of the experiment

Upon the arrival of the subjects in the virtual room, the subjects are instructed that only one of the two subjects will participate in a working task and that the second subject will receive a lump-sum payment and has to leave the room. First, the experimenter tosses a virtual coin to determine the size of the lump-sum payment. In this way, subjects are randomly assigned to either a *High treatment* (lump-sum payment = £7.10) or *Low treatment* (lump-sum payment = £2.90). After both subjects observe the coin toss and are informed about the size of the lump-sum payment, the experimenter tosses another coin to determine which subject participates in the working task and which subject has to leave with the lump-sum payment. In the experiment we referred to these roles as Player A and Player B, for the remainder of the paper we will refer to them as worker and peer.

We randomize the size of the lump-sum payment to the peer in each experimental session because we want to ensure that the lump-sum payment does not convey different information about earnings to be expected from the working task. If the lump-sum payment conveys any information in our experiment, this information would be conveyed by the average of the two possible lump-sum payments which is constant across treatments.

As the working task, we chose the same task as used in Gagnon et al. (2020) where workers are asked to copy lines into a text box. Each worker receives £0.06 per correctly entered line. In addition, workers received a fixed payment of £2. This task does not require any prior knowledge and the performance of the worker is easily measurable. The task is clearly artificial and it should be obvious that the performance of a subject is of no value to the experimenter. Consequently, this minimizes any tendency for workers to use effort in the experiment to reciprocate the experimenter.

The task gets increasingly harder over time, such that the length of the lines increases with the number of completed lines whereas the piece-rate remains unchanged. Consequently, engaging in the task should also become less attractive for the workers over time. The workers are instructed that they can work at most for 90 minutes but they can stop working at any point in time. This is identical to Gagnon et al. (2020) with the exception that workers have slightly more time and we introduced a higher maximum number of solved lines. We do this to allow the majority of subjects to work as much as they want. The workers were unaware that there was a maximum number of lines they could solve during the experiment. We use the number of lines entered correctly as the measure of effort provision. As the subject works in isolation, other sources of peer effects, other than earnings comparisons, are mitigated (e.g. imitation, learning, or peer pressure). Then, the worker answers a short questionnaire and the experiment is over.

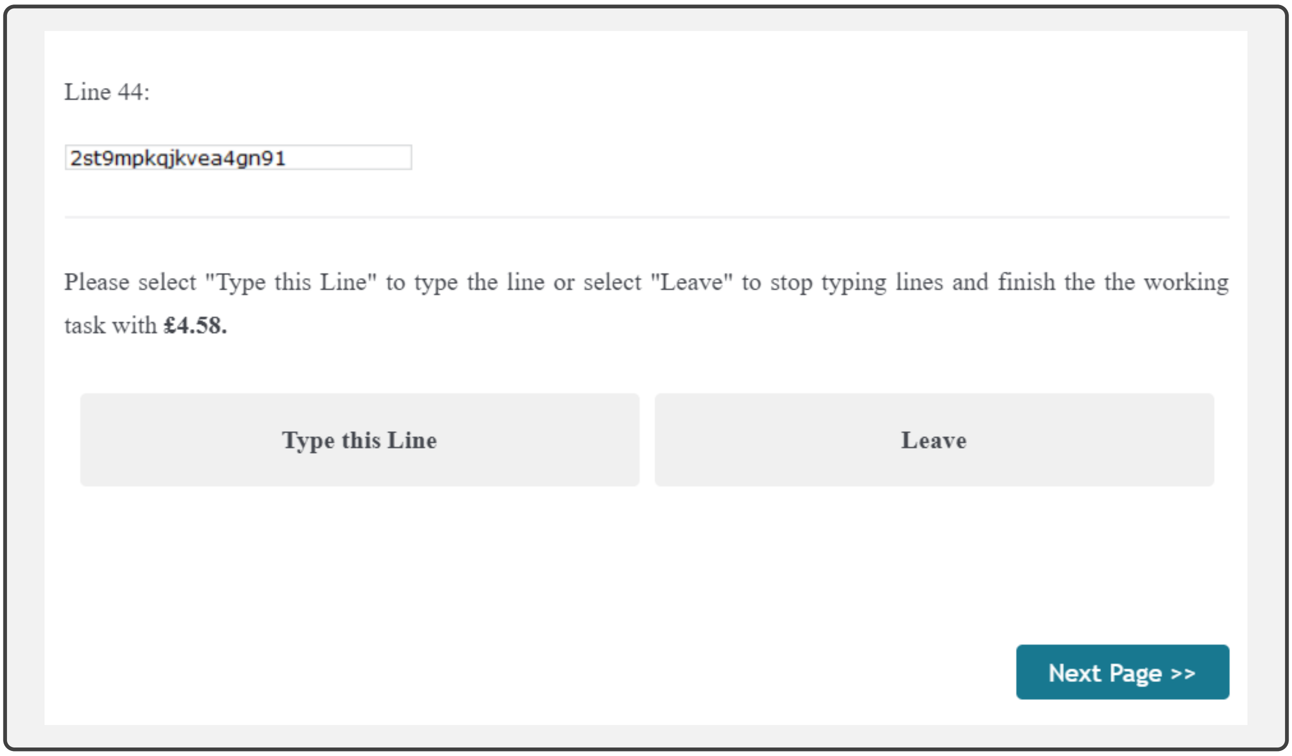


Figure 2.2: Screenshot of the working task

The questionnaire includes questions about whether the subject remembers the size of the lump-sum payment, regarding their gender, the gender of their peer, the perceived social closeness to their peer (Gächter et al., 2015), and self-reported competitiveness. Moreover, we elicit (without incentives) parameters of a Fehr-Schmidt inequality aversion model following Blanco et al. (2011). To elicit the aversion against disadvantageous equality α_i we use responder decisions from an ultimatum game.[1] For eliciting the aversion against advantageous inequality β_i we use a modified dictator game.[2] We use non-incentivized measures of α_i and β_i to ensure that the only source of payoff comparisons is the earnings from the work task.

2.4 Theoretical Predictions & Hypotheses

2.4.1 Theoretical Predictions

Consider two individuals, worker i and its peer j. Worker i is engaged in a real-effort task receiving a show-up fee R and piece-rate wages we, whereas its peer j receives a lump-sum payment F. Worker i chooses her optimal effort level e^* given her wage w, the social reference point F and her cost function $c(e; \lambda_i)$ including an individual-specific cost parameter λ_i. For simplicity, we assume the cost function to be quadratic. Worker i chooses her optimal effort level to maximize her utility function:

$$U_i = R + we_i - \alpha_i \max\{(F - R - we_i, 0)\} - \beta_i \max\{(R + we_i - F, 0)\} - \frac{\lambda_i e_i^2}{2} \tag{2.1}$$

The first and second term on the RHS of equation (1) corresponds to the utility of monetary earnings derived from working, the third and fourth term account for a possible disutility in consequence of payoff inequalities and the final term is the utility cost of providing effort. The modeling of disutility created by a payoff inequality follows Fehr and Schmidt (1999). The third term of the RHS measures the disutility from disadvantageous payoff inequality and the fourth term measures the disutility from advantageous payoff inequality, with $\alpha_i \geq \beta_i \geq 0$.

[1] Suppose x_i' is the lowest offer the responder accepts and $x_i' - 1$ is the highest offer the responder rejects, where offers are integers between 0 and 10 out of a total amount of 20. Then a responder with well-behaved preferences is indifferent between accepting some offer $x_i \in [x_i' - 1,\ x_i']$ and both players receiving nothing. From $U_i\,(x_i, 20 - x_i) = x_i - \alpha_i\,(20 - x_i - x_i) = 0$ we can determine the point of indifference $\alpha_i = \frac{x_i}{2(10 - x_i)}$. Following Blanco et al. (2011) we approximate x_i as the average of x_i' and $x_i' - 1$ when there is no more than one switch-point in the responder's strategy. The workers who accept all offers are assigned $\alpha_i = 0$ and the workers who reject all offers are assigned $\alpha_i = 4.5$.

[2] We obtain β_i by finding the egalitarian allocation $(x_i,\ x_i)$ that makes the dictator indifferent between keeping the entire endowment (10, 0) and $(x_i,\ x_i)$. Suppose the dictator switches to the egalitarian outcome at $(x_i',\ x_i')$. That is, he prefers $(10, 0)$ over $(x_i' - 1,\ x_i' - 1)$ but $(x_i',\ x_i')$ over $(10, 0)$. We conclude that he is indifferent between $(10, 0)$ and the (x_i, x_i) egalitarian outcome where $x_i \in [x_i' - 1, x_i']$ and $x_i' \in \{1, \ldots, 10\}$. Thus, $\beta_i = 1 - \frac{x_i}{10}$. Again, we approximate x_i as the average of x_i' and $x_i' - 1$. The workers who always choose the egalitarian outcome are assigned $\beta_i = 1$ and the workers who never choose the egalitarian outcome are assigned $\beta_i = 0$.

Consequently (1) collapses to:

$$U_i = \begin{cases} R + we_i - \alpha_i(F - R - we_i) - \frac{\lambda_i e_i^2}{2} & \text{if } we_i < F - R \\ R + we_i - \beta_i(R + we_i - F) - \frac{\lambda_i e_i^2}{2} & \text{if } we_i \geq F - R \end{cases} \tag{2.2}$$

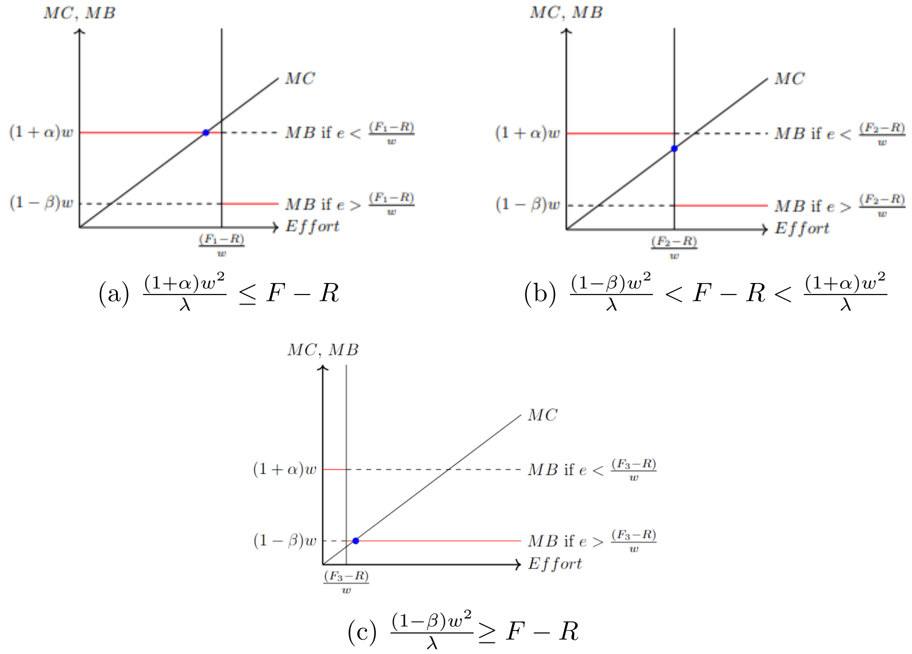
Differentiating (2), we obtain the following FOCs and optimal levels of effort:

$$\frac{\partial U_i}{\partial e} = \begin{cases} w + \alpha_i w - \lambda_i e_i & \text{if } we_i < F - R \\ w - \beta_i w - \lambda_i e_i & \text{if } we_i \geq F - R \end{cases} \tag{2.3}$$

$$e_i^* = \begin{cases} \frac{(1+\alpha_i)w}{\lambda_i} & \text{if } \frac{(1+\alpha_i)w^2}{\lambda_i} \leq F - R \\ \frac{F-R}{w} & \text{if } \frac{(1-\beta_i)w^2}{\lambda_i} < F - R < \frac{(1+\alpha_i)w^2}{\lambda_i} \\ \frac{(1-\beta_i)w}{\lambda_i} & \text{if } \frac{(1-\beta_i)w^2}{\lambda_i} \geq F - R \end{cases} \tag{2.4}$$

The marginal utility in (3) can be viewed as the difference between marginal benefit and marginal cost of entering lines, where the marginal benefit depends on whether earnings exceed or fall short of the reference earnings. The optimal effort in (4) consequently depends on whether the reference earnings are comparatively large or small. This is in contrast to the standard model where the optimal level is $e_i^* = \frac{w}{\lambda_i}$ independently of F.

Figure 2.3 graphically illustrates the three cases to consider for a given w, R, λ, α and β. The reference earnings can be either (a) high, (b) medium or (c) low. Panel (a) describes the case where the individual faces high reference earnings. Here, she chooses the level of effort that corresponds to the intersection between her marginal cost function and her marginal benefit function which yields earnings below the reference earnings. Consequently, if the reference earnings are shifted further to the right, this does not alter her optimal level of effort provision. Panel (b) describes the case where the individual faces an intermediate social reference point. This case illustrates the situation where she chooses her level of effort such that she matches the earnings of her peer. Consequently, if the reference point is shifted further to the right, she increases her level of effort such that it corresponds to the earnings of the new social reference point or up to the point when the marginal cost function and marginal benefit function intersect. Panel (c) describes the case where the individual faces a low social reference point. If the



(a) $\frac{(1+\alpha)w^2}{\lambda} \leq F - R$ (b) $\frac{(1-\beta)w^2}{\lambda} < F - R < \frac{(1+\alpha)w^2}{\lambda}$

(c) $\frac{(1-\beta)w^2}{\lambda} \geq F - R$

Figure 2.3: Optimal individual effort for a given w, R, λ, α and β when the reference earnings are (a) high, (b) medium or (c) low.

reference point is shifted to the right this might have three possible different effects. If the shift is only very slight, the optimal level of effort does not change. If the shift is moderate, she increases her level of effort such that she matches the reference earnings. If the shift is large, she increases her effort level to the point where the marginal cost function and marginal benefit function intersect. Consequently, an increase in the social reference point leads to at least a weakly higher level of effort in all three panels.

Theorem 2.1 *Higher social reference points lead to weakly higher levels of effort.*

Note that the assumption that $\beta_i > 0$ can be relaxed. Our predictions also hold, as long as we assume $\alpha_i > |\beta_i|$. Under this condition, the marginal benefit function $(1+\alpha_i)w$ is always above the marginal benefit function $(1-\beta_i)w$ and consequently intersects the marginal cost function at a higher level of effort.

2.4.2 Hypotheses

Given our theoretical analysis, we arrive at the following testable hypotheses for the comparison between the two treatments:

Main Hypothesis: The distribution of effort in the "high reference point treatment" stochastically dominates the distribution of effort in the "low reference point treatment".

We test this hypothesis using a one-sided Wilcoxon-rank sum test. To determine the appropriate sample size we conducted a power analysis. Based on the results of this we determined that for our experimental parameters a sample of 180 worker-peer pairs, randomly assigned to *High treatment* and *Low treatment* with equal probability would provide adequate power ($> 80\%$) of the test (see Appendix 2A for details).

According to our theoretical analysis, the optimal effort level can be determined by two processes. Either the individual chooses the effort level that corresponds to the intersection between the marginal benefit function and the marginal cost function, or she matches the reference earnings. Assuming that we have cases where individuals choose the effort level that corresponds to the reference earnings, our model suggests not only that the effort levels differ between treatments, but they differ in a very specific way. Consequently, our subsidiary hypothesis predicts a higher probability of matching the earnings of the high reference point in the high reference point treatment compared to the low reference point treatment.

Subsidiary Hypothesis: The probability for the workers to stop at earnings = size of the high reference point is higher in the "high reference point treatment" compared to the "low reference point treatment".

While our theory also predicts a higher likelihood for workers to stop at the low reference point in the low reference point treatment compared to the high reference point treatment, we do not formalize this as a testable hypothesis. The reason is that, based on our experimental parameters and assumptions about the distribution of effort cost and inequality aversion parameters, we do not have sufficient power to test this prediction (see Appendix 2A for details).

From the perspective of expected utility theory as well as models of status-quo-based (Tversky and Kahneman, 1979) and expectations-based (Kőszegi and Rabin, 2006) reference points, no treatment effect on effort provision is predicted since decision-makers faced the same working task across treatments and knew that their earnings were determined solely by their performance. Consequently, we are confident that a difference in effort between the "high reference point treatment" and the "low reference point treatment" can be solely attributed to the effect of peer earnings.

2.5 Results

2.5.1 Main Analysis

In total, 360 subjects participated in the experiment. 180 of whom participated in the work task and 180 of whom were allocated the role of a passive peer. The random assignment to treatments resulted in 102 workers in the *High treatment* and 78 workers in the *Low treatment*. The workers in both treatments are well balanced. There are no significant differences between the two treatments in terms of gender (54.9% in *Low*, 55.1% in *High*). This resulted in 53% of workers paired with a peer of the same gender (52.9% in *Low*, 53.8% in *High*). Further, the level of competitiveness elicited in the post-experimental questionnaire is very similar (4.92 in *Low*, 5.06 in *High*). All sessions were conducted online with students of the University of Nottingham in June 2021. As we preregistered, we use one-sided tests for our hypotheses and two-sided tests for the more exploratory analysis.

It appears that workers cared about the earnings of their peer. The large majority of the workers could remember the earnings of their peer very precisely. Further and especially interesting, the workers were significantly happier when they were in the *Low treatment* compared to the *High treatment* (4.40 in *Low*, 3.87 in *High*, two-sided Wilcoxon rank-sum test, p-value= 0.030). This shows that workers cared about the income of their peers on a psychological level. However, this did not translate to a change in working behavior in our experimental task.

We do not find support for our main hypothesis in the original experiment. In the *Low treatment* with peer earnings of £2.90, the workers solved on average 110.34 lines correctly. In the *High treatment* with peer earnings of £7.10, the workers solved on average 112.21 lines. The treatment difference of 1.78 correctly solved lines is negligible and corresponds to a marginal effect of 1.02%. To test for the equality of effort provision we use the one-sided Wilcoxon rank-sum test yielding a p-value of 0.495.[3]

[3]In our analysis, we give priority to non-parametric tests because they do not assume that error terms are normally distributed.

Result 1: The workers do not work significantly harder in the *High treatment* compared to the *Low treatment* in the original experiment.

What makes the absence of a treatment effect in terms of effort provision especially interesting is that workers were less happy with their earnings in the *High treatment* compared to the *Low treatment*. The treatment difference in terms of happiness remains significant once we control for the individual performance in the working task in an OLS regression (p-value < 0.01). Together, this suggests that the workers cared about the earnings of their respective peers but did not change their behavior accordingly. A possible reason why we do not observe a treatment effect in effort provision could be that some subjects reduce their effort when they observe a high reference point. Then, we would expect a greater variance in effort provision in the *High treatment* compared to the *Low treatment* which we do not find (two-sided Kolmogorov–Smirnov test, p-value=0.844).

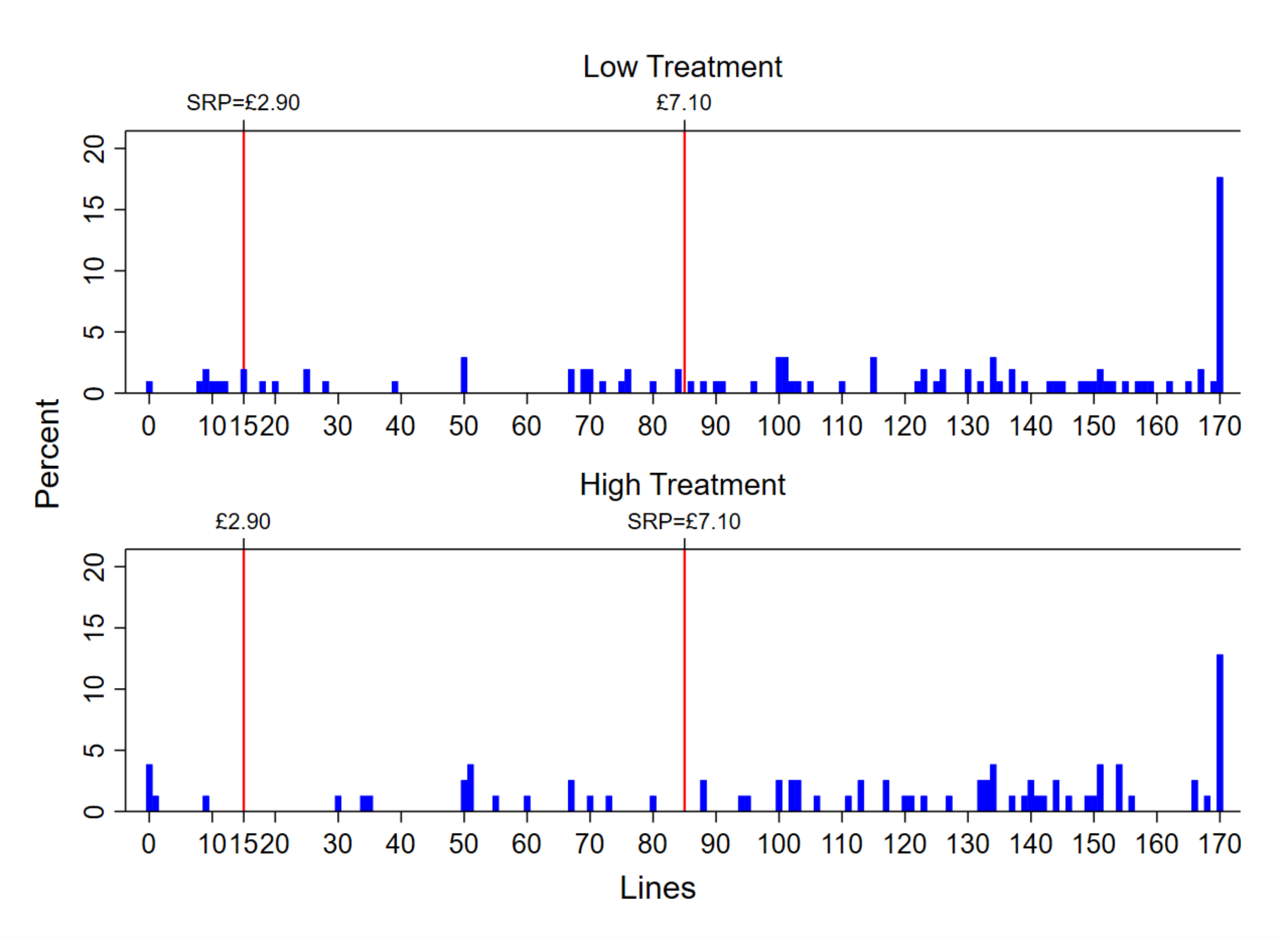


Figure 2.4: Histograms of correctly solved lines by treatment where the social reference points are indicated by the red lines in the original experiment

The modal choice in both treatments is to solve the maximum number of 170 lines. Based on results in Gagnon et al. (2020) we expected the participants to exert considerably less effort and did not expect the ceiling to be binding. This might indicate that the working task has been perceived as too easy and not tedious enough in our experimental setting. It could be that the cost of effort is not only determined by the number of lines solved but also by the actual time workers spend on the working task. We, therefore, consider the time spent working as an alternative measure of effort provision. The average times working in both treatments do also not differ significantly from each other. The workers in the *High treatment* work on average for 53.92 minutes, while the subjects in the *Low treatment* work on average for 52.98 minutes (Wilcoxon rank-sum test, p-value= 0.474).

As shown in Section 4, our model does not only predict treatments to be different but to be different in a very specific way. Our subsidiary hypothesis predicts a higher probability of stopping when the accumulated earnings equal the peer's earnings in the *High treatment.*

Figure 2.4 shows a histogram of correctly solved lines for each treatment (*Low treatment* in the top panel, *High treatment* in the bottom panel). First of all, one can see that stopping decisions are dispersed over a wide range. Some workers stop directly or very early, while others solve all 170 lines. This heterogeneity is what one would expect given that productivity, cost of effort, and opportunity cost differ across workers. We are interested to see whether there are systematic differences in the clustering between treatments in terms of clustering of stopping decisions exactly at the peer's earnings: In neither treatment does a subject stop at exactly 85 lines, which corresponds to the income of the peer in the *High treatment.* This does not allow us to test our hypothesis using our preregistered test. If we compare the number of workers stopping in the range between 80 and 90 correctly solved lines, we find no significant differences between treatments (one-sided Fisher's exact test, p-value= 0.474). This also holds for the ranges between 75 and 95 (one-sided Fisher's exact test, p-value= 0.258) and 70 and 100 correctly solved lines (one-sided Fisher's exact test, p-value= 0.260).

Result 2: The probability for the workers to stop at earnings = size of the high reference points is not higher in the *High treatment* compared to the *Low treatment* in the original experiment.

2.5.2 Discussion of Main Results

Since we do not find a significant difference regarding the levels of effort provision between treatments, we use additional data from our questionnaire to check whether our design induced a salient reference point. To do this, we investigate whether workers remembered the reference point while working on the task. Finally, we discuss whether our results could be related to the degree of behindness aversion and leading aversion of our subject pool.

Salience

Given that we found no differences in effort provision between the two treatments, it is an important question to address whether the workers remembered the reference point while they did the task. In our questionnaire, we asked the participants who took part in the work task whether they could remember the amount of the payment we made to the other participant. 75% (73.5% in *Low*, 77% in *High* of all workers remember the exact size of the reference point). If we allow for a deviation of ± 1 pound, 97.8% of all workers answer correctly. In summary, this means that the workers can remember their peer's income fairly well, but it does not seem to influence their decision to work in a significant way. We also asked the workers in our questionnaire what factors influenced their decision to work. 26.1% of all workers mention their peer's income as an influencing factor for their effort provision. Interestingly, workers are significantly more likely to mention their peer earnings in the *High treatment* (37.1%) compared to the *Low treatment* (17.6%). We do not find a statistically significant difference in effort provision between treatments if we restrict the sample to workers who mentioned their peer earnings as a relevant factor (Wilcoxon rank-sum test, p-value= 0.637). Together with the finding that workers are significantly less satisfied with their earnings when their peer earns relatively more, this yields suggestive evidence that the treatment manipulation itself worked well but did not lead to significant changes in the effort provision of the workers.

Behindness Aversion

In our questionnaire, we elicited a measurement of each subject's level of behindness aversion. One possible reason why our treatment manipulation does not lead to the expected results is that a large fraction of our workers do not care about the income of others which would be reflected in very low levels of behindness aversion. The alphas we elicited in our experiment are on average 0.71. This is very much in line with the literature. For instance, Beranek et al. (2015) elicited alphas and betas at the same university in an incentivized way. Comparing the distribution of alphas between the two studies, we find that workers in our experiment are slightly more likely to report smaller values of alpha. In our experiment, 63.3% of the workers report an alpha below 0.4, whereas in their experiment 54% do so but overall the distributions are very similar (mean in our experiment = 0.71, mean in Beranek et al. (2015) = 0.75). Originally, we expected that workers with a higher degree of behindness aversion would respond more strongly to the treatment manipulation than people with a lower degree. We can find no support for this in our experimental data. We find no significant relationship between the degree of behindness aversion and the number of solved lines using an OLS regression (p-value= 0.733). We would not expect the degree of behindness aversion to have an effect in the *Low treatment*, but only in the *High treatment.* To account for this we add an interaction term between the treatment and the effort provision but the coefficient for alpha remains insignificant and so is the interaction term (p-value= 0.762). We obtain qualitatively similar results if we exclude the extremes on both sides ($\alpha = 0$ and $\alpha = 4.5$). Overall, the proportion of alphas is not far from what we expected and does not explain the lack of a treatment effect.

Leading Aversion

In our questionnaire, we have elicited a measurement of each subject's aversion to earning more than their peer. Initially, we did hypothesize that betas should not play a substantial role in the individual's decision of how much effort to provide. The betas we elicited in our experiment are on average 0.49 (where 6.98% have a beta of 0 and 8.72% have a beta of 1). If anything, we expected that workers with a higher degree of leading aversion should provide less effort. Surprisingly, this is not what we see in the data. First, if we run a simple OLS regression of the betas on the number of correctly solved lines we find a positive significant relationship (p-value= 0.009). However, if we exclude the extreme values ($\beta = 0$ and $\beta = 1$), this relationship vanishes. Comparing the distribution of betas to Beranek et al. (2015), we find that workers in their experiment are slightly more likely to report smaller values of beta but overall the distributions are very similar. In our experiment, 38.95% of the workers report a beta below 0.5, whereas in their experiment 54% do so (mean in our experiment = 0.49, mean in Beranek et al. (2015) = 0.48). Even though our measures for α and β are not incentivized, we find overall very similar results using the same subject pool.

2.5.3 Further results

Table 2.1: OLS Regressions: Effort provision as dependent variable

	(1)	(2)	(3)
Treatment (1=high)	1.774	1.741	-0.170
	(7.642)	(7.524)	(7.713)
Female		20.865***	22.464***
		(7.545)	(7.569)
Same gender		-1.638	-6.599
		(7.524)	(7.551)
Competitiveness			9.100***
			(2.703)
Oneness			-1.969
			(4.189)
α_i			-1.581
			(2.888)
β_i			29.246*
			(14.899)
Constant	110.431***	99.843***	45.0376**
	(5.030)	(7.329)	(17.229)
Observations	180	180	171

The table reports estimates from OLS regressions. The dependent variable is the number of solved lines in the real-effort task. No workers are excluded from the analysis in columns (1) and (2). In column (3) workers with multiple switchpoints for our measures of α or β are excluded. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Here we report some additional analyses that go beyond our pre-registered main analysis. In our experiment, females tend to solve on average 20 lines more than males. This difference is large and statistically significant (Wilcoxon rank-sum test=0.029). We also find that more competitive workers solve more lines using an OLS regression (p-value= 0.005). It might be that more competitive people react more strongly to the treatment. To account for this we add an interaction term between the treatment and the degree of competitiveness but the interaction term is insignificant (p-value= 0.298). The gender composition of each pair does not seem to play a substantial role. There is no statistically significant difference in effort provision between a worker who is matched with someone of their gender and someone who is not (Wilcoxon rank-sum test=0.982). Similarly, the degree to which the workers feel connected does not seem to have a substantial effect. If we regress effort provision on the oneness, we find a weak and statistically insignificant relationship (p-value= 0.820).

Our results are supported by the OLS regressions in 2.1. We regress the number of solved lines for each subject on a treatment dummy (see Table 2.1, column 1). The treatment difference is insignificant and stays insignificant when we control for observable characteristics (see Table 2.1, column 2) and the non-observable characteristics we elicited in the questionnaire (see Table 2.1, column 3). We control for the subject's gender, the gender composition of the group, the subject's degree of competitiveness, the perceived oneness of the group and our measures of behindness and leading aversion. The only highly significant control variables are gender and the level of competitiveness. This again suggests that females and more competitive people worked harder in our experiment.

We checked for gender differences in the reaction to the treatment manipulation. Men reacted positively to the treatment manipulation as they increased their average level of effort provision by 12.63 lines, women on the other hand decreased their average level of effort by 7.22 lines. However, the differences between the *High treatment* and the *Low treatment* are statistically insignificant for both subgroups (two-sided Wilcoxon rank-sum tests, p-value= 0.303 respectively 0.369).

2.6 Discussion & Recalibrated Experiment

2.6.1 Discussion

The workers in our experiment clearly solved more lines than expected given the findings of Gagnon et al. (2020). Whereas in Gagnon et al. (2020), workers solved on average 44 lines correctly, they solved more than 2.5 times as many in our experiment. There are different potential reasons why this may have happened. First, we used different subject pools. Whereas Gagnon et al. (2020) conducted their study on Prolific, we used the student subject pool of CeDEx. Another and potentially more important reason than the subject pool itself is that workers on Prolific can enter Prolific and search for experiments that are currently online. In contrast to that, we invited workers via Orsee and told them that the experiment could take up to two hours. It is likely that some workers on Prolific would need to leave the experiment at some point due to time restrictions whereas this is less likely for our experiment. Further, some workers on Prolific may have the possibility to leave the experiment and instead participate in another one which makes it more likely for them to quit if they expect higher earnings from switching.

The difference in effort levels between our study and the study of Gagnon et al. (2020) is problematic because we assume similar levels of effort for the calibration of our experiment. For this reason, we re-calibrated the working task by making the lines grow faster in the course of the experiment (now they increase by two characters instead of one character every five correctly solved lines). This way we are able to observe a larger fraction of our workers that are below the high social reference point.

2.6.2 Recalibrated experiment

Main Analysis

We ran the identical experiment again with the exception that the lines increased now by two characters instead of one character for each five correctly solved lines. In total, the experiment consisted of 250 subjects of whom 125 subjects participated in the work task and 125 were passive peers. We collected 67 observations for the *High treatment* and 58 observations for the

Low treatment respectively. The randomization to treatments lead to 63.8% females in the *High treatment* and 55.2% in the *Low treatment*. In both treatments, roughly 50% of the workers were paired with a peer of the same gender (50% in *Low* and 50.7% in *High*). All sessions were conducted online with students of the University of Nottingham in July 2021.

As in the original experiment, we do not find statistically significant differences between the treatments in terms of effort provision. In the *Low treatment*, the workers solved on average 57.65 lines correctly. In the *High treatment*, the workers solved on average 62.41 lines. Again, we use a one-sided Wilcoxon rank-sum test to test for the equality of effort provision between treatments yielding a p-value of 0.270. As in the original experiment, we do not find this effect even though workers in the *High treatment* are significantly less happy (two-sided Wilcoxon rank-sum test, p-value= 0.053). The treatment difference in terms of happiness is even stronger once we control for the individual performance in the working task in an OLS regression (p-value=0.019).

Result 3: The workers do not work significantly harder in the *High treatment* compared to the *Low treatment* in the recalibrated experiment.

Figure 2.5 shows a histogram of correctly solved lines for each treatment (*Low treatment* in the top panel, *High treatment* in the bottom panel). Again, one can see that the stopping decisions are dispersed over a wide range. Some workers stopped directly or very early, however, this time no subject was able to solve all 180 lines. It is easy to see that the distribution of effort is shifted to the left compared to the original experiment. Again, the distributions of effort are very similarly dispersed between treatments (two-sided Kolmogorov–Smirnov test, p-value=0.506).

Our subsidiary hypothesis predicts a higher probability of stopping when the accumulated earnings equal the peer's earnings in the *High treatment*. We are interested to see whether there are systematic differences in terms of clustering of stopping decisions exactly at the peer's earnings: In neither treatment does a subject stop at exactly 85 lines, which corresponds to the income of the peer in the *High treatment*. If we compare the number of workers stopping in the range between 80 and 90 correctly solved lines, we find no significant differences between treatments (one-sided Fisher's exact test, p-value= 0.442). This also holds true for the ranges between 75 and 95 (one-sided Fisher's exact test, p-value= 0.172) and 70 and 100 correctly

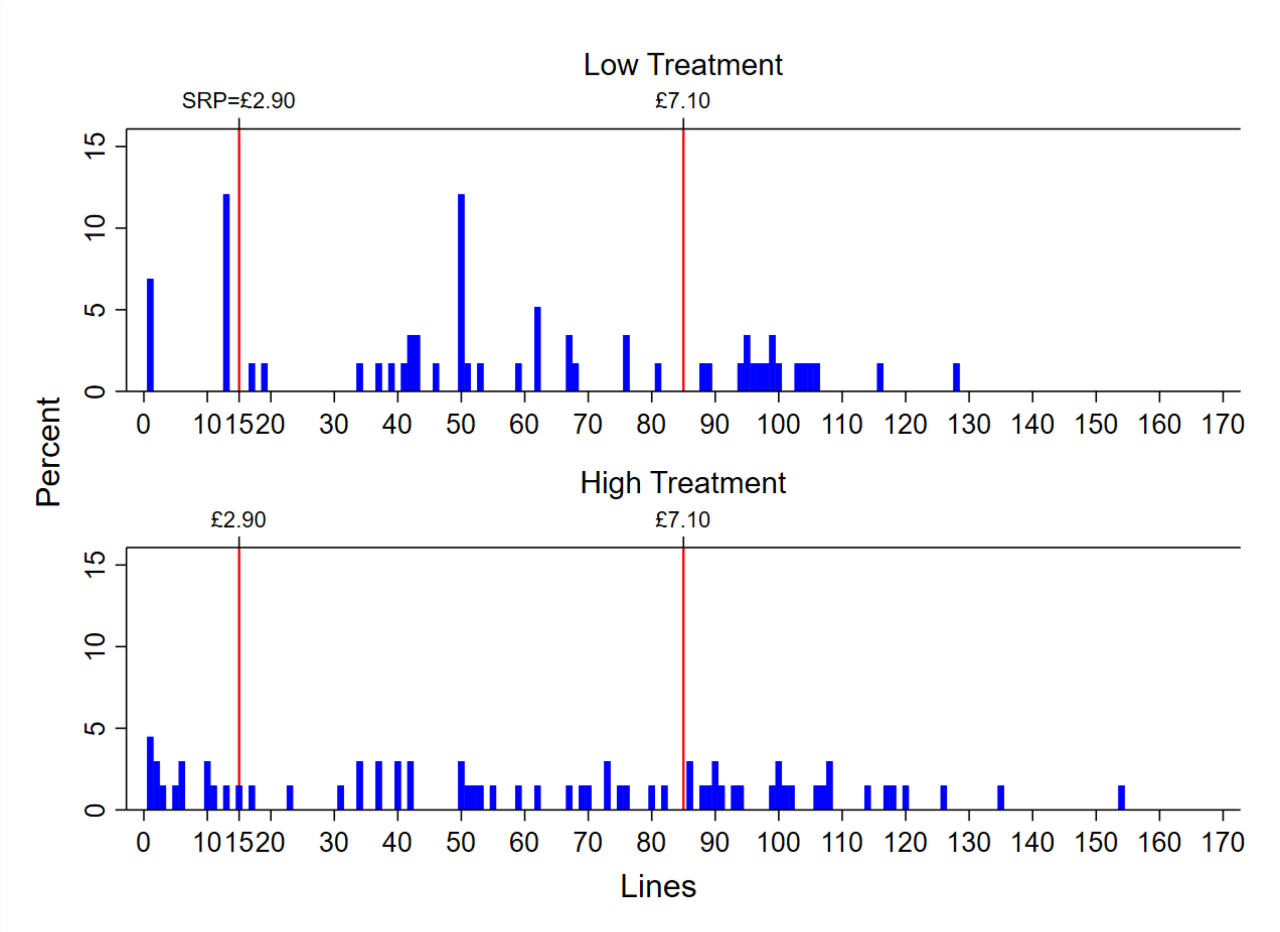


Figure 2.5: Histograms of correctly solved lines by treatment where the social reference points are indicated by the red lines in the re-calibrated experiment

solved lines (one-sided Fisher's exact test, p-value= 0.509).

Result 4: The probability for the workers to stop at earnings = size of the high reference points is not higher in the *High treatment* compared to the *Low treatment* in the recalibrated experiment.

Further Results

As in our original experiment, subjects remembered the earnings of their peer fairly well - 80% of the workers remember the exact size of their peer's earnings and 98.4% if we allow for a deviation of ± 1. Again, we find no effect of the degree of behindness aversion on the level of effort provision using a simple OLS regression (p-value= 0.779) and this lack of treatment effect continues once we add an interaction term (p-value= 0.423). In contrast to the original experiment, we also find no effect of individual betas on the number of correctly solved lines (p-value= 0.376). Again, females work significantly harder than males using a two-sided

Wilcoxon rank-sum test (p-value= 0.005). Similarly to our original experiment we find that more competitive workers supply more effort (p-value= 0.050) and the degree of oneness does not seem to play a substantial role in determining effort provision using OLS regressions (p-value= 0.634). In contrast to our original experiment, workers who have been matched with a person of the same gender work significantly harder (Wilcoxon rank-sum test, p-value= 0.018). These results are confirmed by the regression in Table 2.2 .

Table 2.2: OLS Regressions: Effort provision as dependent variable

	(1)	(2)	(3)
Treatment (1=high)	4.763	5.990	3.266
	(6.774)	(6.443)	(6.851)
Female		15.797**	14.947**
		(6.755)	(7.092)
Same gender		16.955**	16.983**
		(6.614)	(7.067)
Competitiveness			4.777**
			(2.270)
Oneness			0.298
			(3.818)
α_i			-0.240
			(2.713)
β_i			5.814
			(13.374)
Constant	57.655***	39.100***	16.463
	(4.960)	(6.662)	(15.500)
Observations	125	125	116

The table reports estimates from OLS regressions. The dependent variable is the number of solved lines in the real-effort task. No workers are excluded from the analysis in columns (1) and (2). In column (3) workers with multiple switchpoints for our measures of α or β are excluded. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Again we check for heterogeneous treatment effects for men and women. In contrast to our original experiment, women solved on average 11.64 lines more when being allocated to the *High treatment*. As in our original experiment, this difference is not statistically significant (two-sided Wilcoxon rank-sum test, p-value= 0.251). Also in contrast to our original experiment, men worked slightly less in the *High treatment* compared to the *Low treatment* and again this difference is not statistically significant (two-sided Wilcoxon rank-sum test, p-value= 0.592).

2.7 Conclusion

Using a simple laboratory experiment, we test whether social reference points influence real-effort provision in a working task. Our experiment is based on a simple model of social reference points, following Fehr and Schmidt (1999). We design an experimental paradigm that allows us to measure the level of effort provision precisely while exogenously manipulating a reference point. This allows us to identify the effect of a social reference point by comparing the level of effort provision between treatments.

In our experiment, only two subjects participate in each session. Each subject is assigned one of two roles worker or peer, by a publicly observed coin toss. The peer receives a fixed payment, whereas the worker participates in a real-effort task. The fixed payment to the peer is varied between the two treatments. This allows us to trace any difference in the average level of effort provided by the workers across treatments back to social reference points.

We find that the workers in our experiment care about the earnings of their respective peers, i.e. they remember the earnings of their peers and are less happy with their earnings when their peer earns relatively more. Despite this, we cannot observe a change in working behavior. The workers in our experiment do not increase their level of effort when faced with relatively high peer earnings compared to low reference earnings. This suggests that people care about income differences but this does not necessarily translate to a change in behavior in incentivized environments.

To test the robustness of our findings we conducted a follow-up experiment with a modified and more difficult working task. The results are in line with our original experiment. Again, workers were significantly less happy when their peer earned relatively more. As in our original experiment, the difference in satisfaction did not lead to a significant treatment effect in terms of real-effort provision in the work task.

We now turn to the issue of why we do not observe a treatment effect. Festinger (1954) discusses what makes a reference point salient. He argues that people who are most like you are the most influential point of comparison. While our participants are very similar in many ways, they differ in how they can earn their income. Whereas workers earn their income by providing effort, the income of peers is determined by luck. One possible explanation for why we do not find differences in effort provision between the treatments could be that workers are less likely to use the income of a peer as a relevant point of comparison if it does not result from the same work task. However, it remains puzzling that we do find that workers care about the income of their peers but it did not translate to changes in economic behavior.

It is interesting to compare our results with the literature on wage inequalities that suggests that others' earnings might modify behavior. For instance, Bracha et al. (2015) and Gagnon et al. (2020) find that workers work less when they earn less than others for the same work task. However, this effect can not be attributed to relative income comparisons but must be interpreted as a net effect of wage discrimination. It is possible that there is a positive effect of income targeting on effort provision hidden under a stronger negative effect of discrimination. Our experiment shuts down several channels that might influence effort provision and cleanly tests the effect of relative income comparisons on effort provision.

One possible interpretation of our results is that fixed peer outcomes do not serve as reference points. However, there is some evidence that salient peer outcomes in the way we implement them in our experiment can serve as reference points and modify behavior. Schwerter (2023) finds that people are more willing to take risks when they face higher social reference points using a similar design. The experiments of Schwerter (2023) and us differ in that workers in his experiment make only one payoff-relevant decision about a lottery, while in our experiment they work for up to 90 minutes. Social reference points as used in our experiments may have short rather than long-term effects on economic behavior. A possible explanation would be that social reference points might be a transient thing where the duration and size of the effect depend on how strong the point of comparison is. However, the workers in our experiment remembered the social reference point by the end of the experiment and cared about it at least on a psychological level, so it is hard to view the effects of the social reference point in our experiment as transient.

For future research, it would be interesting to investigate in more detail where and why relative incomes serve as social reference points, and how this affects labor supply behavior.

2.8 References

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). Reference points and effort provision. *American Economic Review 101*(2), 470–92.

Akerlof, G. A. and J. L. Yellen (1990). The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics 105*(2), 255–283.

Austin, W., N. C. McGinn, and C. Susmilch (1980). Internal standards revisited: Effects of social comparisons and expectancies on judgments of fairness and satisfaction. *Journal of Experimental Social Psychology 16*(5), 426–441.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review 84*(2), 191.

Beranek, B., R. Cubitt, and S. Gächter (2015). Stated and revealed inequality aversion in three subject pools. *Journal of the Economic Science Association 1*(1), 43–58.

Blanco, M., D. Engelmann, and H. T. Normann (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior 72*(2), 321–338.

Bracha, A., U. Gneezy, and G. Loewenstein (2015). Relative pay and labor supply. *Journal of Labor Economics 33*(2), 297–315.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review 102*(6), 2981–3003.

Clark, A. E. and A. J. Oswald (1996). Satisfaction and comparison income. *Journal of Public Economics 61*(3), 359–381.

Cohn, A., E. Fehr, B. Herrmann, and F. Schneider (2014). Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association 12*(4), 877–898.

Falk, A. and A. Ichino (2006). Clean evidence on peer effects. *Journal of Labor Economics 24*(1), 39–57.

Fehr, E. and S. Gächter (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives 14*(3), 159–181.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Ferreri-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics 89*(5-6), 997–1019.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations 7*(2), 117–140.

Gächter, S., L. Huang, and M. Sefton (2018). Disappointment aversion and social comparisons in a real-effort competition. *Economic Inquiry 56*(3), 1512–1525.

Gächter, S., C. Starmer, and F. Tufano (2015). Measuring the closeness of relationships: a comprehensive evaluation of the'inclusion of the other in the self'scale. *PLOS ONE 10*(6), e0129478.

Gächter, S. and C. Thöni (2010). Social comparison and performance: Experimental evidence on the fair wage–effort hypothesis. *Journal of Economic Behavior & Organization 76*(3), 531–543.

Gagnon, N., K. Bosmans, and A. Riedl (2020). The effect of unfair chances and gender discrimination on labor supply. *Available at SSRN 3519540*.

Genicot, G. and D. Ray (2017). Aspirations and inequality. *Econometrica 85*(2), 489–519.

Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review 102*(1), 469–503.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Kőszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics 121*(4), 1133–1165.

Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly Journal of Economics 120*(3), 963–1002.

Marmot, M. (2005). *Status syndrome: How your social standing directly affects your health.* A&C Black.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–45.

Perez-Truglia, R. (2020). The effects of income transparency on well-being: Evidence from a natural experiment. *American Economic Review 110*(4), 1019–54.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Schwerter, F. (2023). Social reference points and risk taking. *Management Science*.

Tversky, A. and D. Kahneman (1979). Prospect theory: An analysis of decision under risk.

Econometrica 47(2), 263–291.

Appendices

2.A Power Analysis

We conducted a power analysis based on the following structural model. We assume subject i gets utility from:

$$U_i = \begin{cases} R + we_i - \alpha_i(F - R - we_i) - \frac{\lambda_i e_i^2}{2} & \text{if } R + we_i < F \\ R + we_i - \frac{\lambda_i e_i^2}{2} & \text{if } R + we_i \geq F \end{cases} \tag{2.5}$$

where R is the show-up fee and w the piece-rate wage. With this specification optimal effort is,

$$e_i^* = \begin{cases} \frac{(1+\alpha_i)w}{\lambda_i} & \text{if } R + \frac{(1+\alpha_i)w^2}{\lambda_i} \leq F \\ \frac{F-R}{w} & \text{if } R + \frac{w^2}{\lambda_i} < F < R + \frac{(1+\alpha_i)w^2}{\lambda_i} \\ \frac{w}{\lambda_i} & \text{if } \frac{w^2}{\lambda_i} + R \geq F \end{cases} \tag{2.6}$$

In the absence of reference points, or if $a_i = 0$ optimal effort is simply $e_i = \frac{w}{\lambda_i}$. However, if $a_i > 0$ the worker gets disutility from earning less than the paired subject, and this results in higher effort. Moreover, the model predicts higher effort in the high reference point treatment. We calibrate the model as follows. First, using the results from Gagnon et al. (2020) we assume $\frac{w}{\lambda_i} \sim \mathcal{N}(44.03, 28.93)$. We use a very similar real-effort task as them and use the same piece-rate parameters to their treatment which results in an average effort of 44.03 lines with a standard deviation of 28.93 lines (see Table 4 of Gagnon et al. (2020)).[4] If anything, we expect that the workers solve slightly more lines in our experiment. To account for this we will present power estimates for the results from Gagnon et al. (2020) and for the case that the average effort and standard deviation increase by 10% compared to Gagnon et al. (2020), i.e. $\frac{w}{\lambda_i} \sim \mathcal{N}(48.43, 31.82)$.

Our task parameters are $w = 0.06$ per line and $R = 2$. Next, we take a slightly conservative approach based on Fehr and Schmidt (1999) and assume that one-third of the population has

[4]Gagnon et al. (2020) impose an upper bound of 85 correctly solved lines. Unlike Gagnon et al (2020), we increase the upper bound to 170 lines. Also, we increased the amount of time for the working task from 70 to 90 minutes.

$a_i = 0$, one-third has $a_i = 0.5$ and one-third has $a_i = 1$.[5] Further we assume all betas to be zero.

To estimate the power of the test we conducted a Monte Carlo simulation with 10000 replications for varying sample sizes ($n = 60, 90, 120, 150, 180$) with the reference earnings $((F_H, F_L)) =$ (7.10,2.90). The sample size refers to the total number of workers completing the task in both treatments. We chose the reference points so that reference earnings are exactly attainable given our task parameters (i.e. $(F - 2)/0.06$ is an integer) and workers taking part in the experiment but leaving with a fixed sum receive a minimum payment and expected payment per hour consistent with our lab standards.[6]

In each replication, we drew n observations, where each observation corresponds to an experimental subject and each subject is assigned λ_i and α_i parameters by independent draws from the distributions described above. Each observation is then assigned $F = F_H$ or $F = F_L$ with equal probability to form two samples corresponding to workers assigned to complete the task with reference earnings F_H or F_L respectively. For each observation, effort is calculated using the optimal effort function.

To test our main hypothesis, we then conduct a one-sided Wilcoxon Mann-Whitney rank-sum test at the 5% significance level. The proportions of rejections in 10000 replications are our estimates of power. Table 2.A1 reports estimated power for $w/\lambda_i \sim \mathcal{N}(44.03, 28.93)$ as in Gagnon et al. (2020) and $w/\lambda_i \sim \mathcal{N}(48.43, 31.82)$.

n	$w/\lambda_i \sim \mathcal{N}(44.03, 28.93)$	$w/\lambda_i \sim \mathcal{N}(48.43, 31.82)$
60	0.5	0.44
90	0.64	0.58
120	0.75	0.68
150	0.84	0.77
180	0.88	0.83

Table 2.A1: Estimated Power for Main Hypothesis

[5] Fehr and Schmidt (1999) initially assume that 0.3 of the population has $a_i = 0$, 0.3 has $a_i = 0.5$ and 0.3 has $a_i = 1$ and 0.1 of the population has $a_i = 4$.

[6] We expect an average hourly wage of £8.10 for participating in the working task.

Our subsidiary hypothesis is that more workers match the high reference earnings in the high reference point treatment compared to the low reference point treatment. For our subsidiary hypothesis, we conduct a one-sided Fisher's exact test at the 5% significance level. Again, the proportion of rejections in 10000 replications are our estimates of power. Table 2.A2 reports the estimated power for $w/\lambda_i \sim \mathcal{N}(44.03, 28.93)$ and $w/\lambda_i \sim \mathcal{N}(48.43, 31.82)$. There is another subsidiary hypothesis that more workers match the earnings of the social reference point in the low reference point treatment. However, given our parameters, it is unlikely that this test will be rejected, e.g. for $n = 180$ the probability of rejection is less than 20%.

n	$w/\lambda_i \sim \mathcal{N}(44.03, 28.93)$	$w/\lambda_i \sim \mathcal{N}(48.43, 31.82)$
60	0.86	0.88
90	0.98	0.98
120	1	1
150	1	1
180	1	1

Table 2.A2: Estimated Power for Subsidiary Hypothesis

Based on this analysis we concluded that the proposed test procedure has adequate power ($> 80\%$) with 180 workers completing the task. Note that this implies recruiting a total of 360 subjects (recall there is a paired subject for each subject completing the task).

2.B Recruitment message

Dear fname,

You are registered with CeDEx to participate in experiments. We would like to invite you to take part in our upcoming experiment. The experiment will take place online and can take from 5 to 100 minutes. All payments will be done with PayPal.

For this experiment, you will need Microsoft Teams and your camera on for the first 10 minutes of the experiment. We will videocall you at the time you have registered for.

You will need to use a computer for this experiment. Please, do not try to use your mobile phones or tablets instead.

We are planning to run many sessions this week and more sessions might be added.

If you would like to participate, please click on the link below to sign up for the session of your choice. Please note: people that sign up for a session and do not turn up cause us problems. Please sign up only if you are sure you can attend it and if you sign up please do attend. We operate a policy of removing participants from our database who sign up but then fail to turn up.

#link#

(If the link does not work, copy it and paste it into the address field of your internet browser.)

Best regards,

CeDEx Team

2.C Welcome Speech

At the beginning of each experimental session, the following welcome speech will be given by the experimenter:

Welcome to our experiment today! The experiment consists of a simple work task! However, only one of you will participate in this task - Player A. The second person, Player B, will receive a fixed amount of money and will not participate in the work task. Both the amount of money and which one of you will participate in the work task will depend on luck! For this purpose, I will toss two coins. Both coin tosses will be visible for you. The first coin determines the amount of money, Player B will receive. If the coin shows £7.10, Player B will receive £7.10. If the coin shows £2.90, Player B will receive £2.90. Then I will toss another coin. If the coin shows green, [first name of second participant] will be Player A and will take part in the work task. If the coin shows blue, [first name of second participant] will be Player A and will take part in the work task. After that, Player A will receive a link that leads him to the instructions of the work task and both of you can leave the videocall.

2.D Instructions for Player A

Instructions

Welcome to this economic experiment! In this experiment you earn money by participating in a working task. These instructions will describe the task and how you earn money, so please read the instructions carefully.

The working task

The working task consists of entering lines of text on your computer. You will see one line at a time. Each time you see a new line, you can decide whether to type this line or leave the experiment.

If you decide to type the line you must type it correctly before going on to the next page to see the next line. In case you make a mistake when entering the line, the software will tell you so and you will have to type it again.

If you decide to leave the experiment this will end the working task, you will be asked to fill out a short questionnaire (which will take about five minutes), and you will be informed of your final payment. Note that if you decide to leave, you will not be able to start working again. That is, once you leave the working task you cannot go back.

Each time you are presented a new line you will have the option of typing it or leaving the experiment. The length of the lines will increase as you complete more lines. After each five lines are correctly entered, the length of a line increases by two characters.

You have up to 90 minutes to work on the task. However, you can finish earlier if you want by choosing to leave the experiment when you are presented with a new line.

How you earn money

When you leave the experiment according to the described procedure you will be informed of your payment. You will receive a payment of £0.06 for each line you entered correctly. In addition, you will receive a fixed amount of £2.00, irrespective of the number of lines entered.

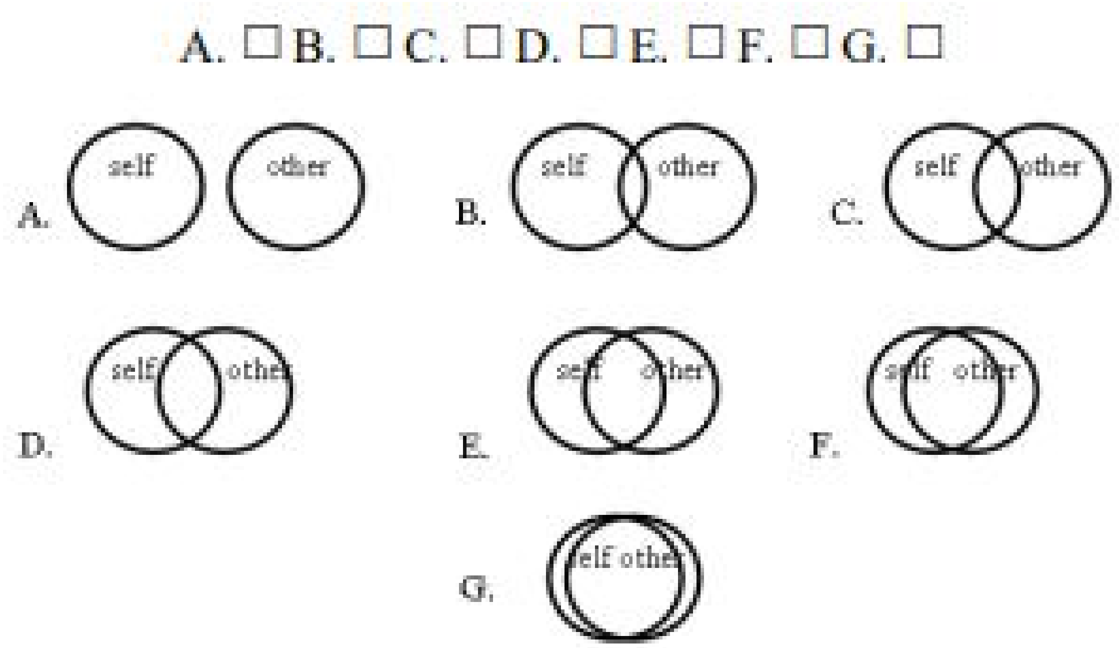
2.E Questionnaire

1. What is your gender? [male, female, neutral]

2. What is the gender of player B? [male, female, neutral, I don't know]

3. How much money did Player B receive?

4. Please, look at the circles diagram provided on your desk. Then, consider which of these pairs of circles best represents your connection with this person before this experiment. By selecting the appropriate letter below, please indicate to what extent you and Player B were connected.



5. How competitive do you consider yourself to be?" [1 (not competitive at all) -7 (very competitive)]

6. How happy are you with your earnings in today's experiment? [1 (not happy at all) -7 (very happy)]

Imagine the following hypothetical scenario: You are asked to choose between two possible allocations of money between you and another person in eleven different decision problems as presented below. Please indicate for each row which decision you prefer.

What do you prefer?			
You receive £0, the other person receives £0	○	○	You receive £0, the other person receives £20
You receive £0, the other person receives £0	○	○	You receive £1, the other person receives £19
You receive £0, the other person receives £0	○	○	You receive £2, the other person receives £18
You receive £0, the other person receives £0	○	○	You receive £3, the other person receives £17
You receive £0, the other person receives £0	○	○	You receive £4, the other person receives £16
You receive £0, the other person receives £0	○	○	You receive £5, the other person receives £15
You receive £0, the other person receives £0	○	○	You receive £6, the other person receives £14
You receive £0, the other person receives £0	○	○	You receive £7, the other person receives £13
You receive £0, the other person receives £0	○	○	You receive £8, the other person receives £12
You receive £0, the other person receives £0	○	○	You receive £9, the other person receives £11
You receive £0, the other person receives £0	○	○	You receive £10, the other person receives £10

Imagine the following hypothetical scenario: You are asked to choose between two possible allocations of money between you and another person in eleven different decision problems as presented below. Please indicate for each row which decision you prefer.

What do you prefer?			
You receive £10, the other person receives £0	○	○	You receive £0, the other person receives £0
You receive £10, the other person receives £0	○	○	You receive £1, the other person receives £1
You receive £10, the other person receives £0	○	○	You receive £2, the other person receives £2
You receive £10, the other person receives £0	○	○	You receive £3, the other person receives £3
You receive £10, the other person receives £0	○	○	You receive £4, the other person receives £4
You receive £10, the other person receives £0	○	○	You receive £5, the other person receives £5
You receive £10, the other person receives £0	○	○	You receive £6, the other person receives £6
You receive £10, the other person receives £0	○	○	You receive £7, the other person receives £7
You receive £10, the other person receives £0	○	○	You receive £8, the other person receives £8
You receive £10, the other person receives £0	○	○	You receive £9, the other person receives £9
You receive £10, the other person receives £0	○	○	You receive £10, the other person receives £10

Chapter 3

Peer earnings and giving

3.1 Introduction

The traditional economic theory of philanthropy treats charitable giving as a voluntary contribution to a public good. Early models found that the total contributions of self-interested parties are insensitive to what others have (Warr, 1983). Over time models of philanthropy grew more sophisticated and realistic by adding terms for warm glow, fairness considerations and joy-of-giving that capture utility from one's own contributions to a good (Andreoni, 1990; Fehr and Schmidt, 1999; Ribar and Wilhelm, 2002). It might also be that social comparisons in terms of peer earnings could affect to what degree people are willing to give. If peer earnings are a relevant source of comparison, it might be that higher peer earnings could motivate individuals to change their behavior to avoid unfavorable social comparisons by being less generous.

Early research from social psychology already suggested that people care about their relative status and that social comparisons might influence our behavior (Festinger, 1954; Bandura, 1977). Afterward, social comparisons have been found to be important determinants for various aspects of our lives. Evidence from survey studies shows that relative income concerns and perceived happiness are negatively correlated at the state level, in the labor market and in social life (Clark and Oswald, 1996; Alesina et al., 2004; Luttmer, 2005). Similar findings have also been documented for subjective well-being (Ferreri-Carbonell, 2005; Perez-Truglia, 2020) and health (Marmot, 2005).

Derin-Güre and Uler (2010) show that people with higher incomes give more to charity when they are concerned about inequality, while people with lower incomes give less to charity when they are more averse to inequality. On the other hand, substantial empirical literature has found that inequality and giving are negatively associated. Lab experiments with voluntary within-experiment redistribution find that greater inequality of incomes generally decreases contributions (Chan et al., 1996; Buckley and Croson, 2006; Côté et al., 2015). Duquette and Hargaden (2021) randomize whole income distributions in a lab experiment to identify the causal effect of inequality on giving behavior. They find fewer donations both on the intensive and the extensive margin when income distributions are more unequal.

We investigate the effect of peer earnings on giving in a controlled experiment and aim to address the question of whether and how peer earnings affect donation decisions. Do higher peer earnings cause people to donate less? And if so, is this because they want to avoid unfavorable social comparisons? The main challenge to this identification is that relevant peer earnings are usually hard to observe and it is difficult to generate exogenous variation over them outside of the lab. Consequently, we address this research question using an experiment that allows us to exogeneously manipulate peer earnings between treatments.

In the social version of our experiment, two subjects participate in each session. Each subject is assigned one of the two roles, decision-maker or peer, by a publicly observed coin toss. The peer receives a fixed payment, whereas the decision-maker faces a real donation situation. Peer earnings, hence, serve as a natural source of comparison for decision-makers. In a between-subject design, the peer earnings were exogenously varied between £5 and £2 and each decision-maker was endowed with £6. As the donation task, we gave the subjects the possibility to allocate money from their initial endowment to WWF's Climate Relief Fund. By introducing a real donation decision to an outside charity, we abstract away from within-lab re-distributive concerns (Côté et al., 2015; Sheremeta and Uler, 2021). Our experiment is designed such that any difference in the average level of donations to the charity by decision-makers between treatments can be attributed to the amount of peer earnings.

In the non-social treatments, the social nature of peer earnings and, hence, donation incentives based on social comparisons were removed, whereas potential effects based on feelings of disappointment, experimenter demand, anchoring, and reference points based on counterfactual information remained untouched. The non-social treatments replicate the social treatments with a crucial change: only one subject participated in each session. A coin toss assigned each single subject per session to make a donation decision or leave with pre-determined earnings. The donation decision was exactly the same as in the social treatments. Prior to making their decision, decision-makers were exposed to the amount of money they would have earned if their coin had landed on the other side. Consequently, the donation decision differs between the social and non-social treatments only in the fact that exposure to counterfactual earnings referred to peer earnings in the social treatments and not obtained earnings in the non-social treatments.

To derive the main hypothesis that our experiment is designed to test, we use a simple Fehr-Schmidt model of inequality aversion around the peer's earnings. The manipulation of the peer's earnings changes the incentives for decision-makers that want to avoid unfavorable social comparisons with their peers. Consequently, our main hypothesis is that decision-makers in the social treatments give more to charity when the earnings of their peer are relatively low compared to when they are relatively high. We assume that the subjects' status quo determines their reference points in the non-social treatments and, hence, predict no treatment difference in the non-social treatments.

In the social treatments, we find that decision-makers donate more to charity when the peer earnings are relatively low compared to when they are high. When peer earnings were low, decision-makers donated on average £2.29 and when peer earnings were high they donated £1.91. The treatment effect is substantial in size and statistically significant. In the non-social treatments, we find that decision-makers chose fairly similar donations on average. While decision-makers chose slightly higher donations, when the reference earnings were high (£2.17) compared to low (£2.10), this effect is not statistically significant. Further, the pairwise comparison of the high and low peer earnings conditions between the social and the non-social treatments indicates that reference earnings of a peer might be more relevant when they are unfavorable compared to when they are favorable. We can corroborate this by showing that the willingness to donate is significantly greater in the non-social version of the experiment than in the social version when peer earnings are high, but not when peer earnings are low. Together, this suggests that people care about income differences and change giving behavior accordingly but more so if they are to their disadvantage. Finally, when checking for heterogeneous treatment effects, we find that our results are entirely driven by the behavioral changes of men while women behave similarly across the four decision situations.

In terms of the experimental setup, our study is the most closely related to experiments that manipulate pre-determined peer earnings in different settings such as risk-taking or real-effort provision (Schwerter, 2023; Maus et al., 2023) and differs from the majority of studies on inequality and giving across many dimensions. First, in related studies, decision-makers often observe the outcomes and behavior of others and expect others to observe their own behavior and corresponding outcomes. As a consequence, this does not allow us to distinguish between relative concerns and additional peer effects such as social pressure and imitation. By randomizing peer earnings but by keeping the decision environment constant, we differentiate our research from related studies on social pressures and giving behavior (DellaVigna et al., 2012; Andreoni et al., 2017; List, 2011). Our experiment is designed to test a specific channel through which peer effects could work, namely that people compare their income to that of a relevant peer and adjust their behavior accordingly while closing the door for potential confounding effects. Second, related studies do not implement pre-determined peer earnings. For example, decision-makers in Duquette and Hargaden (2021) are aware that their peers have a higher or lower endowment before their donation decision but not necessarily after. This is because the peers themselves participate in the same task and their income changes as a result. Accordingly, decision-makers may form beliefs about the decisions of their group members, which in turn influences their own decisions. Finally by using social and non-social treatments, our design allows us to isolate the effect of income comparisons from accompanying effects such as feelings of disappointment about missing out on higher initial endowments.

The remainder of the paper is organized as follows. Section 2 explains the experimental design in detail. Section 3 discusses the behavioral predictions and derives our main hypothesis. In Section 4 we present the results of our experiment. Section 5 concludes.

3.2 Experimental Design

The experiment is designed to create an environment that allows for a precise measurement of a donation decision and in which we can exogenously manipulate the earnings of a peer and its existence. Between the two social treatments, we manipulate predetermined peer earnings exogenously. A between-subject comparison of donations across treatments tests whether peer earnings influence donation decisions to a charity. In order to control for potential confounds in form of disappointment, the two social treatments are accompanied by two non-social treatments. Our study was preregistered on AsPredicted.[1]

The experiment was conducted in virtual rooms on Microsoft Teams with students of the University of Nottingham in June and July 2023. Each experimental session consisted of one or two subjects only. The subjects were recruited via Orsee (Greiner, 2015). The experimental materials are reproduced in Appendix 3B to Appendix 3E.

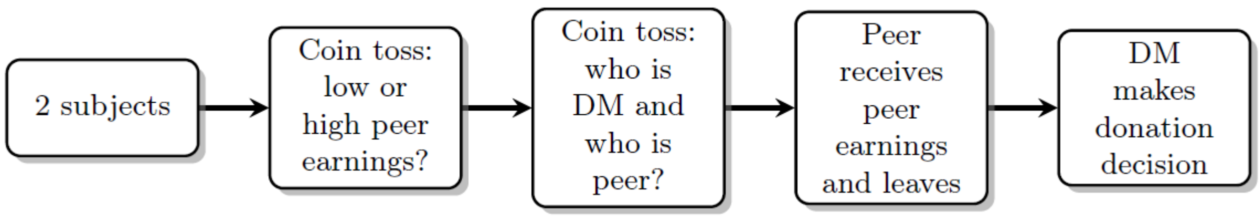


Figure 3.1: Procedural overview of the social treatments

Figure 3.1 shows the procedural overview of the social treatments. In the social treatments, upon the arrival of the subjects in the virtual room, the subjects are instructed that only one of the two subjects will make a decision that influences her payment and that the second subject will receive a lump-sum payment and has to leave the room. First, the experimenter tosses a virtual coin to determine the size of the lump-sum payment. In this way, subjects are randomly assigned to either a high peer earnings condition (lump-sum payment = 5) or a low peer earning condition (lump-sum payment = £2). After both subjects observe the coin toss and are informed about the size of the lump-sum payment, the experimenter tosses another coin to determine which subject makes a donation decision and which subject has to leave with the lump-sum payment. In the experiment we referred to these roles as Person A and Person B, for the remainder of the paper we will refer to them as decision-maker and peer.

We randomize the size of the lump-sum payment to the peer in each experimental session because we want to ensure that the lump sum payment does not convey different information about expected donations across treatments. If the lump-sum payment conveys any information

[1]The preregistration can be found on https://aspredicted.org/N9M_MXN.

in our experiment, this information would be conveyed by the average of the two possible lump-sum payments which is constant across treatments.

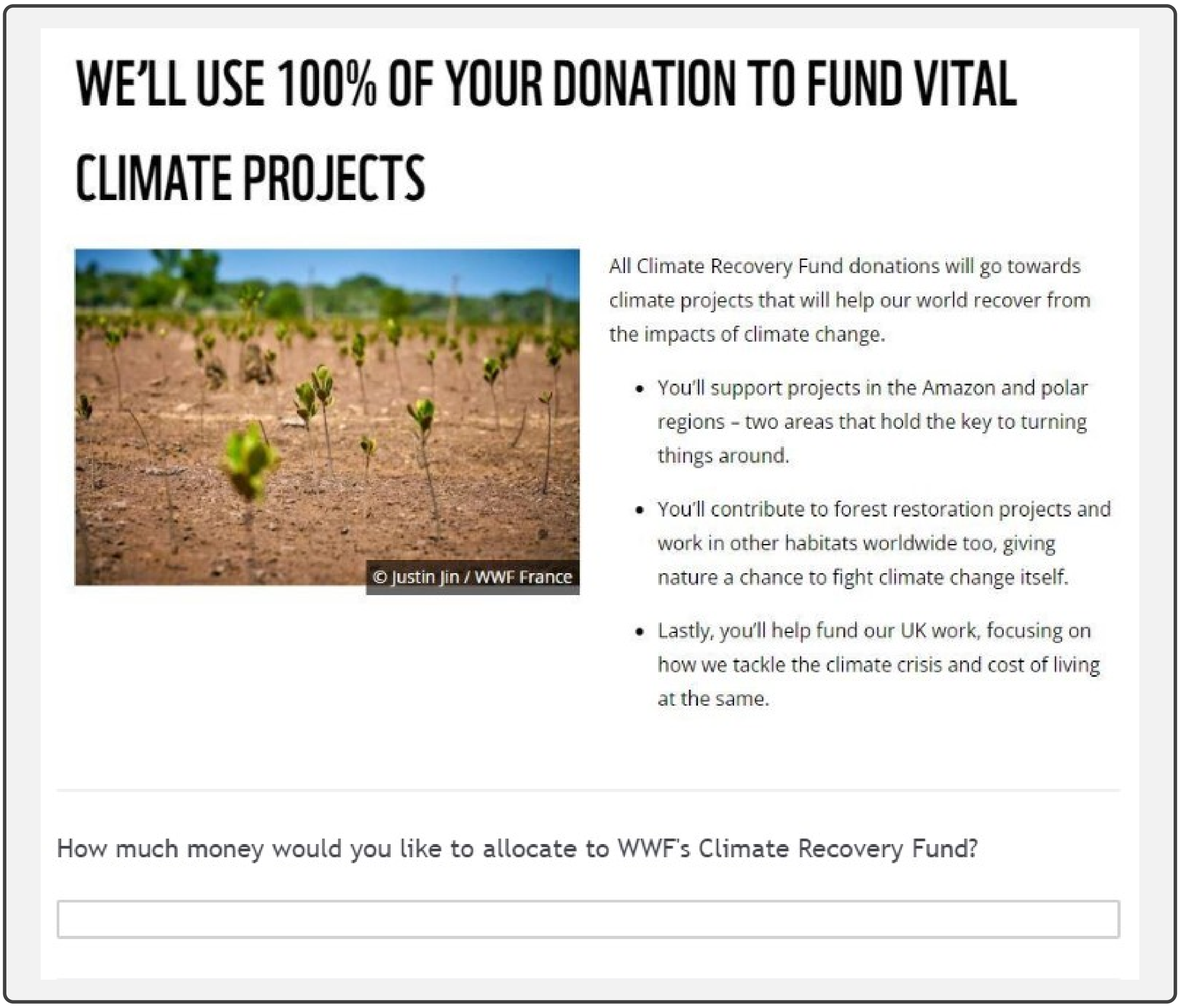


Figure 3.2: Screenshot of Donation Decision

Figure 3.2 illustrates the donation task decision-makers faced in our experiment. In our experiment, each decision-maker was endowed with £6 and was asked to make a donation decision. As the donation task, we gave the subjects the possibility to allocate money from their initial endowment to WWF's Climate Relief Fund. All donations could be made in full or half amounts. To incentivize donations, we told the decision-makers that the experimenters will multiply the amount they allocate to WWF by 1.5. An important feature of this setup is that it focuses on a real, external charity, unlike related literature that often looks at how the within-experiment distribution affects within-experiment redistribution. That is, by giving subjects the opportunity to donate to an actual cause rather than to a within-lab game, the real-charity design separates the decision to contribute from within-group social pressures. As the subject makes its donation decision in isolation, other sources of peer effects, other than earnings comparisons, are mitigated (e.g. imitation or peer pressure). After making the donation decision, the decision-maker answers a short questionnaire and the experiment is over.

The questionnaire includes questions about whether the decision-maker remembers the amount of the peer earnings, regarding their gender, the gender of their peer, the perceived social closeness to their peer (Gächter et al., 2015), self-reported competitiveness, and how much they trust the charity and how socially appropriate they perceive different donation decisions. Moreover, we elicit (without incentives) parameters of a Fehr-Schmidt inequality aversion model following Blanco et al. (2011). To elicit the aversion against disadvantageous equality α_i we use responder decisions from an ultimatum game.[2] For eliciting the aversion against advantageous inequality β_i we use a modified dictator game.[3] We use non-incentivized measures of α_i and β_i to ensure that the only source of payoff comparisons are the earnings from the donation decision.

Our two social treatments are accompanied by two non-social treatments. In the non-social treatments, we remove the social nature of the peer earnings, while keeping counterfactual earnings and hence potential feelings of disappointment due to losing out on a high income, potential experimenter demand and anchoring effects in place. By maintaining counterfactual earnings the non-social treatments are designed to test whether potential differences due to feelings of disappointment, anchoring and experimenter can account for differences in donation differences across treatments.

Figure 3.3 shows the procedural overview of the non-social treatments. The non-social treatments replicate the social treatments with a crucial change: only one subject participated in

[2]Suppose x_i' is the lowest offer the responder accepts and $x_i' - 1$ is the highest offer the responder rejects, where offers are integers between 0 and 10 out of a total amount of 20. Then a responder with well-behaved preferences is indifferent between accepting some offer $x_i \in [\mathrm{x}_i' - 1,\ x_i']$ and both players receiving nothing. From $U_i\left(x_i, 20 - x_i\right) = x_i - \alpha_i\left(20 - x_i - x_i\right) = 0$ we can determine the point of indifference $\alpha_i = \frac{x_i}{2(10-x_i)}$. Following Blanco et al. (2011) we approximate x_i as the average of x_i' and $x_i' - 1$ when there is no more than one switch-point in the responder's strategy. The decision-makers who accept all offers are assigned $\alpha_i = 0$ and the decision-makers who reject all offers are assigned $\alpha_i = 4.5$.

[3]We obtain β_i by finding the egalitarian allocation $(x_i,\ x_i)$ that makes the dictator indifferent between keeping the entire endowment (10, 0) and $(x_i,\ x_i)$. Suppose the dictator switches to the egalitarian outcome at $(x_i',\ x_i')$. That is, he prefers $(10, 0)$ over $(x_i' - 1,\ x_i' - 1)$ but $(x_i',\ x_i')$ over $(10, 0)$. We conclude that he is indifferent between $(10, 0)$ and the (x_i, x_i) egalitarian outcome where $x_i \in [x_i' - 1, x_i']$ and $x_i' \in \{1, \ldots, 10\}$. Thus, $\beta_i = 1 - \frac{x_i}{10}$. Again, we approximate x_i as the average of x_i' and $x_i' - 1$. The decision-makers who always choose the egalitarian outcome are assigned $\beta_i = 1$ and the decision-makers who never choose the egalitarian outcome are assigned $\beta_i = 0$.

each session. The experimenter tossed a coin in front of the subject to randomly assign one of two roles: Person A or Person B. Subjects that have been assigned the role of Person A took the identical donation task as decision-makers in the social treatments. Due to implementing the same procedure of coin tosses as in the social treatments, all decision-makers know how much they would have earned if they would have been assigned the role of Person B. Consequently, a decision-maker makes its donation decision knowing that they would have earned either £5 or £2 if the coin would have assigned her the role of Person B. Subjects that have been assigned the role of Person B left with either £2 or £5 without making a donation decision as peers in the social treatments did.

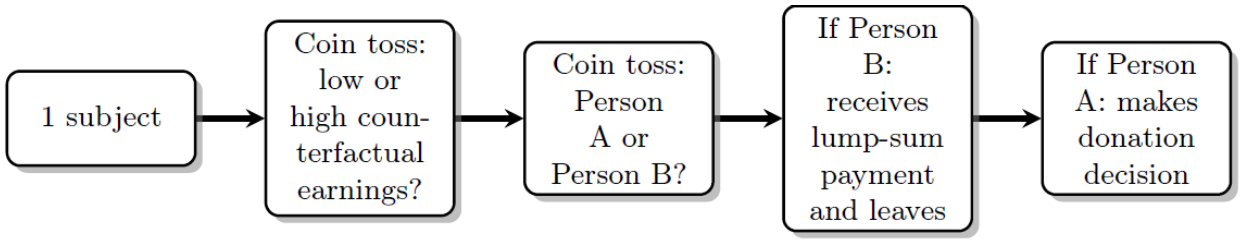


Figure 3.3: Procedural overview of the non-social treatments

In the non-social treatments, the social nature of peer earnings and, hence, the incentives to donate due to avoiding unfavorable social comparisons are removed, while keeping the remaining decision environment constant. The crucial feature is that both decision-makers in the social and non-social treatments chose their donation in light of the same counterfactual earnings. However, only for decision-makers in the social treatments this constitutes a social comparison in earnings. Therefore, motives based on social comparison between participants were removed from their donation decision in the non-social treatments. Significantly different donations in the social treatments compared to the non-social treatments would hence provide clear evidence that social comparison effects on giving go beyond feelings of losing out on high earnings, anchoring and demand effects. Overall, the social and the non-social versions of our experiment give us a 2 x 2 design that allows us to test the causal effect of higher peer/reference earnings both when a peer is present and absent separately. Further, it allows us to compare the donations between the social version of the experiment and the non-social version of the experiment.

3.3 Theoretical Predictions & Hypotheses

3.3.1 Theoretical Predictions

This section examines how subjects in our experiment are predicted to behave. Consider two individuals, decision-maker i and its peer j. Decision-maker i makes the decision about how much to donate to a charity in the interval from £0 to her total endowment of £6, whereas its peer j receives a lump-sum payment F. The decision-maker i chooses her optimal donation D_i^* given her endowment of £6, the peer earnings F, and a function of warm glow $\lambda_i \ln(D_i)$ that includes an individual-specific parameter λ_i which describes how much an individual enjoys giving to the charity. For simplicity, we assume the function of warm glow to be $\ln(D_i)$. Accordingly, the decision-maker i chooses her optimal donation to maximize her utility function:

$$U_i = 6 - D_i - \alpha_i \max\{(F - (6 - D_i), 0)\} - \beta_i \max\{((6 - D_i) - F, 0)\} + \lambda_i \ln(D_i) \tag{3.1}$$

The first and second term on the RHS of equation (1) correspond to the utility of monetary earnings, the third and fourth term account for a possible disutility as a consequence of payoff inequalities and the final term represents the warm glow from a donation. The modeling of disutility due to payoff inequality follows Fehr and Schmidt (1999) where the third term of the RHS measures the disutility from disadvantageous payoff inequality and the fourth term measures the disutility from advantageous payoff inequality, with $\alpha_i \geq \beta_i \geq 0$. Consequently (1) collapses to:

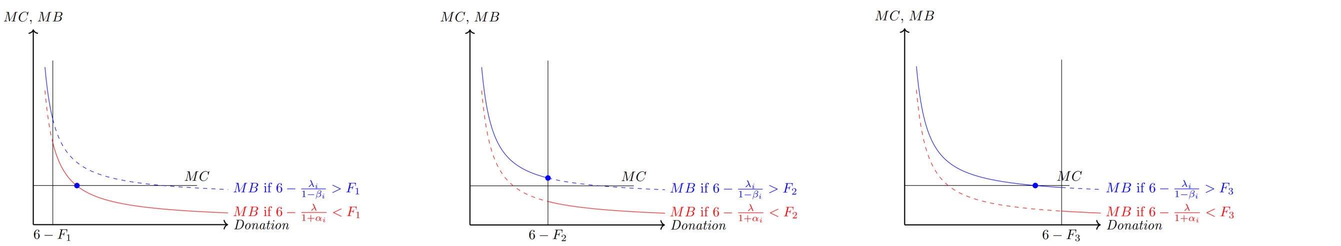
$$U_i = \begin{cases} 6 - D_i - \alpha_i \max\{(F - (6 - D_i), 0)\} + \lambda_i \ln(D_i) & \text{if } 6 - D_i < F \\ 6 - D_i - \beta_i \max\{((6 - D_i) - F, 0)\} + \lambda_i \ln(D_i) & \text{if } 6 - D_i \geq F \end{cases} \tag{3.2}$$

Differentiating (2), we obtain the FOCs and the optimal levels of donation depending on the reference earnings.

$$\frac{\partial U_i}{\partial D} = \begin{cases} -1 - \alpha_i + \frac{\lambda}{D_i} & \text{if } 6 - D_i < F \\ -1 + \beta_i + \frac{\lambda}{D_i} & \text{if } 6 - D_i \geq F \end{cases} \tag{3.3}$$

$$D_i^* = \begin{cases} \frac{\lambda}{1+\alpha_i} & \text{if } 6 - \frac{\lambda_i}{1+\alpha_i} \leq F \\ 6 - F & \text{if } 6 - \frac{\lambda_i}{1-\beta_i} < F < 6 - \frac{\lambda_i}{1+\alpha_i} \\ \frac{\lambda_i}{1-\beta_i} & \text{if } 6 - \frac{\lambda_i}{1-\beta_i} \geq F \end{cases} \tag{3.4}$$

The marginal utility in (3) can be viewed as the difference between the marginal benefit of a donation and the marginal monetary cost of the donation, where the marginal benefit of a donation depends on whether earnings exceed or fall short of the peer earnings, i.e. it takes on the value $\frac{\lambda_i}{D_i} - \alpha_i$ for the case of unfavorable social comparisons with j and $\frac{\lambda_i}{D_i} + \beta_i$ for the case of favorable social comparisons with j. Consequently, the optimal level of donation in (4) depends on whether the peer earnings are comparatively large or small. This is in contrast to the standard model where the optimal level is λ which does not depend on the amount of the peer earnings.



(a) $6 - \frac{\lambda_i}{1+\alpha_i} \leq F$ (b) $6 - \frac{\lambda_i}{1-\beta_i} < F < 6 - \frac{\lambda_i}{1+\alpha_i}$ (c) $6 - \frac{\lambda_i}{1-\beta_i} \geq F$

Figure 3.4: Optimal donations for given λ_i, α_i and β_i when peer earnings are (a) high, (b) medium or (c) low.

Figure 3.4 graphically illustrates the three cases to consider for a given F, λ_i, α_i and β_i. The peer earnings can be either (a) high, (b) medium or (c) low. Graphically, this means that the higher the peer earnings F are, the further to the left $6 - F$ is located. Panel (a) describes the case where the individual faces high peer earnings. Here, she chooses her optimal donation such that it corresponds to the intersection between her marginal cost function and her marginal benefit function which yields earnings below the reference earnings. Consequently, if peer earnings are increased and $6 - F$ is shifted further to the left, this does not alter her optimal donation decision. Panel (b) describes the case where the individual faces intermediate peer earnings. This case illustrates the situation where she chooses her donation such that she matches the earnings of her peer. Therefore, if the reference earnings are increased and $6 - F$

is shifted further to the left, she decreases her donations such that it corresponds to the new peer earnings or up to the point when the marginal cost function and marginal benefit function intersect. Panel (c) describes the case where the individual faces low peer earnings. If the peer earnings are increased and $6 - F$ is shifted further to the left, this might have three possible different effects. If the shift is only very slight, the optimal donation does not change. If the shift is moderate, she chooses her donation such that she matches the reference earnings. If the shift is large, she increases her donation to the point where the marginal cost function and marginal benefit function intersect. Consequently, an increase in peer earnings leads to at least a weakly higher level of donations in all three panels.

Theorem 3.1 *Higher peer earnings lead to weakly lower levels of donations.*

Note that the assumption that $\beta_i > 0$ can be relaxed. Our predictions also hold, as long as we assume $\alpha_i > |\beta_i|$. Under this condition, the marginal benefit function $\frac{\lambda_i}{D_i} + \beta_i$ is always above the marginal benefit function $\frac{\lambda_i}{D_i} - \alpha_i$ and consequently intersects the marginal cost function at a higher optimal donation.

3.3.2 Hypotheses

Given our theoretical analysis, we arrive at the following testable hypotheses for the comparison between the two treatments:

Hypothesis 3.1: In the social treatments, the distribution of donations when peer earnings are low stochastically dominates the distribution of donations when peer earnings are high.

We do not expect a similar effect for the non-social treatments. According to our theoretical analysis, the optimal level of donations in the social treatments can be determined by two processes. Either the individual chooses the donation level that corresponds to the intersection between the marginal benefit function and the marginal cost function, or she matches the peer earnings. Assuming that we have cases where individuals choose the donation that corresponds to the peer earnings, our model suggests not only that the donations differ between treatments, but they differ in a very specific way. Consequently, our subsidiary hypotheses predict a higher probability of choosing the donation that leads to matching the high earnings when the peer earnings are high and a higher probability of choosing the donation that leads to matching the low earnings when the peer earnings are low.

Hypothesis 3.2a: In the social treatments, more decision-makers donate exactly £1 when the peer earnings are high compared to when they are low.

Hypothesis 3.2b: In the social treatments, more decision-makers donate exactly £4 when the peer earnings are low compared to when they are high.

From the perspective of expected utility theory as well as models of status-quo-based (Tversky and Kahneman, 1979) and expectations-based (Kőszegi and Rabin, 2006) reference points, no treatment effect on donation is predicted since decision-makers faced the same donation decision across treatments. Consequently, we are confident that a difference in donations between the high and the low peer earnings condition in the social treatments can be solely attributed to the effect of peer earnings. Given that we anticipate a negative effect of relatively high peer earnings on donations in the social treatments compared to the standard case and we do not anticipate the same effect for the non-social treatments, we also expect that donations in the social treatments are lower compared to the non-social treatments when peer earnings are high.

Hypothesis 3.3: When reference earnings are high, the distribution of donations in the non-social treatment stochastically dominates the distribution of donations in the social treatment.

3.4 Results

In total, 360 decision-makers participated in the experiment. 180 of whom made their donation decision in the social treatments and 180 in the non-social treatments. The random assignment to treatments resulted in 93 decision-makers in the high peer earnings condition and 87 decision-makers in the low peer earnings condition for the social treatments and 86 decision-makers with high counterfactual earnings and 94 decision-makers with low counterfactual earnings in the non-social treatments respectively.

Table 3.1 shows the sample characteristics regarding demographics and survey answers. As one can see, decision-makers in the four treatments are well balanced across most variables. The female share of participants is almost identical across the four treatments ranging from 55.3% to 56.9% and so is the reported monthly available budget (ranging from £267 to £277). Further, decision-makers show similar levels of competitiveness and social comparison concerns. Moreover, they share similar opinions regarding the trustworthiness of the charity and the impact of a donation. No pairwise comparison between variables across treatments yields statistically significant differences using two-sided Wilcoxon rank-sum tests.

Table 3.1: Descriptive statistics

Variables	Social				Non-social			
	High		Low		High		Low	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Demographics								
Share of females	0.55	0.50	0.57	0.50	0.55	0.50	0.56	0.50
Same gender as peer	0.52	0.50	0.61	0.49	-	-	-	-
Number of siblings	1.41	1.54	1.53	1.28	1.40	1.33	1.42	1.20
Income	270.5	307.6	277.1	264.5	268.8	257.0	266.7	259.0
Personal Characteristics								
Competitiveness	4.77	1.48	4.81	1.52	5.08	1.37	5.01	1.40
Oneness	1.47	1.22	1.46	1.01	-	-	-	-
Behindness aversion	0.88	1.44	0.82	1.34	0.66	1.08	0.72	1.32
Leading aversion	0.60	0.32	0.63	0.31	0.67	0.29	0.59	0.29
Relation to charity								
Impact	5.05	1.66	5.01	1.78	5.40	1.39	5.32	1.53
Trustworthiness	4.51	1.57	4.76	1.56	5.30	1.40	5.03	1.47

All experimental sessions were conducted online with students of the University of Nottingham in June and July 2023. The decision-makers earned on average £3.88 and spent on average 10.2 minutes to make a donation decision and answer the post-experimental survey. Across all treatments, more than 95% remembered the peer/counterfactual earnings their peer received which makes us confident that our treatment manipulation worked. As we preregistered our experiment, we use one-sided tests for our directed hypotheses and two-sided tests for the more exploratory analysis. Our analysis prioritizes non-parametric tests because they do not assume that error terms are normally distributed which they are not (Shapiro-Wilk Test, $p < 0.01$).

3.4.1 Main Analysis

Our first result supports Hypothesis 3.1. In the social version of the experiment with relatively low peer earnings of £2 (*So-Low*) decision-makers allocated on average £2.28 to the charity, while decision-makers with relatively high peer earnings of £5 (*So-High*) donated on average £1.90. Comparing *So-Low* to *So-High*, the treatment difference is £0.37 and the marginal effect of having lower peer earnings compared to higher peer earnings on donations is roughly 19.3%. This treatment difference is statistically significant using a one-sided Wilcoxon rank-sum test (p-value=0.020). While peer earnings affected the donation decisions in the social treatments, the donation decisions in the non-social treatments between the high (*NoSo-High*) and the LOW (*NoSo-Low*) counterfactual earnings conditions are fairly similar. Decision-makers in *NoSo-Low* allocated on average £2.10 to WWF while decision-makers in *NoSo-High* donated on average £2.17. We did not expect to see a significant treatment difference in the non-social treatments but if anything we would have expected decision-makers to donate less when they lost out on the relatively high earnings. Overall, the treatment difference in donations is negligible and not statistically significant (two-sided Wilcoxon rank-sum test, p-value=0.42).

Result 3.1: In the social treatments, decision-makers donate more when peer earnings are low compared to when they are high. In the non-social treatments, we can not find a similar effect.

Figure 3.5 shows a histogram of donation decisions in the two social treatments (*So-High* in the top panel, *So-Low*) in the bottom panel). First of all, donation decisions are dispersed over a wide range. Some subjects do not donate anything at all, others donate the maximum amount possible of £6. This is what one would expect given that feelings of worthiness of the cause and warm glow differ across subjects. Interestingly, there are systematic differences in terms of donation decisions between the treatments. Comparing the donation distributions between the two treatments, it appears striking that while there is a similar amount of donating nothing or fully, there are clearly fewer people donating intermediate amounts between £2 and £4 in *So-High* compared to *So-Low*. A one-sided Kolmogorov-Smirnov rejects the equality of distributions between treatments (p-value= 0.04).

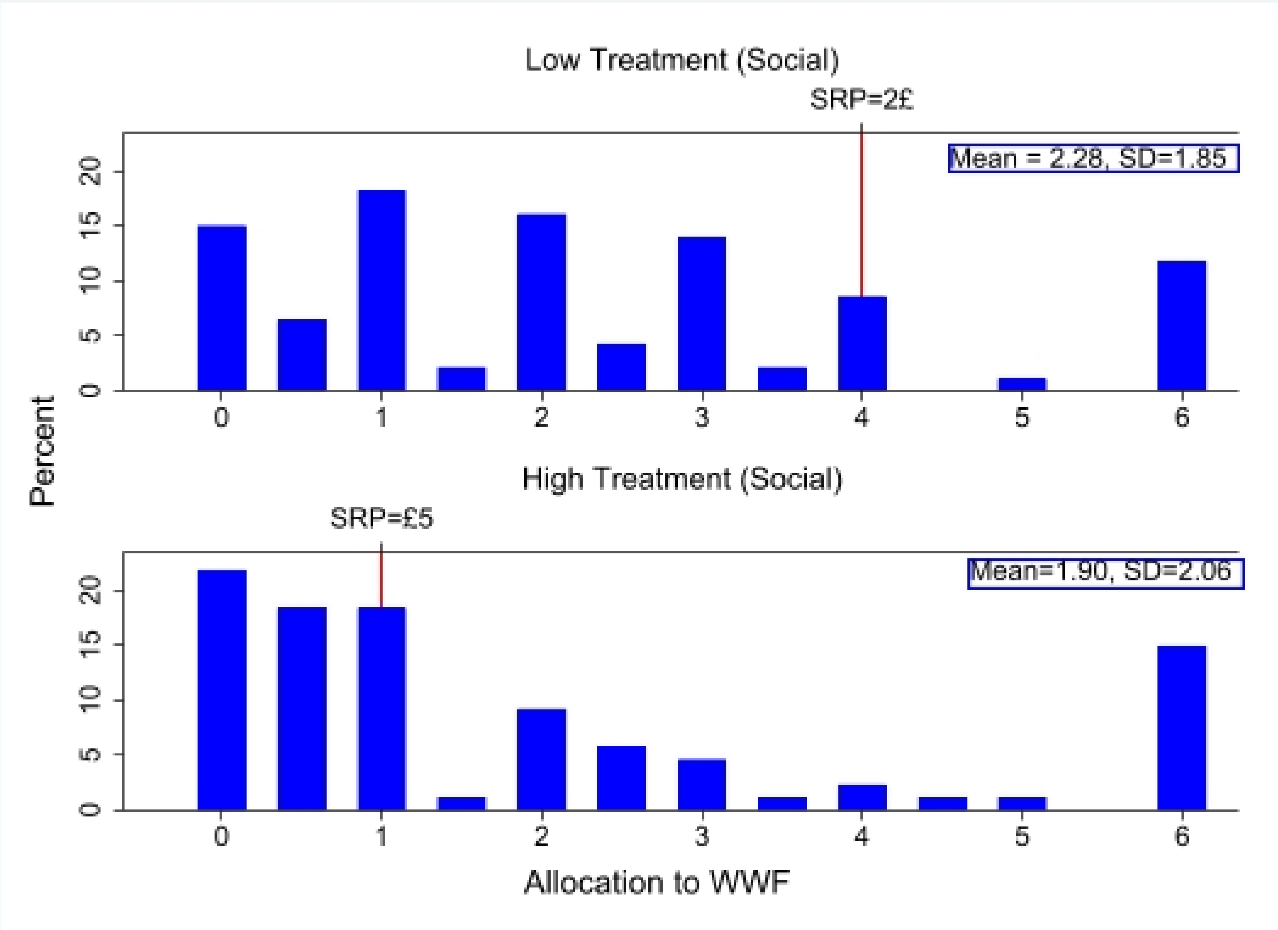


Figure 3.5: Histograms of donation decisions in the social treatments

As shown in Section 3.3, our model does not only predict treatments to be different but to be different in a very specific way. Hypotheses 3.2a and 3.2b predict that in the social treatments, the frequency of subjects choosing the donation that leads to matching the high peer earnings is higher in *So-High* compared to *So-Low* and the frequency of subjects choosing the donation that leads to matching the low peer earnings is higher in *So-Low* compared to *So-High*. Consequently, we would expect more decision-makers to donate exactly £1 when the peer earnings are high and more decision-makers to donate exactly £4 when the peer earnings are low.

In both treatments, many subjects donated exactly £1 (18.27% in *So-Low* and 18.39% in *So-High*) and the difference between treatments is not statistically significant (one-sided Fisher's

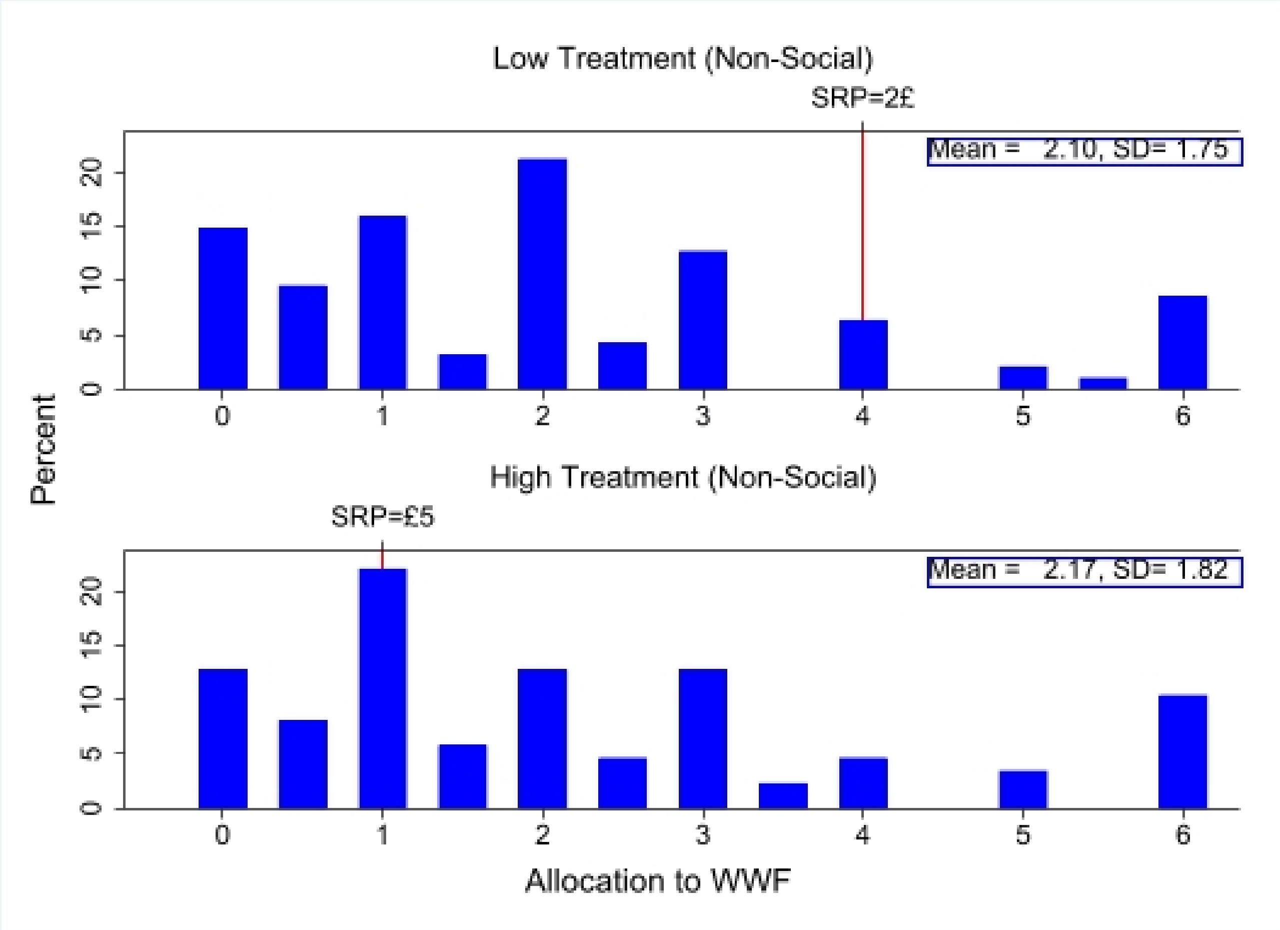


Figure 3.6: Histograms of donation decisions in the non-social treatments

exact test, p-value=0.568). By contrast, more decision-makers in *So-Low* donated exactly £4 compared to *So-High*. While in *So-Low* eight subjects (8.6%) donated exactly £4, only two decision-makers were donating exactly £4 in *So-High* (2.3%). This difference is marginally significant using a one-sided Fisher's exact test (p-value = 0.062).

Figure 3.6 shows the histograms for the donation decisions in the non-social treatments (*NoSo-Low* in the top panel, *NoSo-High* in the bottom panel). There is a similar amount of subjects not donating anything and donating the full endowment compared to the social treatments. However, the donation distributions between the *NoSo-Low* and *NoSo-High* treatments look very similar overall. Strikingly, we can not see the difference between the high and low counterfactual earnings conditions for the donations in the intermediate range in the non-social treatments that we could see for the social treatments. While more decision-makers in the *NoSo-Lo* treatment donated £4 compared to *NoSo-High* and more decision-makers donated £1 in *NoSo-High* compared to *NoSo-Lo*, none of these differences are statistically significant (one-sided Fisher's exact tests, p-value = 0.195 and 0.430 respectively).

Result 3.2a: In the social treatments, the probability to donate exactly £1 is not higher when the peer earnings are high compared to when they are low.

Result 3.2b: In the social treatments, the probability to donate exactly £4 is marginally higher when the peer earnings are low compared to when they are high.

Figure 3.7 reports the comparisons of cumulative distribution functions between the high and low peer earnings conditions for both the social and the non-social treatments. There is a clear reaction to higher peer earnings with the cumulative distribution of donations shifting to the right when peer earnings are relatively low in the social treatments, while we can not see a comparable effect in the non-social treatments. This graphical evidence is perfectly in line with Hypothesis 3.1.

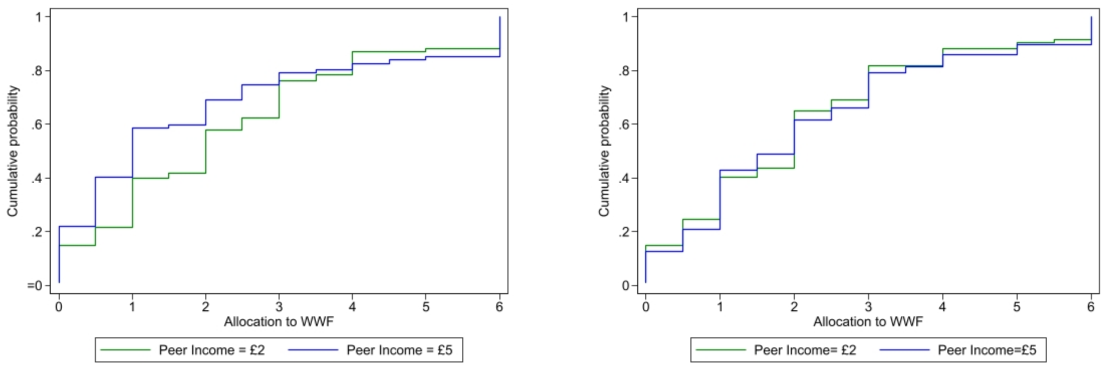


Figure 3.7: Comparison of cumulative distribution functions between the low and high peer earnings treatments (Social on LHS, Non-social on RHS)

Next, we analyse when the social and non-social treatments differ. To do this, we compare donations in the treatments with high and low peer earnings between the social and non-social treatments. The comparison indicates that subjects donate substantially more in the non-social treatments when peer earnings are high (one-sided Wilcoxon rank-sum test, p-value < 0.05), while we can not reject the hypothesis of identical donation behavior when peer earnings are low (p-value=0.26). Figure 3.8 graphically illustrates that the cumulative distributions functions in the social and the non-social treatments are nearly identical when reference earnings are low. When reference earnings are high the cumulative distribution function of donations in the social treatment is shifted to the left compared to the non-social treatments.

Result 3.3: When reference earnings are high, the decision-makers donate less when the reference earnings are obtained by a peer compared to when they are not.

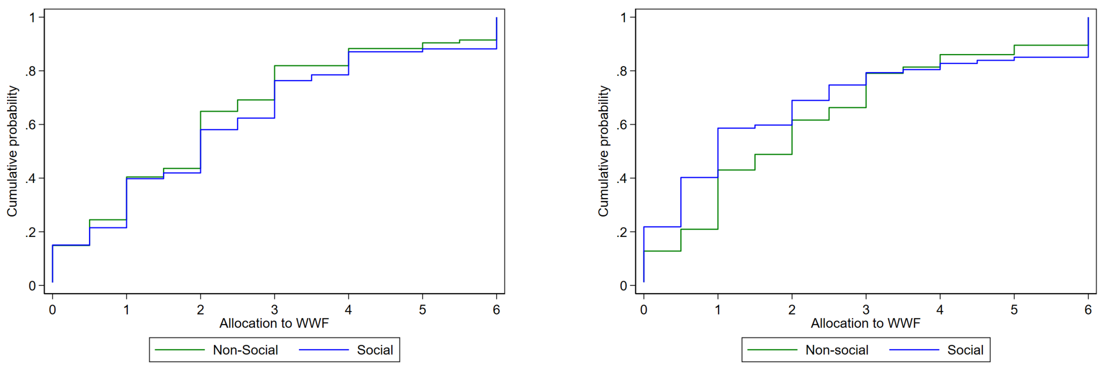


Figure 3.8: Comparison of cumulative distribution functions between the social and the non-social treatments (Low Treatment on LHS, High Treatment on RHS)

3.4.2 Checking for heterogeneous treatment effects

In this part of the analysis, we will use data from the post-experimental questionnaire to investigate the mechanisms behind our findings. To do this, we discuss whether our results could be related to satisfaction with the earnings and the degree of behindness aversion and leading aversion of our subject pool. Finally, we discuss gender heterogeneity in terms of the treatment effect.

Satisfaction

Previous experiments (see Maus et al. (2023) suggest that subjects are happier with their own earnings when those compare favorably to their peer earnings keeping everything else equal. In the social version of the treatments, there is no indication that decision-makers might be happier when they are matched with a low-earning peer (two-sided Wilcoxon rank-sum test, p-value=0.628). This still holds true when we run an OLS regression controlling for the donation that has been made. As expected, we do not find a significant treatment effect in terms of satisfaction in the non-social treatment either (two-sided Wilcoxon rank-sum test, p-value=0.142).

Behindness Aversion

In our questionnaire, we elicited a measurement of each subject's level of behindness aversion. It might be that our treatment manipulation has an especially strong effect on the decision-makers who report high levels of behindness aversion in the post-experimental survey. The average α for the subjects in our experiment was 0.767 and is very similar to the literature.[4] Originally, we expected that the treatment manipulation might have a stronger effect on decision-makers with high degrees of behindness aversion. Table 3.2 reports the average level of behindness aversion by donation and treatment. If the level of behindness aversion is a driving factor behind our main results, we would expect that people who donate less have higher alphas in the social version of the experiment - especially when peer earnings are high. This is what we can observe in the data - for *So-High* we can see a monotonic relationship between the level of donations and the degree of behindness aversion which we can not observe for the remaining treatments.

Table 3.2: Behindness Aversion by Donation and Treatmentt

	Treatments			
	So-High	***So-Low***	**NoSo-High**	**NoSo-Low**
$Donation = 0$	1.194	0.975	0.927	0.935
$1 > Donation > 0$	0.979	1.130	0.740	0.738
$4 > Donation > 1$	0.663	0.646	0.538	0.599
$Donation > 4$	0.516	0.670	0.590	0.901

Leading Aversion

In our questionnaire, we have also elicited a measurement of each subject's aversion to earning more than a peer. Similar to the previous analysis, it might be that decision-makers do not only like to have less than their peer they might also feel an aversion towards having more than their peer. Consequently, while we expected a negative relationship between the measure for alpha

[4]Beranek et al. (2015) measure the degree of behindness aversion in an incentivized way. The means and the distributions in both experiments are very similar (mean in our experiment = 0.71 with 55% reporting an alpha below 0.4, mean in Beranek et al. (2015) = 0.75 with 55% reporting an alpha below 0.4).

and the donation we expected a positive effect between the measure for beta and the donation. The betas we elicited in our experiment are on average 0.625 (where 6.16% have a beta of 0 and 24.63% have a beta of 1). Table 3.3 reports the average level of leading aversion by donation and treatment. If the level of leading aversion is a driving factor behind our main results, we would expect to see a more monotonic relationship in the social treatments, especially for the case when peer earnings are low. In So-Low higher degrees of leading aversion are associated with higher donations. However, the relationship seems to be not perfectly monotonic and is very similar to the three remaining treatments. A potential reason for this may be that people with higher betas are generally more willing to donate.

Table 3.3: Leading Aversion by Donation and Treatment

	Treatments			
	So-High	***So-Low***	**NoSo-High**	**NoSo-Low**
$Donation = 0$	0.461	0.321	0.550	0.396
$1 > Donation > 0$	0.588	0.587	0.602	0.617
$4 > Donation > 1$	0.722	0.742	0.708	0.625
$Donation > 4$	0.660	0.683	0.850	0.615

3.4.3 Further results

Here we report some additional analyses that go beyond our pre-registered main analysis. In our experiment, females donated on average £2.33 whereas males donated on average £1.84. This difference is large and statistically significant (two-sided Wilcoxon rank-sum test, p-value < 0.01). We checked for gender differences in the reaction to the treatment manipulation. Men in the social treatments reacted very strongly to the treatment manipulation, i.e. they donated £1.45 less when their peer received a relatively high lump-sum payment (p-value <0.01, *So-High*= £1.19, *So-Low*= £2.65). The treatment difference is £1.46 and the marginal effect of facing low peer earnings compared to high peer earnings is 44.9%. Importantly, we do not find a similar effect for men in the non-social versions of the experiment (*So-High*= £1.70, *So-Low*= £1.81, two-sided Wilcoxon rank-sum test, p-value=0.761) indicating that men care more about income differentials with others rather than themselves.

We can not observe a similar effect for women. The donations of female decision-makers in the social treatment are not significantly affected by the peer earnings (High Treatment= £2.44, Low Treatment= £2, two-sided Wilcoxon rank-sum test, p-value=0.666). As expected, we find the same for the non-social version of the treatments (two-sided Wilcoxon rank-sum test, p-value=0.74, High Treatment= £2.50, Low Treatment= £2.40). Taken together, this is very interesting, as it suggests that the entire treatment effect observed between *So-High* and *So-Low* stems from the behavior of men, whereas for women neither forgone earnings nor whether a peer obtains them seems to matter.

We also find that more competitive decision-makers donate more to the charity than less competitive decision-makers in the social treatments (two-sided Wilcoxon-rank sum test, p-value= 0.055), while they do not in the non-social version (two-sided Wilcoxon-rank sum test, p-value=0.691). It might also be the case that competitive people react more strongly to a manipulation of the peer earnings in the social treatments. To account for this we add an interaction term between the treatment and the degree of competitiveness but the interaction term is insignificant (p-value= 0.332). Additionally, we find that decision-makers who report a higher income are donating marginally more on average (two-sided Wilcoxon rank-sum test with median split, p-value= 0.055). The gender composition of each pair does not seem to play a substantial role. There is no statistically significant difference in donations between a subject that is matched with someone of their gender and someone who is not (two-sided Wilcoxon rank-sum test=0.31). Similarly, the degree to which the decision-makers feel connected does not seem to have a substantial effect. If we regress the individual donation on oneness, we find a weak and statistically insignificant relationship (p-value= 0.314).

Table 3.4 reports our findings in a regression framework. We regress the donations for each subject on dummies for the reference earnings and the social dimension. Further, we include the interaction term between the two dummies and control for behindness aversion and leading aversion using a median split, as well as a control for gender and its interaction with high reference earnings. We consider a decision-maker behindness averse if its alpha is larger than 0.4 and we consider an individual leading aversion if its beta is larger than 0.55. The first column reports the regression for the whole sample, the second column for the social treatments only and the final column for the non-social treatments only.

Table 3.4: OLS Regressions: Donations as dependent variable

	Pooled	Social	Non-social
High reference earnings	0.366	0.434	-0.018
	(0.314)	(0.357)	(0.344)
Social	0.123		
	(0.261)		
High reference earnings x Social	-0.345		
	(0.377)		
Male	-0.010	0.734*	-0.696*
	(0.266)	(0.382)	(0.361)
Male x High	-0.893**	-1.76***	-0.105
	(0.380)	(0.541)	(0.522)
Behindness averse (dummy)	-0.433**	-0.497*	-0.369
	0.190	(0.273)	(0.260)
Leading averse (dummy)	0.959***	1.28**	0.641***
	(0.199)	(0.285)	(0.272)
Constant	1.723***	1.350***	2.191***
	(0.271)	(0.349)	(0.325)
Observations	360	180	180

The table reports estimates from OLS regressions. The dependent variable are the donations to WWF's Climate Relief Fund. Column (1) reports the regression analysis for all treatments and the whole sample, Column (2) for the social treatments only (3) for the non-social treatments only. No decision-makers are excluded from the analysis in column (1) and (2). In column (3) decision-makers with multiple switch-points for our measures of α or β are excluded. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

For the whole sample, we can see that the coefficients for high reference earnings and the presence of a peer are insignificant. We are mainly interested in the interaction term between the high reference earnings and whether a peer was present, which allows us to test whether the treatment effect is statistically different in the social treatments compared to the non-social treatments. The coefficient is relatively large and goes into the anticipated direction. However, it is not significant. A potential reason for this could be the non-normality of our data. In terms of gender effects we can see that men decrease their donations significantly more compared to women as a reaction to higher reference earnings. Over the whole sample, decision-makers with a higher degree of behindness aversion tend to donate more and decision-makers with a higher degree of leading aversion tend to donate less.

For the social treatments, the coefficient for higher reference earnings remains insignificant. Further, we can see that when reference earnings are low, men tend to donate significantly more than women do but this reverses as a reaction to higher peer earnings. As for the whole sample, decision-makers with a higher degree of behindness aversion tend to donate more and decision-makers with a higher degree of leading aversion tend to donate less.

Also in the non-social treatments, the coefficient for higher reference earnings remains insignificant. However, in the non-social treatments, there is no significantly different treatment effect for men and women. Also, in contrast to the social treatments, the degree of behindness aversion does not predict donation decisions. Given that we found significantly different behavior for men and women in our experiment, we report separate regressions for each of them in Appendix 3A.

3.5 Conclusion

In this paper, we investigate the effect of peer earnings on charitable giving in a controlled environment. The experiment is based on a simple Fehr-Schmidt model of inequality around a peer's earnings and predicts that higher peer earnings should lead to lower levels of donations. To test the predictions of our model, we use an experimental paradigm that allows us to observe precisely how much individuals donate while exogenously manipulating peer earnings.

In the social treatments of our experiment, only two subjects participate in each session. Each subject is assigned one of two roles - a decision-maker or a peer. The peer receives a fixed payment, whereas the decision-maker faces a donation decision to a real charity. Between the two treatments, we vary the fixed payment of the peer. We find that the earnings of a peer are relevant for the behavior of the decision-maker. Decision-makers donated significantly more when the peer earnings were relatively low compared to when they were relatively high.

To test the robustness of our findings, the two social treatments were accompanied by two non-social treatments. In those treatments, we follow the same procedure as in the social treatments with the crucial difference that there is only one subject per session. We do this to make sure that our treatment effect is not confounded by other possible influences such as anchoring or feelings of disappointment. The crucial feature is that both decision-makers in the social and the non-social treatments chose their donations in light of the same counterfactual earnings. As expected, decision-makers did not donate more in the non-social treatments when their reference earnings were relatively low. Consequently, our treatment effect in the social treatments can be fully attributed to social comparison concerns.

Further, the pairwise comparison of the high and low peer earnings conditions between the social and the non-social treatments indicates that reference earnings of a peer might be more relevant when they are unfavorable compared to when they are favorable. Together, this suggests that people care about income differences and change giving behavior accordingly but more so if social comparisons are more likely to be their disadvantage. Finally, when checking for heterogeneous treatment effects, we find that our results are entirely driven by the behavioral changes of men while women behave similarly across the four decision situations.

Our results are in line with Schwerter (2023) who finds that people are more willing to take risks when they face higher social reference points using a similar design. Maus et al. (2023) finds that decision-makers do care about income differentials when performing a real-effort task, however this does not lead to changes in effort provision. In the future, more work should be done to investigate under which circumstances social comparisons are more and less powerful.

3.6 References

Alesina, A., R. Di Tella, and R. MacCulloch (2004). Inequality and happiness: are europeans and americans different? *Journal of Public Economics 88*(9-10), 2009–2042.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The Economic Journal 100*(401), 464–477.

Andreoni, J., J. M. Rao, and H. Trachtman (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy 125*(3), 625–653.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review 84*(2), 191.

Beranek, B., R. Cubitt, and S. Gächter (2015). Stated and revealed inequality aversion in three subject pools. *Journal of the Economic Science Association 1*(1), 43–58.

Blanco, M., D. Engelmann, and H. T. Normann (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior 72*(2), 321–338.

Buckley, E. and R. Croson (2006). Income and wealth heterogeneity in the voluntary provision of linear public goods. *Journal of Public Economics 90*(4-5), 935–955.

Chan, K. S., S. Mestelman, R. Moir, and R. A. Muller (1996). The voluntary provision of public goods under varying income distributions. *Canadian Journal of Economics*, 54–69.

Clark, A. E. and A. J. Oswald (1996). Satisfaction and comparison income. *Journal of Public Economics 61*(3), 359–381.

Côté, S., J. House, and R. Willer (2015). High economic inequality leads higher-income individuals to be less generous. *Proceedings of the National Academy of Sciences 112*(52), 15838–15843.

DellaVigna, S., J. A. List, and U. Malmendier (2012). Testing for altruism and social pressure in charitable giving. *The quarterly journal of economics 127*(1), 1–56.

Derin-Güre, P. and N. Uler (2010). Charitable giving under inequality aversion. *Economics Letters 107*(2), 208–210.

Duquette, N. J. and E. P. Hargaden (2021). Inequality and giving. *Journal of Economic Behavior & Organization 186*, 189–200.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Ferreri-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics 89*(5-6), 997–1019.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations 7*(2), 117–140.

Gächter, S., C. Starmer, and F. Tufano (2015). Measuring the closeness of relationships: a comprehensive evaluation of the'inclusion of the other in the self'scale. *PLOS ONE 10*(6), e0129478.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Kőszegi, B. and M. Rabin (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics 121*(4), 1133–1165.

List, J. A. (2011). The market for charitable giving. *Journal of Economic Perspectives 25*(2), 157–180.

Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *The Quarterly Journal of Economics 120*(3), 963–1002.

Marmot, M. (2005). *Status syndrome: How your social standing directly affects your health.* A&C Black.

Maus, P., M. Montero, and M. Sefton (2023). Cedex discussion paper series issn 1749-3293.

Perez-Truglia, R. (2020). The effects of income transparency on well-being: Evidence from a natural experiment. *American Economic Review 110*(4), 1019–54.

Ribar, D. C. and M. O. Wilhelm (2002). Altruistic and joy-of-giving motivations in charitable behavior. *Journal of Political Economy 110*(2), 425–457.

Schwerter, F. (2023). Social reference points and risk taking. *Management Science*.

Sheremeta, R. M. and N. Uler (2021). The impact of taxes and wasteful government spending on giving. *Experimental Economics 24*, 355–386.

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

Warr, P. G. (1983). The private provision of a public good is independent of the distribution of income. *Economics Letters 13*(2-3), 207–211.

Appendices

3.A Additional Analysis

Table 3.A1: OLS Regressions: Donations as dependent variable

	Pooled	Female	Male
High reference earnings	-0.002	-0.004	0.031
	(0.269)	(0.369)	(0.374)
Social	0.118	-0.473	0.873***
	(0.263)	(0.359)	(0.370)
High reference earnings x Social	-0.340	0.459	-1.376***
	(0.381)	(0.520)	(0.531)
Behindness averse (dummy)	-0.395**	-0.188	-0.762***
	0.191	(0.261)	(0.272)
Leading averse (dummy)	1.026***	0.911***	0.976***
	(0.198)	(0.285)	(0.267)
Constant	1.670***	1.923***	1.482**
	(0.235)	(0.336)	(0.316)
Observations	360	201	159

The table reports estimates from OLS regressions. The dependent variable is the donations to WWF's Climate Relief Fund. Column (1) reports the regression analysis for all treatments and the whole sample, Column (2) for females only and column (3) for males only. No decision-makers are excluded from the analysis in columns (1) and (2). In column (3) decision-makers with multiple switch-points for our measures of α or β are excluded. Standard errors are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

3.B Recruitment message

Dear fname,

You are registered with CeDEx to participate in experiments. We would like to invite you to take part in our upcoming experiment. The experiment will take place online and can take from 5 to 20 minutes. All payments will be done with PayPal.

For this experiment, you will need Microsoft Teams and your camera on for the first 10 minutes of the experiment. We will videocall you at the time you have registered for.

You will need to use a computer for this experiment. Please, do not try to use your mobile phones or tablets instead.

We are planning to run many sessions this week and more sessions might be added.

If you would like to participate, please click on the link below to sign up for the session of your choice. Please note: people that sign up for a session and do not turn up cause us problems. Please sign up only if you are sure you can attend it and if you sign up please do attend. We operate a policy of removing participants from our database who sign up but then fail to turn up.

#link#

(If the link does not work, copy it and paste it into the address field of your internet browser.)

Best regards,

CeDEx Team

3.C Welcome Speech

At the beginning of each experimental session, the following welcome speech is given by the experimenter in the social and non-social treatments respectively.

Social Treatments

Welcome to our experiment today! In this experiment, you have to make an allocation decision. However, only one of you will participate in this task – Person A. The amount of money you earn as Person A depends on your decision during the experiment. The second person, Person B, will receive a fixed amount of money and will not make an allocation decision. Both the amount of money and which one of you will participate in the allocation task will depend on luck! For this purpose, I will toss two coins. Both coin tosses will be visible for you. The first coin determines the amount of money, Person B will receive. If the coin shows £5, Person B will receive £5. If the coin shows £2, Person B will receive £2. Then I will toss another coin. If the coin shows green, [first name of first participant] will be Person A and will take part in the allocation task. If the coin shows blue, [first name of second participant] will be Person A and will take part in the allocation task. After that, Person A will receive a link that leads him to the instructions of the allocation task and both of you can leave the videocall.

Non-social Treatments

Welcome to our experiment today! In this experiment, you have to make an allocation decision. However, you will only participate in this task if you are assigned the role of Person A. If you are assigned the role of Person B, you will receive a fixed amount of money and will not make an allocation decision. Both the amount of money and whether you will participate in the allocation task will depend on luck! For this purpose, I will toss two coins. Both coin tosses will be visible for you. The first coin determines the amount of money, you would receive as Person B. If the coin shows £5, you will receive £5 as Person B. If the coin shows £2, you will receive £2 as Person B. Then I will toss another coin. If the coin shows green, you will be Person A and will take part in the allocation task. If the coin shows blue, you will be Person

B and will not take part in the allocation task. If you are assigned the role of Person A, you will receive a link that leads you to the instructions of the allocation task and you can leave the videocall. If you are assigned the role of Person B, you will receive the fixed amount of money and can leave the videocall.

3.D Instructions for Person A

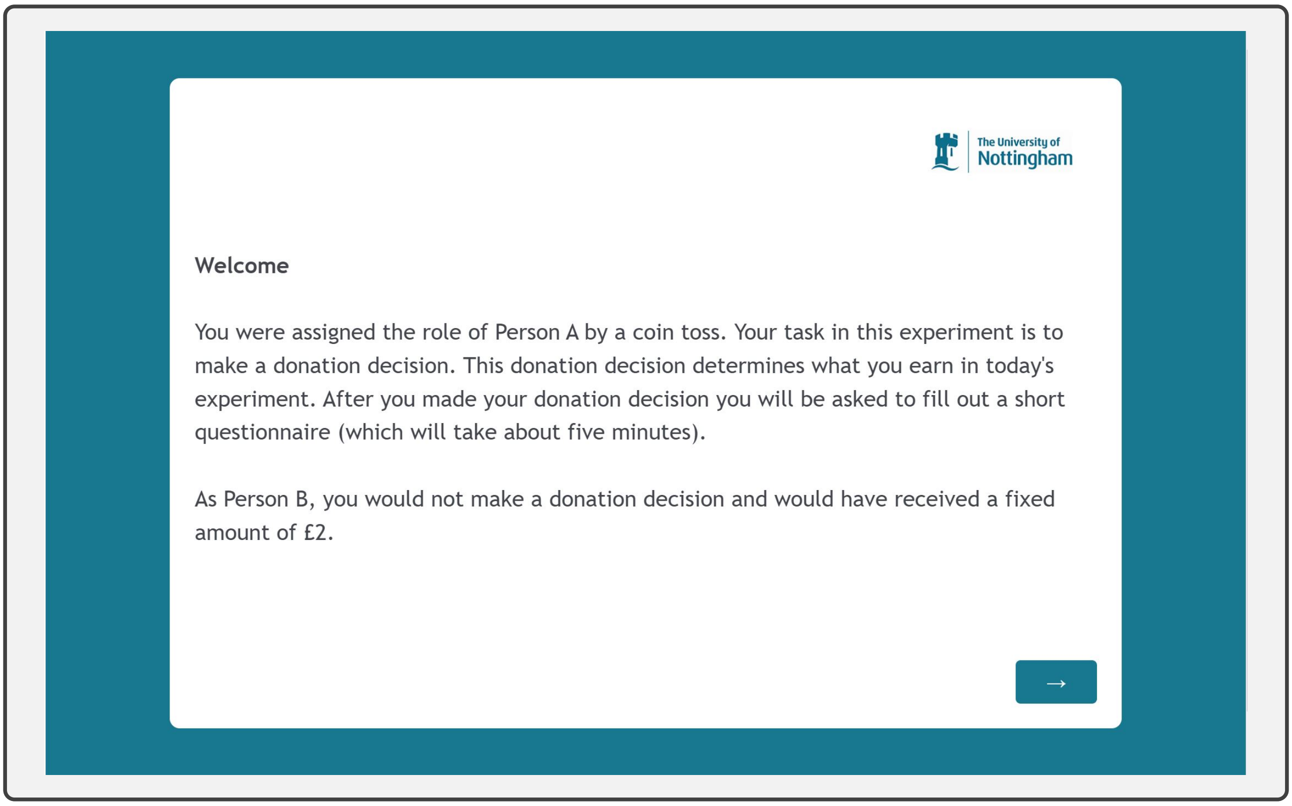
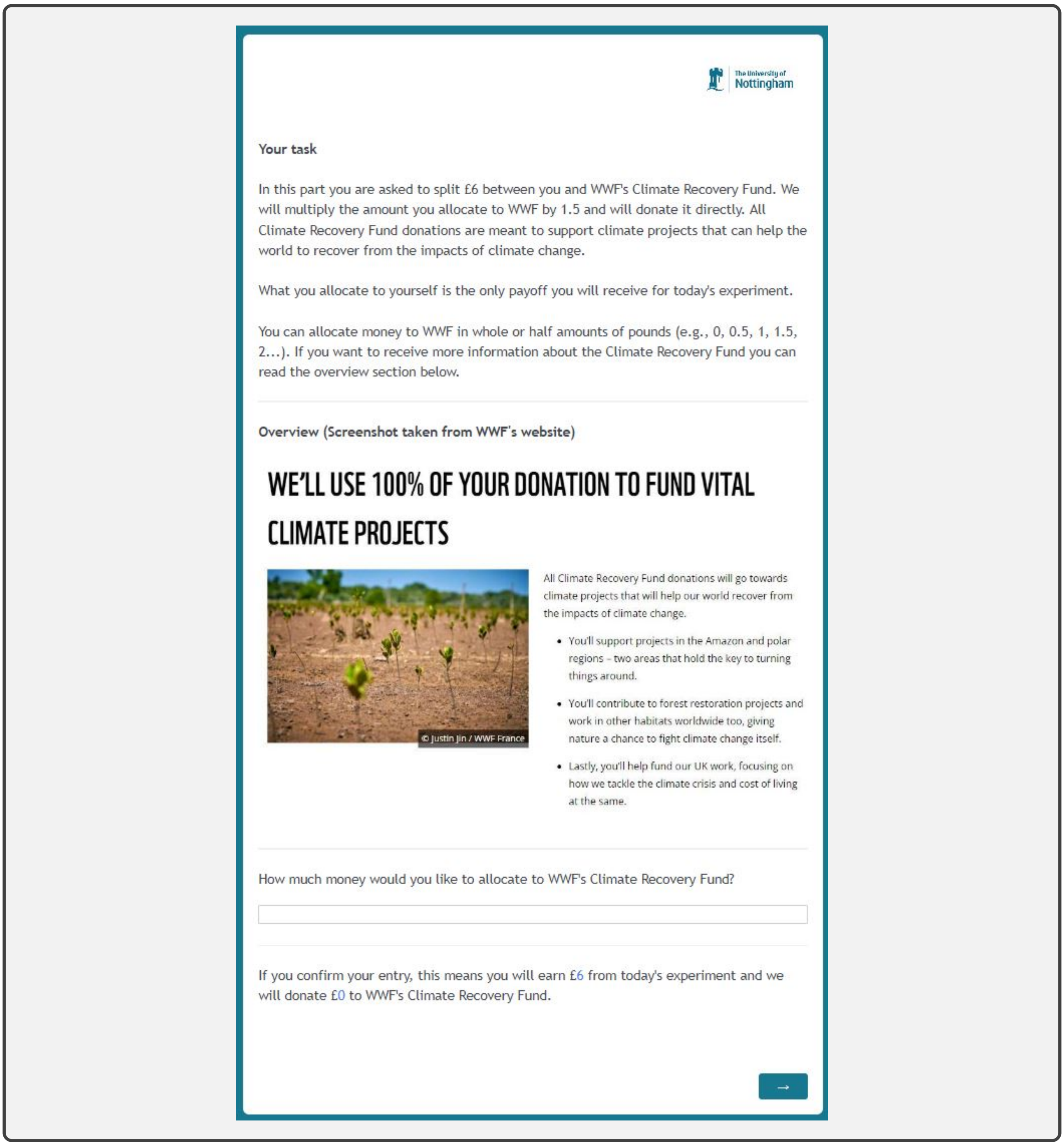


Figure 3.D1: Screenshot of the first screen



The University of Nottingham

Your task

In this part you are asked to split £6 between you and WWF's Climate Recovery Fund. We will multiply the amount you allocate to WWF by 1.5 and will donate it directly. All Climate Recovery Fund donations are meant to support climate projects that can help the world to recover from the impacts of climate change.

What you allocate to yourself is the only payoff you will receive for today's experiment.

You can allocate money to WWF in whole or half amounts of pounds (e.g., 0, 0.5, 1, 1.5, 2...). If you want to receive more information about the Climate Recovery Fund you can read the overview section below.

Overview (Screenshot taken from WWF's website)

WE'LL USE 100% OF YOUR DONATION TO FUND VITAL CLIMATE PROJECTS

© Justin Jin / WWF France

All Climate Recovery Fund donations will go towards climate projects that will help our world recover from the impacts of climate change.

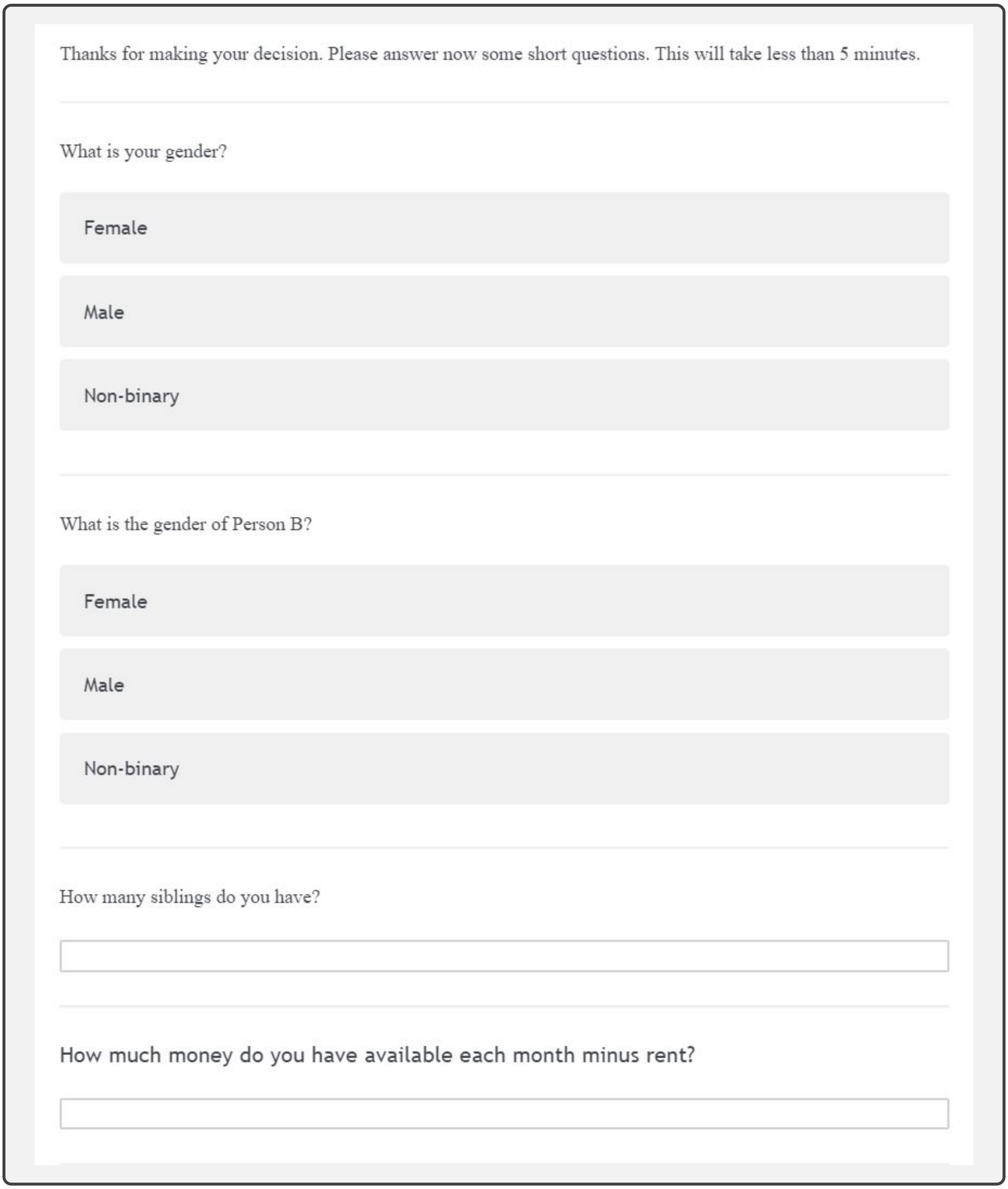
- You'll support projects in the Amazon and polar regions – two areas that hold the key to turning things around.
- You'll contribute to forest restoration projects and work in other habitats worldwide too, giving nature a chance to fight climate change itself.
- Lastly, you'll help fund our UK work, focusing on how we tackle the climate crisis and cost of living at the same.

How much money would you like to allocate to WWF's Climate Recovery Fund?

If you confirm your entry, this means you will earn £6 from today's experiment and we will donate £0 to WWF's Climate Recovery Fund.

→

Figure 3.D2: Screenshot of the second screen


Thanks for making your decision. Please answer now some short questions. This will take less than 5 minutes.

What is your gender?

Female

Male

Non-binary

What is the gender of Person B?

Female

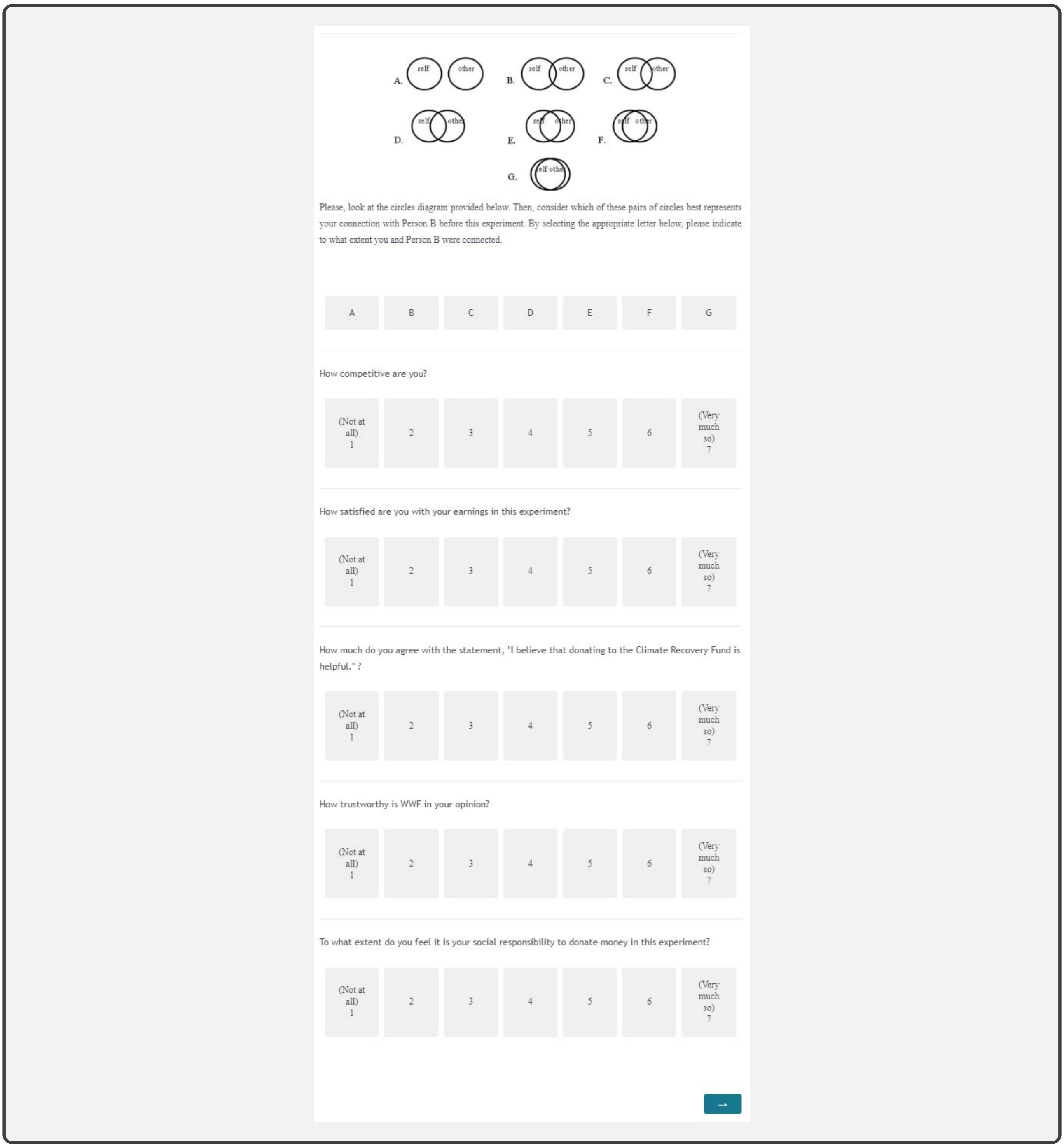
Male

Non-binary

How many siblings do you have?

How much money do you have available each month minus rent?

Figure 3.D3: Screenshot of the third screen



Please, look at the circles diagram provided below. Then, consider which of these pairs of circles best represents your connection with Person B before this experiment. By selecting the appropriate letter below, please indicate to what extent you and Person B were connected.

A	B	C	D	E	F	G

How competitive are you?

(Not at all) 1	2	3	4	5	6	(Very much so) 7

How satisfied are you with your earnings in this experiment?

(Not at all) 1	2	3	4	5	6	(Very much so) 7

How much do you agree with the statement, "I believe that donating to the Climate Recovery Fund is helpful." ?

(Not at all) 1	2	3	4	5	6	(Very much so) 7

How trustworthy is WWF in your opinion?

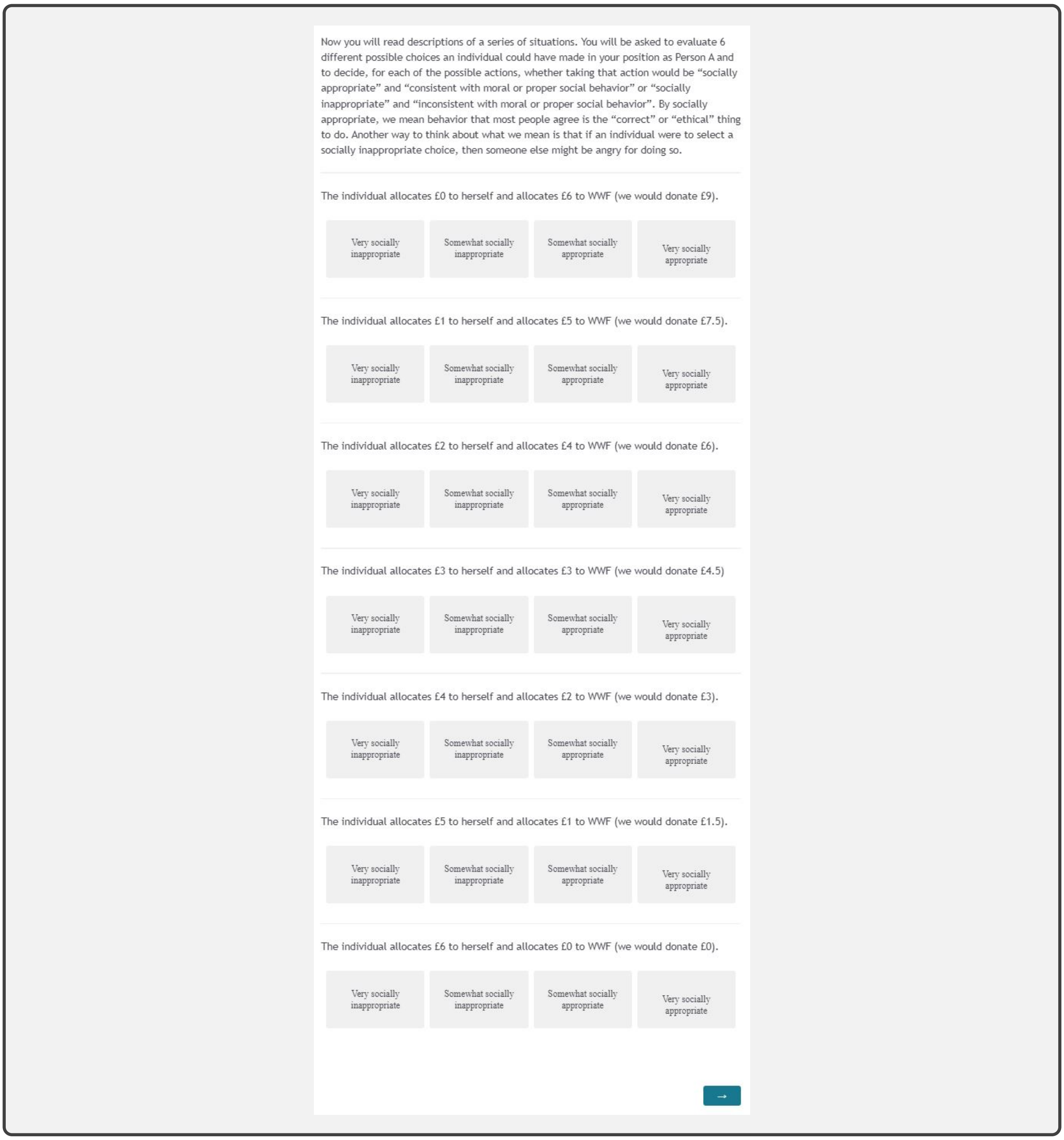
(Not at all) 1	2	3	4	5	6	(Very much so) 7

To what extent do you feel it is your social responsibility to donate money in this experiment?

(Not at all) 1	2	3	4	5	6	(Very much so) 7

→

Figure 3.D4: Screenshot of the third screen ctd.



Now you will read descriptions of a series of situations. You will be asked to evaluate 6 different possible choices an individual could have made in your position as Person A and to decide, for each of the possible actions, whether taking that action would be "socially appropriate" and "consistent with moral or proper social behavior" or "socially inappropriate" and "inconsistent with moral or proper social behavior". By socially appropriate, we mean behavior that most people agree is the "correct" or "ethical" thing to do. Another way to think about what we mean is that if an individual were to select a socially inappropriate choice, then someone else might be angry for doing so.

The individual allocates £0 to herself and allocates £6 to WWF (we would donate £9).

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate

The individual allocates £1 to herself and allocates £5 to WWF (we would donate £7.5).

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate

The individual allocates £2 to herself and allocates £4 to WWF (we would donate £6).

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate

The individual allocates £3 to herself and allocates £3 to WWF (we would donate £4.5)

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate

The individual allocates £4 to herself and allocates £2 to WWF (we would donate £3).

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate

The individual allocates £5 to herself and allocates £1 to WWF (we would donate £1.5).

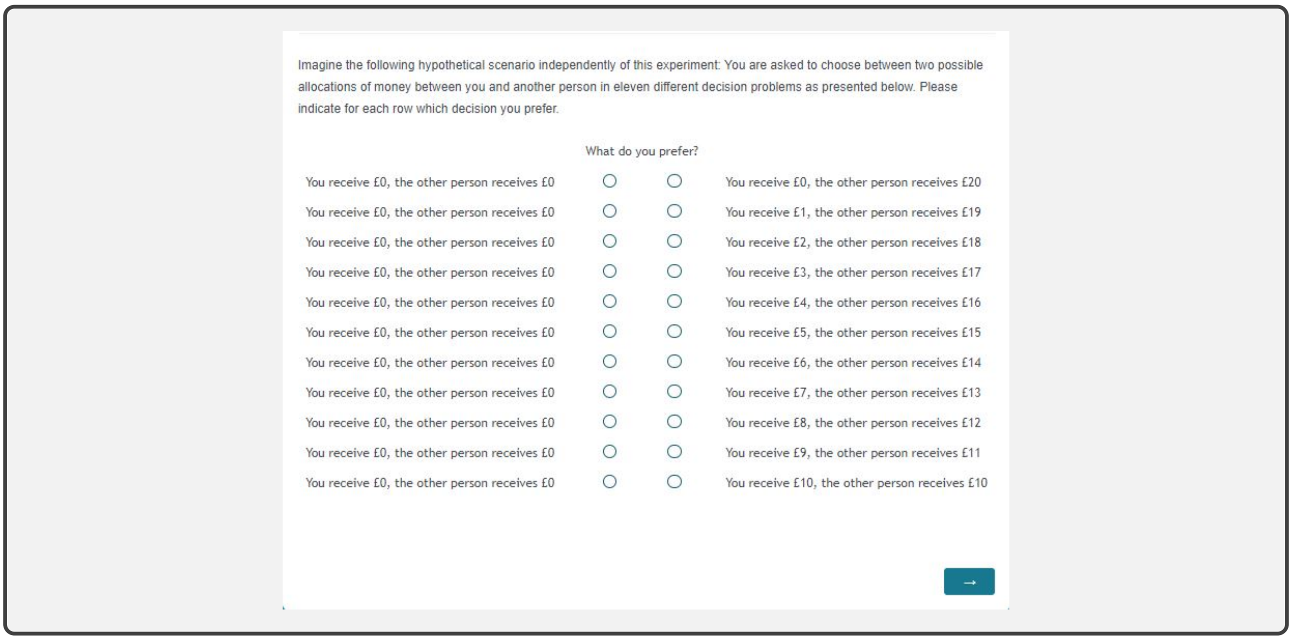
Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate

The individual allocates £6 to herself and allocates £0 to WWF (we would donate £0).

Very socially inappropriate	Somewhat socially inappropriate	Somewhat socially appropriate	Very socially appropriate

→

Figure 3.D5: Screenshot of the fourth screen.



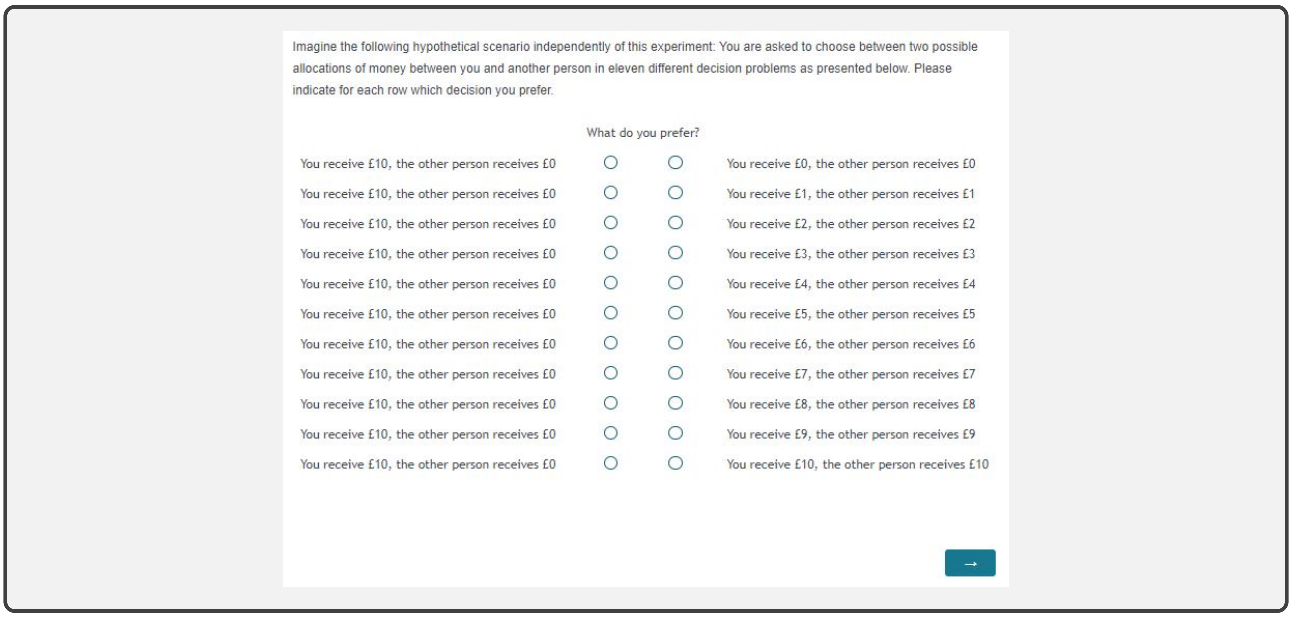
Imagine the following hypothetical scenario independently of this experiment: You are asked to choose between two possible allocations of money between you and another person in eleven different decision problems as presented below. Please indicate for each row which decision you prefer.

What do you prefer?

You receive £0, the other person receives £0	You receive £0, the other person receives £20
You receive £0, the other person receives £0	You receive £1, the other person receives £19
You receive £0, the other person receives £0	You receive £2, the other person receives £18
You receive £0, the other person receives £0	You receive £3, the other person receives £17
You receive £0, the other person receives £0	You receive £4, the other person receives £16
You receive £0, the other person receives £0	You receive £5, the other person receives £15
You receive £0, the other person receives £0	You receive £6, the other person receives £14
You receive £0, the other person receives £0	You receive £7, the other person receives £13
You receive £0, the other person receives £0	You receive £8, the other person receives £12
You receive £0, the other person receives £0	You receive £9, the other person receives £11
You receive £0, the other person receives £0	You receive £10, the other person receives £10

→

Figure 3.D6: Screenshot of the fifth screen



Imagine the following hypothetical scenario independently of this experiment: You are asked to choose between two possible allocations of money between you and another person in eleven different decision problems as presented below. Please indicate for each row which decision you prefer.

What do you prefer?

You receive £10, the other person receives £0	You receive £0, the other person receives £0
You receive £10, the other person receives £0	You receive £1, the other person receives £1
You receive £10, the other person receives £0	You receive £2, the other person receives £2
You receive £10, the other person receives £0	You receive £3, the other person receives £3
You receive £10, the other person receives £0	You receive £4, the other person receives £4
You receive £10, the other person receives £0	You receive £5, the other person receives £5
You receive £10, the other person receives £0	You receive £6, the other person receives £6
You receive £10, the other person receives £0	You receive £7, the other person receives £7
You receive £10, the other person receives £0	You receive £8, the other person receives £8
You receive £10, the other person receives £0	You receive £9, the other person receives £9
You receive £10, the other person receives £0	You receive £10, the other person receives £10

→

Figure 3.D7: Screenshot of the sixth screen

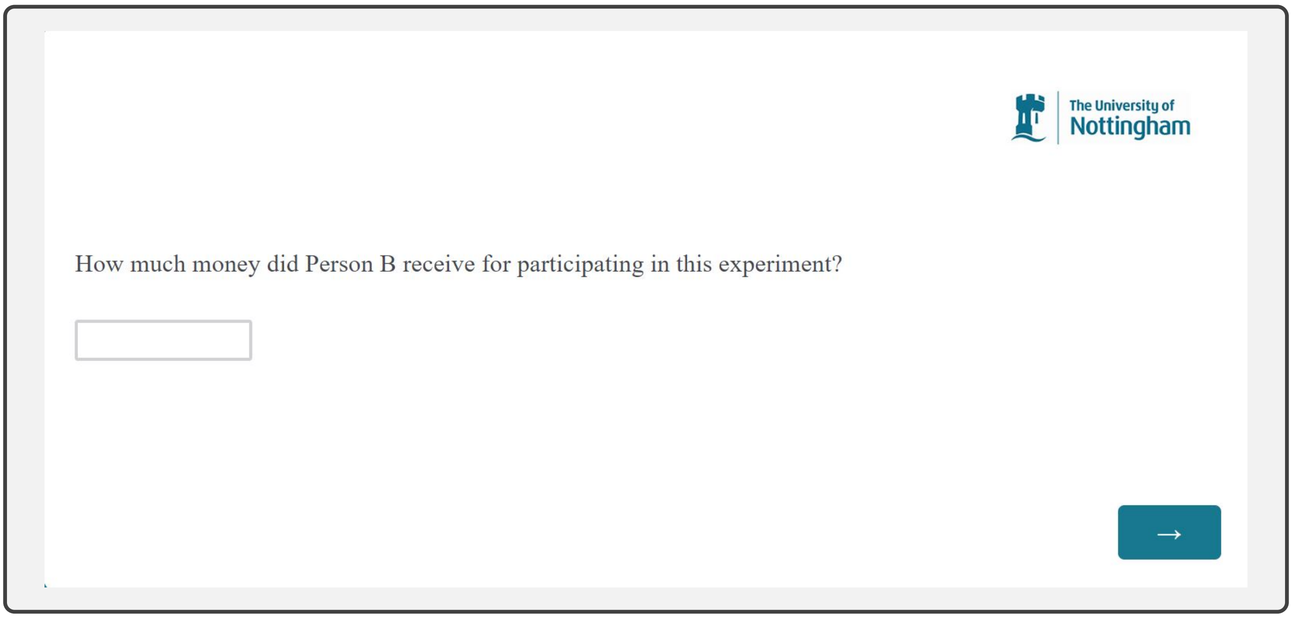


Figure 3.D8: Screenshot of the seventh screen

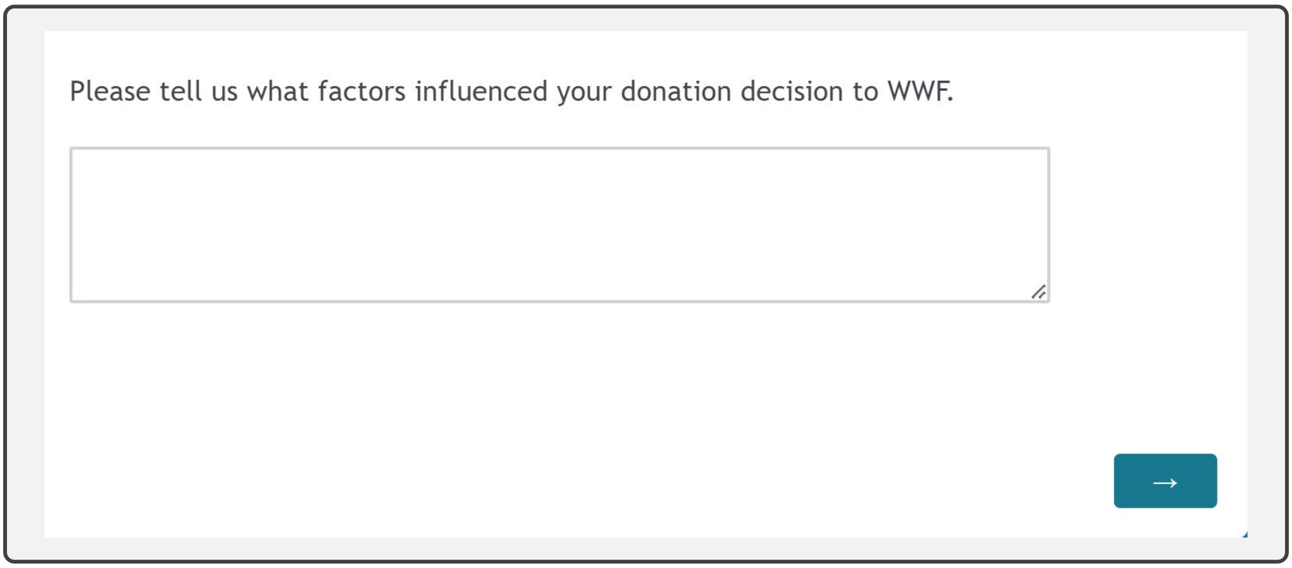


Figure 3.D9: Screenshot of the eighth screen

Chapter 4

Compromise Effects in Bargaining games: Experimental Evidence

4.1 Introduction

A crucial assumption of rational choice theory is that decision-makers have well-defined preferences over all available options. It implies that a change in the composition of the choice set should not affect the decision maker's preferences over two alternatives. However, a large body of work in psychology and economics shows that this assumption is frequently violated. In many situations, the choice between two alternatives depends on which other options are available. In this paper, we study the compromise effect which refers to a tendency to choose non-extreme options (Simonson, 1989). To illustrate this suppose that a person must choose between two cars, A and B, that differ in two salient attributes, such as safety and speed. The choice is nontrivial since A is safer than B but also slower than B. Consider now we add a third car C that is even faster and more unsafe than B - this makes B a compromise between A and C as it is the second best on both attribute dimensions. The compromise effect (CE) occurs when the decision-maker is more likely to choose B when the set of options is $\{A, B, C\}$ than when it is $\{A, B\}$, a violation of regularity (Luce, 1977).

The compromise effect has been found to significantly influence choice in a variety of situations.[1] However, there is scant evidence on the role of the compromise effect in bargaining.

[1] Such as product choice (Doyle et al. (1999), Munro and Popov (2013)), contingent valuation (Bateman et al. (2008), job candidate selection (Highhouse (1996), Slaughter et al. (2006), Kuncel and Dahlke (2020)),

The only studies that we are aware of are Galeotti et al. (2019) and Galeotti et al. (2021) which investigated the compromise effect in unstructured two-player bargaining games. In their experiments, players can send messages to one another and make proposals from a menu of contracts, where a contract specifies two attributes, how much Player 1 receives and how much Player 2 receives. When a proposal is accepted, it is implemented.

Galeotti et al. (2019) considered the trade-off between equality and efficiency in a bargaining setting. Their games vary the severity of the trade-off between equality and total-earnings maximization. Some of their games are two-sided, consisting of one equal contract and two symmetrically unequal contracts, whereas others are one-sided, consisting of one equal and one unequal contract. In comparing one-sided and two-sided games, they observed that the equal contract is more likely to be agreed on in two-sided games, a finding they interpret as a compromise effect. Similarly, Galeotti et al. (2021) observe some significant compromise effects, but these are sensitive to whether or not an exactly equal payoff is available.

Due to the complexity of bargaining in unstructured situations it is difficult to determine the mechanism behind the compromise effect. When a contract becomes less extreme, is it proposed more often by one of the players? Does it become more acceptable to the other player? And what is the reason it is proposed more often and/or becomes more acceptable?

In this paper, we examine compromise effects and the mechanism behind them by varying the set of feasible contracts in simple structured bargaining games. Specifically, we use Ultimatum Games as they have a very straightforward structure, where a proposer makes a take-it-or-leave-it proposal about how to split a pie and a receiver can either accept or reject the proposal. For more detailed information on the acceptability of contracts, receivers indicate whether they would accept or reject each feasible contract (i.e. we use the strategy method, see Brandts and Charness (2011)). For further insight into proposer/responder behavior, we use two more, even simpler, protocols. Using Dictator Games, where the receiver cannot reject a proposal, we can observe whether proposer preferences are affected by adding a menu option. Finally, we use Random Proposal Games, in which the receivers indicate the acceptability of proposals when the proposal is chosen by Nature. This allows us to observe whether the acceptability of

sampling decision (Trueblood et al. (2013), Noguchi and Stewart (2014)), elections (O'Curry and Pitts (1995), Herne (1997)), choices among gambles (Wedell (1991); Herne (1999); De Haan and Van Veldhuizen (2015); Beauchamp et al. (2020), legal judgments (Kelman et al. (1996)), menu choices in restaurants in a restaurant (Pinger et al. (2016)) and partner choice (Sedikides et al. (1999)).

contracts is subject to 'pure' menu effects, independent of the intentions of proposers. In each of these games, we vary the set of feasible contracts and observe how this changes behavior and outcomes.

In our base scenarios, there is a Greedy contract (80-20) alongside a Fair contract. We examine two base scenarios where the Fair contract is either exactly equal (50-50) or nearly equal (60-40). In our other scenarios, we add a third Decoy contract. In some of these the Decoy is (90-10), making the Greedy contract a compromise, and in others it is (20-80), making the Fair contract a compromise.

We find that the addition of a third contract in the Ultimatum Games does matter for both responders and proposers but only when the Fair contract is unequal. Responders are significantly more likely to accept the Greedy contract once an even greedier Decoy contract is added to the set of feasible contracts. Similarly, proposers are more likely to switch to offering a more unfair contract once an even greedier contract is added to the set of feasible contracts. However, since in our experiment many proposers chose the greediest contract when it became available, we cannot detect an increase in the frequency of a contract when it becomes a compromise. When the Fair contract is equal, we cannot find any statistically significant differences for the behavior of proposers or respondents and for the results of the Ultimatum Games as a whole.

In the Dictator Games, we also see that a large proportion of proposers switch to the greediest contract when it becomes available. However, in contrast to the Ultimatum Games, we do not observe that proposers are more likely to switch from the Fair to the Greedy contract once an even greedier contract is added to the set of feasible contracts. This suggests that proposers in the Ultimatum Game anticipate changes in the responders' rejection rates and that their change in behavior cannot be explained by pure menu effects.

In the Random Proposal Games, we see a similar picture as in the Ultimatum Games. Responders are significantly more likely to accept the Greedy contract once an even greedier Decoy is added to the set of feasible contracts for the base Scenario if the Fair contract is unequal. This suggests that the effects observed in the Ultimatum Games survive even when possible reasons for reciprocity are removed.

Ours is not the first experiment to identify menu effects in Ultimatum Games. Falk et al. (2003) use a similar framework and show that given offers in Ultimatum Games generate systematically different rejection rates depending on the other offers that are available to the proposer. They interpret this as evidence of reciprocity. Their design is different from ours in numerous ways, for example, they always consider games where the proposer has two options to choose from, while we consider the effects of adding options to the proposer's choice set. Moreover, our design allows us to isolate the role of reciprocity considerations and observe pure compromise effects. We find that the acceptance behavior of responders toward a given proposal changes when another proposal option is added, even when reciprocity considerations are absent.

The remainder of the paper is organized as follows. In Section 2, we describe our experimental design. In Section 3, we outline our hypotheses. In Section 4, we report our experimental findings and Section 5 concludes.

4.2 Experimental Design

We design an experiment to explore the compromise effect in a structured bargaining situation. The experiment is designed to create an environment where we can exogenously manipulate the menu of contracts, observe how this influences bargaining outcomes, and provide insights into the underlying behavior.

4.2.1 The Bargaining Games

Our experiment uses a within-subject design where subjects play several different bargaining games with no feedback between games. Specifically, we use Ultimatum Games as they have a particularly simple structure, where a proposer makes a take-it-or-leave-it proposal and a receiver can either accept or reject the proposal. In our baseline scenarios, we restrict the set of feasible proposals to just two - a Greedy contract which splits the pie 80-20 in favor of the proposer, or a Fair contract. In fact, we conduct two baseline scenarios varying whether the Fair contract offers an exactly 50-50 split or a near 50-50 split, since Galeotti et al. (2021) find that the availability of an exactly equal division reduces compromise effects. We were interested to see if this would be observed in an Ultimatum Game environment, and, if so, get a better understanding of what drives it.

We compare outcomes in these baseline scenarios with outcomes in other scenarios where we add an additional feasible contract. In one scenario we add an extremely Greedy 90-10 contract, making the Greedy contract a compromise. In another scenario we add a generous 20-80 contract, making the Fair contract a compromise. In all scenarios, subjects split a £20 pie, so that, for example, the Greedy contract gives the proposer £16 and the receiver £4. The scenarios are summarised in Table 4.1.

Table 4.1: Table of feasible Contracts

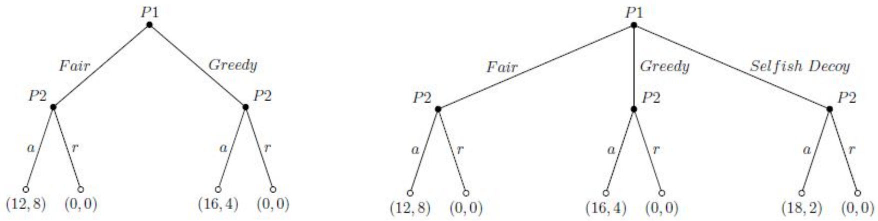
Game	**Feasible Contracts**			
	Generous Decoy	**Fair**	**Greedy**	**Selfish Decoy**
Scenario 1	-	(12,8)	(16,4)	-
Scenario 2	-	(12,8)	(16,4)	(18,2)
Scenario 3	(4,16)	(12,8)	(16,4)	-
Scenario 4	-	(10,10)	(16,4)	-
Scenario 5	-	(10,10)	(16,4)	(18,2)
Scenario 6	(4,16)	(10,10)	(16,4)	-

The Ultimatum Games are very simple and share the same structure. In each scenari, the proposer is asked to divide £20 between himself and the responder who can either accept or reject the proposal. Accepting the proposal leads to a payoff distribution according to the proposer's offer. A rejection implies zero payoffs for both players. We employ the strategy method as we want to see how receivers would respond to any offer. This means that every responder has to indicate his action at all decision nodes, i.e., for the case that the Fair contract, the Greedy contract and the Decoy contract has been offered, without knowing the decision of the proposer.[2]

As an example, Figure 4.1 illustrates Scenarios 1 and 2. In both scenarios, the proposer can propose two contracts, the Fair contract and the Greedy contract. The Fair and the Greedy contracts are the same across the first three scenarios. In the base scenario, there is no third contract available. In the second scenario, we add the Selfish Decoy contract (18,2) that makes the Greedy contract a compromise. We say a contract becomes a compromise if, for each player,

[2]Further, it helps with the no feedback feature of our design.

its money payouts are the second highest, whereas each of the other contracts gives one player its highest payoff and the other its lowest. The only difference between the base scenario and the compromise effect scenario is whether the Selfish Decoy contract is feasible or not.



(a) Base Game (b) Compromise Effect Game

Figure 4.1: Panel (a) shows the base scenario. In panel (b), we introduce a Selfish Decoy that makes the Greedy contract a compromise between the Fair contract and the Selfish Decoy.

Similarly, the only difference between Scenario 1 and Scenario 3 is whether the Generous Decoy contract is feasible or not. A large literature shows that fair proposals are typically observed in Ultimatum Games, and so it may be the case that adding a generous proposal has no impact on outcomes. Nevertheless, we wanted to see whether the frequency of the Fair contract could still increase if the Generous Decoy results in making the Greedy outcome less attractive to the proposer and/or less acceptable to the receiver.

In the Ultimatum Games, the behavior of the proposers might be influenced by beliefs about differing rejection rates of the responder depending on the available feasible contracts. To test whether proposer behavior is changing due to a change in proposer preferences over outcomes or due to a change in beliefs about responder behavior, we additionally conduct Dictator Games in which the scope for strategic thinking about the responder's behavior is removed. In the Dictator Games, the proposer chooses from the menus described in Table 4.1, the receiver has no choice, and the proposal is implemented. Given that responders can not reject a proposal, the behavior of the proposer reflects which allocation of money he prefers the most. If the change in behavior of the proposers in the Ultimatum Games is due to a change in proposer preferences, we would expect to find the qualitatively same results for the behavior of proposers in the Dictator Games as in the Ultimatum Games. The Dictator Games give us a clear indication of whether proposers' preferences over outcomes are sensitive to menu effects.

It might be that responders do not only take the distributive consequences of the action by the proposer into account but also the intention that is signaled by the action itself. For example, a receiver might reject the Greedy contract to punish the proposer for offering this contract. To test whether rejection rates are purely driven by reciprocity considerations, we additionally conduct Random Proposal Games in which the scope for reciprocal behavior to praise or punish the choice of the proposer is removed. This allows us to distinguish whether differing rejection rates are purely a reaction to proposer intentions or are menu effects. If the change in behavior of the responders is at least partly due to a menu effect, we would expect to find the qualitatively same results of responders in Random Proposal Games and in Ultimatum Games.

4.2.2 Experimental Procedures

In total we ran 6 sessions, resulting in 110 experimental subjects. Each subject participates in 6 different blocks of games: As a proposer in Ultimatum Games, in Dictator Games and in Random Proposal Games, and as a receiver in each of the games. The feasible proposals we ask the subjects to choose from or reject respectively are the same across blocks (see Table 4.1 for a table that includes all scenarios).[3]

When arriving at the lab, subjects were allocated a desk. They received printed instructions which were read out aloud by the experimenter (for Instructions, see Appendix 4B). Subjects were informed that they would make decisions in six 'scenarios' in each of the six blocks, where in each scenario they would be matched with another subject, with random re-matching between scenarios. To avoid subjects' behavior being influenced by the observed behavior of other subjects, we do not give feedback on the decision of their partners and the outcome of a scenario during the experiment. At the beginning of the experiment, subjects are randomly assigned the proposer or the responder role and they keep it through the first 3 blocks. For blocks 4-6, the roles are reversed (see Figure 4.2).

The subjects are informed about the bargaining protocol to be used in each block at the beginning of the block and are not informed about the content of the next blocks in advance. This means that the subjects know which game they have to play at the beginning of a new block, but they have no information about the upcoming blocks. Similarly, subjects are informed of

[3]For every Scenario in every block, we elicit beliefs about their matched partner's behavior. Since we want to avoid hedging effects and the relevance of the decisions in each game should be as clear as possible, we do not incentivize the elicitation of beliefs.

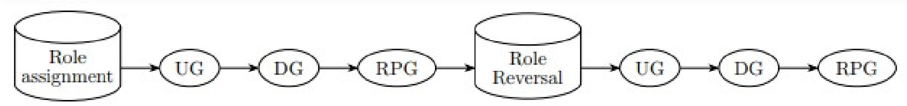


Figure 4.2: Procedure overview of the experimental structure, where UG stands for Ultimatum Game, DG for Dictator Game and RPG for Random Proposal Game.

the set of feasible contracts in a scenario when they are confronted with the scenario, but they have no information about the upcoming scenarios. Subjects are also not informed that they will face a role reversal for blocks 4-6.

The order of the blocks and the order for the scenarios in each block of Ultimatum Games, Dictator Games and Random Proposal Games is varied between sessions. We vary the order of the blocks such that two of the six sessions start with the same game and we never have the same order of games, while keeping the order of the games before and after the role reversal constant.[4] We vary the order of the scenarios such that each of the six blocks within a session starts with a different scenario. Between the six experimental sessions, each block starts and ends with a different scenario. The order in which subjects faced the scenarios in each session can be found in Appendix 4.A. Since we balance the order in which subjects face the scenarios, any learning or order effects in the data would not systematically bias our comparisons across scenarios.

At the end of an experimental session, one scenario in one block is randomly selected for payment. Then, subjects receive a show-up fee of £5 plus their earnings from the randomly chosen scenario. The experiment was computerized and conducted with the program Lioness (Arechar et al., 2018) and subjects were invited from the CeDEx subject pool consisting of students at the University of Nottingham using ORSEE ((Greiner, 2015).

4.3 Hypotheses

Given the existing literature on compromise effects, we expect that subjects are more likely to agree on a contract in the Ultimatum Game when it is made a compromise. Consequently, when the Greedy contract is targeted as a compromise by adding the Selfish Decoy contract

[4]Consequently we have six sessions with the following orders: (UG, DG, RPG), (UG, RPG, DG) (DG, UG, RPG), (DG, RPG, UG), (RPG, UG, DG), (RPG, DG, UG).

(18,2), we expect proposers and responders to be more likely to agree on the Greedy contract compared to the base scenario. When the Fair contract is targeted as a compromise by adding the Generous Decoy contract (4,16), we expect proposers and responders to be more likely to agree on the Fair contract compared to the base scenario.

H4.1: The frequency of agreeing on the Greedy contract is greater when the Greedy contract is made a compromise compared to the base scenario.

H4.2: The frequency of agreeing on the Fair contract is greater when the Fair contract is made a compromise compared to the base scenario.

There are two separate issues when we determine how likely a feasible contract is going to be implemented in the Ultimatum Games: 1) How likely the proposer is to offer a contract and 2) how likely the receiver is to accept the contract. For the hypothesized compromise effect to occur, it must be the case that the introduction of the Decoy contract makes the target contract more likely to be offered, more likely to be accepted, or both. We formulate separate hypotheses about each of these possibilities.

First, regarding the acceptability of the target contract, we expect that when the Greedy contract is targeted as a compromise by adding the Selfish Decoy contract (18,2), responders will be less likely to reject the Greedy contract compared to the base scenario. When the Fair contract is targeted as a compromise by adding the Selfish Decoy contract (4,16), we expect responders to be more likely to reject the Greedy contract compared to the base scenario.

H4.3: The Greedy contract is rejected less frequently when the Selfish Decoy is added to the list of feasible contracts.

H4.4: The Greedy contract is rejected more frequently when the Generous Decoy is added to the list of feasible contracts.

We expect that the Fair contract is already always (or nearly always) accepted, so any change in responder behavior would be in rejecting the Greedy contract more often.

Second, in terms of the propensity for proposers to propose the target contract, we expect to observe a higher frequency of proposers offering the Greedy contract when the Selfish Decoy contract (18,2) is added to the list of feasible payoffs. When the Fair contract is targeted as a compromise by adding the Generous Decoy contract (4,16), we expect proposers to be more likely to offer the Fair contract compared to the base scenario.

H4.5: The Greedy contract is offered more frequently when the Selfish Decoy (18,2) is added to the list of feasible contracts.

H4.6: The Fair contract is offered more frequently when the Generous Decoy (4,16) is added to the list of feasible contracts.

Hypothesis 4.3 to Hypothesis 4.6 arise from the basic idea of a compromise effect. However, it may be very difficult to observe a compromise effect if the additional contract is very attractive to the proposer. While our hypotheses for the responders are straightforward, the hypotheses for the behavior of the proposers are not. This is because the Selfish Decoy contract (18.2), when added to the list of possible contracts, gives the highest possible payout to the proposer. Naturally, we would expect some proposers who offered the Greedy contract in the scenarios where the Selfish Decoy is not available to switch to the Selfish Decoy once it is available. If a significant proportion of proposers switch to the Selfish Decoy contract when it becomes available, we would not expect to see a compromise effect in the original sense.

The advantage of using a within-subject design is that we can observe exactly how an individual's behavior changes between scenarios. This allows us to hypothesize not only how proposal rates differ between scenarios, but also how an individual's decision changes based on their choices in a different scenario. If proposers' preferences about contracts and proposers' beliefs about response behavior are unchanged, we would not expect proposers to switch from the Fair contract to the Greedy contract after a decoy becomes available. Such a switch would be a violation of revealed preference.

The compromise effect would imply that some subjects choose the Fair contract over the Greedy contract in the base scenario, but the Greedy over either the Fair contract or the Selfish Decoy contract in the scenario where three contracts are available. Note that a violation of revealed preference could go in a direction consistent with the compromise effect (i.e., choosing Greedy

in the base scenario and Fair when the Selfish Decoy is added) or the opposite direction. This observation motivates our next set of hypotheses: the proportion of subjects violating revealed preference by switching to the target contract when the Decoy is added is greater than the proportion of subjects violating revealed preference in the opposite direction. If a subject violates revealed preference, are violations in one direction more likely than another?
Consequently, we would expect to see that out of all the switches we observe after the Selfish Decoy is added, switches from the Fair to the Greedy contract are more likely than switches from the Greedy to the Fair contract. Similarly, we would expect to see that out of all the switches we observe after the Generous Decoy is added, switches from the Greedy to the Fair contract are more likely than switches from the Fair to the Greedy contract.
In other words: We expect that, out of the proposers who switched between the Fair and the Greedy contract when the Selfish Decoy is added, there are more proposers offering the Fair contract in Scenario 1 (4) and the Greedy contract in Scenario 2 (5) than proposers offering the Greedy contract in Scenario 1 (4) and the Fair contract in Scenario 2 (5). Similarly, we expect that, out of the proposers who switched between the Fair and the Greedy contract when the Generous decoy is added, there are more proposers offering the Greedy contract in Scenario 1 (4) and the Fair contract in Scenario 3 (6) than proposers offering the Fair contract in Scenario 1 (4) and the Greedy contract in Scenario 3 (6).

H4.7: Proposers are more likely to switch from the Fair contract to the Greedy contract compared to the reverse after the Selfish Decoy is added.

H4.8: Proposers are more likely to switch from the Greedy contract to the Fair contract compared to the reverse after the Generous Decoy is added.

For each of the hypotheses H4.1-H4.8, we conduct two separate tests, one for each of the base scenarios. We anticipate compromise effects to be stronger in the case where the Fair contract in the base Scenario is slightly unequal.
In the Ultimatum Game, the behavior of the proposers might be influenced by beliefs about differing rejection rates of the responder depending on the available feasible contracts. To test, whether proposals are changing due to a change in beliefs, we additionally conduct Dictator

Games in which the scope for strategic thinking about the responder's behavior is removed. If the change in behavior of the proposers is at least partly driven by a change in proposer preferences, we would expect the behavior of proposers in the Dictator Games to be qualitatively the same as in the Ultimatum Games.

H4.9: In the Dictator Games, the Greedy contract is offered more frequently when the Selfish Decoy (18,2) is added to the list of feasible contracts.

H4.10: In the Dictator Games, the Fair contract is offered more frequently when the Generous Decoy (4,16) is added to the list of feasible contracts.

Similarly to the predictions for the Ultimatum Games, we would expect to see more individual switches in the direction of the compromise effect than in the opposite direction.

H4.11: In the Dictator Games, proposers are more likely to switch from the Fair contract to the Greedy contract compared to the reverse after the Selfish Decoy is added.

H4.12: In the Dictator Games, proposers are more likely to switch from the Greedy contract to the Fair contract compared to the reverse after the Generous Decoy is added.

In the Ultimatum Game, it might be that responders do not only take into account the distributive consequences of the action by the proposer but also the intention that is signaled by the action itself. To test whether rejection rates are changing due to reciprocity considerations, we additionally conduct Random Proposal Games in which the scope for reciprocal behavior to praise or punish the choice of the proposer is removed. If the change in behavior of the responders is not exclusively driven by reciprocity, we would expect to find the qualitatively same results for the behavior of responders in the Random Proposal Games and in the Ultimatum Games.

H4.13: In the Random Proposal Games, the Greedy contract is rejected less frequently when the Selfish Decoy (18,2) is added to the list of feasible contracts.

H4.14: In the Random Proposal Games, the Greedy contract is rejected more frequently when the Generous Decoy (4,16) is added to the list of feasible contracts.

We test our hypotheses using Fisher's exact tests for the outcomes of the Ultimatum Games and Binomial tests for our remaining hypotheses accounting for our within-subject design.

4.4 Results

As the compromise effects are directional, our hypotheses predict an effect in a particular direction. Hence, all statistical tests reported to test our directed hypotheses (H4.1-H4.14) are one-sided (this is specified when we report the results). In all the other cases where we do not have a priori or directional hypotheses, the tests are two-sided. We evaluate significance at the 5% level, unless otherwise specified.

4.4.1 Outcomes

Table 4.2: Outcomes in the Ultimatum Games

Scenario	**Feasible contracts**				
	Generous (4,16)	**Fair (12,8)/ (10,10)**	**Greedy (16,4)**	**Selfish (18,2)**	**Disagree**
Scenario 1	-	76.36%	10.91%	-	12.73%
Scenario 2	-	67.27%	9.09%	3.64%	20%
Scenario 3	0%	77.06%	13.76%	-	10%
Scenario 4	-	80.00%	10.00%	-	10.00%
Scenario 5	-	76.36%	8.18%	3.64%	11.82%
Scenario 6	0%	82.73%	10.91%	-	6.36%

Table 4.2 shows the ex-post outcomes of the six Ultimatum Games, pooling the outcomes from the ultimatum games before and after the role reversal. The feasible contracts for each scenario are shown in the "Feasible contracts column", followed by the percentage of bargaining pairs who agreed on the Generous Decoy when available, the Fair contract, the Greedy contract,

the Selfish Decoy when available and the percentage of bargaining pairs who did not reach an agreement across all scenarios.

In general, in our experiment, the rate of disagreement is not very high across all scenarios (the average rate is below 12%), with the lowest being 6.36% in Scenario 6 and the highest being 20% in Scenario 2.

Initially, we expected to see a higher likelihood of bargaining pairs agreeing on the Greedy contract once we add the Selfish Decoy (18,2). Consequently, we would expect a larger share of the bargaining pairs agreeing on the Greedy contract in Scenario 2 compared to Scenario 1 and in Scenario 5 compared to Scenario 4. However, we can not observe this in the outcomes of our experiment. We find for both base scenarios that the agreement rates on the Greedy contract go down slightly once we introduce the Selfish Decoy to the games - therefore not providing support for the existence of a compromise effect in our structured bargaining setup (one-sided Fisher's exact tests, p-value=0.589 for the comparison between S1 and S2 and p-value=0.684 for the comparison between S4 and S5).

Result 4.1: The frequency of agreeing on the Greedy contract is not greater when the Greedy contract is made a compromise compared to the base scenario.

While we can not observe a compromise effect, there are some differences in the agreement rates between Scenario 1 and Scenario 2. The likelihood to agree on the Fair contract is around 10% higher when the Selfish Decoy is absent. While this is not necessarily surprising (it is natural that adding more options will reduce the frequency of the existing options) it suggests that some proposers have switched from proposing the Fair contract in Scenario 1 to proposing either the Greedy contract or the Selfish Decoy in Scenario 2.[5] However, this difference is not statistically significant (two-tailed Fisher's exact test, p-value=0.384).

Further, the probability to not arriving at an agreement was around 7% higher when there was a Selfish Decoy. It might be that the proposers tend to offer the Selfish Decoy (18,2) frequently, which may lead to a higher probability of a proposal being rejected or that responders are less likely to accept a given offer when the Selfish Decoy is added to the set of feasible proposals. But

[5]We will refer to "switching" after the addition of a Decoy, but note that some subjects will have played Scenario 2 before Scenario 1.

again, this effect is not statistically significant (two-sided Fisher's exact test, p-value=0.202). But we will go back to this when we discuss individual behavior.

Our second hypothesis predicts, that there is a higher likelihood to agree on the Fair contract compared to the base scenario once we add the Generous Decoy (16,4) to the set of feasible proposals. Consequently, we would expect a larger share of the bargaining pairs agreeing on the Fair contract in Scenario 3 compared to Scenario 1 and in Scenario 6 compared to Scenario 4. While the agreement rates on the Fair contract are higher for both Scenario 3 compared to Scenario 1 and Scenario 6 compared to Scenario 4, this effect is very small and not significantly different. (one-sided Fisher's exact test, p-value=0.500 for comparison between S1 and S3 and p-value=0.365 for comparison between S4 and S6).

Result 4.2: The frequency of agreeing on the Fair contract is not greater when the Fair contract is made a compromise compared to the base Scenario.

4.4.2 Responder behavior in the Ultimatum Games

While we find no evidence that adding a Decoy contract leads to a compromise effect in terms of outcomes, it is possible that there is an effect on the behavior of proposers or responders. In this section we analyze how the addition of a Decoy affects the responder behavior in our Ultimatum Games.

Table 4.3: Rejection rates in the Ultimatum Games

Game	**Feasible Contracts**			
	Generous (4,16)	**Fair (12,8)/(10,10)**	**Greedy (16,4)**	**Selfish (18,2)**
Scenario 1	-	2.7%	57.36%	-
Scenario 2	-	2.7%	48.2%	70%
Scenario 3	8.1%	3.6%	53.6%	-
Scenario 4	-	0%	52.7%	-
Scenario 5	-	0%	50.09%	72.7%
Scenario 6	11.8%	0%	54.5%	

Table 4.3 shows the rejection rates for each of the contracts in all six Ultimatum Games. Given that we employ the strategy method, we obtain a response for each proposal from every experimental subject. Across all games, we can see that the more favorable the proposal is towards the responder the more likely it is to get accepted with the exception of proposals that give more to the responder than to the proposer. In all scenarios, the Fair contract is nearly always accepted with a rejection rate ranging from 0% to 3.6%. While there are some responders rejecting even the Fair contract in Scenario 1 to Scenario 3, no responder rejects the equal split in Scenario 4 to Scenario 6. A potential reason why this could be the case is that there are some responders in our sample with very strong other-regarding concerns that may have a reason to reject a nearly equal-split proposal but do not have a reason to reject an equal-split proposal.

Comparing the behavior of responders between all scenarios, we find that only 40 out of 110 responders (36.4%) behave consistently. This means they either always accept or reject the Greedy and the Fair Contract. This is a violation of revealed preferences as in all cases they are choosing either between (16, 4) and (0, 0), (12, 8) and (0, 0) or (10,10) and (0, 0) . In the following, we will investigate whether we can explain this violation with the compromise effect. Our hypotheses would predict that because adding (18,2) to the set of feasible contracts makes the Greedy contract a compromise, we should observe more responders rejecting B when C is not available compared to when it is. Comparing Scenario 2 to its Base Scenario 1 the likelihood of rejecting (16,4) decreased from 57.36% to 48.2% when the Selfish Decoy is available. This effect is large in size and statistically significant (one-sided Binomial test, p-value < 0.001). Similarly, we would expect that the rejection rate of the Greedy contract in Scenario 5 is higher compared to Scenario 4. We can observe a slight decrease of 2.6% in the rejection rate of the Greedy contract, however, this effect is not statistically significant (one-sided Binomial test p-value=0.395).

Result 4.3: The Greedy contract is rejected less frequently when the Selfish Decoy (18,2) is added to the list of feasible proposals but only when the Fair contract is unequal.

We expect the equal split to be always acceptable and if there is a change for (12,8) when the Generous Decoy is introduced it would be towards increasing rejection rates, so any contribution

of responder behavior to the compromise effect must consist of rejecting (16,4) more often.
For the unequal split base scenario, we can observe the rejection rate of the Fair contract is the same for Scenario 1 and Scenario 2. Once, the Generous Decoy is added the rejection rate of the Greedy Contract is slightly higher. However, this effect is small and statistically insignificant (one-sided Binomial test, p-value= 0.254). We do also not find significantly higher rejection rates of the Greedy contract when we add the Generous Decoy compared to the base scenario (one-sided Binomial test, p-value=0.875).
For the equal split base Scenario, in neither Scenario, the Fair Contract is rejected. Similarly to the unequal split base Scenario, we also do not find significantly higher rejection rates of the Greedy contract when we add the Generous Decoy compared to the base Scenario.

Result 4.4: The Greedy contract is not rejected more frequently when a Generous Decoy is added to the set of feasible proposals.

4.4.3 Proposer behavior in the Ultimatum Games

We now turn to the question of whether adding a Decoy influences proposer behavior, specifically to whether proposers become more likely to propose a contract after it becomes a compromise. Table 4.4 shows the proposal rates for each of the contracts in all six Ultimatum Games. Across all scenarios, we can see that the majority of the proposers offered the fairest available contract ranging from 69.09% in Scenario 2 to 82.73% in Scenario 6.
In Scenario 2 and Scenario 5, the proposers could offer the very unfair contract (18,2) compared to the base scenario and 15.45% and respectively 10.91% chose to do so.
Hypothesis 4.5 predicts that we should observe more proposers offering the Greedy contract when the Selfish Decoy is available compared to when it is not because the Selfish Decoy makes the Greedy contract a compromise. Clearly, this is not what we can observe in our data. Comparing Scenario 2 to Scenario 1 the likelihood of proposing the Greedy contract decreased from 20.91% to 15.45%. Similarly in in Scenario 4 20.00% of the proposers offered the Greedy Contract but only 10.91% did so in Scenario 5 when a more unfair contract was available. Consequently, we do not find any support Hypothesis 4.5.

Table 4.4: Proposal Rates in the Ultimatum Games

Game	Feasible Contracts			
	Generous (4,16)	**Fair (12,8)/(10,10)**	**Greedy (16,4)**	**Selfish (18,2)**
Scenario 1	-	79.09%	20.91%	-
Scenario 2	-	69.09%	15.45%	15.45%
Scenario 3	5.50%	79.09%	15.45%	-
Scenario 4	-	80.00%	20.00%	-
Scenario 5	-	78.18%	10.91%	10.91%
Scenario 6	1.82%	82.73%	15.45%	-

Result 4.5: The Greedy contract is not offered more frequently when the Decoy (18,2) is added to the list of feasible contracts.

Similarly, Hypothesis 4.6 would predict that when the Fair contract is made a compromise we should observe more proposers offering the Fair contract when the Generous Decoy is available compared to when it is not. Again, this is not what we observe in our data. Comparing Scenario 3 to its base scenario we see that the rate of proposing the Fair contract remains unchanged (one-sided Binomial test, p-value=0.598). When we compare Scenario 6 to its Base Scenario 4, we observe a slight increase in the proposal rate of the Fair contract but this is not statistically significant (one-sided Binomial test, p-value= 0.29). Together, this means we co not find any support for Hypothesis 4.6.

Result 4.6: The Fair contract is not offered more frequently when the Decoy (18,2) is added to the list of feasible contracts.

While we do not observe a compromise effect, it might be that proposers who would like to propose the very unfair contract move from (16,4) to (18,2) when the unfair contract is available but subjects who proposed the fair options (10,10) or (12,8) respectively in the base game are now moving to the intermediate option. If proposer preferences over contracts and proposer beliefs over responder behavior are unchanged, we would not expect proposers to switch away from the Fair to the Greedy contract after the Decoy becomes available. This is because a proposer who prefers the Fair to the Greedy contract will also prefer the Fair contract to the Selfish Decoy under standard models of social preferences. Comparing Scenario 2 and Scenario 1, the difference in the fairest proposal offered is 10% and is statistically significant (two-sided Binomial test, p-value < 0.01). However, we do not find the same effect comparing Scenario 5 and Scenario 4 (two-sided Binomial test, p-value=0.77).

Another approach is to look at the pattern of individual changes of proposers between the Fair and the Greedy contract when the Decoy becomes available. Table 4.5 contains the contingency tables for the contracts offered in Ultimatum Games in Scenario 1 and Scenario 2 (LHS) as well as those for Scenario 4 and 5 (RHS). Hypothesis 4.7 would predict that proposers are more likely to switch from the Fair contract to the Greedy contract compared to the reverse after the Selfish Decoy is added.

Table 4.5: Contingency tables for Proposer behavior in the Ultimatum Games when the Selfish Decoy is added to the set of feasible contracts.

	Scenario 2					Scenario 5			
Scenario 1	Fair	Greedy	Decoy	Total	Scenario 4	Fair	Greedy	Decoy	Total
Fair	75	9	3	87	Fair	81	6	1	88
Greedy	1	8	14	23	Greedy	5	6	11	22
Total	76	17	17	110	Total	86	12	12	110

Out of the 87 proposers that offered the Fair contract in Scenario 1, twelve switched to proposing another contract when the Selfish Decoy was available (nine to the Greedy contract, three to the Selfish Decoy). Out of the 76 proposers that offered the Fair contract in Scenario 2, only one proposer switched to the Greedy Contract when the Selfish Decoy was not available in Scenario 1. The difference in switches in terms of directions is large and statistically significant (one-sided Binomial test, p-value=0.011)

Out of the 88 proposers offering the Fair contract in Scenario 4, seven switched to offering another contract in Scenario 5 (six proposers moved to the Greedy contract, one proposer moved to the Selfish Decoy) and five proposers who offered the Fair contract in Scenario 5 switched away from it in Scenario 4. This difference is not statistically significant (one-sided Binomial test, p-value= 0.500). It might be that an equal split causes different emotional reactions compared to a near-equal split (Güth et al., 2001; Galeotti et al., 2021). Together, this means we find support for Hypothesis 4.7 but only when the Fair contract is unequal.

Result 4.7: Proposers are more likely to switch from the Fair contract to the Greedy contract compared to the reverse after the Selfish Decoy but only when the Fair contract is unequal.

Table 4.6: Contingency tables for Proposer behavior in the Ultimatum Games when the Generous Decoy is added to the set of feasible contracts.

	Scenario 3					Scenario 6			
Scenario 1	Fair	Greedy	Decoy	Total	Scenario 4	Fair	Greedy	Decoy	Total
Fair	79	3	5	87	Fair	83	4	1	87
Greedy	8	14	1	23	Greedy	8	13	1	23
Total	87	17	6	110	Total	91	17	2	110

Table 4.6 contains the contingency tables for the contracts offered in Ultimatum Games in Scenario 1 and Scenario 3 (LHS) as well as those for Scenario 4 and 6 (RHS). Hypothesis 4.7 would predict that proposers are more likely to switch from the Greddy contract to the Fair contract compared to the reverse after the Generous Decoy is added.

Out of the 22 proposers who offered the Greedy contract in Scenario 1, six offered the Fair contract in Scenario 3. On the other hand, out of the 17 proposers who offered the Greedy contract in Scenario 3, only 3 switched to proposing the Fair contract in Scenario 1. However, this difference is not statistically significant (one-sided Binomial test, p-value = 0.113).

Out of the 21 proposers who offered the Greedy contract in the Scenario 4, eight proposers offered the Fair contract in Scenario 6. On the other hand, out of the 17 proposers who offered the Greedy contract in Scenario 6, only four proposers offered the Fair contract in Scenario 4. This difference is also not statistically significant (one-sided Binomial test, p-value=0.194). Consequently, we do not find support for Hypothesis 4.8, independent of whether the Fair contract is equal or unequal.

Result 4.8: Proposers are not more likely to switch from the Greedy contract to the Fair contract compared to the reverse after the Generous Decoy is added.

It might be that the behavior of the proposers is not only affected by distributional concerns but also due to a change in the belief about differing acceptabilities of the contracts once the Decoy contract is available. Then we would expect to see this in a significant change in beliefs.

Table 4.7: Beliefs about acceptance rate in the Ultimatum Games

Game	**Feasible Contracts**			
	Generous (4,16)	**Fair (12,8)/(10,10)**	**Greedy (16,4)**	**Selfish (18,2)**
Scenario 1	-	81.86%	31.77%	-
Scenario 2	-	81.62 %	33.37%	17.35%
Scenario 3	57.85%	79.28%	24.045%	-
Scenario 4	-	91.96 %	30.08%	-
Scenario 5	-	90.35 %	31.67%	18.13%
Scenario 6	70.49%	76.54 %	23.05%	-

Table 4.7 reports the beliefs we elicited from the proposer about the acceptance rates of the responders. If the proposer would anticipate that (16,4) is more likely to be accepted when a more unfair contract is available, we should see significantly different beliefs about the acceptance rate between Scenario 2 and Scenario 1 (and potentially Scenario 5 and Scenario 4). Comparing Scenario 1 to Scenario 4, as anticipated, the fairer the contract is, the more likely proposers think that the responder will accept the contract. The beliefs about the acceptance rate of the Greedy Contract are almost identical for Scenario 1 and Scenario 2 as well as Scenario 4 and Scenario 5. We do not see a statistically significant effect between the beliefs about acceptance rate between Scenario 1 and Scenario 2 (two-sided sign rank test, p-value=0.36) as well as Scenario 4 and Scenario 5 (two-sided sign rank test, p-value=0.30). However, we do see a decrease in the belief about the acceptability of the greedy contract when the Generous Decoy is added both in the case of the equal split and near equal split scenarios (sign rank tests, both p-values < 0.01).

4.4.4 Dictator Games

Next, we turn to the behavior in the Dictator Games. In the Ultimatum Game, the behavior of the proposers might be influenced by beliefs about differing rejection rates of the responder

Table 4.8: Implemented contracts in the Dictator Games

Game	**Feasible Contracts**			
	Generous (4,16)	**Fair (12,8)/(10,10)**	**Greedy (16,4)**	**Selfish (18,2)**
Scenario 1	-	53.64%	46.34%	-
Scenario 2	-	45.45%	16.36%	38.18 %
Scenario 3	0.91%	51.82%	47.27%	-
Scenario 4	-	52.73%	47.27%	-
Scenario 5	-	46.36%	11.82%	41.82 %
Scenario 6	0%	56.36%	43.64%	-

depending on the available feasible contracts. To test whether proposals are changing due to a change in proposer preferences over contracts, we additionally conduct Dictator Games in which the scope for strategic thinking about the responder's behavior is removed. Table 4.8 reports the implemented contracts in all six Dictator Games. Across all scenarios, we can see that many of the proposers implemented the Fair contract ranging from 45.45% in Scenario 2 to 56.36% in Scenario 6. In Scenario 2 and Scenario 5 the proposers could implement the very unfair contract (18,2) and large shares of 38.18% and 41.82% respectively chose to do so. Hypothesis 4.9 would predict that because the Selfish Decoy (18,2) makes the Greedy contract a compromise, we should observe more proposers implement the Greedy contract when the Selfish Decoy is available compared to when it is not. Perhaps, not very surprisingly, this is not what we can observe given that so many proposers chose to implement the most unequal contract available. Comparing Scenario 2 to Scenario 1, the likelihood of proposing the Greedy Contract decreased from 46.36% to 16.36%. Comparing Scenario 5 to Scenario 4, we see a similar picture. The likelihood to propose the Greedy contract decreased from 47.27% to 11.82%.

Result 4.9: The Greedy contract is not offered more frequently in the Dictator Games when the Decoy (18,2) is added to the list of feasible contracts.

Similarly, Hypothesis 4.10 would predict that when the Fair contract is made a compromise we should observe that more proposers implement the Fair contract when the Generous Decoy is available compared to when it is not. Again, this is not what we can observe in our data. Comparing Scenario 3 to its base we see that the rate of proposing the Fair contract decreases very slightly and the effect is not statistically significant (one-sided Binomial test, p-value=0.387). When we compare Scenario 6 to its Base Scenario 4, we observe a slight increase in the proposal rate of the Fair contract but this difference is also not significantly different (one-sided Binomial test, p-value=0.240).

Result 4.10: The Fair contract is not offered more frequently in the Dictator Games when the Generous Decoy (4,16) is added to the list of feasible contracts.

Also in the Dictator Games, our design allows us to follow how proposers change their behavior between scenarios. If proposer preferences over contracts are independent of what other contracts are available, we would expect proposers who propose the Fair contract in Scenario 1 (or 4 respectively) to also propose the Fair contract in Scenario 2 (or 5 respectively). This is because the Fair contract has been revealed preferred to the Greedy contract, and under standard models of social preferences any proposer who prefers the Fair contract to the Greedy contract would also prefer the Fair contract to the Selfish Decoy. There may be some movements between contracts due to noise, but if significantly more proposers move from the Fair to the Greedy contract rather than the reverse when the Decoy is added, this would suggest a context effect.

Table 4.9: Contingency Tables for Proposer behavior in the Dictator Games when the Selfish Decoy is added to the set of feasible contracts.

	Scenario 2					Scenario 5			
Scenario 1	Fair	Greedy	Decoy	Total	Scenario 4	Fair	Greedy	Decoy	Total
Fair	46	9	4	59	Fair	48	7	3	58
Greedy	4	9	38	51	Greedy	3	6	43	52
Total	50	18	42	110	Total	51	13	46	110

Table 4.9 contains the contingency tables for the contracts implemented in the Dictator Games in Scenario 1 and Scenario 2 (LHS) as well as those for Scenario 4 and 5 (RHS). Hypothesis 4.11 would predict that proposers are more likely to switch from the Fair contract to the Greedy contract compared to the reverse after the Selfish Decoy is added.

Nine subjects offered the Fair contract in Scenario 1 and the Greedy contract in Scenario 2 when the Selfish Decoy was made available, compared with four subjects who offered the Greedy contract in Scenario 1 but the Fair contract in Scenario 2. However, this difference is not statistically significant (one-sided Binomial test, p-value=0.133). Comparing Scenario 5 and Scenario 4, we see a similar picture. Once the Selfish Decoy is added, 7 subjects switch from the Fair to the Greedy contract and only three switch from the Greedy to the Fair contract but the effect is not statistically significant (one-sided Binomial test, p-value=0.172). In summary, this means that we cannot support Hypothesis 4.11.

Result 4.11: Proposers are not more likely to switch from the Greedy contract to the Fair contract compared to the reverse after the Selfish Decoy is added in the Dictator Games.

Table 4.10: Contingency tables for Proposer behavior in the Dictator Games when the Generous Decoy is added to the set of feasible contracts.

	Scenario 3					Scenario 6			
Scenario 1	Fair	Greedy	Decoy	Total	Scenario 4	Fair	Greedy	Decoy	Total
Fair	52	7	0	59	Fair	51	7	0	58
Greedy	5	45	1	51	Greedy	11	41	0	52
Total	57	52	1	110	Total	62	0	48	110

Table 4.10 reports the contingency tables for the contracts implemented in the Dictator Games in Scenario 1 and Scenario 3 (LHS) as well as those for Scenario 4 and 6 (RHS). Hypothesis 4.11 would predict that proposers are more likely to switch from the Fair contract to the Greedy contract compared to the reverse after the Selfish Decoy is added.

We now look at the effect of introducing a Generous Decoy in the unequal split base scenario. Hypothesis 4.12 would predict that proposers are more likely to switch from the Greedy contract to the Fair contract compared to the reverse when the Generous Decoy becomes available.

When the Fair contract was unequal, five proposers switched from the Greedy to the Fair contract (that is, in the direction of the compromise effect) while seven proposers switched in the opposite direction. Consequently, even the direction of the effect is not as anticipated between Scenario 1 and Scenario 3.

For the equal split base scenario, the introduction of the Generous Decoy led to eleven proposers switching from the Greedy to the Fair contract and seven proposers from the Fair to the Greedy contract. However, the difference is not statistically significant (one-sided Binomial test, p-value=0.240).

Result 4.12: Proposers are not more likely to switch from the Greedy contract to the Fair contract compared to the reverse after the Generous Decoy is added in the Dictator Games. In summary, this means that we cannot support Hypothesis 4.12.

4.4.5 Random Proposal Games

In the Ultimatum Game, the behavior of the proposers might be influenced by reciprocity considerations. To test whether rejection rates are changing due to a change of responder preferences over contracts or reciprocity concerns over proposer behavior, we additionally conducted Random Proposal Games in which the scope for reciprocity considerations was removed. Next, we look at the behavior of responders in the Random Proposal Games.

Table 4.11: Rejection rates in Random Proposal Games

Game	**Feasible Contracts**			
	Generous (4,16)	**Fair (12,8)/(10,10)**	**Greedy (16,4)**	**Selfish (18,2)**
Scenario 1	-	2.7%	40.1%	-
Scenario 2	-	3.6%	35.5%	56.4%
Scenario 3	7.3%	6.3%	39.3%	-
Scenario 4	-	0%	39.0%	-
Scenario 5	-	0%	37.3%	56.4%
Scenario 6	7.3%	0%	40%	-

Table 4.11 shows the rejection rates for each of the proposals in all six Random Proposal Games. Across all scenarios, we can see that the more favorable the contract is towards the responder the more likely it is to get accepted with the exception of proposals that give more to the responder than to the proposer. In all scenarios, the Fair contract is nearly always accepted

with a rejection rate ranging from 0% to 6.3%. While there are some responders rejecting even the Fair contract in Scenario 1 to Scenario 3, no responder rejects the equal split in Scenario 4 to Scenario 6. This is very similar to the behavior we could observe in the Ultimatum Game. Hypothesis 4.13 would predict that adding the Selfish Decoy to the set of feasible contracts is associated with lower rejection rates of the Greedy contract. Comparing Scenario 1 to Scenario 2, we can see the likelihood of rejecting (16,4) decreased from 40.1% to 35.5% when the Selfish Decoy is available and this effect is statistically significant using a one-sided Binomial test (p-value=0.035). Similarly, we would expect that the rejection rate of the Greedy contract in Scenario 5 is higher compared to Scenario 4. We can observe a slight decrease in the rejection of 1.7%, however, this effect is not statistically significant (one-sided Binomial Test, p-value=0.387).

Result 4.13: In the Random Proposal Games, the Greedy contract is rejected less frequently when the Selfish Decoy (18,2) is added to the list of feasible proposals but only when the Fair contract is unequal.

Hypothesis 4.14 would predict that adding the Generous Decoy to the set of feasible contracts is associated with higher rejection rates of the Greedy contract in Scenario 2 compared to Scenario 1 as w. Similarly, even though adding the Generous Decoy makes the Greedy contract more extreme compared to the base Scenario, its rejection rate is unaffected (Scenario 3 versus Scenario 1 (p-value= 0.051), the Fair contract is never rejected in Scenario 4 or Scenario 6).

Result 4.14: In the Random Proposal Games, the Greedy contract is not rejected more frequently when a Generous Decoy is added to the set of feasible proposals.

In summary, we find mixed support for H13 in our Random Proposal Games. Greedy contracts are more acceptable when they propose a compromise, but only when the Fair contract is slightly unequal. When the Fair contract offers an equal split, the Greedy contract is rejected at the same rate whether or not the selfish Decoy is present. We do not find any support for Hypothesis 4.14.

4.5 Conclusion

Galeotti et al. (2021) find significant compromise effects across a wide variety of payoff conditions in an unstructured bargaining situation. Due to the complexity of bargaining in unstructured situations, it is difficult to determine the mechanism behind the compromise effect. The goal of this paper is to unpack the compromise effect in bargaining situations.

To do this, we explore the compromise effect in a structured bargaining situation. Specifically, we use Ultimatum Games as they have a particular simple structure that allows us to observe whether changing the menu of contracts makes a given contract more or less acceptable (relative to the acceptance rate in the base Scenario), or whether it makes a given contract more likely to be proposed.

In all we examine six scenarios using an ultimatum bargaining protocol. In addition, we examine each Scenario using two more, even simpler, protocols. Using Dictator Games, where the receiver cannot reject a proposal, we are able to observe whether proposer preferences are affected by adding a menu option. Finally, we use Random Proposal Games, in which the receivers indicate the acceptability of proposals when the proposal is chosen by nature. This allows us to observe whether the acceptability of contracts is subject to 'pure' menu effects, independent of the intentions of proposers.

Unlike Galeotti et al. (2021), we do not observe significant compromise effects in our data. This may be due to the difference in bargaining protocol or to the difference in the underlying situations. Our games do not have an equality/efficiency trade-off. It may be the case that context effects are stronger in situations with several relevant criteria (e.g. the degree of equality of payoffs and the total payoffs). We do observe some differences in individual behavior in the direction of the compromise effect.

We find that the addition of a third contract in the Ultimatum Games does matter for both responders and proposers but only when the Fair contract is unequal. Responders are significantly more likely to accept the Greedy contract once an even greedier Decoy is added to the set of feasible contracts. The findings on the individual level are consistent with Galeotti et al. (2021) finding that context effects are easier to observe in the absence of an equal contract.

The effects on the responders are also persistent in the Random Proposal Games, indicating that the observed differences in rejection rates in the UG can not only be attributed to reciprocal behavior. Interestingly, the effects on the proposer behavior do not translate to Dictator Games which may indicate that the effects in the Ultimatum Game are at least partly driven by strategic considerations.

4.6 References

Arechar, A. A., S. Gächter, and L. Molleman (2018). Conducting interactive experiments online. *Experimental Economics 21*, 99–131.

Bateman, I. J., A. Munro, and G. L. Poe (2008). Decoy effects in choice experiments and contingent valuation: Asymmetric dominance. *Land Economics 84*(1), 115–127.

Beauchamp, J. P., D. J. Benjamin, D. I. Laibson, and C. F. Chabris (2020). Measuring and controlling for the compromise effect when estimating risk preference parameters. *Experimental Economics 23*, 1069–1099.

Brandts, J. and G. Charness (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics 14*, 375–398.

De Haan, T. and R. Van Veldhuizen (2015). Willpower depletion and framing effects. *Journal of Economic Behavior & Organization 117*, 47–61.

Doyle, J. R., D. J. O'Connor, G. M. Reynolds, and P. A. Bottomley (1999). The robustness of the asymmetrically dominated effect: Buying frames, phantom alternatives, and in-store purchases. *Psychology & Marketing 16*(3), 225–243.

Falk, A., E. Fehr, and U. Fischbacher (2003). On the nature of fair behavior. *Economic Inquiry 41*(1), 20–26.

Galeotti, F., M. Montero, and A. Poulsen (2019). Efficiency versus equality in bargaining. *Journal of the European Economic Association 17*(6), 1941–1970.

Galeotti, F., M. Montero, and A. Poulsen (2021). The attraction and compromise effects in bargaining: Experimental evidence. *Management Science*.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Güth, W., S. Huck, and W. Müller (2001). The relevance of equal splits in ultimatum games. *Games and Economic Behavior 37*(1), 161–169.

Herne, K. (1997). Decoy alternatives in policy choices: Asymmetric domination and compromise effects. *European Journal of Political Economy 13*(3), 575–589.

Herne, K. (1999). The effects of decoy gambles on individual choice. *Experimental Economics 2*, 31–40.

Highhouse, S. (1996). Context-dependent selection: The effects of decoy and phantom job candidates. *Organizational Behavior and Human Decision Processes 65*(1), 68–76.

Kelman, M., Y. Rottenstreich, and A. Tversky (1996). Context-dependence in legal decision making. *The Journal of Legal Studies 25*(2), 287–318.

Kuncel, N. R. and J. A. Dahlke (2020). Decoy effects improve diversity hiring. *Personnel Assessment and Decisions 6*(2), 5.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of mathematical psychology 15*(3), 215–233.

Munro, A. and D. Popov (2013). A portmanteau experiment on the relevance of individual decision anomalies for households. *Experimental Economics 16*(3), 335–348.

Noguchi, T. and N. Stewart (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition 132*(1), 44–56.

O'Curry, Y. P. S. and R. Pitts (1995). The attraction effect and political choice in two elections. *Journal of Consumer Psychology 4*(1), 85–101.

Pinger, P., I. Ruhmer-Krell, and H. Schumacher (2016). The compromise effect in action: Lessons from a restaurant's menu. *Journal of Economic Behavior & Organization 128*, 14–34.

Sedikides, C., D. Ariely, and N. Olsen (1999). Contextual and procedural determinants of partner selection: Of asymmetric dominance and prominence. *Social Cognition 17*(2), 118–139.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of consumer research 16*(2), 158–174.

Slaughter, J. E., J. Bagger, and A. Li (2006). Context effects on group-based employee selection decisions. *Organizational Behavior and Human Decision Processes 100*(1), 47–59.

Trueblood, J. S., S. D. Brown, A. Heathcote, and J. R. Busemeyer (2013). Not just for consumers: Context effects are fundamental to decision making. *Psychological Science 24*(6), 901–908.

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition 17*(4), 767.

Appendices

4.A Order of Scenarios

Table 4.A1: Order of scenarios by Session

Ultimatum Game					
Session 1	Session 2	Session 3	Session 4	Session 5	Session 6
S1	S2	S3	S4	S5	S6
S6	S4	S5	S3	S2	S1
S2	S1	S6	S5	S4	S3
S5	S6	S1	S2	S3	S4
S3	S5	S4	S6	S1	S2
S4	S3	S2	S1	S6	S5
Dictator Game					
S6	S1	S4	S5	S3	S2
S2	S4	S1	S3	S6	S5
S5	S3	S2	S6	S1	S4
S3	S5	S6	S2	S4	S1
S4	S2	S3	S1	S5	S6
S1	S6	S5	S4	S2	S3
Random Proposal Games					
S4	S3	S2	S1	S6	S5
S1	S4	S5	S6	S3	S2
S6	S2	S3	S5	S4	S1
S3	S5	S6	S2	S1	S4
S5	S6	S1	S4	S2	S3
S2	S1	S4	S3	S5	S6
Ultimatum Game (RR)					
S3	S5	S6	S2	S4	S1
S6	S2	S3	S4	S1	S5
S1	S4	S2	S6	S5	S3
S2	S6	S1	S5	S3	S4
S5	S3	S4	S1	S6	S2
Dictator Game (RR)					
S2	S4	S5	S6	S1	S3
S4	S2	S1	S3	S6	S5
S1	S3	S6	S5	S2	S4
S5	S6	S3	S1	S4	S2
S3	S1	S2	S4	S5	S6
S6	S5	S4	S2	S3	S1
Random Proposal Games (RR)					
S5	S6	S1	S3	S2	S4
S1	S2	S5	S6	S4	S3
S2	S4	S3	S1	S5	S6
S6	S3	S4	S5	S1	S2
S3	S5	S2	S4	S6	S1
S4	S1	S6	S2	S3	S5

4.B Experimental Instructions

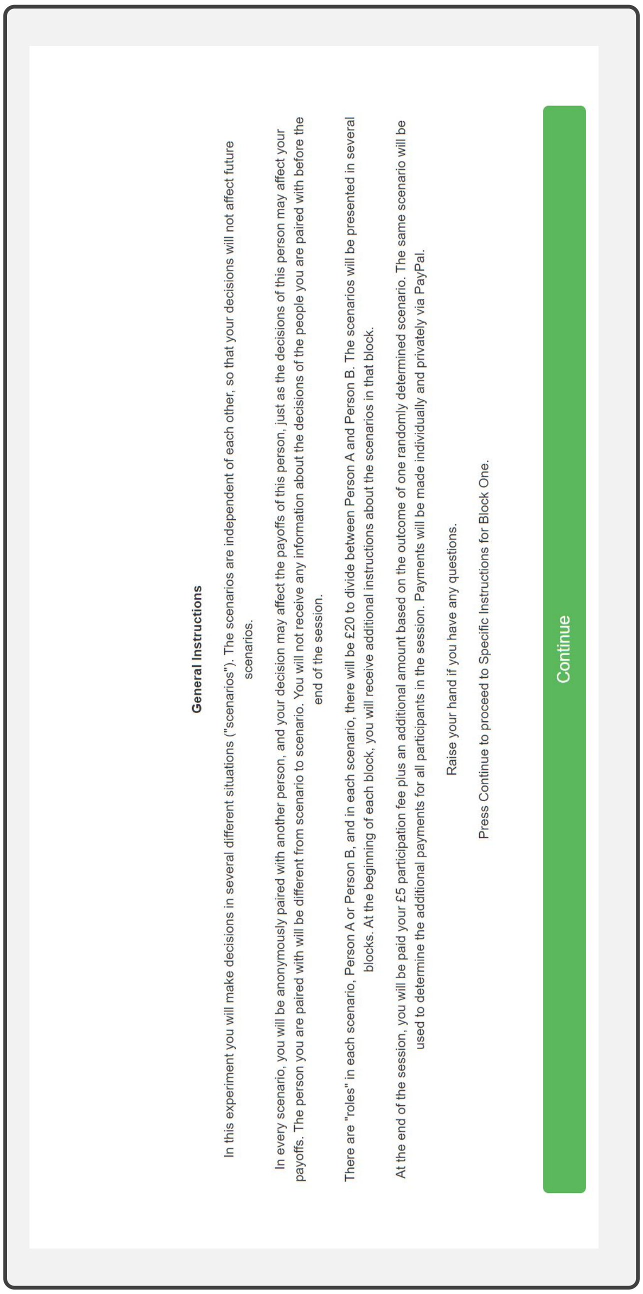


Figure 4.B1: General Instructions

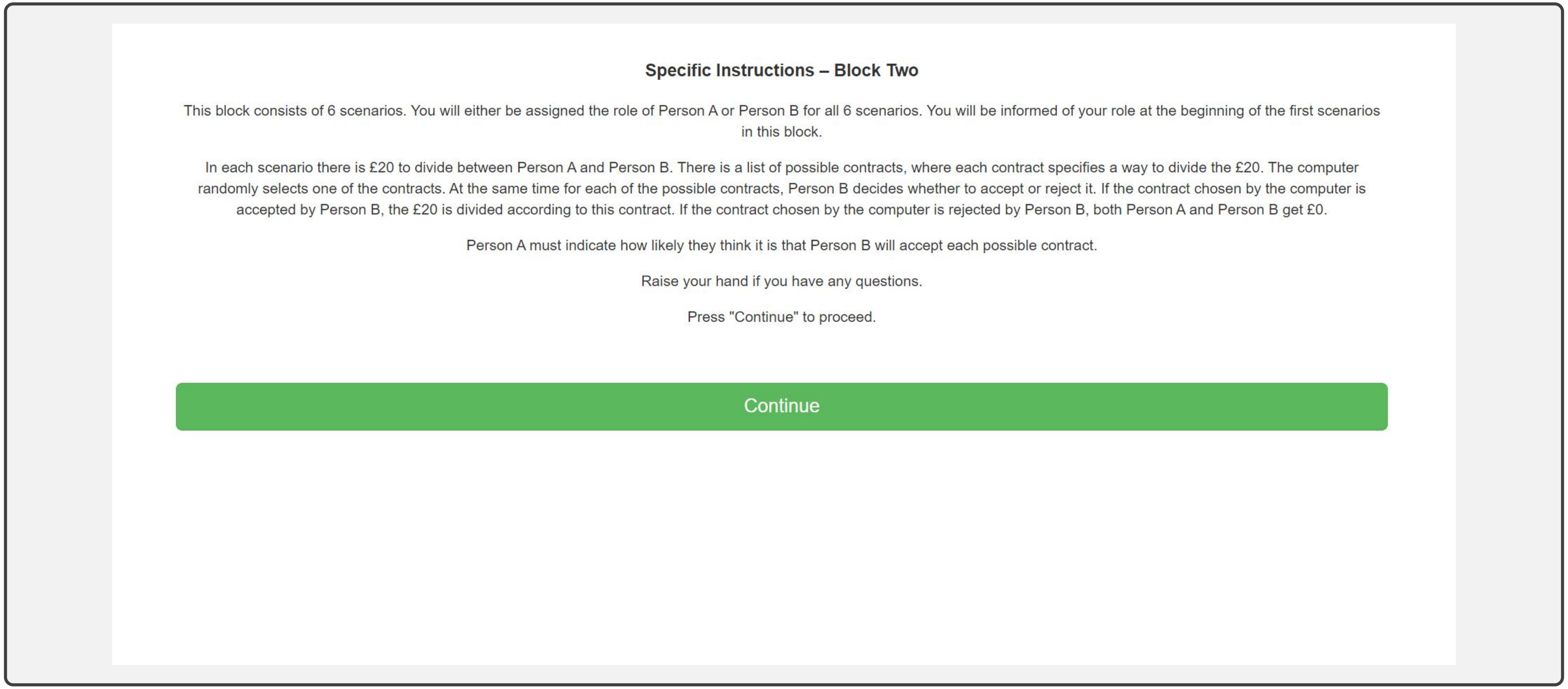


Figure 4.B2: Specific Instructions in the Ultimatum Game

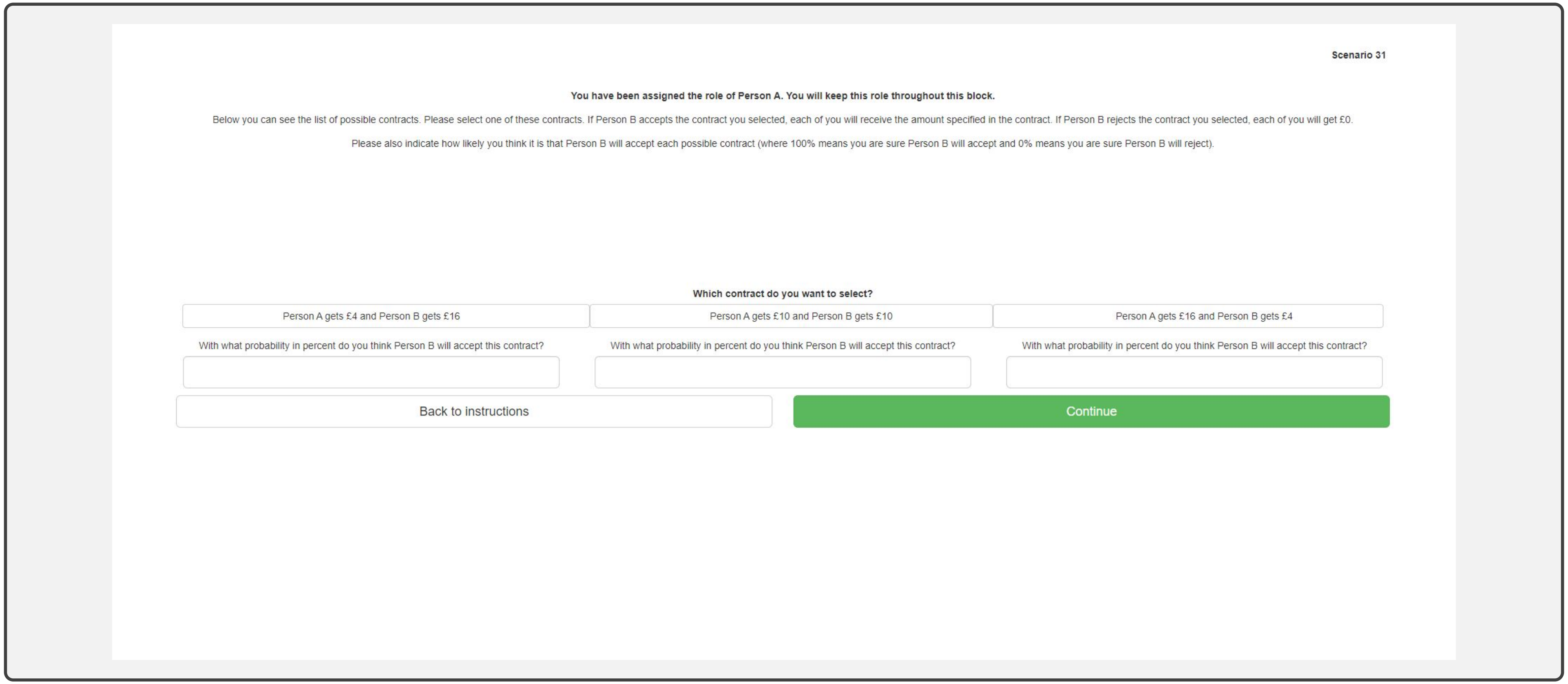


Figure 4.B3: Proposer Decision in the Ultimatum Game

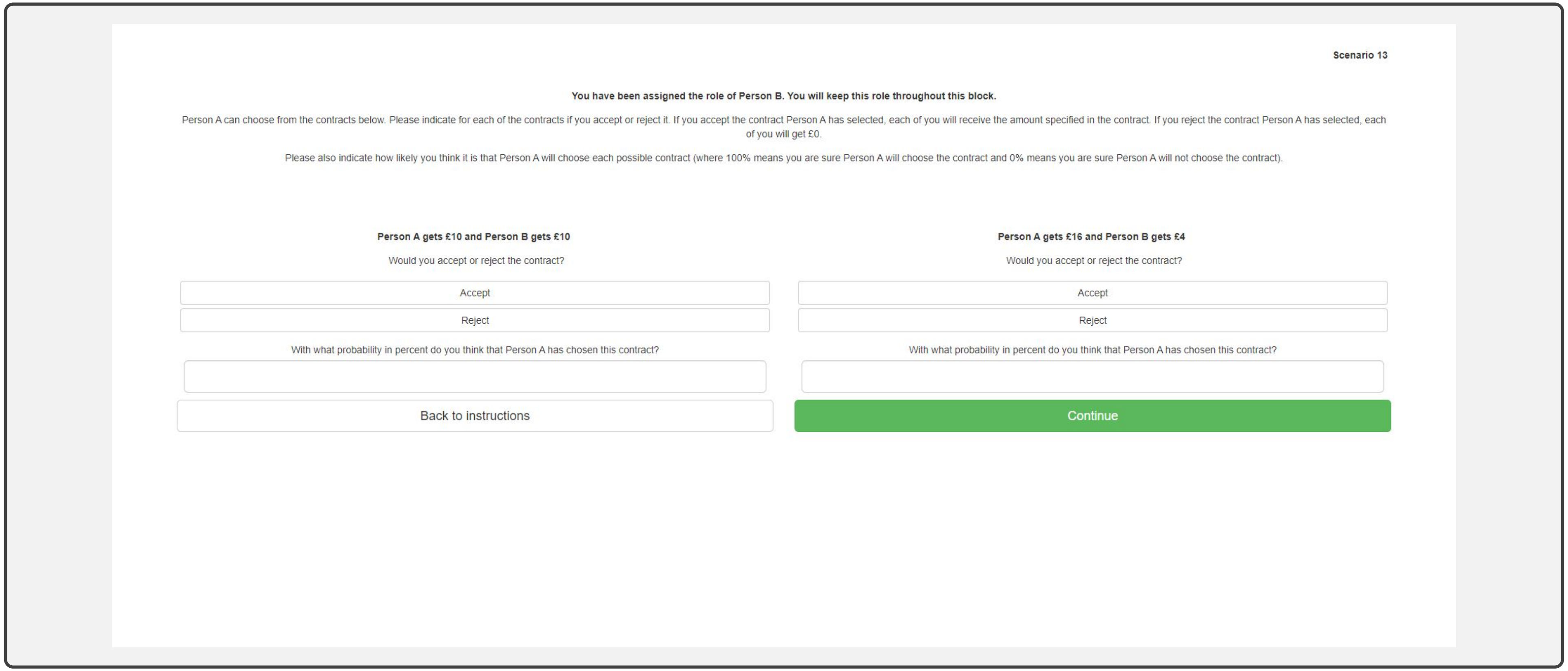


Figure 4.B4: Responder Decision in the Ultimatum Game

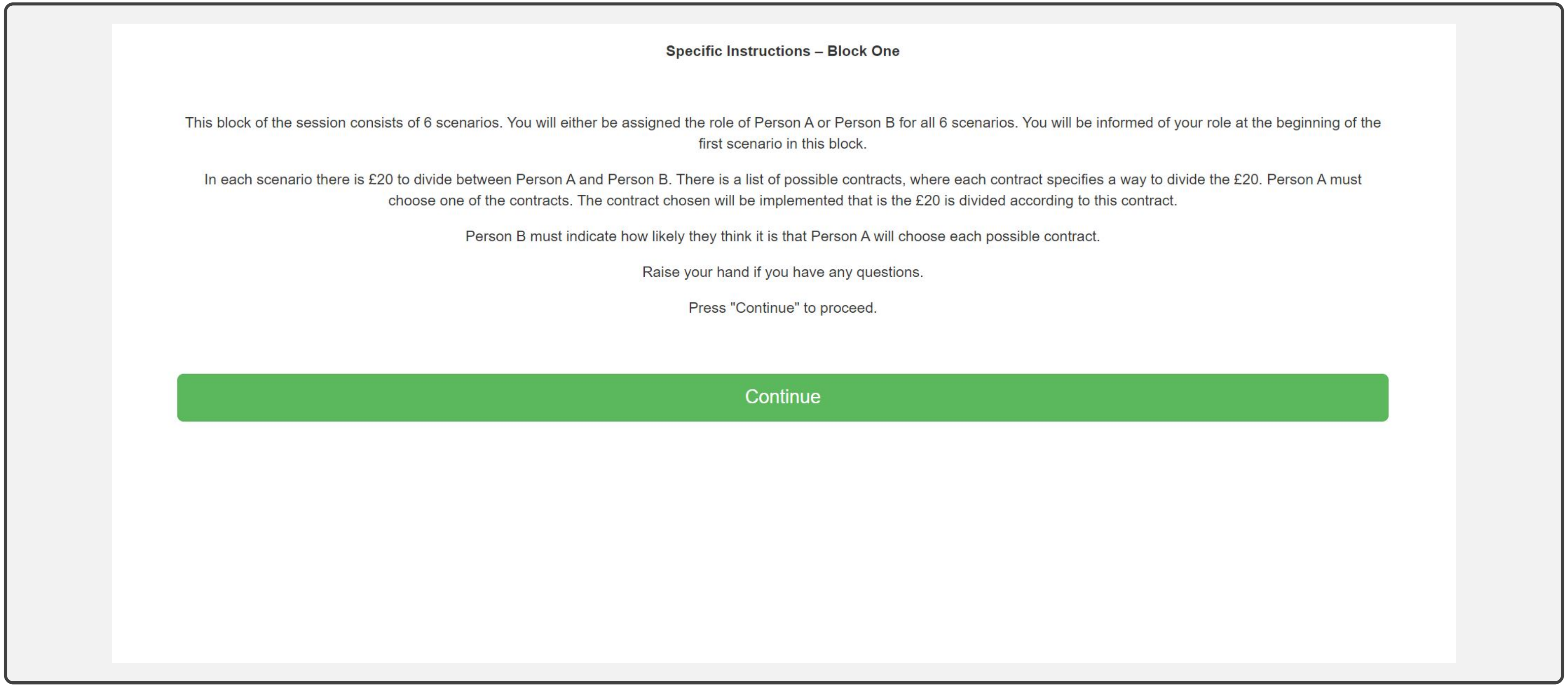


Figure 4.B5: Specific Instructions for the Dictator Games

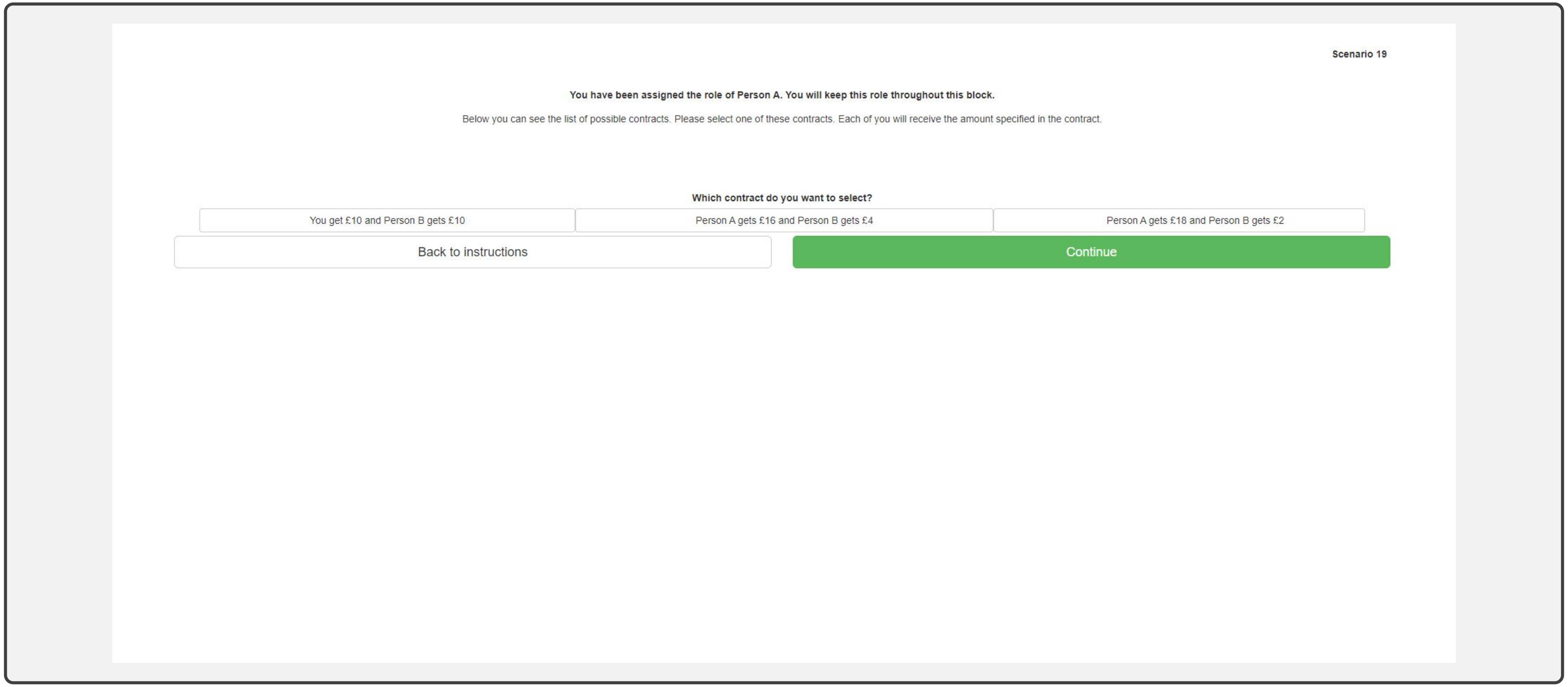


Figure 4.B6: Decision Screen for the Proposer in the Dictator Game

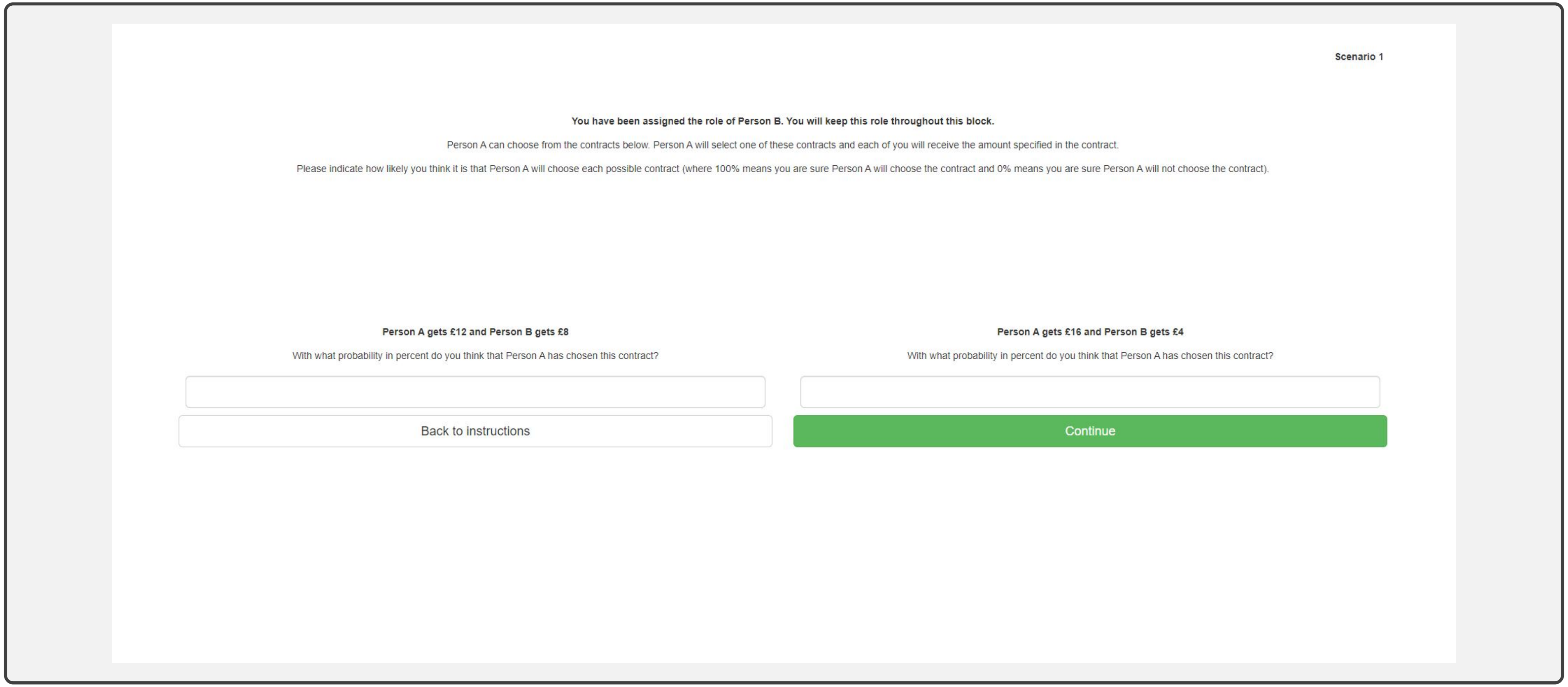


Figure 4.B7: Decision Screen for the Responder in the Dictator Game

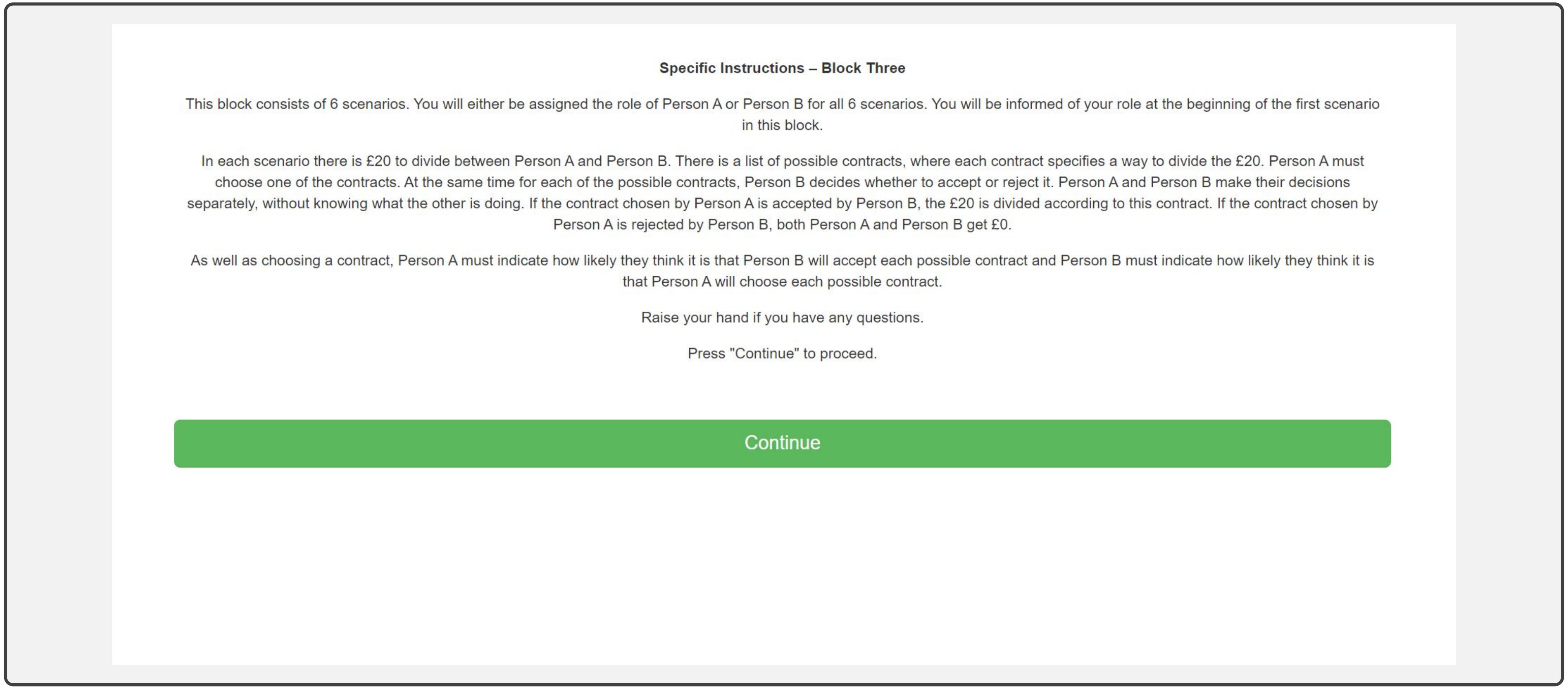


Figure 4.B8: Specific Instructions for the Random Proposal Game

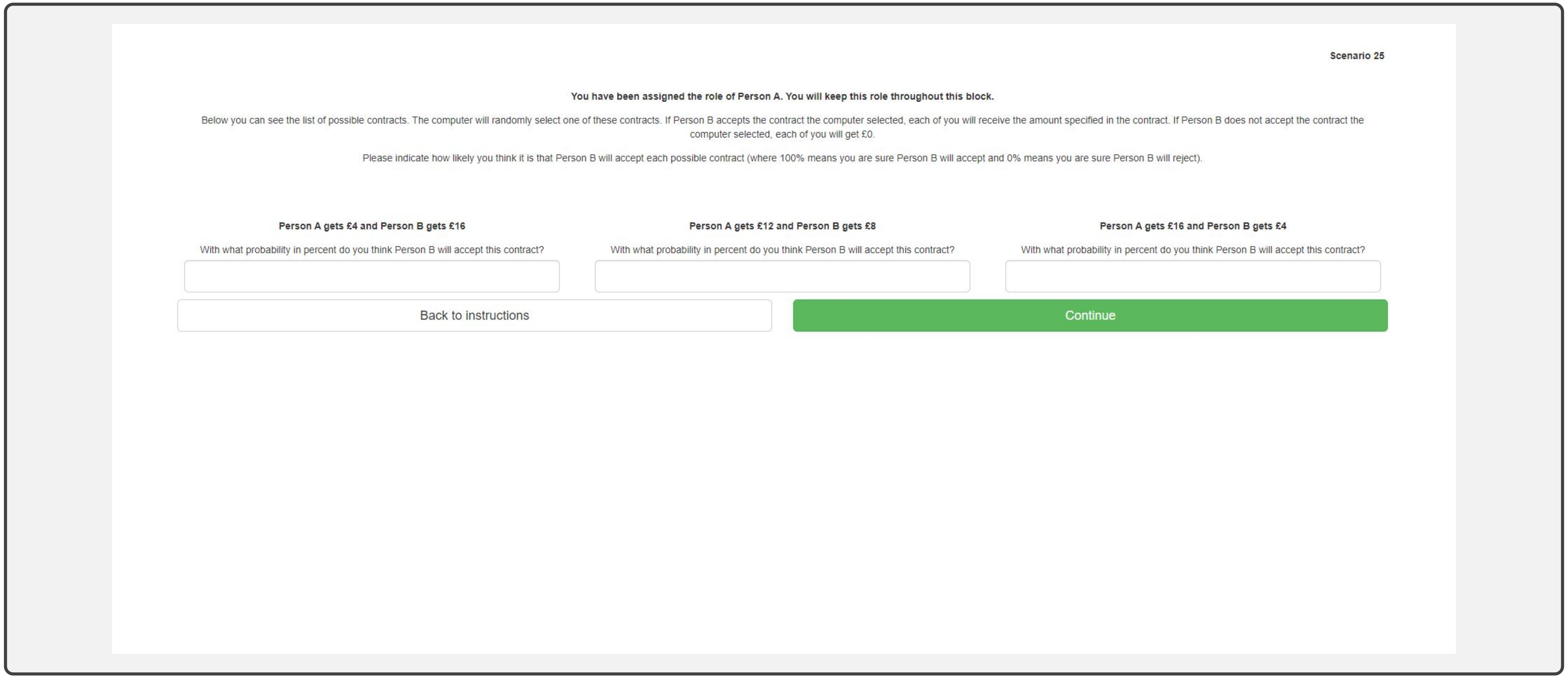


Figure 4.B9: Decision Screen for the Proposer in the Random Proposal Game

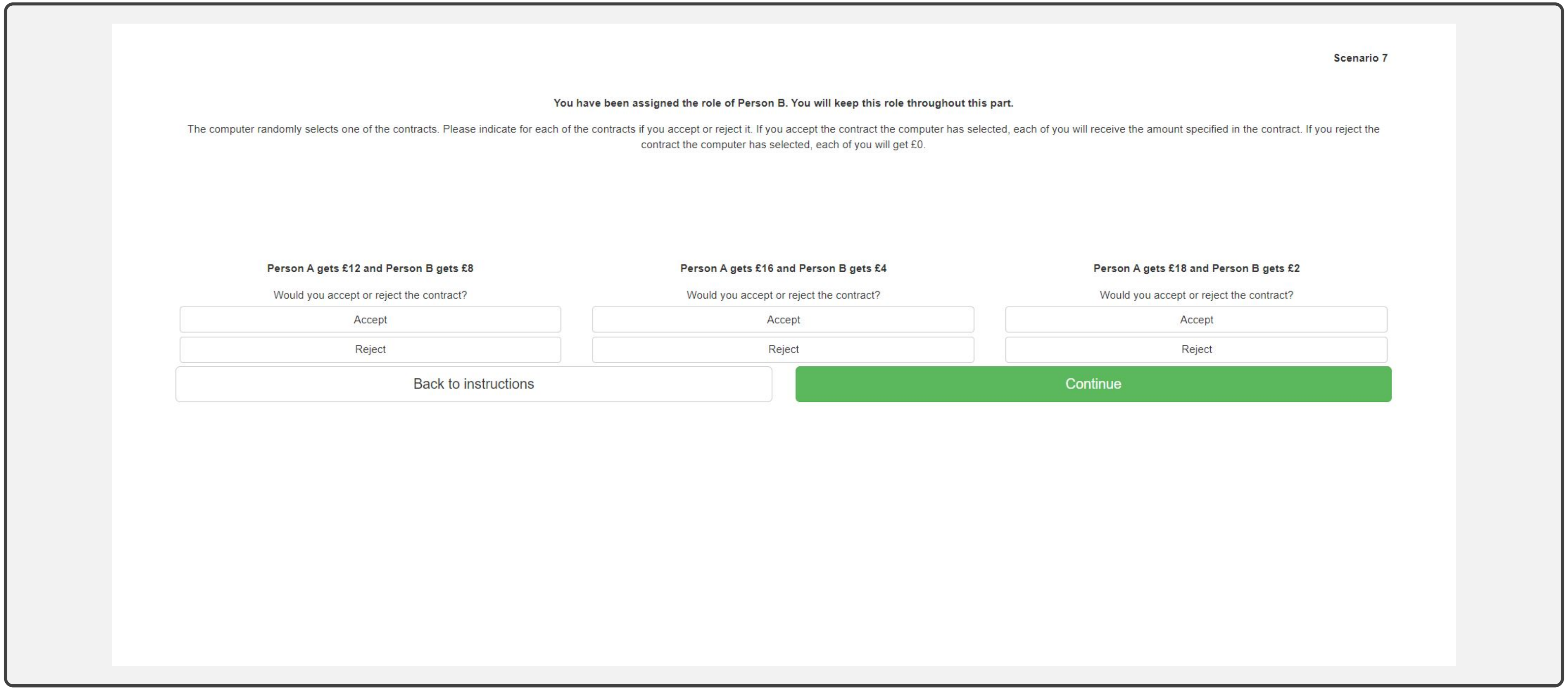


Figure 4.B10: Decision Screen for the Responder in the Random Proposal Game