STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
5350 Master's thesis in economics
Academic year 2018–2019

Who gets it?
An experimental study of non-monetary outside options in the ultimatum game

Joakim Niklas Karlsson
23002

Abstract

This thesis seeks to explore whether non-monetary outside options and a concern for the experimenter influence both proposer and responder behaviour in the ultimatum game. With pre-registered design and analysis plans, I run an experiment with students at the Stockholm School of Economics assigned to one control and two treatment groups, using a between subject design. In the first treatment, participants are informed that the rejection of the offer by the proposer leads to the destruction of the money. In the second treatment, players are informed that the rejected money will be given to a charity. My results show that responders are significantly affected by the treatments, reflecting that social welfare concerns for the experimenter do exist in an ultimatum game setting, contrary to popular belief. However, I cannot prove that non-monetary outside options have a significant effect on proposer behaviour. Overall, my results could have an impact on the way we interpret previous ultimatum game studies.

Keywords: Ultimatum Game, Outside Options, Rationality, Social Welfare, Fairness
JEL: C7, C91, D03, D80

Supervisor: Magnus Johannesson
Date submitted: 13 May 2019
Date examined: 28 May 2019
Discussant: Aylin Shawkat
Examiner: Mark Sanctuary

Acknowledgements

I would like to give my sincerest thanks to my supervisor, Magnus Johannesson, who from the beginning to the end provided me with invaluable feedback and advice. I also want to thank Lina Thomas, Pietro Galardi, and Markus Jury, who could not have provided me with a firmer foundation on which to stand in order to pull off this experiment. I also want to thank my family and friends for their continuous support and guidance.

Table of Contents

1. Introduction

Bargaining is a fundamental facet of daily life and attempting to understand why people act the way they do is central within Behavioural Economics. The ultimatum game (UG), introduced by Güth, Schmittberger, and Schwarze (1982), seeks to provide insights into people's economic decision-making. In the game, two players bargain anonymously to split a given amount of money. The first player proposes how to divide the 'pie' between the other player and themselves. The second player responds to this offer by either rejecting or accepting the offer. If the responder chooses to accept, the money is split accordingly between the two. However, if the responder rejects, neither player receives any money.

Classical rational-actor economic theory prescribes that the responder should accept any positive offer. The rational proposer should foresee this and thus, in order to maximize their utility, offer the smallest positive amount possible (Selten, 1975; Güth & Kocher, 2014). However, this prediction is not in line with empirical evidence from experiments, where the literature shows that responders are likely (probability 0.4-0.6) to reject offers sharing less than 20% of the amount, and proposers typically end up offering divisions in the 40-50% range (Güth & Kocher, 2014).

Numerous studies have sought to explain this divergence between classical theory and practice, manipulating all sorts of variables in the UG. Variations include running one-shot games (Güth, Schmittberger, & Schwarze, 1982) or repeated games (Ochs & Roth, 1989; Slonim & Roth, 1998), as well as making changes with regards to social distance (Bolton & Zwick, 1995; Eckel & Grossman, 1996), entitlement (Hoffman, Van Dijk, & De Cremer, 1996), information asymmetry (Straub & Murnighan, 1995; Camerer & Thaler, 1995), degrees of anonymity (Hoffman, McCabe, Shachan, & Smith, 1994), framing (Forsythe, Horowitz, Savin, & Sefton, 1994), nationality & culture (Roth, Prasnikar, Okuno-Fujiwara, & Zamir, 1991), stakes (Andersen, Ertac, Gneezy, Hoffman, & List, 2011) and so forth.

In the ultimatum game literature, an alternative payoff in case of rejection is called an outside option. These outside options have also been analysed to some degree and have been shown to play a role in influencing behaviour (Abbink, Bolton, Sadrieh, & Tang, 2001; Schmitt 2004; Knez & Camerer, 1995; Pillutla & Murnighan, 1996; Blount & Bazerman, 1996). However, little focus has been paid to the effect of outside options that do not provide an alternative positive or negative payment for the players. I call these outside options non-monetary outside options.

The results of all these diverse studies suggest that the ultimatum game is highly complex and influenced by numerous factors; a full understanding of the ultimatum game has yet to be reached.

My thesis idea was born out of the simple idea that a reason why responders reject the proposed amount is because by rejecting, they are actually saving the whole stake for the experimenter. While there have been numerous studies on the role of outside options, there has only been one study, by Frank (1998) that has examined the effect of non-monetary outside options in the same manner as in my experiment. Frank designed an experiment studying the role of non-monetary outside options in the ultimatum game. He examines whether participants care about the experimenter's welfare. If this concern for the experimenter is something that participants truly care about then this diminishes the role of fairness concerns in the ultimatum game. Fairness is typically cited as one of the leading explanations why responder's reject small proposals.

Importantly, my experiment differs from Frank as it includes another treatment group with a different type of outside option (i.e. charity). This study also has bigger sample sizes, a bigger "pie", and stronger statistical power. Nonetheless, the experiment is designed with the same goal, that is, in order to investigate whether changing the context of rejected offers in a non-payoff fashion could influence decisions made by either party in an ultimatum game.

The experiment seeks to answer two questions: first, does changing the default result of rejection from a neutral to an efficiency-losing (i.e. destroying money) or to a charitable outside option (i.e. giving money to charity) scenario lead to different offers from proposers? Second, do responders significantly change their rate of rejection depending on those changes to the default?

The hypotheses I am testing are that there will be a difference between the three treatments in terms of the proposals as well as the rejection rates for unfair offers. I expect to reject the null hypothesis that all treatments are equal for both proposals and rejection rates. I also expect to reject the hypothesis that the destruction and charity treatments are equal. I believe that this will be the case based on previous research that suggests that people care about more than only their own monetary payoff.

For this study, 307 subjects were recruited from the Stockholm School of Economics (contacted via their school email accounts). This academic audience (i.e. Economics undergraduate and graduate students) were all asked to fill in an online survey (estimated time of 5 minutes to complete) that contained a standard Ultimatum Game with 1,000 SEK as the total sum to split. The participants were randomly selected to one of the three treatment groups and were asked to act in both roles, as proposers and responders in the UG. A strategy vector mode was chosen which implied that the participants had to choose first how much they would propose, and then make eleven acceptance decisions as the responder to all the possible proposals.

In order to strengthen the credibility of my findings, and increase distinction between hypothesis generation and hypothesis testing, I pre-registered my analysis plan, including experimental design, hypotheses and empirical strategy on the Open Science Framework

(OSF) prior to carrying out the experiment.[1] Pre-registration is an important step towards improving the validity of research and is strongly recommended by among others Nosek, Ebersole, DeHaven, and Mellor (2018). In the pre-registration, I made it clear that I expected to find that both treatments would have a significant effect on the participants. I envisioned that the destruction treatment would lower rejection rates due to efficiency concerns which would thus, through backward induction, induce proposers to offer smaller proposals. The charity treatment would instead increase rejection rates, and thus also through backward induction increase proposals. The results from the experiment, presented in Section 5, partly proved my hypothesis. On the one hand, they showed, contrary to Frank (1998) that non-monetary outside options have a significant effect on responder behaviour in the ultimatum game. This was shown to be true for both the treatment groups, destruction and charity. On the other hand, the proposer results showed that I could not reject the null hypothesis of no difference in proposals between treatments.

In this paper I start with a review, in Section 2, of existing research on the ultimatum game and similar public good games, with a focus on the main areas of interest: efficiency, fairness (conditional and unconditional), spitefulness, outside options, and social welfare concerns. Following this review, in Section 3, I explain the design of the experiment including discussions of subjects, procedure, treatments, and the statistical tests used. In Section 4, I discuss what results I expect to find based on the literature review, as well as the direction of the results for each of the tests. I then proceed to show the actual results of my analysis in Section 5, in accordance with the pre-analysis plan. I then move on to Section 6, where I answer the main research questions and discuss the implications of my results. Section 6 also contains a critical discussion of the experimental results and their validity. I conclude with a short summary and an opening up to further avenues of research in Section 7.

[1] I have attached a full copy of my pre-registration in the Appendix. My pre-registration plan as well the raw experiment data can be accessed at https://osf.io/zj5em

2. Previous research

In this section I will dive into the behavioural economics literature relevant to the ultimatum game, which is necessary in order to understand the results of my experiment. The following literature review will only provide a brief overview of the literature and I highly recommend readers to read more comprehensive and detailed reviews such as the one by Fehr and Gächter (2006) on the economics of fairness, reciprocity and altruism.

2.1. General ultimatum game

The ultimatum game, introduced by Güth, Schmittberger, and Schwarze (1982), provides insights into people's economic decision-making behaviour. In the game, two players bargain anonymously to split a given amount of money. The first player, the 'proposer', offers a division of the 'pie'. The second player, the 'responder', accepts or rejects this offer. If the responder chooses to accept, both players get their agreed upon shares. However, if the responder rejects, both players receive nothing. According to economic theory, the responder should accept any offer greater than zero. With this in mind, the proposer should offer the smallest amount possible, as this strategy would maximize their utility.

Formally, the proposer, X, chooses shares (x, y) for themselves and their responder, Y, respectively, with $x, y \geq 0$ and $x + y = p$. Then responder Y either accepts, $\delta(x,y)=1$, or rejects, $\delta(x,y)=0$. The payoffs are $\delta(x,y)x$ for X and $\delta(x,y)y$ for Y.

Material opportunism implies $\delta(x,y)=1, \forall y>0$. That is, responder Y accepts any actual positive offer y. Anticipating this, a rational proposer X should offer the smallest positive amount possible, as implied by subgame perfectness of equilibria (Güth & Kocher, 2014; Selten, 1975).

Contrary to rational decision-making and the subgame perfectness of equilibria, experimental economists show that participants systematically reject positive offers in the ultimatum game. These results challenge the idea of pure material opportunism. Clearly these anomalies are difficult to attribute to solely noise, mistakes, or confusion. The consistent rejection of certain types of positive offers in all ultimatum game experiments since 1982 implies that there is more going on. In a literature review by Güth and Kocher (2014), proposers tend to offer about 40-50% of the total amount in general ultimatum games. This is an extremely robust phenomenon. They also show how responders respond to these proposals. Responders rarely refuse such offers of 40-50% of the pie, but the rejection rate increases as proposals decrease. Typically, proposals offering the responder below 20% of the pie are rejected half the time, with probability 0.4 to 0.6 (Fehr & Schmidt, 2006). These results suggest that an equal or near-equal split (40-50%) maximizes the proposer's expected payoff (Camerer, 2003). On the other side of equal

offers, Güth, Schmidt, and Sutter (2003) find that although most responders have monotonic rejection strategies, a small number of "super fair" offers (i.e. those offering a share greater than 50% for the responder) are consistently rejected.

The evidence from empirical studies has two implications. First, contrary to what the standard economic theory suggests, responders care about more than their own monetary payoff. Second, proposers are aware of this, as they offer fair shares for the responder (Güth & Kocher, 2014).

Indeed, according to Cox (2004), monetary incentives are insufficient for judging player's utility in regard to making a decision. Cox's paper shines light on how and why subjects play in different variations of the trust game[2], which gives us insight into how these same players would play the ultimatum game. Results regarding beliefs about altruism as described in Andreoni and Miller (2002), inequality-aversion (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) and reciprocity should be incorporated in social preference models as they help explain the outcomes of economic games and economic behaviour more generally.

Another explanation for the rejection of low offers is provided by the pride/spite model, which states that responders, feeling angry, may react to small ultimatum offers spitefully (Pillutla & Murnighan, 1996). Less intuitive results may also contain clues to players' behaviour in the ultimatum game. Perhaps the most common explanation referred to in the literature is fairness and inequity aversion. Participants may treat the 'pie' given to them as a windfall gain, and therefore feel that they should share it equally (Güth & Kocher, 2014). Güth and Tietz (1985) and Hoffman, Mccabe, and Smith (1996) provide further evidence to support this idea. When proposers are made to feel entitled to an endowment through an auction or performance on a quiz, the proposers rarely make equal splits but rather only offer around one third of the pie.

2.2. Fairness

The notion that self-regarding preferences are sufficient to explain all participant behaviour in ultimatum games has been dispelled. I turn now to a specific set of theories that aim at explaining anomalous behaviour in ultimatum games, namely the effect that arises from the comparison between someone's own payoff and the payoff of the other player. This subsection will show that behaviour has been successfully expressed in terms of well-behaved preferences where fairness, in the form of altruism and inequity aversion, plays a central role.

[2] In a trust game, both players receive an endowment. The first player has to decide how much of their endowment they want to send to the second player, knowing that whatever they send will be tripled. Then, the second player decides how much of this increased endowment to send back to the first player.

The concept of fairness is intricate and multifaceted, but it is extremely relevant in order to understand the ultimatum game and the results of the experiment. I will divide fairness into two parts: unconditional and conditional fairness. On one side, I have unconditional feelings of fairness or what I will call altruism. I will analyse what altruism implies and what previous research within behavioural economics believes to be its leading motivators. On the flip side, I examine unconditional fairness, including a discussion on factors such as inequity aversion, reciprocity, and intentions.

2.2.1. Unconditional fairness

Unconditional fairness has been thoroughly analysed by researchers such as Andreoni (1989), Andreoni and Miller (2002), Cox, Sadiraj and Sadiraj (2001), and Charness and Rabin (2002). The concept of altruism has mostly been explored in the context of dictator games.

The dictator game works similarly to the ultimatum game, but the offer made by the proposers is immediately executed and responders passively receive their share. This allows experimenters to explore the altruism of a proposer (the dictator) without responder effects because it does not involve the backward induction that leads to strategic behaviour. Literature on dictator games has repeatedly found that on average dictators do not keep the whole pot for themselves. Instead, they will share at least some of it with the passive receiver (Camerer, 2003). Camerer shows that it is common to find that 50% of participants donate a part of the pot.

The question that researchers seek to answer is why people behave in this non-payoff maximizing manner? Why would a participant donate a part of their participation fee to somebody else? Is this altruism and how is altruism defined?

Andreoni (1989) divides altruism into two parts: intrinsic and extrinsic or pure and impure. Frey (1997) suggests that individuals derive utility from observing other people receive some sort of benefit. This is what is known as intrinsic or pure altruism. Andreoni (1989) calls impure altruism the warm-glow effect, originating from Becker (1974), who suggests that people feel good when they give. Andreoni suggests that the warm-glow effect is the dominating form of altruism with regards to charitable donations. Extrinsic motivation is, by contrast, any material reward or benefit associated with giving, even when strictly speaking the material reward is worth less than the altruistic giving. Rewards span from kind gestures to monetary benefits such as tax breaks for charitable giving.

Another form of extrinsic motivation is image motivation or signalling motivation. Signalling, as the name implies, refers to an individual's tendency to be motivated partly by others' perceptions, which captures the role of opinion in utility (Fehr & Schmidt, 2003; Meier, 2007). A field experiment carried out by Andreoni, Rao, and Trachtman (2017) sheds lights on this facet of altruism. The authors studied how people behave if

they can choose to avoid walking by charity organizations that ask for money. By varying the presence of a fundraiser at different entrances to a mall, they discover that people prefer to avoid being asked to give money for charity. This provides evidence in favour of the hypothesis that people are partly extrinsically altruistic. When people were asked verbally rather than non-verbally to donate, they tended to donate more, supporting the signalling hypotheses. Ariely, Bracha, and Meier (2009) also investigated the effects of prosocial incentives and publicity on the effort exerted to perform a given task. In their experiment, subjects could donate to a charitable organization by clicking repeatedly on two keys on a computer keyboard. Participants were randomly assigned to have personal monetary incentives, to have their donations publicized or kept private, and to have donations go to a "good" or "bad" charity. They found that in public, people tend to behave prosocially, meaning that they exert more effort if their effort is public information and leads to a prosocial outcome. On the other hand, people are more driven by monetary incentives if the setting is private.

The framing of both dictator games and ultimatum games plays a significant role in what results are found (Eckel & Grossman, 1996; Bohnet & Frey, 1999; Haley & Fessler, 2005). For example, Bardsley (2008) suggests that the dictator game is not as representative of altruism as one could think. Bardsley manipulates the dictator game by adding a taking game, whereby dictators cannot only choose to give a part of their pot but can also take a part of the other player's pot. Bardsley shows that donations are significantly diminished through this change. The underlying mechanism is the "range effect", implying that not giving anything is no longer the worst action. The range effect is based on the Parducci and Wedell (1986) model, according to which the participant's choice to donate is dependent on the range of the possible donation and taking alternatives.

2.2.2. Conditional fairness

Inequity aversion is one of the leading explanations as to why subjects reject unfair offers. This theory is built on research by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) among others. Inequity aversion means that people are willing to give up some material payoff to which they would be entitled, in order to move in the direction of more equitable outcomes. Fehr and Schmidt (1999) present a model in which some people behave in a self-interested way only as long as payoffs between players are similar. Formally, the utility of an inequity averse individual "i" is dependent on his payoff "x_i" and the payoff of the other player "x_j":

$$U_i(x) = x_i - \alpha_I \, max[x_j - x_i, 0] - \beta_I \, max[x_i - x_j, 0], \; i \neq j \qquad (1)$$

Where $\beta_i \leq \alpha_i$, implying that people are more averse to inequity if they are worse-off, $0 \leq \beta_i$, implying that inequity averse subjects do not derive utility from being better-off than

others, and $\beta_i < 1$, to avoid implausible outcomes. The acceptance threshold of responders is then described by $s'(\alpha) = \alpha / (1 + 2\alpha)$.

An important factor in Fehr and Schmidt (1999) is that people are heterogeneous. There are some purely selfish players that only care about their own monetary payoff, while there are other participants who always propose fair proposals, even though they realistically increase their monetary pay-out by proposing a lower sum to the other player. Andreoni and Miller (2002) find that 25% of participants have a purely selfish payoff-maximizing strategy.

Another popular model of conditional altruism is the social utility model. The social utility model has been commonly referred to as one of the best for describing the behaviour of players in the ultimatum game by researchers such as Blount (1995), Messick and Sentis (1985), and Van Dijk and Vermunt (2000). It is a simple model consisting of only two parameters: an absolute payoff factor and a comparative factor (Handgraaf, Van Dijk, & De Cremer, 2003). The absolute payoff factor pushes players to maximize their payoff, independent of the other player. The comparative factor is a relative measure where the player's utility is derived from the relation of his/her payoff with that of the other player. This factor is partly based on factors such as fairness, which I have previously mentioned.

In a dictator game experiment, Eckel and Grossman (1996) find that players are also sensitive to the reputation of the receiver. A receiver who is deemed more "deserving", for example a charity (they use the American Red Cross in their experiment), receives significantly more from the proposer.

Charness and Rabin (2002) build on this model of inequity aversion and altruism and combine it with intention-based reciprocity. Their research suggests through their novel approach that players want to punish those who behave unfairly and that players are less willing to sacrifice a part of their pay-off in order to help others when the others do not show the same eagerness.

2.3. Spitefulness, envy, and the joy-of-destruction

The opposite behaviour to altruism is envy or spitefulness (Bolton, 1991; Kirchsteiger, 1994; Mui Vai-Lam, 1995). A participant with these characteristics would derive utility from the other person having less utility. They would even be prepared to take personal cuts in monetary payoff in order to decrease the other player's pay-out. Kirchsteiger (1994) builds a model fitting ultimatum game data and claims that the reason ultimatum game proposals differ from dictator game offers are not fairness concerns, but instead a fear from the proposer's side that the responder will reject the proposal out of envy or spitefulness.

Abbink and Sadrieh (2009) introduce a joy-of-destruction game in which players each received an endowment and simultaneously could destroy any proportion of the others' endowment; their results suggest that with some anonymity involved (i.e. no fear of retaliation) money was destroyed in almost 40% of games played. There was no evidence of inequity aversion contributing to this destruction; possible explanations offered include utility from 'being nasty' or possibly pre-emptive retaliation. They also note that if there is an 'excuse' present, such as a chance that nature will destroy the other player's money anyway, destruction rates rise significantly (Abbink & Sadrieh, 2010).

Spitefulness is not always just about deriving pleasure from the destruction of surplus for the other player. In fact, it can also be an efficient tactic. Avery and Zemsky (1994) discuss how Rubinstein bargaining[3] is vulnerable to players threatening to destroy surplus in order to achieve bargaining power. The concepts described may underlie some of general ultimatum game behaviour when receiving parties reject low offers (at least before the last round), the prospect of this threat may be especially salient when the rejected surplus will be utterly destroyed rather than recycled.

Spitefulness is naturally based on a notion of fairness (or unfairness). It is critical to understand that this notion of fairness is not the same for everybody. In fact, it varies widely between different populations and individuals. Andreoni and Miller (2002) clearly state that this extensive diversity in opinions is essential to understanding the ultimatum game and the concept of fairness and altruism. They show that there is a wide range of preferences within tested subjects, from Rawlsian to Utilitarian.

Zhang and Ortmann (2016) discuss and further investigate the joy-of-destruction game, building on dictator game research by List (2007). They note that player responses in joy-of-destruction as well as dictator games also seemed to depend on the choice sets that were offered. In their joy-of-destruction games they ran a classic only-destruction joy-of-destruction as a control and a treatment where players could instead increase the other players' endowment (at no cost to their own pay-out), which introduced the possibility of countervailing efficiency concerns. If players were given the choice to instead increase total welfare, this option was chosen significantly more often. Zhang and Ortmann (2016) also remark on the importance of experimental conduct, noting that robustness tests such as controlling for social distance, order effects, increased stakes and asset legitimacy can be vital to get consistent results from a joy-of-destruction game. They suggest that this is because there is a considerable level of context dependence when it comes to players' decisions to act pro- or anti-socially.

Schröder and Sadrieh (2012) cover in more detail the effect of experimenter demand and full information on behaviour in a similar give-and-take joy-of-destruction game in "The Desire to Influence Others". While they find a social effect that increases the extent of

[3] In Rubinstein bargaining, two players take turns making offers, until one player accepts the other's offer. However, a rejection leads to a delay. This is accounted for by making use of discount factors.

pro-social behaviour, it does not change people who were not already inclined to do so, nor does it influence the prevalence of anti-social behaviour. Zizzo and Fleming (2011) reach similar conclusions.

2.4. Efficiency

Another important aspect for interpretation of the ultimatum game is the concept of efficiency. In the ultimatum game, Güth and Kocher (2014) claim that responders care more about fairness than efficiency when deciding to accept or reject a proposal. This implies that they are essentially "burning money" when rejecting.

However, in behavioural economics literature, efficiency plays a key role in explaining how people act. In Charness and Rabin's (2002) two-person dictator game experiments, they show that subjects are willing to take cuts in personal gain for the good of efficiency. The experiments involved giving players a list of binary choices in the style of dictator game experiments. They conclude that inequity aversion does not play as large a role as previously thought and that instead, what they call social welfare preferences are more telling when it comes to player behaviour in these games. Charness and Rabin (2002) suggest that social welfare preferences are a mix of efficiency and altruism.

Cartwright and Stepanova (2017) studied the conditions under which forced contributions lead to high efficiency in binary threshold public good games (PGG). A threshold public goods game involves a project only being started once a certain funding level is reached for example for a new community school (Andreoni, 1998). They found that there are ways of forcing or at least helping people contribute to these public goods. Fischbacher, Gächter, and Fehr (2001) show that in the classic public goods game cooperation tends to unravel in repeated games. In order to increase contributions, subjects can be forced to commit to their investment in the threshold game once the threshold sum is reached. Additionally, experimenters can alleviate subjects' risk-aversion by telling them that they would get a refund if the threshold is not met (Dawes & Thaler, 1988). In the PGG, punishment (monetary and non-monetary) can improve cooperation and efficiency (Ostrom, 1999; Fehr & Gächter, 2000; Masclet, Noussair, Tucker, & Villeval, 2003). The classic public goods game is structured in such a way that subjects can choose to contribute to a common cause, but that the rewards are split equally among the whole population. The way this game is structured incentivizes players to free-ride. What Ostrom (1999) and Fehr and Gächter (2002) show is that if players can punish free riders, then players will generally contribute more. This would suggest that participants have some proven concern for public welfare or efficiency.

2.5. Stakes

More recent studies have focused on the importance of stakes. One significant issue in the analysis of stakes' role in the UG is the lack of information on how responders react to low proportional offers in high stakes games. Some studies that have presented results on whether stakes matter (Slonim & Roth, 1998; Cameron, 1999) show that stakes have a small significant effect. However, the authors counted only on very few offers below 20 percent, the threshold that typically marks a very unfair offer. Nonetheless, they still managed to show that rejection rates decreased with an increase in stakes. In order to address the issue of having too few unfair offers, Andersen, Ertaç, Gneezy, Hoffman and List (2011) run an experiment in Northeast India with high stakes and designed in a way that encouraged low offers. They found the similar result that stakes matter, in the sense that when stakes are high, relatively low offers are less likely to be rejected. Amir, Rand, and Kobi Gal (2012) run multiple economic games with very low stakes (1 USD) over the internet and find their results mirror the general literature, excepting the case of extremely high stakes such as Andersen's experiment.

2.6. Outside options

One of the most relevant strands of ultimatum game research for my experiment is that of subject responses to outside options. An outside option is when subjects have a nonzero outcome to compare to offered pay-outs.

In a review of ultimatum game research, Handgraaf, Van Dijk, and De Cremer (2003) mention outside options within a group of factors within game characteristics that can influence the social utility model. Their findings suggest that an outside option can significantly change the way people perceive offers in the ultimatum game. The main mechanism through which they believe outside options play a role is that they decrease the interpersonal, or comparative factor, while increasing the intrapersonal, or the absolute personal payoff. Without an outside option, players have a higher tendency to compare their payoff with the proposer's payoff. Through suddenly having another clear positive outcome presented to them, players are more likely to compare the payoff they would receive if they accept the ultimatum game offer from the proposer, with the outside option payoff. They conclude that adding a non-zero positive outside-option leads to higher acceptance levels of low offers.

Blount and Blazerman (1996) also show that players pay less attention to relative (and unfair) pay-outs when they are given a choice between two outcomes. They show this by analysing take-up rates between two treatment groups. There were two experiments, one where all participants were paid $7, and the second where the participants were paid $8, while some other participants were paid $10. Blount and Blazerman showed through the take-up rates of the experiments that the participants who could choose between both

picked the $8 experiment, while those who could only choose the $8 treatment were more likely to refuse to participate because of fairness concerns (others would be paid $10 for the task). This suggests that outside options play a big role in responders' rejection rates and perceptions.

Schmitt (2004) shows that outside options have an impact on both proposer and responder behaviour. In several experiments, he varies the amount of information that participants have and whether they have positive non-zero outside options. Schmitt demonstrates that outside options significantly increase the minimal acceptable offers of responders. The outside option thus becomes a reference point that subjects can use to increase their demands. Hennig-Schmidt, Irlenbusch, Rilke, and Walkowitz (2017) also validate this claim in a laboratory experiment where they adjust the magnitude of the outside option. Clearly the size of the outside option plays a significant role in the bargaining power of the players, be they proposer or responders.

Abbink, Bolton, Sadrieh, and Tang's (2001) experiment is very much evident proof of the relevance of outside options. The experiment involves an ultimatum game with two treatments. In one group, the proposer has a positive pay-off if the proposal is rejected, a "reward", and in the other, a negative pay-off, a "punishment". However, the proposer does not know whether he has one or the other. Only the responder is aware of this outside option. The researchers show that responders have significantly higher rejection rates of proposals when they are aware of the outside option that they can punish the proposer. This research suggests that the type of outside option can influence participant behaviour and that a punishment is apparently more potent than a reward.

Although much research has been done on the ultimatum game, the literature with regards to outside options is relatively sparse and there is a particular gap in the literature when looking at outside options without a monetary payoff for the participants, or what I call non-monetary outside options.

2.7. Social welfare concerns for the experimenter

Altruism and other-regarding preferences have been shown to play a significant role in how participants act (see previous section on altruism and efficiency). Tjøtta (2019) finds that there may be more to the social welfare model than was previously thought. In an effort to analyse why participants leave money on the table in certain experiments, Tjøtta claims that his findings could compromise the interpretations of many other experiments.

It is a common norm among behavioural experiments that participants choose more money. However, participants typically do not keep the whole surplus for themselves. For example, in the dictator game players tend to give a small part of the total sum to the other player, reflecting a concern for the other player. Now, how would players act in an experiment where they do not have another player to take into account? In Tjøtta's

experiment, the game is designed in such a way that the participants could play a game anonymously where they had to choose between receiving a lower sum or a higher sum of money. Surprisingly, even with the double-blinded design and the lack of another participant who would be affected, nearly 29% of participants chose to receive less money. One of the reasons he identifies that could explain this astonishing result is that participants have a social preference towards the experimenter (Tjøtta, 2019).

This social preference towards the experimenter is what Frank (1998) attempted to study 20 years earlier. He tested this in an ultimatum game setting by telling the participants that he would burn the total sum to be split if the proposal was rejected. In order to test this new model, Frank designed a lab experiment. One of the groups, the control group, plays a standard ultimatum game where the participants play as both the responder and proposer, and where they are told that in case of rejection, none of the participants will receive any money. In the treatment group, the participants are additionally told that in the case of rejection, the money will be burned. However, Frank's results cannot reject the hypothesis that the results from the treatment groups are equal. It is important to note that Frank also had a small sample and low power in his study with only approximately 50 participants in each treatment group. He concludes that participants do not care about the experimenter's payoff in the ultimatum game since he did not find a significant change in responder behaviour due to this treatment. Nonetheless, he asserts that his experiment and his results are not sufficient to reject the hypothesis that participants have a concern for the experimenter's wealth or welfare.

2.8. Research questions

Previous research has assumed that players do not care much about efficiency in the context of ultimatum games and are happy to 'burn money' (Güth & Kocher, 2014). However, this does not account for the fact that if a responder rejects a proposal, the money is not truly burnt - it generally is simply not spent and stays in the experimental budget.

My experiment seeks to answer two questions: first, does changing the default rejection scenario from a neutral to a strictly efficiency-lowering (i.e. actually destroying money) or to a positive non-monetary outside option (i.e. giving money to charity) scenario lead to different offers from proposers? Secondly, do responders significantly change their rejection rates in response to low offers, depending on those changes to the default?

The literature review suggests that there is a gap in the literature with regards to understanding the relationship between efficiency, fairness, and altruism concerns in the ultimatum game. The two treatments will help illuminate these competing concerns and lead to a better understanding of the ultimatum game and economic behaviour. In the following section I go through the design of the experiment.

3. Design

I investigate whether changing the context of rejected offers in a fashion that does not affect monetary payoffs could influence decisions made by either party in an ultimatum game. In order to study this very specific mechanism, I designed an experiment where I controlled for every other factor other than the treatment effect that I am interested in, that is, the effect of non-monetary outside options on behaviour of proposers and responders in an ultimatum game. The design as well as the hypotheses I tested were all decided in advance and locked-in through the use of a pre-registration on the Open Science Framework (OSF).

In this section I introduce the design of the experiment, including a description of the treatments, subjects, procedure, and statistical tests, as well as a discussion on why I made these particular choices. I begin with the treatments.

3.1. Treatments

The study involves three treatment groups:

> **Treatment 0:** Control Group. Standard ultimatum game
>
> **Treatment 1:** Destruction Group. Standard ultimatum game + "if rejected, the money will be destroyed"
>
> **Treatment 2:** Charity Group. Standard ultimatum game + "if rejected, the money will be donated to charity"

Treatment 0 corresponds to the control group. The control group plays a standard ultimatum game: the proposer makes an offer on how to split the given sum (1000 SEK), and the responder either accepts or rejects. Acceptance leads to a split of the money as suggested by the proposer. In case of rejection, the participants does not receive the money. It is important to note that in this case the instructions do not specify what will happen to the money in case of a rejection. The money actually goes back to experimenter (i.e. me).

The participants in this experiment get to play the strategy method of the ultimatum game, where they act as both proposer and responder. First the player submits the proposal they would make if they were the proposer, and then answers which proposals they would accept if they were the responder (11 different proposals, intervals of 100 SEK).

Treatment 1 is similar to Treatment 0. The only difference is that participants are now informed that, in case of a rejection by the responder, the money is destroyed.

Treatment 2 is another variation along the lines of Treatment 1. In this treatment participants are informed that if the responder rejects the offer, the whole pot is given to a charity, Save the Children Sweden (Rädda Barnen).

3.2. Subjects

The participants in this study are students as the survey was only distributed by email to students enrolled in the Stockholm School of Economics (SSE).

The experiment is written in English and is based on the assumption that everybody who studies at SSE has a fluent level of English and should thus be able to perfectly understand the questions being asked. The study was designed primarily for these international and Swedish students at SSE also through the choice of the chosen charity: Save the Children Sweden (Rädda Barnen). The chosen money to be divided was in Swedish crowns (SEK) and the destruction treatment is specifically for a Swedish population since destroying currency is legal in Sweden.

3.2.1. Exclusion criteria

In contrast with the pre-registration plan I choose to include two exclusion criteria. The reasoning for the inclusion of these criteria are the following. The ultimatum game is based on making decisions on your payoff. If this payoff is hypothetical, then this would weaken the validity of the results. The only manner in which participants could be paid was by providing their contact details, including name, and email address. I could not pay those who did not fill in the survey in its entirety. On account of this being common knowledge, I believe that participants who did not complete the survey had a lower probability of answering the questions truthfully.

Multiple participation could also contaminate the results of the experiment. Given that nobody filled in the survey twice with the same name and/or email address the multiple participation criteria did not end up having an effect on this experiment.

I run a robustness check in order to verify that my new exclusion criteria do not have a repercussion on the outcome of my experiment. I find the same results with and without these exclusion criteria. Adding the incomplete observations does not change the significance and direction of the tests that I run. The results of these robustness checks can be found at the end of section 5 and the tests themselves can be found in the Appendix.

1) **Non-completion**: I decided to exclude those participants that did not complete the survey. The survey was designed in such a way that the participants had to answer every question in order to go to the next page. Participants had to submit their name

and email address on the last page in order to complete the survey and in order for me to be able to randomly select the winners and contact them so that they could receive a participation fee. Also, so called "Attention checks" were included in the survey in the form of questions that the participant had to answer correctly. These questions had only one possible correct answer and were intended as proof that the participants had understood the choices they would make.

2) **Multiple participation**: I did not find any participants that filled in the survey twice, but I would have excluded both of the responses if this had been the case. It was relatively straightforward to check who had participated since I had access to names, email addresses, telephone numbers, as well as IP addresses collected by Qualtrics

As pre-specified in the pre-registration, a stopping rule was not implemented, and the experiment ran during the whole month of March until the end date (31st March 2019). The data is available on the Open Science Foundation (OSF) so that my results can be replicated by those who wish to do so. The raw data directly taken from Qualtrics is also available, but in order to respect the anonymity of the participants their names and personal details have been redacted.

3.2.2. Sample size

My target sample size was 300 participants. With 300 total participants and a well-functioning randomization software through Qualtrics, this implies approximately 100 participants in each treatment group. This in turn, due to the strategy method, implies 100 proposals and 100 sets of responder answers for each treatment (Responders have to specify if they would accept or reject 11 different proposals, ranging from a 0/1000 division in favour of the proposer to a 1000/0 division in favour of the responder).

3.2.3. Experimental power

A possible important internal limitation to note is that of experimental power. To make sure that the tests can correctly reject the null hypotheses, I have chosen large sample sizes per treatment group. I hope that this will help avoid type M and type S errors (Gelman & Carlin, 2014). It may still be the case that the effect sizes of the treatments are not large enough to give significant results. Frank (1998) shows that low power due to a low effect size can be a fundamental issue in the experimental design, excluding the limitations discussed above. Considering that Frank (1998) used a sample size of approximately 50 in each treatment and did not find a significant effect on responder behaviour, I double the treatment group sample sizes to at least 100 participants in each treatment group. There is therefore still a possibility of a false null or a type 2 error.

3.3. Experimental procedure

A copy of the survey and the pre-registration plan is attached in the Appendix.

An email was sent out to all enrolled students at SSE starting from the 25th of February asking them to participate in the experiment. Students were also asked not to discuss the experiment with others once they had finished. People were able to take the survey up until the 31st of March when the survey was made inaccessible.

The randomization to each treatment group happens after the second page of the survey is completed. You can see the complete survey in the Appendix. The first page of the survey is an introduction to the experiment and the second page is a consent form. No detailed information on the exact nature of the experiment is provided to the participants on these two pages.

Figure 1: First Page of the Survey

Welcome to the Ultimatum Game 1000 Experiment !

You are being asked to participate in an experiment about people's decision-making, conducted as part of a master's thesis work at the Stockholm School of Economics.

The experiment will take approximately 5 minutes. Your decisions in this game will be anonymous, and your identity will be anonymous to the other participants. Please do not discuss this experiment with others after your participation.

Figure 2: Second Page of the Survey

Online Survey Consent Form

Answer a short decision-making study

Purpose of research: To examine decision-making.

What you will do in this research: You will participate in a simple decision-making study and then answer a short survey.

Time required: Participation will take about 5 minutes.

Risks: We believe there are no known risks associated with this research study. Your participation in this study will remain anonymous to other participants.

Benefits: 6 participants will be randomly selected to have the opportunity to receive up to 1,000 SEK depending on the choices they made.

Participation and withdrawal: Your participation in this study is completely voluntary and you can withdraw at any time.

To contact the researcher: This study is being done by Joakim Karlsson from the Stockholm School of Economics. If you have questions about this research, please contact Joakim at 23002@student.hhs.se.

By clicking the link to start the experiment you are indicating that you have read and understood this consent form and agree to participate in this research study.

The participants were told that six participants in this experiment were to be randomly selected and then randomly assigned to one of the two roles. Thus, when somebody was selected, their actions in this experiment were matched with the actions of another player. If a pair agreed on the proposal, the money was awarded to the participants. However, if the responder rejected the proposal then the money was either destroyed, donated to charity, or given back to the experimenters depending on the treatment group.

3.3.1. Pilot study

A pilot study was carried out on 10 participants on the 14th February 2019, two weeks before the planned release of the proper survey was sent out. The pilot study was intended to help find inaccuracies in the survey and make sure that all the questions would be understood by the participants and that the software was working as it should as well as other minor adjustments.

3.4. Measured variables

One outcome variable is the proposal made by the proposer. I measure this by asking participants "Given that I would be selected as the Proposer, I would like to submit the following proposal:" Participants are asked to choose one of the 11 alternatives going from "The Proposer receives 0 (1000 for the Responder)", "The Proposer receives 100 (900 for the Responder)", step-wise on a 100 SEK scale, all the way to "The Proposer receives 1000 (0 for the Responder)".

The second set of outcome variables are the answers to whether the participant would accept or reject the above 11 proposals. I measure this by asking participants "Given that I would be selected as the Responder, I would accept the following offer(s):" Participants pick either "Accept" or "Reject" for each of the 11 possible proposals. In the charity

Treatment, or Treatment 2, the first proposal is "Proposer receives 0, Responder receives 1000" which the participant can either "Accept", or "Reject (Proposer receives 0, Responder receives 0, Save the Children Sweden (Rädda Barnen) receives 1,000)".

Many variables are transformed. The output variables from Qualtrics give independent outcome variables for each of the treatments. In order to collect all of the answers for the proposal by the proposer, I create a new variable called "Proposal" where I collect all of the answers from the different treatment groups. I create the same sort of variables for the Responder answers to each of the Proposals which I call "Responder 0/1000", "Responder 100/900", "Responder 200/800", "Responder 300/700", "Responder 400/600", "Responder 500/500", "Responder 600/400", "Responder 700/300", "Responder 800/200", "Responder 900/100", and "Responder 1000/0". The answer for each of these variables are Accept or Reject which are recoded to 1 or 0. I create a new variable called "Treatment" for each observation so that I can track which participant had which treatment. I call the treatments 1, 2, and 3 for the control, destruction, and charity treatments respectively.

In the end I have 12 outcome variables for each Treatment Group.

3.5. Statistical tests

In this study, two hypotheses are being tested. Firstly, I expect that the offers made by proposers differ across the three treatment groups, which are (0) control i.e. the classic ultimatum game, (1) destroying the money following a rejection, (2) giving the money to charity following a rejection. Secondly, I predict that the rejection rate for unfair offers will not be the same across these three treatment groups.

I start first with testing the hypothesis that the offers (ranging from 0 SEK to 1000 SEK, in 100 SEK increments) made by the proposers differ over the treatment groups. Hence, I seek to reject the *H1 null hypothesis* that offers are equal over the groups in favour of the *H1 alternative hypothesis* that offers of different groups differ. After collecting the data on the 300 proposals made (100 proposals per treatment group), I rank them in a table.

The null is rejected if I find a p-value below 0.05. When this happens, I will perform pairwise Mann-Whitney U tests post-hoc to identify which treatment group exhibits stochastic dominance.

H1:
Null hypothesis: Proposals within all three treatment groups are equal
Alternative hypothesis: Proposals within all three treatment groups are not equal

The second hypothesis that I seek to test is that the rejection rate for unfair offers differs over the treatment groups. As I am only analysing the rejection rate of unfair offers, I look at the response to proposed shares equal or below 40% for the responder, as supported by the literature (Güth & Kocher, 2014). I then perform the Kruskal-Wallis H test corrected for ties as described above. In particular, I expect the rejection rate to be lower in the destruction treatment, because of efficiency concerns on the part of the participants. In the charity treatment, I expect higher rejection rates for unfair proposals. I will explain in further detail why I expect to find these results in section 4.

If I am able to reject the null at the 5%-significance level, I can conclude that the rejection rate differs among the different treatment groups. To identify which treatment group differs from the control group, I again perform pairwise Mann-Whitney U tests.

H2:
Null hypothesis: Rejection rates for unfair proposals within all three treatment groups are equal
Alternative hypothesis: Rejection rates for unfair proposals within all three treatment groups are not equal

I perform the Kruskal-Wallis H test for each of the two hypotheses. This test is an extension of the Mann-Whitney U test to more than two groups. There are three conditions that must be met in order to be able to appropriately use the Kruskal-Wallis H-test (Siegel & Castellan, 1988):

1) A continuous or discrete dependent variable
2) At least 3 independent groups in the independent variable
3) Independence of observations. Different participants in each group and no relationship between observations in each group.

The Kruskal-Wallis H test is a non-parametric method, and thus does not require an underlying normal distribution of the residuals (unlike the one-way ANOVA test). The test was chosen because of this factor since the results of ultimatum game experiments are not necessarily normally distributed.

Kruskal and Wallis (1952) write that:

> "Thus, what H really tests is a tendency for observations in at least one of the populations to be larger (or smaller) than all the observations together, when paired randomly. In many cases, this is practically equivalent to the mean of at least one population differing from the others, (p. 598)"

> "in practice the H test may be fairly insensitive to differences in variability, and so may be useful in the important 'Behrens-Fisher problem' of comparing means without assuming equality of variances (p. 599)."

The Kruskal-Wallis H-test is commonly used to test whether the different treatment groups come from the same distribution. The null hypothesis in this experiment for H1 is thus that the three treatment groups come from the same distribution. The alternate hypothesis is that at least one of the groups is different from the others.

If I am able to reject the null at the 5%-significance level for the Kruskal-Wallis H test, I can conclude that the proposals (H1) or rejection rates (H2) differs among the different treatment groups. To identify which treatment group differs from the control group, I perform pairwise Mann-Whitney U tests.

The pairwise Mann-Whitney U tests are performed only if I can reject the null hypothesis for the Kruskal-Wallis H-test. Thus, the following 2 groups of 3 hypotheses for proposals and rejection rates will only be relevant if the afore-mentioned condition is met.

H3: Proposals between the control treatment and the destruction treatment are equal
H4: Proposals between the destruction treatment and the charity treatment are equal
H5: Proposals between the control treatment and the charity treatment are equal

H6: Rejection rates between the control treatment and the destruction treatment are equal
H7: Rejection rates between the destruction treatment and the charity treatment are equal
H8: Rejection rates between the control treatment and the charity treatment are equal

3.5.1. Exploratory analysis

In the pre-analysis plan I included an exploratory analysis which consisted of a Chi-squared test for each responder accept/reject ratio, that is, 11 tests for each pairwise comparison. In the pre-registration I did not specify my expectations. Nonetheless, the direction of this exploratory analysis for the rejection rates of unfair proposals should be in line with the previous tests where I expect that unfair rejection rates will be significantly different as a whole. Thus, the majority of the unfair proposals should have significantly different rejection rates on an individual level. I also expect the direction of the results to be similar, where the destruction treatment would have the lowest rejection rates for unfair proposals compared to the others, and where the charity treatment would have the highest rejection rates for unfair proposals.

On the other hand, I did not have any expectations for the pairwise comparison of the rejection rates for the fair and super-fair proposals. These results will only be studied ex-post.

4. Expected results and direction

In this section, I describe in more detail, using the research literature, what results I expected to find before running the experiment. I have divided this discussion into three parts, one for each of the treatment groups.

4.1. Control group

Checking whether the baseline results are in line with the empirical literature on the general ultimatum game is an important first step towards verifying if the results are reasonable.

As mentioned by Güth and Kocher (2014), previous research reveals that a clear majority of proposers offer approximately 40-50% of the total amount to responders. This is an extremely robust phenomenon in the classic ultimatum game. Responders typically reject approximately half of proposals when shares are below 20% of the pie with an exception in the case of very high stakes (Fehr & Schmidt, 2006).

A failure to replicate these results indicates that the design of the experiment may be flawed. Having a different distribution of proposals and rejection could be driven by various factors. Although the research on the subject of the impact of different student populations is ambiguous, I cannot rule out that students attending the Stockholm School of Economics (SSE) who participate in the experiment are a particularly abnormal sample. Given that most experiments are conducted on student populations the results should be very similar, although SSE students are perhaps more inclined to think from a game theoretical point of view than a student who has not studied microeconomics. Studies by Kahneman (1986) and Carter and Irons (1991) support this hypothesis that economists behave differently.

Another reason that results could differ with other ultimatum games is that experimental conditions could contaminate the sample. For instance, since the experiment takes place continuously over a month, students might discuss the experiment after their participation and possibly influence the decisions taken by subjects who have chosen to participate in the experiment on following days. I have tried to limit this effect by stating explicitly in the instructions given to the subjects that communication regarding the experiment after participation is discouraged, but it is impossible to perfectly control for this.

The experimenter demand effect is likewise a possible factor that could influence the participants' choices. The experimenter effect refers to behavioural adjustments by participants due to participants anticipating what the experimenter wants them to do (Zizzo, 2010). The participants' choices were anonymous to each other, but they were also aware that I would be able to see their choices since I had to have their contact details and answers in order to pay-out the participation fee to the lottery winners. It is possible

that the fact that I could see their results and that this was common knowledge, had an effect on the answers they gave. Hoffman, McCabe, Shachat, and Smith (1994) identified this experimenter effect in a dictator game study where they varied anonymity conditions. This effect could also be strengthened through the fact that many of the participants personally knew who I was, given that I am a student at the school.

4.2. Destruction treatment

In the first treatment, both proposers and responders are given clear instructions that, in case of rejection, the 1000 SEK that they could have split would be destroyed. The potential differences between proposal and rejection rates in the control and the destruction treatment should provide new information regarding the relevance of non-monetary outside options, efficiency and other-regarding preferences. As explained in the design section, I test whether the magnitude of proposals and rejection rates for unfair offers are significantly different between treatments. Here, I explain why I believe that the responder and proposer behaviour will be significantly different between the destruction treatment and the other treatments.

4.2.1. Lower rejection rates

I start off with analysing the expected effect of this treatment on responder behaviour. I believe that rejection rates of unfair proposals will be higher for the destruction treatment than for the control and charity treatments.

Efficiency concerns is one of the main channels through which I think the results will be impacted. Charness and Rabin (2002) claim that social welfare preferences are a mix of efficiency and altruism. In their experiments, they show that participants care more about total payoffs than just their own payoff. They show that these players are ready to take cuts in their pay-out in order to give more money to the other player. Thus, a participant in the destruction treatment should be more inclined to accept a lower offer in order to guarantee that no money is destroyed.

Frank's (1998) study predictions are also based on social welfare preferences. In his study, Frank (1998) starts out with a simple fairness model, inspired by Bethwaite and Tompkinson (1996), where the responder's utility (U_R) is derived from a mix between her payoff (P_R) and a measure of perceived unfairness ($c \geq 0$, and $\Delta = 0$ if the proposal is rejected):

$$U_R = P_R - C * \Delta \qquad (2)$$

This model is similar to the Fehr and Schmidt (1999) model I described earlier. It implies that a participant will reject an offer if their disutility from unfairness is greater than the utility from their payoff. Frank (1998), then considers the possibility that the participant could gain some utility from the experimenter "receiving" (or keeping) the money in case

of rejection. He calls this the concern for the experimenter's welfare (CEW) hypothesis. The new CEW model (see below), includes a new factor called P_{exp}, which is the "payoff" of the experimenter. The factor c represents a weight for the concern for the experimenter's welfare. This new model is thus:

$$U_R = P_R + c * P_{exp} \quad (3)$$

In the control treatment, all the money goes back to the experimenter in case of rejection. However, in the destruction treatment, $P_{exp} = 0$, since the money is destroyed if the proposal is rejected. Frank expects that rejection rates will be lower in the destruction treatment compared to the control because the participant does not receive any extra utility from rejecting in the destruction treatment (since $P_{exp} = 0$).

Fairness concerns have been widely accepted as one of the leading explanatory variables in describing behaviour in experimental economics. In the ultimatum game, outside options have been shown to lower rejection rates, by diminishing these fairness concerns (Schmitt, 2004). The social utility model, similar to Fehr and Schmidt (1999) and Bethwaite and Tompkinson (1996), is based on two factors, the interpersonal and the interpersonal. Non-zero outside options have also been shown to shift participant focus from the interpersonal to the intrapersonal, lowering the minimum acceptable offer for responders. Although I am dealing with non-monetary outside options, I believe that this research on outside options strengthens the idea that rejection rates will be lower in the destruction treatment than in the control treatment in my experiment.

4.2.2. More unfair proposals

I now turn to proposer behaviour. I expect proposers in the destruction treatment, compared to the control treatment, mainly through the channel of backward induction, to offer more unfair proposals. Proposers should understand that rejection rates will decrease for unfair proposals, and thus in order to maximise their pay-outs, send lower proposals. Research shows that the players with more bargaining power tend to have higher demands (Hennig-Schmidt, Irlenbusch, Rilke, & Walkowitz, 2018), which strengthens this idea of proposers offering lower amounts.

4.2.3. Additional considerations

It is possible that I will not be able to reject the hypotheses that proposals and rejection rates are equal in the first treatment as in the control treatment. There is a possibility that telling players that the total sum of money will be destroyed in case of rejection may not have any effect on offers and rejections. A lack of change in behaviour could be due to several factors.

One factor to be considered is that players may not believe or realize the gravity of destroying money and its implications. For example, it may be hard to believe as a subject

that experimenters would destroy money for experimental purposes. If a responder does not believe the threat then they would consider that the money will not be wasted, and that it will instead go back to the experimenters, and thus would play as under normal circumstances. However, this should be accounted for by assuring the subjects that the experiment does not involve deception.

Another factor that could play a role is that players may still value inequity aversion far higher than efficiency even when a rejection literally implies burning money. Theory suggests fairness is the most important aspect in ultimatum game decision-making for responders (Güth & Kocher, 2014).

4.3. Charity treatment

Participants assigned to treatment 2 were told that, in the case of rejection, the 1000 SEK would be donated to a charity organization. As explained in the design section, I test whether the magnitude of proposals and rejection rates for unfair offers are significantly different between treatments. Here, I explain why I believe that the responder and proposer behaviour will be significantly different between the charity treatment and the other treatments. I expect responders to have higher rejection rates of unfair proposals in the charity treatment than the control treatment. Through backward induction, I expect that proposers will increase the size of the proposals that they make.

4.3.1. Higher rejection rates

I expect responders to increase their rate of rejections of unfair proposals. Using Frank's (1998) model, I assume that an outside option of a donation to charity is similar to the outside option of the experimenter receiving the money. Both of these provide positive utility to the participant. I even make the claim that a donation to charity will provide even more utility to participants compared to giving that same amount of money back to the experimenter. This result can be expected due partly to the fact that people derive utility from being altruistic, implying that they increase their rejection rate because they want to donate to charity. Since the experiment is carried out in semi-private setting, in the sense that only the instructor would know about the decision made by the responder, I cannot establish in this experiment whether such altruism is driven by the intrinsic motivation, warm-glow, signalling, or the experimenter demand effect. Examples from the literature would suggest that the warm-glow effect tends to dominate in these settings (Andreoni, Rao & Trachtman, 2017; Ariely, Bracha, & Meier, 2009).

My new outside option model (4), adapted for the charity treatment, includes a new factor called $P_{charity}$, which is the "payoff" of the charity. I keep the other factors from the previous models, including the measure of perceived unfairness (Δ), and the concern for the experimenter's welfare (P_{exp}). The factor $c_{charity}$ represents a weight for the concern for the charity's welfare. This model represents my predictions that the utility of the

responder depends on a combination of unfairness concerns, concern for the experimenter's welfare, and concern for the charity's welfare. This new model is thus:

$$U_R = P_R + c_\Delta * \Delta + c_{exp} * P_{exp} + c_{charity} * P_{charity} \quad (4)$$

It is also important to note that the charity effect could be magnified due to the fact that an additional 1000 SEK represents a relatively small amount of money for most participants. On the other hand, I assume that people believe that charities would derive much more utility from the same amount.

The choice of charity is naturally an important factor as well. Donations to charity depend very much on the problem that each charity is tackling and how well that charity is perceived. Misallocation of funds or other forms of corruption can have an impact on the image of the charity itself as well as of the whole sector (Sargeant, Ford & West, 2000; Gaskin, 1999). A relevant Swedish example would be the trial and sentencing of Johan af Donner, who swindled more than a million dollars from both the Swedish Red Cross and the Swedish Cancer Society ("Red Cross Swindler Gets Five Years", 2010). In my pilot study, concerns were raised by participants about my choice of the Swedish Red Cross due to this incident. This led me to change my choice of charity to Save the Children Sweden (Rädda Barnen) because I wanted to have a charity that for one, participants would recognise, and two, that they would have a positive reaction to.

An increase in responders' rejection rate might also arise from a magnification of the spite effect (Pillutla & Murninhan, 2016). Responders then decide to reject offers that are close to fair (around 40%) more often than in control group because the slightly selfish behaviour of the proposer is perceived as even more unfair in this setting due to the presence of the charitable alternative, or that the spite is perceived as more acceptable due to being 'rewarded' by the additional payment to charity.

4.3.2. Higher proposals

For proposers, once again using the backward induction theory, and following the assumption that most players are payoff-maximising, I expect proposers to offer fairer proposals, which would be higher than the proposals in the control treatment.

I expect proposers to increase the amount offered because they expect the responders to have a higher rejection rate of unfair offers. This is supported by the backward induction argument suggested by Thaler (1988).

4.3.3. Additional considerations

The existence of the outside option might trigger some comparison mechanism that pushes players to make more rational decisions, hence accepting lower offers more frequently. This result might be comparable to Andersen, Ertaç, Gneezy, Hoffman, and List (2011), who found evidence that higher stakes increased rationality. Alternatively, it

could be that responders derive disutility from donating to charity for some reason, and hence would avoid such an outcome more often. However, I believe that while such reasoning could apply to a few outliers, it is unlikely in the general population.

It might also be possible that some participants derive a higher utility from giving to charity rather than splitting the pie fairly and receiving a direct payment. This would confirm the standard derivation of utility from prosocial contributions (Camerer, 2003).

Another important factor to note is the name of the chosen charity in this study. Supposedly, the name of the charity including "Children" could strongly influence the decision-making of the subjects in the treatment group.

4.4. Further limitations

The study that I plan to carry out contains some limitations.

The destruction treatment requires that participants believe the threat to destroy the money in the case of rejection, and while I can provide documentation that supports that threat, it may still be disbelieved by some subjects. If this is the case, the treatment will function similarly to the control portion of the experiment for those subjects, leading to untrustworthy and biased results.

The charity treatment may also not have the desired effect of operating as a 'positive efficiency' outside option: some subjects may not have heard of the charity I chose, and that unfamiliarity may influence their decision, even after they have been briefed on what the charity does. Other students may even think negatively of the charity I chose for various reasons (such as suspicions of inefficient charitable giving or disliking the cause the charity is working towards).

A final possible important internal limitation to note is that of experimental power. To make sure that the tests can correctly reject the null hypotheses, I have chosen relatively large sample sizes (>100) per treatment group. I hope that this will help avoid type M and type S errors (Gelman & Carlin, 2014). As in Frank (1998), it could still be the case that the effect sizes of the treatments are not large enough to give significant results. Low power due to a low effect size could be a fundamental issue in the experimental design, excluding the limitations discussed above.

Amir, Rand, and Kobi Gal (2012) find some significant effects in the basic ultimatum game when dealing with 1 USD stakes online, and with the stakes being an order of magnitude higher, I expect that the size of the stakes will not be an issue for this experiment.

5. Results and analysis

In this section, I show the results of the experiment, including descriptive statistics and statistical tests of the main hypotheses that I am interested in, as specified in the pre-analysis plan.

In summary, the results show that both treatments have a significant impact on rejection rates for the responders. As predicted, the charity treatment increases rejection rates, while the destruction treatment lowers rejection rates. However, I am unable to reject the null hypothesis that proposer behaviour is equal in all treatment groups.

5.1. Summary statistics

In total 307 people participated in the experiment. There was a grand total of 460 responses, but 153 of those were incomplete submissions. Participants had to click through 5 different pages and could only continue to the next page once they had answered every question on the previous page. As previously mentioned, I adjusted the exclusion criteria, choosing to only keep complete responses. Given that I made this change to the pre-registration analysis plan, I have added the results of the experiment with the incomplete observations added, which can be seen in the robustness check at the end of this section. This robustness check shows that keeping these additional observations or not does not change the results and implication of the experiment.

Table 1 provides summary statistics for the control variables, divided by treatment groups, as well as ANOVA statistics for differences between treatment groups. Mutz et al. (2019) believe that balance tables should be avoided in experimental studies where there is a clean randomization process. However, in this experiment, the number of observations differs greatly between treatments, especially with regards to the charity treatment. The randomization in this study was performed independently by the Qualtrics software. The results shown in Table 1 imply that the participants were randomly distributed into treatment groups and are thus an indication of balanced control variables. The results of the ANOVA tests can be found in the appendix (Table A.3.6.).

Table 1. Summary statistics: Control variables

	Control	Destruction	Charity	ANOVA
Age	22.721	23.103	22.545	1.24
	(2.458)	(2.677)	(2.521)	(0.291)
Sex	1.369	1.403	1.442	0.49
(Male=1,	(0.484)	(0.492)	(0.500)	(0.611)
Female=2)				
Observations	111	117	77	307

In column (1)-(3) mean estimates, standard deviations in parenthesis.
In column (4) F-statistics with p-values from one-way ANOVA in parentheses.

As per the pre-registration plan and the instructions of the experiment, 6 participants were randomly chosen by a Professor at the Stockholm School of Economics (Anna Dreber Almenberg). One pair was chosen in each treatment group. The results were the following:

- All of the randomly chosen participants succeeded in reaching an agreement with regards to splitting the 'pie'.
- Two pairs agreed on a "500/500" deal, while the third group agreed on a "600/400" deal in favour of the proposer. No money was therefore either destroyed or donated to charity.
- The most frequent Proposal was "500/500", an equal split.

In total, only 153 (33%) out of 460 participants did not complete the survey. Only 26 (17%) of those 153 were randomly provided a treatment group. 13 (50%) out of 26 filled in answers to the survey except for their personal details on the last page which included name, email address, telephone number, and age. Since the majority of the participants were not sorted into one of the treatment groups then I can safely assume that there was not a causal relationship between a given treatment group and the completion rate and thus that there is no attrition problem in this study.

I also analysed whether the completion rate for the 26 individuals who proceeded further into the survey (although without finishing it completely) is similar between the treatment groups. 7 (6%) out of 118 dropped out in the control group. 10 (8%) out of 129 dropped out in the destruction treatment. 9 (10%) out of 86 did not complete the experiment in the charity treatment. Table 2 shows some descriptive statistics of the frequency of participation.

Table 2. Frequency of participation by treatment group

Medium	Control	Destruction	Charity	Total
Complete	111	119	77	307
Incomplete (84%)	4	6	3	13
Incomplete (>71%)	7	10	9	26
Incomplete (<71%)	n/a	n/a	n/a	127
Total	n/a	n/a	n/a	460

84% completion implies participants who answered the main experimental questions but did not fill in their personal details. 71% completion implies that the participant was sorted into one of the treatment groups, but only answered the two control questions. Below 71% completion implies that the participant only saw the introductory pages.

5.1.1. Proposals

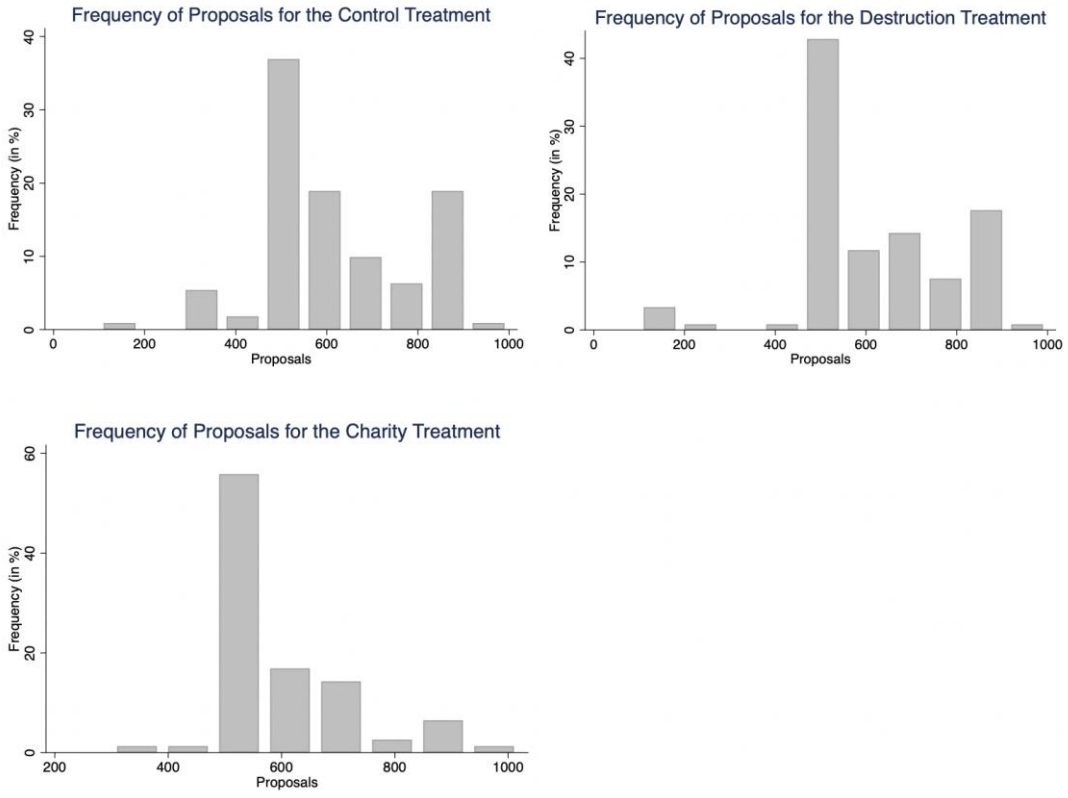
First, some quick observations can be made from the data collected from the experiment. Table 3 includes descriptive statistics such as the frequency of the different proposals across treatment groups. Figure 3 shows the same results in an alternate, slightly more intuitive, design.

Table 3. Summary statistics: Proposals

	Observations	Mean	Min	Max
Control	111	621.62	100	1000
		(182.61)		
Destruction	119	621.01	100	1000
		(189.50)		
Charity	77	581.81	300	1000
		(132.51)		

Mean estimates, standard deviations in parenthesis

Figure 3. Distribution of proposals across treatments



A couple of observations can be made right off the bat. Firstly, the total sample size is over 300 which is the minimum sample size I was looking for. Secondly, there is a clear difference in sample sizes between treatment groups. On one hand, two treatment groups, the control group and the destruction group, have sample sizes above 110; the control group has 111 observations (36.16%) and the destruction group has 119 (38.76%) On the other hand, the charity treatment only has 77 (25.08%) participants. This difference in treatment group size, in particular the much lower size of the charity treatment is not due to attrition, but due to the randomization software built into Qualtrics.

With regards to proposer behaviour, the control treatment is very interesting. There is a clear double peak in the control group and destruction group. As the literature on the ultimatum game predicts, most participants offer proposals of 500/500. What is unusual is the second peak at 900/100 whereby the proposer would receive 900 and the responder 100. The 900/100 offer is interesting because it is considered to be the game-theoretic optimal play according to the classical understanding of game theory where players only derive utility from monetary payoffs. This theory implies, given that responders are rational, that they should accept any offer greater than 0. Through backward induction a proposer should thus offer the lowest amount of money that the responder would be willing to accept. In this experiment, this proposal is 900/100. It is uncommon to see such a high number of participants choosing this proposal of 900/100 (19% and 18% respectively for the control and destruction treatments) according to previous literature.

Although the charity treatment looks different to the other treatments at a first glance, the Kruskal-Wallis H-test fails to reject the null hypothesis ($p>0.05$, H-test) that the proposals are equal between treatments (See table A.2.1.). In the pre-registration plan I make it clear that if the Kruskal-Wallis H-test fails to reject the hypothesis then I would not continue with the pairwise Mann-Whitney U-tests comparing each of the treatment groups. I will discuss the effect size in the Discussion section as well as why the Kruskal-Wallis H-test did not go as I expected.

I will also discuss whether this result could potentially be a false null. The results do show a slight movement of proposals in the direction I had predicted. The proposals are extremely similar in the destruction group compared to the control group, while the charity group's proposals show a sign of moving in the hypothesized direction. Regardless, without a significant difference in the Kruskal-Wallis H-test I cannot perform pairwise Mann-Whitney U-tests according to the pre-analysis plan.

5.1.2. Rejection rates

In terms of rejection rates, table 4 shows some summary statistics of acceptance rates of unfair proposals including the mean acceptance rate for each of the treatment groups for

proposals equal or lower than 600/400. To calculate the acceptance rate, I counted the ratio of unfair proposals that each participant accepted.

Table 4. Summary statistics: acceptance rates of unfair proposals

	Observations	Mean	Min	Max
Control	111	62.16	0	100
		(28.58)		
Destruction	119	71.26	0	100
		(28.24)		
Charity	77	41.04	0	100
		(33.38)		

Acceptance rates in percent of unfair proposals. Unfair proposals are defined as proposals under 500/500, that is, 600/400, 700/300, 800/200, 900/100, and 1000/0. Mean estimates, standard deviations in parenthesis.

In terms of responder behaviour, table 5 shows the acceptance rates for the different proposals for each of the different treatment groups. In total the most commonly accepted proposal is 500/500, which is accepted 98,70% of the time in aggregate. It seems intuitive that the equal split is the most accepted. The least accepted proposal is the 1000/0 proposal which implies that the responder would receive 0 SEK if they accept. This proposal is accepted 17% of the time in aggregate. Again, the pay-off maximising game-theoretic optimal solution is to accept all offers if utility is solely derived from monetary gain, and the 900/100 proposal is accepted in aggregate 55% of the time. This is higher than what results of previous ultimatum game experiments have found. I will go deeper into the implications of this in the discussion section below, but this seems to imply that the participants in my experiment are playing more game-theoretic optimally - according to the payoff maximization model - than in other studies of the ultimatum game.

Another interesting observation is that there is a monotonic decline in acceptance rates from the peak at 500/500 (100% acceptance rate) to 1000/0 (11.71% acceptance rate) for the control group. This is relatively similar in all of the treatment groups. At the same time, there are also rejections of so-called "super-fair" proposals, where the responder would receive more money than the proposer. This is a commonly found result and is not irrational in the complete sense of the word since it can reflect a sense of inequity aversion. This is typically called non-monotonic behaviour and indicates a strong aversion from these participants against more or less biased payoffs even when they receive more than the other participant. A monotonic strategy on the other hand means accepting all proposals greater than or equal to the participants minimum acceptance level.

Table 5. Acceptance rates of all proposals across treatments

	Control	Destruction	Charity	Total
0/1000	94,59%	92,44%	84,42%	91,21%
100/900	99,10%	97,48%	90,91%	96,42%
200/800	100.00%	98,32%	90,91%	97,07%
300/700	100,00%	98,32%	94,81%	98,05%
400/600	100,00%	99,16%	93,51%	98,05%
500/500	100,00%	100,00%	94,81%	98,70%
600/400	94,59%	94,96%	76,62%	90,23%
700/300	83,78%	88,24%	53,25%	77,85%
800/200	63,96%	77,31%	35,06%	61,89%
900/100	56,76%	68,91%	29,87%	54,72%
1000/0	11,71%	26,89%	10,39%	17,26%
Total	82,23%	85,64%	68,60%	80,13%

Acceptance rates in percent of all proposals.

Using the same data, Figure 4 shows in graph form how responders reacted to the different proposals. There is clearly an overall lower acceptance rate for the charity treatment than for the others, as well as a sharper drop after the 500/500 proposal.

Figure 4. Acceptance rates for all proposals across treatments

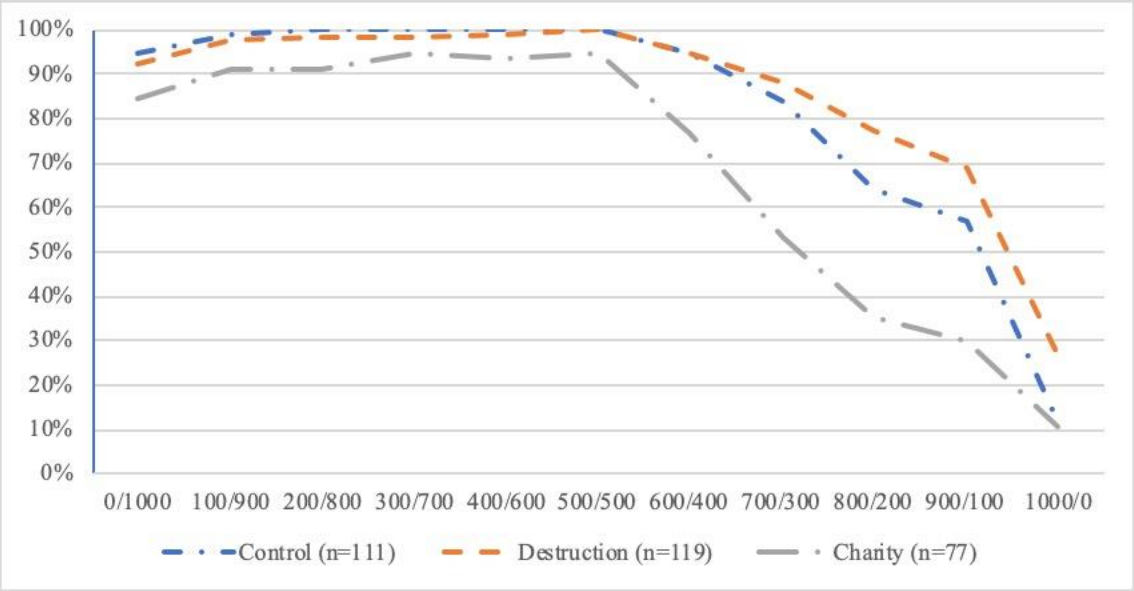


Table 6. Summary statistics: frequency of rejection rates of unfair proposals (in %)

	Control	Destruction	Charity
0	5.41	5.04	23.38
20	10.81	6.72	20.78
40	19.82	10.08	19.48
60	7.21	9.24	7.79
80	45.05	42.86	20.78
100	11.71	26.05	7.79

Frequency of each rejection rate of unfair proposals for each treatment group. 0 implies that that 0 out of 5 unfair proposals were accepted. 40 implies that 2 out of 5 unfair proposals were rejected.

I perform a Kruskal-Wallis H-test to first test whether the rejection rates for unfair proposals are significantly different from one another. As can be seen in table A.2.1. the test is significant at a 5% and even at a 1% level with a p-value of 0.0001 ($p<0.05$, H-test). This means that I can reject the null hypothesis that the three treatment groups are equal and thus proceed with the pairwise Mann-Whitney U-tests.

I then proceed with pairwise Mann-Whitney U-tests to test whether the rejection rates for unfair proposals in the treatment groups are individually different from one another. The tables showing the exact results of the tests can be found in the Appendix. The Mann-Whitney U-test for between the control and destruction treatment yields a p-value of 0.0065 (table A.2.2.). The Mann-Whitney U-test for between the destruction and charity treatment yields a p-value of 0.0000 (table A.2.2.). The Mann-Whitney U-test for between the control and charity treatment yields a p-value of 0.0000 (table A.2.2.). All of the tests reject the null hypothesis at a 1% level.

This implies that all of the treatment groups have significantly different rejection rates for unfair proposals. This would mean that there is a significant effect of non-monetary outside options on responder behaviour. There is also a clear movement in the direction I predicted.

5.1.3. Optimal play

In the ultimatum game, given a set responder behaviour, proposers can find a profit-maximizing proposal. In table 7, I show the profit maximizing proposal across treatments. It is interesting to note the difference between the actual mean proposals by the participants compared to the optimal play if they only cared about their own monetary payoff. The optimal proposal given expected payoffs is 700/300 for the control group, 900/100 for the destruction group, 500/500 for the charity group, and 700/300 on average across all treatment groups. However, it is important to note that the difference in expected payoff is minimal between certain proposals. For example, the expected payoff

is virtually the same for the 700/300, 800/200, and 900/100 proposals in the destruction treatment.

According to many studies, such as Andreoni and Miller (2002) and Fehr and Schmidt (1999), as seen in the literature review, not all participants are purely profit maximizing. This is reflected in the fact that the actual proposals are not significantly different from one another across treatment groups. As seen in table 3, the average proposal are 621.6, 621.0, and 581.8, respectively for the control, destruction, and charity treatments. The only average proposal that seems to be far apart from the profit-maximizing proposal that the participants could have offered, seems to be the charity treatment.

Table 7. Expected proposer payoff given rejection rates across treatments

	Control	Destruction	Charity	Total
0/1000	0	0	0	0
100/900	99.1	97.5	90.0	96.4
200/800	200.0	196.6	181.8	194.1
300/700	300.0	295.0	284.4	294.1
400/600	400.0	396.6	374.0	382.2
500/500	500.0	500.0	**474.0**	493.5
600/400	567.6	569.7	459.7	541.4
700/300	**586.5**	617.6	372.7	**545.0**
800/200	511.7	618.5	280.5	495.1
900/100	510.8	**620.2**	268.8	492.5
1000/0	117.1	268.9	103.9	172.6

Expected payoffs for proposer (in SEK) given responder rejection rates across treatments. The profit-maximizing proposal for each group is marked in bold.

5.1.4. Exploratory analysis

In the pre-analysis plan, I specify that I will perform chi-square tests to compare the rejection rates per individual offer across treatments.

Table 8 shows these results in a summarized format. In the pre-registration I did not specify what results I expected to find. These findings are thus purely exploratory and require further research.

Table 8. Chi-square pairwise comparisons of rejection rates for each proposal

	Control + Destruction	Destruction + Charity	Control + Charity
0/1000	0.508	0.076	**0.020**
100/900	0.348	**0.041**	**0.006**
200/800	0.170	**0.015**	**0.001**
300/700	0.170	0.163	**0.015**
400/600	0.333	**0.025**	**0.007**
500/500	1.000	**0.012**	**0.015**
600/400	0.901	**0.000**	**0.000**
700/300	0.330	**0.000**	**0.000**
800/200	**0.026**	**0.000**	**0.000**
900/100	0.056	**0.000**	**0.000**
1000/0	**0.004**	**0.005**	0.777

Cells in bold are cells whose p-values are under 0.05 for each $\chi 2$ test

There are 3 conditions that must be met in order for a chi-square test of independence to be appropriate:

1. The sampling method is simple random sampling
2. The variables are categorical
3. The expected frequency count for each cell of the table is at least 5

The 2 first conditions are fulfilled since the participants were randomly allocated to each treatment group and the chosen variables are categorical. However, the 3rd condition is not met for some of the proposals. Table 9 shows that the frequency of rejections are nearly all under 5 for the 500/500, 400/600, 300/700, 200/800 and 100/900 proposals. All the other proposals, including the 0/1000, can thus be appropriately analysed using the chi-squared test of independence.

Table 9. Frequency of rejections across treatments

	0/1000	100/900	200/800	300/700	400/600	500/500	600/400	700/300	800/200	900/100	1000/0
Control	6	1	0	0	0	0	6	18	40	48	98
Destruction	9	3	2	2	1	0	6	14	27	37	87
Charity	12	7	7	4	5	4	18	36	50	54	69

To demonstrate how misleading the chi-squared test can be, I take the chi-squared test of the 500/500 proposal comparing the control and the charity treatment. I see that there are 0 rejections of 500/500 proposals for the control treatment, while there are 4 rejections for the charity treatment. Although the p-value of the chi-squared test is 0.015, which should imply that there is a significant difference between the two treatments, this result is not appropriate given that the frequency of rejections was under 5 for both of the treatments.

Taking this limitation of the exploratory analysis into account, I have the following findings (5% significance level) as can be seen in table A.2.3.:

- I find a significant difference between control and destruction for the 800/200 ($p<0.05$, $\chi 2$ test) and 1000/0 proposals ($p<0.05$, $\chi 2$ test).
- There is also a high significant difference in rejections between the destruction and charity treatments for all proposals equal to or under 600/400 ($p<0.05$ for each proposal, $\chi 2$ test).
- There is similarly a high significant difference in rejections between the control and charity treatments for the proposals 600/400, 700/300, 800/200, 900/100 ($p<0.05$, $\chi 2$ test). However, I fail to reject that the rejections are equal for the 1000/0 proposal. I can, on the other hand, claim that there is a significant difference for the 0/1000 proposal. This is something that I did not expect to find. It would imply that the participants derive a higher utility from giving all the money to charity than keeping it for themselves.
- I discard all chi-squared calculations for the "super-fair" proposals where the frequencies are lower than 5 since the chi-squared test is inappropriate for judging the statistical significance of these differences. However, it is possible that an appropriate significant result could be found with a higher sample size. It would be interesting for further research to analyse whether this is only noise or actually proof of utility preferences for participants.

5.1.5. Robustness check

Given my change of exclusion criteria from having chosen not to exclude any observations in the pre-registration plan, to excluding incomplete observations, I have to check the robustness of the results of the previous subsections. The exclusion of observations could have caused the observed differences between treatment groups. As it turns out, responder behaviour still differs between treatment groups ($p<0.05$, H-test), and proposer behaviour still cannot be significantly distinguished between treatment groups. I will show the results of these tests below. The tests themselves can be found in the Appendix.

In my main analysis I chose to exclude observations based on two criteria: non-completion and multiple participation. No observations had multiple participation based on the names and email addresses provided in the survey. However, there were many participants who did not complete the experiment: 153 participants, to be precise. As seen in table 2, out of those 153, 26 started the section that included the main experiment instructions. Only 13 of these 26 actually answered the whole section, including their choices acting as proposer and responder. The other 13 only answered the two control questions. I have chosen to include the 13 observations that answered the experimental questions in this robustness check so as to make sure that the changes that I made with regards to the exclusion criteria did not impact the validity of my experiment.

Of these 13 observations, 4 were in the control group, 6 were in the destruction group, and 3 were in the charity group. The total treatment group sizes thus became 115, 125, and 80 respectively for the control, destruction, and charity treatments.

The Kruskal-Wallis H-test for differences between the proposals by treatment group gave a p-value of 0.0971 (table A.3.2), which is still insignificant at the 5% level, implying that I cannot reject the hypothesis that the proposals are equal between treatment groups (This test gave a p-value of 0.1471, see table A.2.1., with the chosen exclusion criteria). Since the test is insignificant at the 5% level, I do not perform the Mann-Whitney U-tests for pairwise comparisons between treatments.

The Kruskal-Wallis H-test for differences between the responder rejection rates for unfair proposals gives a p-value of 0.0001 (table A.3.2.), which is in line with our previous results. It also implies that I reject the null hypothesis at a 5% level that the responder behaviour is equal between treatment groups. Since this test is significant at the 5% level, I perform Mann-Whitney U-tests for pairwise comparisons between treatments.

These Mann-Whitney U-tests yield p-values of 0.0119 (Control and Destruction, table A.3.4.), 0.0000 (Control and Charity, table A.3.4.), and 0.0000 (Destruction and Charity, table A.3.4.). All of these tests show that there is a significant difference at the 5% level between treatment groups for rejection rates of unfair proposals. This is also in line with the previous findings, where I could also reject all the null hypotheses at the 5% level.

Finally, I also check the robustness of the explanatory analysis tests. I find the same results when including the incomplete observations than in the previous tests without them.

5.2. Analysis

These results allow me to conclude that non-monetary outside options have an impact on participant behaviour in the ultimatum game. I claim that an outside option with positive effects for society, through a donation to a charity, significantly increases rejection rates of unfair proposals. Additionally, I can conclude that an outside option with negative effects on social welfare, the destruction of money, significantly decreases rejection rates of unfair proposals. However, I cannot conclude that non-monetary outside options have an impact on proposer behaviour.

This result is very relevant for the literature on the ultimatum game. It is true that I have based my assumptions and predictions on previous research, but nobody has put this research together to show that concern for the experimenter's welfare actually exists in an ultimatum game setting. This simple idea that there is a concern for the experimenter's welfare can have a far-reaching impact. Previously it was believed that the main considerations of players are a mix of their own monetary payoff as well as fairness concerns (the interpersonal factor). The fact that players change their actions based on

where the money will go in case of rejection would imply that a supplementary factor should be added to the standard utility function of responders in the ultimatum game: that of concern for the experimenter in a classic ultimatum game.

Also, while I was not able to reject the null hypothesis that proposals were equal between all treatment groups, there might still be a possibility that this is mainly due to a smaller effect size and insufficient power.

In the following discussion I will dive deeper into whether and why my predictions were in line with the results.

6. Discussion

This study contributes to the literature on the ultimatum game by studying whether non-monetary outside options matter and their possible effects. I seek to understand whether participants actually care about more than just their own and the other player's monetary payoff. I break new ground in ultimatum game research by creating an experiment that studies both the effect of positive and negative non-monetary outside options.

My study adds to experimental economics research by showing that players have a concern for the experimenter's welfare in a standard ultimatum game. The fact that responders' rejection rates change depending on who receives the rejected money implies that social welfare concerns towards the experimenter exists. A secondary finding highlights that non-monetary outside options do not have a significant impact on proposer behaviour, and I will develop why this finding went against my predictions.

In light of my results, I first discuss the theoretical implications of my findings on responder behaviour and rejection rates, with an added discussion on the exploratory analysis. I then analyse in particular the case of proposals and look at the explanations behind the null result while also considering the possibility that it could be a false null. Lastly, I acknowledge limitations to both the internal and external validity of the study.

6.1. Theoretical implications

My results are clearly surprising in more ways than one. Previous research has pointed to a wide range of reasons why people act the way they do. The current leading hypotheses to why participants reject positive proposals are inequity aversion, fairness considerations, reciprocity, non-zero outside options, the joy of destruction, pride/spite, stakes, the experimenter demand effect, and efficiency. These explanations are still insufficient to completely understand behaviour in the ultimatum game. Nobody, except Frank (1998), thought of the possible impact of concern for the experimenter's welfare (CEW). Unfortunately, Frank's results were unable to prove his theory. 21 years later, my experiment shows that Frank was on the right track. As a first in experimental economics, I show that adding a non-monetary outside option actually has an impact on responder behaviour.

I show that the average rejection rate is lower in the destruction treatment compared to the control and charity treatment according to a Mann-Whitney U-test at a 1% level. Additionally, the charity treatment is significantly different from the control treatment at a 1% level from both the other treatments, also according to Mann-Whitney U-test. The fact that there is a difference between participants in these treatment groups when the only change that I made was mentioning that the money would be destroyed or donated

to charity, implies that participants care about this factor. I derive the interpretation of these results from research on altruism and efficiency as well as social welfare concerns.

Why do people care about the welfare of the experimenter? As mentioned in section 4, social welfare preferences are a strong contender to explain this effect. Players have been shown to choose lower personal pay-outs in order to increase the pay-outs of other players (Charness & Rabin, 2002). If players in ultimatum games can be virtuous enough to provide a higher payoff to the other player, then it is within reason to suggest that they could act upon similar feelings towards society by changing their behaviour. I also believe that outside options, both monetary and non-monetary, provide a reference point to participants, thus shifting their focus from the unfair interpersonal proposal to an intrapersonal comparison. When the problem is framed in this intrapersonal way, players are more likely to change their rejection rate of unfair proposals.

The strongest support for the shift of responder behaviour in the charity treatment is that participants provide more utility from giving to charity than they do giving a share of that money to the other player and keeping only a small share for themselves. This increase in rejection rates can also depend on the spite effect being magnified (Pillutla & Murnighan, 2016). The unfair proposals will be deemed more unfair due to the outside option being perceived as kind.

In order to answer the main research question, I conclude that my study shows that a concern for the experimenter's welfare plays a central role on the acceptance rate of unfair proposals of responders in the ultimatum game. Despite this, I do not reject all the previous research. I regard the ultimatum game, like human behaviour in general, as being dependent on a wide range of factors. Concern for the experimenter's welfare would just be an additional factor to add to the list. Further research would be needed to show how strong this effect actually is compared to all the others and whether it potentially plays one of the bigger roles as Frank suggests.

6.2. Interpreting the null result

The most surprising result is that I am unable to reject the null hypothesis that proposer behaviour is the same between all treatments. My prediction was based on the assumption that the rejection rates would be significantly different from one another and that through backward induction, the proposers would also shift their proposals in order to maximise their payoff. Due to heterogeneity concerns, I argue that payoff-maximization is not as vital as one could imagine, playing a smaller role than I imagined in my prediction for proposer behaviour. I claim that the rigidity of proposer behaviour can mainly be explained by population heterogeneity and the amalgamation of opposite effects.

Population heterogeneity is one of the main reasons why my results differed from my predictions. Some players care about optimal play and playing as game-theoretically as

possible, while others have stronger fairness concerns. Fehr and Schmidt (1999) talk about this in detail in their model. They use a Beta parameter that parametrizes the distaste of a participant for advantageous inequality. A beta greater than 0.5 implies that the participant would offer an equal split. While a beta of 0 implies that the participant does not care about inequality concerns. They assume a distribution of different levels of beta within a set population that corresponds to the results of their experiment as well as three other contract games. Their suggested distribution is a 60/40 split between, respectively, players of type a ($\alpha=0$, $\beta=0$), and players of type b ($\alpha=2$, $\beta=0.6$). They call this a "simplification of the distribution" in their own words (Fehr & Schmidt, 2010). According to Fehr and Schmidt (2010), there is a substantial majority of actors who would play in a pay-off maximising way in ultimatum games. However, there are also some 40% share of participants who, even if given the chance to increase their payoff at the cost of the other player, would rather offer an equal split. In terms of my experiment, in the destruction treatment the proposers should rationally be able to predict that their own shift in rejection rate would be mirrored by the other participants and thus that they could offer lower proposals in order to win more money. However, only about half of these proposers who understand this will actually act on this information and offer smaller proposals.

Another reason why I did not find a significant difference between proposals of the three treatment groups is that predicting the optimal proposal is not an easy calculation to make. As table 12 shows, there is a very small difference in expected pay-out between the 700/300, 800/200, and 900/100 proposals in the destruction treatment. Similarly, in the charity treatment, it is difficult to predict which proposal would give the biggest expected payoff.

One of the main explanations for the lack of difference between treatments is the concern for the experimenter's welfare. What I failed to take into account was that most proposers would, similarly to responders, due to the concern for the experimenter's welfare, want to avoid the destruction of money. The reasoning is the following. If proposers offer an unfair proposal in the control treatment and it is rejected, then the proposer would derive some utility from the experimenter receiving the money, similarly to the responder. When the experimenter is no longer in the picture, both the responders and the proposers are more likely to accept proposals. Although smaller proposals could theoretically increase the proposer's expected pay-out, it also implies a higher risk of the money being destroyed if the responder rejects. Even pay-off maximising participants seem to be taking this into account.

Similarly, Frank (1998) warns that the effect of the destruction treatment on proposals is hard to determine, even when using his rather straightforward CEW hypothesis. He explains that there are two opposite effects that cancel each other if the CEW hypothesis is true.

> "On the one hand, if the CEW hypothesis were true, they (the proposers) might offer more in the experimental group, thereby increasing the likelihood of acceptance because rejections of low offers no longer save the experimenter's money. On the other hand, they might offer less in the experimental group because they anticipate that the receivers demand less if rejected stakes are burnt. Due to these opposite effects which the CEW would have, the rejection of this hypothesis cannot be based on the comparison of the offers under the two treatments, which on average turn out to be not significantly different."

The fact that there might be opposite effects acting on the proposers in my experiment is not equal to there not being an effect of non-monetary outside options on participants in an ultimatum game setting. Finally, it is possible that this result might be a false null (i.e. type 2 error). The reasons supporting this are laid out below.

Looking at the distribution of proposals between treatment groups, there seems to be a hint that they actually are different. The mean of the destruction is extremely similar to that of the control treatment, but the charity treatment shows a push towards fairer proposals. One relevant observation is the fact that the charity treatment group is much smaller than the other two (77 versus 111 and 119). This leads us to the other main reason why this result may be a false null: the effect size is small, and our sample sizes are too small to show a significant result. Given the seemingly strong differences between the charity treatment and the other two, I would expect to find a significant effect with a slightly bigger sample size. It is unfortunate that the randomization did not equally distribute the participants between all the treatment groups, since it would probably have allowed me to have stronger proof of this being the case.

6.3. Validity

The study contains some limitations. I discuss in this section how the choice of population and the experimental procedure have an effect on the internal and external validity of the study and its findings.

6.3.1. Internal validity

In order to verify that the experiment was conducted correctly, I first check whether the baseline results are in line with the empirical literature on the general ultimatum game. The results show that the control groups rejection rates are relatively similar to what is typically expected in the theoretical literature. I find, in line with classical literature, that responders on average reject about 20% of all proposals, with acceptance rates close or equal to 100% for proposals higher or equal to the equal split. What is slightly different from standard ultimatum game experiments is the acceptance rate of unfair proposals. For example, offers of 900/100 were accepted 57% of the time by responders in the control treatment. Even the 1000/0 proposal has an acceptance rate of 11%. Typically, one would

find unfair proposals (≤800/200) being accepted approximately half the time, according to Fehr and Schmidt (2006). Although these results are slightly more excessive than under normal circumstances, I do not think that they are significant in showing a flaw in my experiment. Instead, I believe that these rejection rate statistics are indicative of the relatively unique nature of the students at the Stockholm School of Economics, who like many other economics students have been taught to change their behaviour towards the payoff-maximizing game-theoretic solution. With regards to standard proposer behaviour, research shows that proposers typically offer 40-50% of the total amount to responders and a small minority of players propose shares under 20% of the pie. Similarly, in this experiment, over 55% of participants in the control treatment propose equal or near equal splits. However, there is a unique second peak of proposals at 900/100 which is uncommon within other studies but is in line with the assumption that this population is one who contains more players that play game-theoretically . Since all participants come from this same population, this does not impact the internal validity of the experiment.

This experiment was done through an online survey which has some clear internal validity considerations that should be taken into account. Typically, online experiment participants are noted for their lack of limited attention as to the choices they are making, which leads to an increase of noise due to a higher probability of a lack of understanding. However, I included a test in the survey in order to verify that the instructions were properly understood by the participants. Due to these two control questions, participants could not just randomly press different alternatives in the survey. The key experiment questions were also asked after the control questions so that they would be better understood.

One factor that could lead to a limited internal validity is that of limited incentives. The pay-outs were decided through a lottery. Although the total sum was 1000 SEK, the ultimatum game typically leads to an equal or near-equal split, which would imply that technically participants selected would receive on average 500 SEK each. Since only 6 winners were selected, this lottery implied that that the expected value of participating was only 10 SEK. Since these incentives are so small, a possible internal validity limitation is that participants would not take the experiment as seriously as they would with higher stakes.

I chose a between subject design instead of within subject design, which I believe strengthens the internal validity of my experiment. The proper randomization into treatment groups becomes central in the case of a between subject design. The distribution into treatment groups was independently done by the Qualtrics software and I had no say whatsoever in controlling which group each participant would be placed in. I cannot see any possible way the randomization was not independent of any other factor. This legitimate randomization together with no clear attrition rate for a specific treatment suggests a stronger internal validity.

Additionally, it is vital that the participants understood that the money would be paid, destroyed, or donated to charity. If participants did not have faith in that I, the experimenter, would actually carry out what I said I was going to do then this could have an impact on the results and the conclusions drawn from the experiment. I am confident that participants truthfully and seriously answered the questions given to them since I made sure to explicitly state that economics experimenters were not allowed to mislead participants in such a way. My exclusion criteria also attempt to increase the amount of serious answers by eliminating the incomplete responses.

Another important possible limitation that should be mentioned is that students are typically close to each other and have a high likelihood of talking about the experiment. This high level of social interaction, coupled with the fact that the experiment was sent to all students' email addresses and completed through an online survey, is likely to have led to some student talking about the experiment among themselves. I have no way of verifying that this was the case, but it is something that I acknowledge as a possible internal limitation to my results.

6.3.2. External validity

Regarding the external validity of the results, I recruited participants that do not necessarily represent the broader population. The sample is sourced entirely from students at SSE, a Swedish university, focused primarily on business and economics studies. While western student bodies are becoming more diverse, the great majority of the sample is still comprised of WEIRD (Western, Educated, Industrialized, Rich and Democratic) subjects, meaning that the external validity of the conclusions is limited (Henrich, Heine, & Norenzayan, 2010).

With regards to the generalisability to other settings, which is a key part of the external validity of a study, it must be noted that the ultimatum game is a very theoretical game, with strict controls and a limited applicability to other settings. It is difficult to identically extend the results to day-to-day situations where ultimatum situations occur. Nonetheless, my result strengthens the idea that people are motivated by other factors than just their own payoff.

The main focus of this study is however not on generalisability to other settings but instead on improving the understanding of ultimatum game experiments and experimental economics. My goal is to shine a light on an area of the ultimatum game that has not been sufficiently studied. I show that researchers have missed a key aspect, which is the importance of non-monetary outside options in decision-making. I show the existence of a new other-regarding preference, here called the concern for experimenter's welfare, that plays a key part in explaining why participants act the way they do. The fact that this new factor exists has a significance for conclusions drawn from previous ultimatum game experiments, as well as other behavioural economics experiments where the participants

have a say in how much money the experimenters could keep for themselves. This study suggests that previous research has overestimated the effects of their treatments on rejection rates. If a concern for the experimenter exists, then researchers who have not included it into their models, have most likely overestimated the effects of their treatments on rejection rates.

I acknowledge that there are limitations to this study. Even with these limitations in mind, I believe that this study still manages to successfully identify a concern for the experimenter's welfare through non-monetary outside options.

7. Concluding remarks

I designed an online experiment, consisting of a variation of the ultimatum game, and recruited 307 subjects from the Stockholm School of Economics in order to investigate the impact of non-monetary outside options. My results suggest that non-monetary outside options have a significant effect on responder behaviour in the ultimatum game. In particular, I show new evidence that the reason why rejection rates of unfair proposals are significantly different between treatments is that there exists a concern for the experimenter's welfare. In a classic ultimatum games, if the proposal is refused, then neither the proposer nor the receiver, collect a pay-out. However, the money does not just disappear. Typically, a third party actually receives the rejected sum of money: the experimenter. My findings suggest that the concern for the experimenter is a significant factor in deciding how responders behave to unfair proposals.

Previous literature suggests that people are driven by efficiency and fairness concerns such as inequity aversion and altruism. However, these models have not been sufficient in completely predicting participant behaviour. As far as I am aware, I am the first to successfully prove a robust link between non-monetary outside options and participant behaviour in the ultimatum game. My findings suggest that this additional factor, concern for the experimenter's welfare, should be taken into account in ultimatum game and behaviour economics research. An additional implication of this study is the importance of outside factors such as the experimenter demand effect in understanding participant behaviour. A caveat to this study, however, is that only responder behaviour is affected, while I do not find any evidence of significant differences between proposer behaviour across treatment groups. The evidence indicates that proposals are not swayed by changes in non-monetary outside options.

As such, these findings have some implications for the literature on the ultimatum game as well as on behavioural economics as a whole. My experiment shows that a concern for the experimenter's welfare exists and that it decreases rejection rates of unfair proposals for the responders. In standard ultimatum games, the rejected money goes back to the experimenters, and my experimental results imply that participants take the experimenter's welfare into account. This implies that there is a high likelihood that previous ultimatum game research has overestimated the effect sizes of their treatments. The concern for the experimenter's welfare increases acceptance rates in my study and should be included as one of many important factors in the research literature.

This study provides several new avenues for further experimental research within behavioural economics. Since Frank's (1998) study 20 years ago, few researchers, if any, have considered the possibility of the concern for the experimenter's welfare hypothesis being correct. The importance of replication of previous studies and experimental designs has been a hot debate in recent times (Camerer et al., 2016). Despite not being a complete

replication of Frank (1998), this study adds to a long list of research that fails to replicate the findings of previous literature (Open Science Collaboration, 2015). In turn, I count on future researchers to test my methodology and attempt to replicate my findings.

Finally, my study shows the need for further research to be done on the concern for the experimenter's welfare and non-monetary outside options. There are many components of this experimenter welfare model that I do not test such as the reasoning behind the choices that participants make. In particular, I suggest that further research focus on replicating my results with regards to the destruction treatment, but with greater power, so as to truly test whether non-monetary outside options have a significant effect on both proposer and responder behaviour. It would also be of great interest to investigate further why these effects exist. One possible avenue of research is to use information asymmetry to isolate responder and proposer behaviour. Another would be to use different outside options, to control for possible confounding effects due to the specifics of the treatments. Since non-monetary outside options have an impact on behaviour, then it would be interesting to study where on the scale one should place the destruction of money and the donation to charity as well as the control treatment.

8. References

Abbink, K., & Sadrieh, A. (2009). The pleasure of being nasty. *Economics Letters, 105*, 306-308.

Abbink, K., & Sadrieh, A. (2010). The moral costs of nastiness. *Economic Inquiry, 49*, 631-633.

Abbink, K., Bolton, G.E., Sadrieh, A., & Tang, F. (2001). Adaptive learning versus punishment in ultimatum bargaining. *Games and Economic Behaviour, 37*, 1-25.

Amir, O., & Rand, D. G. (2012). Economic games on the internet: The effect of $1 stakes. *PloS one*, 7(2), e31461.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., & List, J.A. (2011). Stakes matter in ultimatum games. *American Economic Review, 101*, 3427-3439.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of political Economy*, 97(6), 1447-1458.

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

Andreoni, J. (1995). Warm-glow versus cold-prickle: the effects of positive and negative framing on cooperation in experiments. *Quarterly Journal of Economics, 110*, 1- 21.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: an experimental test of the consistency of preferences for altruism. *Econometrica, 70*, 737-753.

Andreoni, J., Erard, B., & Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature, 36*, 818-860.

Andreoni, J., Rao, J.M., & Trachtman, H. (2017). Avoiding the ask: a field experiment in altruism, empathy, and charitable giving. *Journal of Political Economy, 125*, 625-653.

Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *American Economic Review, 99*, 544-555.

Avery, C., & Zemsky, P.B. (1994). Money burning and multiple equilibria in bargaining. *Games and Economic Behaviour, 7*, 154-168.

Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11(2), 122-133.

Bethwaite, J., & Tompkinson, P. (1996). The ultimatum game and non-selfish utility functions. *Journal of Economic Psychology*, 17(2), 259-271.

Blount, S. & Bazerman, M.H. (1996). The inconsistent evaluation of absolute versus comparative payoffs in labour supply and bargaining. *Journal of Economic Behaviour and Organization, 30*, 227-240.

Bohnet, I., & Frey, B. S. (1999). Social distance and other-regarding behaviour in dictator games: Comment. *American Economic Review*, 89(1), 335-339.

Bolton, G. E. (1991). A comparative model of bargaining: Theory and evidence. The *American Economic Review*, 1096-1136.

Bolton, G.E., & Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *The American Economic Review, 90*, 166-193.

Bolton, G.E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic Behaviour, 10*, 95-121.

Camerer, C. (2003). *Behavioural Game Theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... & Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433-1436.

Camerer, C., & Thaler, R.H. (1995). Anomalies: ultimatums, dictators and manners. *Journal of Economic Perspectives, 9*, 209-219.

Cameron, L. (1999). Raising the stakes in the ultimatum game. *Economic Inquiry, 37*(1), 47-59.

Carter, J. R., & Irons, M. D. (1991). Are economists different, and if so, why?. *Journal of Economic Perspectives*, 5(2), 171-177.

Cartwright, E., & Stepanova, A. (2017). Efficiency in a forced contribution threshold public good game. *International Journal of Game Theory, 46*, 1163-1191.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics, 117*, 817-869.

Cox, J.C. (2004). How to identify trust and reciprocity. *Games and Economic Behaviour, 46*, 260-281.

Dawes, R.M., & Thaler, R.H. (1988). Anomalies: cooperation. *Journal of Economic Perspectives, 2*, 187-197.

Eckel, C.C., & Grossman, P.J. (1996). Altruism in anonymous dictator games. *Games and Economic Behaviour, 16*, 181-191.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*, 980-994.

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, 1, 615-691.

Fehr, E., & Schmidt, K. M. (2010). On inequity aversion: A reply to Binmore and Shaked. *Journal of economic behavior & organization*, 73(1), 101-108.

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

Fehr, E., & Schmidt, K. M. (2005). The economics of fairness, reciprocity and altruism–Experimental evidence and new theories. *Economics*, 20, 59.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters, 71*, 397-404.

Forsythe, R., Horowitz, J.L., Savin, N.E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behaviour, 6*, 347-369.

Frank, B. (1998). Good news for experimenters: subjects do not care about your welfare. *Economics Letters*, 61(2), 171-174.

Frey, B.S. (1997). *Not just for the money: An economic theory of personal motivation*. Cheltenham, UK: Edward Elgar Publishing.

Gaskin, K. (1999). Blurred vision: Public trust in charities. *International Journal of Nonprofit and Voluntary Sector Marketing*, 4(2), 163-178.

Gelman, A., & Carlin, J. (2014). Beyond power calculations: assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science, 9*, 641-651.

Güth, W., & Kocher, M.G. (2014). More than thirty years of ultimatum bargaining experiments: motives, variations, and a survey of the recent literature. *Journal of Economic Behaviour & Organization, 108*, 396-409.

Güth, W., Schmidt, C. & Sutter, M. (2003). Fairness in the mail and opportunism in the internet: a newspaper experiment on ultimatum bargaining. *German Economic Review, 4*, 243-265.

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behaviour & Organization, 3*, 367-388.

Güth, W., & Tietz, R. (1985). *Strategic power versus distributive justice: an experimental analysis of ultimatum bargaining*. Professur für Volkswirtschaftslehre, insbesondere Verhaltensforschung, Fb Wirtschaftswiss. Johann Wolfgang Goethe-Univ..

Haley, K. J., & Fessler, D. M. (2005). Nobody's watching?: Subtle cues affect generosity in an anonymous economic game. *Evolution and Human behaviour*, 26(3), 245-256.

Handgraaf, M.J.J., Van Dijk, E., & De Cremer, D. (2003). Social utility in ultimatum bargaining. *Social Justice Research, 16*, 263-283.

Hennig-Schmidt, H., Irlenbusch, B., Rilke, R.M., & Walkowitz, G. (2017). Asymmetric outside options in ultimatum bargaining: a systematic analysis. *International Journal of Game Theory, 47*, 301-329.

Henrich, J., Heine, S.J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioural and Brain Sciences, 33*, 61-135.

Hoffman, E., McCabe, K., & Smith, V.L. (1996). Social distance and other- regarding behaviour in dictator games. *American Economic Review, 86*, 653-660.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V.L. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behaviour, 7*, 346-380.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of business*, S285-S300.

Kirchsteiger, G. (1994). The role of envy in ultimatum games. *Journal of economic behaviour & organization*, 25(3), 373-389.

Knez, M.J., & Camerer, C.F. (1995). Outside options and social comparison in three-player ultimatum game experiments. *Games and Economic Behaviour, 10*, 65-94.

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260), 583-621.

List, J.A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy, 115*, 482-493.

Masclet, D., Noussair, C., Tucker, S., & Villeval, M. (2003). Monetary and nonmonetary punishment in the voluntary contributions mechanism. *American Economic Review, 93*, 366-380.

Meier, S., & Stutzer, A. (2008). Is volunteering rewarding in itself? *Economica, 75*(297), 39-59.

Mui, V. L. (1995). The economics of envy. *Journal of Economic Behaviour & Organization*, 26(3), 311-336.

Mutz, D. C., Pemantle, R., & Pham, P. (2019). The perils of balance testing in experimental design: Messy analyses of clean data. The American Statistician, 73(1), 32-42.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600-2606.

Ochs, J., & Roth, A.E. (1989). An experimental study of sequential bargaining. *American Economic Review, 79*, 355-384.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.

Ostrom, E. (1999). Coping with the tragedies of the commons. *Annual Review of Political Science, 2*, 493-535.

Parducci, A., & Wedell, D. H. (1986). The category effect with rating scales: Number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: human perception and performance*, 12(4), 496.

Pillutla, M.M., & Murninhan, J.K. (1996). Unfairness, anger, and spite: emotional rejections of ultimatum offers. *Organizational Behaviour and Human Decision Processes, 68*, 208-224.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review, 83*, 1281-1302.

Red Cross Swindler Gets Five Years (26-05-2010), Sveriges Radio (Radio Sweden). Retrieved from https://sverigesradio.se/sida/artikel.aspx?programid=2054&artikel=3727551 (Accessed 2019-04-29)

Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., & Zamir, S. (1991). Bargaining and market behaviour in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. *American Economic Review, 81*, 1068-1095.

Sargeant, A., Ford, J., & West, D. C. (2000). Widening the appeal of charity. *International Journal of Nonprofit and Voluntary Sector Marketing*, 5(4), 318-332.

Schmitt, P.M. (2004). On perceptions of fairness: the role of valuations, outside options, and information in ultimatum games. *Experimental Economics, 7*, 49-73.

Schröder, M., & Sadrieh, A. (2014). The desire to influence others. Paper presented at the *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2014: Evidenzbasierte Wirtschaftspolitik - Session: Attention and Influence*, No. E20-V1.

Selten, R. (1975). Re-examination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory, 4*, 25-55.

Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.) New York: McGraw-Hill.

Slonim, R., & Roth, A.E. (1998). Learning in high stakes ultimatum games: an experiment in the Slovak Republic. *Econometrica, 66*, 569-596.

Straub, P.G., & Murninghan, J.K. (1995). An experimental investigation of ultimatum games: information, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behaviour & Organization, 27,* 345-364.

Thaler, R.M. (1988). Anomalies: the ultimatum game. *Journal of Economic Perspectives, 2,* 195-206.

Tjøtta, S. (2019). More or Less Money? An Experimental Study on Receiving Money. *Journal of Behavioural and Experimental Economics*.

Zhang, L., & Ortmann, A. (2016). Pro-social or anti-social, or both? A within- and between- subjects study of social preferences. *Journal of Behavioural and Experimental Economics, 62,* 23-32.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75-98.

Zizzo, D.J., & Fleming, P. (2011). Can experimental measures of sensitivity to social pressure predict public good contribution? *Economics Letters, 111,* 239-242.

9. Appendix

9.1. Experiment instructions

The instructions below are the full text that each participant would have read when participating in the survey. Extra information and graphs have been added in order to show the exact structure of the Qualtrics survey. A link to the full survey in PDF format is also available on the OSF platform within the pre-registration plan.

Email sent out to every enrolled student at the Stockholm School of Economics:

Need an excuse to take a break from studying?

I am conducting an experiment for my master's thesis at SSE and I am looking for participants. The experiment is straightforward (hopefully fun too) and your help would be greatly appreciated.

The experiment only takes about **5 minutes** and you will have the opportunity to **win up to 1,000 SEK for participating.**

The experiment is anonymous to other participants and is done individually in front of a computer or on your mobile phone.

You can click on the following link in order to participate in the experiment.

https://hhs.qualtrics.com/jfe/form/SV_25GwZu8zUQ47Jxb

Thank you!

Kind regards,

Joakim Karlsson

23002@student.hhs.se

When the participant clicked on the link, they were sent to the Qualtrics website and the survey containing the experiment.

[Page 1]

[Introduction]

Welcome to the Ultimatum Game 1000 Experiment !

You are being asked to participate in an experiment about people's decision-making, conducted as part of a master's thesis work at the Stockholm School of Economics. The experiment will take approximately 5 minutes.

Your decisions in this game will be anonymous, and your identity will be anonymous to the other participants. Please do not discuss this experiment with others after your participation.

[Page 2]

[Consent]

Online Survey Consent Form

Answer a short decision-making study

Purpose of research: To examine decision-making.

What you will do in this research: You will participate in a simple decision-making study and then answer a short survey.

Time required: Participation will take about 5 minutes.

Risks: We believe there are no known risks associated with this research study. Your participation in this study will remain anonymous to other participants.

Benefits: 6 participants will be randomly selected to have the opportunity to receive up to 1,000 SEK depending on the choices they made.

Participation and withdrawal: Your participation in this study is completely voluntary and you can withdraw at any time.

To contact the researcher: This study is being done by Joakim Karlsson from the Stockholm School of Economics. If you have questions about this research, please contact Joakim at 23002@student.hhs.se.

By clicking the link to start the experiment you are indicating that you have read and understood this consent form and agree to participate in this research study.

After Page 2, the randomization software in Qualtrics randomly selects one of the three treatments for each of the participants.

[Page 3]

[Control treatment]

First the Participant is asked to read the following scenario, which explains how this ultimatum game works. They will then be asked to answer two questions in order to verify that they have correctly understood the experiment.

[insert pic]

The scenario is the following:

Two people have to decide how to share a gifted sum of 1,000 SEK. One person, the Proposer, submits a proposal, and the other, the Responder, decides whether to accept or reject the proposal. If the Responder accepts the proposal, then they share the money according to the proposal. However, if the Responder rejects then nobody receives any money.

You will first be asked two questions in order to make certain that you have understood the concept of this experiment and how the game is to be played. Then you will fill out what actions you would take if you were to be selected as Proposer and Responder.

Six participants in this experiment will be randomly selected and then randomly assigned to one of the two roles. Thus, if you are selected, your actions in this experiment will be matched with the actions of another participant. If a pair agrees on the proposal, the money will be awarded to the participants.

An independent third party from the Stockholm School of Economics will certify that the correct actions have been taken. There is a strong norm within Economics that you cannot lie to participants in an experiment.

On March 31st the prize winners will be contacted via email or telephone. Additionally, an email will be sent out to every participant with the results of the experiment with a certification that the correct actions have been taken.

In order to verify that you have understood the game, we would like you to answer the following questions based on the scenarios below.

The participants are now asked to answer two questions that will allow me to make sure that they have correctly understood the game. The participant was not able to continue to the next page if their responses were false. If they pressed the "next" button on the bottom of the page while having chosen a wrong answer then a prompt appeared, reading: "Wrong Answer: Please read the description carefully and try again".

Scenario 1: The Proposer submits a proposal of 800/200, whereby the Proposer would receive 800 and the Responder 200. The Responder rejects this proposal.

What will be the outcome of this scenario?

- 800 SEK for the Proposer, 200 SEK for the Responder
- 600 SEK for the Proposer, 400 SEK for the Responder
- 400 SEK for the Proposer, 600 SEK for the Responder
- 200 SEK for the Proposer, 800 SEK for the Responder
- 0 SEK for the Proposer, 0 SEK for the Responder

Scenario 2: The Proposer submits a proposal of 800/200, whereby the Proposer would receive 800 and the Responder 200. The Responder accepts this proposal.

What will be the outcome of this scenario?

- 800 SEK for the Proposer, 200 SEK for the Responder
- 600 SEK for the Proposer, 400 SEK for the Responder
- 400 SEK for the Proposer, 600 SEK for the Responder
- 200 SEK for the Proposer, 800 SEK for the Responder
- 0 SEK for the Proposer, 0 SEK for the Responder

[Page 4]

[Continuation of the control treatment]

In this section, the participant has to make their choice with regards to the actions they would take, first acting as the proposer, then acting as the responder.

First, the instructions of the experiment are repeated so that the participant can reread them in case they had forgotten them.

The scenario is the following:

Two people have to decide how to share a gifted sum of 1,000 SEK. One person, the Proposer, submits a proposal, and the other, the Responder, decides whether to accept or reject the proposal. If the Responder accepts the proposal, then they share the money according to the proposal. However, if the Responder rejects then nobody receives any money.

Now we ask you to answer the two questions below, regarding what actions you would take if you were to be selected as Proposer and Responder.

Given that I would be selected as the Proposer, I would like to submit the following proposal:

- The Proposer receives 0 (1000 for the Responder)
- The Proposer receives 100 (900 for the Responder)
- The Proposer receives 200 (800 for the Responder)
- The Proposer receives 300 (700 for the Responder)
- The Proposer receives 400 (600 for the Responder)
- The Proposer receives 500 (500 for the Responder)
- The Proposer receives 600 (400 for the Responder)
- The Proposer receives 700 (300 for the Responder)
- The Proposer receives 800 (200 for the Responder)
- The Proposer receives 900 (100 for the Responder)
- The Proposer receives 1000 (0 for the Responder)

Given that I would be selected as the Responder, I would accept the following proposal(s):

- Proposer receives 0, Responder receives 1000 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 100, Responder receives 900 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 200, Responder receives 800 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 300, Responder receives 700 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 400, Responder receives 600 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 500, Responder receives 500 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 600, Responder receives 400 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 700, Responder receives 300 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 800, Responder receives 200 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 900, Responder receives 100 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]
- Proposer receives 1000, Responder receives 1000 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK)]

[Page 3]

[Destruction treatment]

First the Participant is asked to read the following scenario, which explains how this ultimatum game works. They will then be asked to answer two questions in order to verify that they have correctly understood the experiment.

[insert pic]

The scenario is the following:

Two people have to decide how to share a gifted sum of 1,000 SEK. One person, the Proposer, submits a proposal, and the other, the Responder, decides whether to accept or reject the proposal. If the Responder accepts the proposal, then they share the money according to the proposal. However, if the Responder rejects then nobody receives any money. Moreover, if the proposal is rejected then the money, 1,000 SEK, will be destroyed instead of being given to the participants.

You will first be asked two questions in order to make certain that you have understood the concept of this experiment and how the game is to be played. Then you will fill out what actions you would take if you were to be selected as Proposer and Responder.

Six participants in this experiment will be randomly selected and then randomly assigned to one of the two roles. Thus, if you are selected, your actions in this experiment will be matched with the actions of another participant. If a pair agrees on the proposal, the money will be awarded to the participants.

An independent third party from the Stockholm School of Economics will certify that the destruction of the money will be carried out in case of rejection. There is a strong norm within Economics that you cannot lie to participants in an experiment.

On March 31st the prize winners will be contacted via email or telephone. Additionally, an email will be sent out to every participant with the results of the experiment with a certification that the correct actions have been taken.

In order to verify that you have understood the game, we would like you to answer the following questions based on the scenarios below.

The participants are now asked to answer two questions that will allow me to make sure that they have correctly understood the game. The participant was not able to continue to the next page if their responses were false. If they pressed the "next" button on the bottom of the page while having chosen a wrong answer then a prompt appeared, reading: "Wrong Answer: Please read the description carefully and try again".

Scenario 1: The Proposer submits a proposal of 800/200, whereby the Proposer would receive 800 and the Responder 200. The Responder rejects this proposal.

What will be the outcome of this scenario?

- 800 SEK for the Proposer, 200 SEK for the Responder (0 SEK destroyed)
- 600 SEK for the Proposer, 400 SEK for the Responder (0 SEK destroyed)
- 400 SEK for the Proposer, 600 SEK for the Responder (0 SEK destroyed)
- 200 SEK for the Proposer, 800 SEK for the Responder (0 SEK destroyed)
- 0 SEK for the Proposer, 0 SEK for the Responder (1,000 SEK destroyed)

Scenario 2: The Proposer submits a proposal of 800/200, whereby the Proposer would receive 800 and the Responder 200. The Responder accepts this proposal.

What will be the outcome of this scenario?

- 800 SEK for the Proposer, 200 SEK for the Responder (0 SEK destroyed)
- 600 SEK for the Proposer, 400 SEK for the Responder (0 SEK destroyed)
- 400 SEK for the Proposer, 600 SEK for the Responder (0 SEK destroyed)
- 200 SEK for the Proposer, 800 SEK for the Responder (0 SEK destroyed)
- 0 SEK for the Proposer, 0 SEK for the Responder (1,000 SEK destroyed)

[Page 4]

[Continuation of the destruction treatment]

In this section, the participant has to make their choice with regards to the actions they would take, first acting as the proposer, then acting as the responder.

First, the instructions of the experiment are repeated so that the participant can reread them in case they had forgotten them.

The scenario is the following:

Two people have to decide how to share a gifted sum of 1,000 SEK. One person, the Proposer, submits a proposal, and the other, the Responder, decides whether to accept or reject the proposal. If the Responder accepts the proposal, then they share the money according to the proposal. However, if the Responder rejects then nobody receives any money. Moreover, if the proposal is rejected then the money, 1,000 SEK, will be destroyed instead of being given to the participants.

Now we ask you to answer the two questions below, regarding what actions you would take if you were to be selected as Proposer and Responder.

Given that I would be selected as the Proposer, I would like to submit the following proposal:

- The Proposer receives 0 (1000 for the Responder)
- The Proposer receives 100 (900 for the Responder)
- The Proposer receives 200 (800 for the Responder)
- The Proposer receives 300 (700 for the Responder)
- The Proposer receives 400 (600 for the Responder)
- The Proposer receives 500 (500 for the Responder)
- The Proposer receives 600 (400 for the Responder)
- The Proposer receives 700 (300 for the Responder)
- The Proposer receives 800 (200 for the Responder)
- The Proposer receives 900 (100 for the Responder)
- The Proposer receives 1000 (0 for the Responder)

Given that I would be selected as the Responder, I would accept the following proposal(s):

- Proposer receives 0, Responder receives 1000 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 100, Responder receives 900 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 200, Responder receives 800 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 300, Responder receives 700 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 400, Responder receives 600 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 500, Responder receives 500 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 600, Responder receives 400 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 700, Responder receives 300 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 800, Responder receives 200 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 900, Responder receives 100 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]
- Proposer receives 1000, Responder receives 1000 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK, 1,000 SEK is destroyed)]

[Page 3]

[Charity treatment]

First the Participant is asked to read the following scenario, which explains how this ultimatum game works. They will then be asked to answer two questions in order to verify that they have correctly understood the experiment.

[insert pic]

The scenario is the following:

Two people have to decide how to share a gifted sum of 1,000 SEK. One person, the Proposer, submits a proposal, and the other, the Responder, decides whether to accept or reject the proposal. If the Responder accepts the proposal, then they share the money according to the proposal. However, if the Responder rejects then nobody receives any money. Moreover, if the proposal is rejected then the money, 1,000 SEK, will be donated to Save the Children Sweden (Rädda Barnen) instead of being given to the participants.

You will first be asked two questions in order to make certain that you have understood the concept of this experiment and how the game is to be played. Then you will fill out what actions you would take if you were to be selected as Proposer and Responder.

Six participants in this experiment will be randomly selected and then randomly assigned to one of the two roles. Thus, if you are selected, your actions in this experiment will be matched with the actions of another participant. If a pair agrees on the proposal, the money will be awarded to the participants.

An independent third party from the Stockholm School of Economics will certify that the donation to Save the Children Sweden (Rädda Barnen) will be carried out in case of rejection. There is a strong norm within Economics that you cannot lie to participants in an experiment.

On March 31st the prize winners will be contacted via email or telephone. Additionally, an email will be sent out to every participant with the results of the experiment with a certification that the correct actions have been taken.

In order to verify that you have understood the game, we would like you to answer the following questions based on the scenarios below.

The participants are now asked to answer two questions that will allow me to make sure that they have correctly understood the game. The participant was not able to continue to the next page if their responses were false. If they pressed the "next" button on the bottom of the page while having chosen a wrong answer then a prompt appeared, reading: "Wrong Answer: Please read the description carefully and try again".

Scenario 1: The Proposer submits a proposal of 800/200, whereby the Proposer would receive 800 and the Responder 200. The Responder rejects this proposal.

What will be the outcome of this scenario?

- 800 SEK for the Proposer, 200 SEK for the Responder (0 SEK to Save the Children Sweden (Rädda Barnen))
- 600 SEK for the Proposer, 400 SEK for the Responder (0 SEK to Save the Children Sweden (Rädda Barnen))
- 400 SEK for the Proposer, 600 SEK for the Responder (0 SEK to Save the Children Sweden (Rädda Barnen))
- 200 SEK for the Proposer, 800 SEK for the Responder (0 SEK to Save the Children Sweden (Rädda Barnen))
- 0 SEK for the Proposer, 0 SEK for the Responder (1,000 SEK to Save the Children Sweden (Rädda Barnen))

Scenario 2: The Proposer submits a proposal of 800/200, whereby the Proposer would receive 800 and the Responder 200. The Responder accepts this proposal.

What will be the outcome of this scenario?

- 800 SEK for the Proposer, 200 SEK for the Responder (0 SEK to Save the Children Sweden (Rädda Barnen))
- 600 SEK for the Proposer, 400 SEK for the Responder (0 SEK to Save the Children Sweden (Rädda Barnen))
- 400 SEK for the Proposer, 600 SEK for the Responder (0 SEK to Save the Children Sweden (Rädda Barnen))
- 200 SEK for the Proposer, 800 SEK for the Responder (0 SEK to Save the Children Sweden (Rädda Barnen))
- 0 SEK for the Proposer, 0 SEK for the Responder (1,000 SEK to Save the Children Sweden (Rädda Barnen))

[Page 4]

[Continuation of the charity treatment]

In this section, the participant has to make their choice with regards to the actions they would take, first acting as the proposer, then acting as the responder.

First, the instructions of the experiment are repeated so that the participant can reread them in case they had forgotten them.

The scenario is the following:

Two people have to decide how to share a gifted sum of 1,000 SEK. One person, the Proposer, submits a proposal, and the other, the Responder, decides whether to accept or reject the proposal. If the Responder accepts the proposal, then they share the money according to the proposal. However, if the Responder rejects then nobody receives any money. Moreover, if the proposal is rejected then the money, 1,000 SEK, will be donated to Save the Children Sweden (Rädda Barnen) instead of being given to the participants.

Now we ask you to answer the two questions below, regarding what actions you would take if you were to be selected as Proposer and Responder.

Given that I would be selected as the Proposer, I would like to submit the following proposal:

- The Proposer receives 0 (1000 for the Responder)
- The Proposer receives 100 (900 for the Responder)
- The Proposer receives 200 (800 for the Responder)
- The Proposer receives 300 (700 for the Responder)
- The Proposer receives 400 (600 for the Responder)
- The Proposer receives 500 (500 for the Responder)
- The Proposer receives 600 (400 for the Responder)
- The Proposer receives 700 (300 for the Responder)
- The Proposer receives 800 (200 for the Responder)
- The Proposer receives 900 (100 for the Responder)
- The Proposer receives 1000 (0 for the Responder)

Given that I would be selected as the Responder, I would accept the following proposal(s):

- Proposer receives 0, Responder receives 1000 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 100, Responder receives 900 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 200, Responder receives 800 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 300, Responder receives 700 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]

- Proposer receives 400, Responder receives 600 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 500, Responder receives 500 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 600, Responder receives 400 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 700, Responder receives 300 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 800, Responder receives 200 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 900, Responder receives 100 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]
- Proposer receives 1000, Responder receives 1000 [Accept / Reject (Proposer receives 0 SEK, Responder receives 0 SEK , Save the Children Sweden (Rädda Barnen) receives 1,000)]

[Page 5]

[Personal information]

Please fill in the following information so that we can reach you once the winners of the lottery have been randomly selected.

Full name: *[Text Entry]*

Email address: *[Text Entry]*

Telephone number: *[Text Entry]*

Age: *[Text Entry]*

Once the last "next" had been pressed and all of the above information was included then the survey and the experiment were over.

Once I could access the results in the beginning of April, Anna Dreber Almenberg randomly picked out the 6 winners of the experiment who were matched with one another.

These winners were then subsequently contacted and paid according to their choices. All winners ended up getting paid. Two of the ultimatum games ended up being 500/500 splits that were accepted. The third was a 600/400 split that was also accepted.

9.2. Additional results and robustness checks

9.2.1. Descriptive statistics

Table A.1.1. Summary statistics: Proposals

Control	Freq.	Percent	Cum.
100	1	0.90	0.90
200	0	0	0.90
300	6	5.41	6.31
400	2	1.80	8.11
500	41	36.94	45.05
600	21	18.92	63.96
700	11	9.91	73.87
800	7	6.31	80.18
900	21	18.92	99.10
1000	1	0.90	100.00

Destruction	Freq.	Percent	Cum.
100	4	3.36	3.36
200	1	0.84	4.20
300	0	0	4.20
400	1	0.84	5.04
500	51	42.86	47.90
600	14	11.76	59.66
700	17	14.29	73.95
800	9	7.56	81.51
900	21	17.65	99.16
1000	1	0.84	100.00

Charity	Freq.	Percent	Cum.
100	0	0	0
200	0	0	0
300	1	1.30	1.30
400	1	1.30	2.60
500	43	55.84	58.44
600	13	16.88	75.32
700	11	14.29	89.61
800	2	2.60	92.21
900	5	6.49	98.70
1000	1	1.30	100.00

Table A.1.2. Summary statistics: Age of participants

Variable	Obs	Mean	Std. Dev.	Min	Max
Age	305	22.82	2.561378	18	35

Table A.1.3. Summary statistics: Time to complete survey

	Percentiles	Smallest
1%	119	85
5%	165	110
10%	188	113
25%	219	119
50%	295	
		Largest
75%	398	15691
90%	593	19626
95%	1155	81178
99%	15691	197991

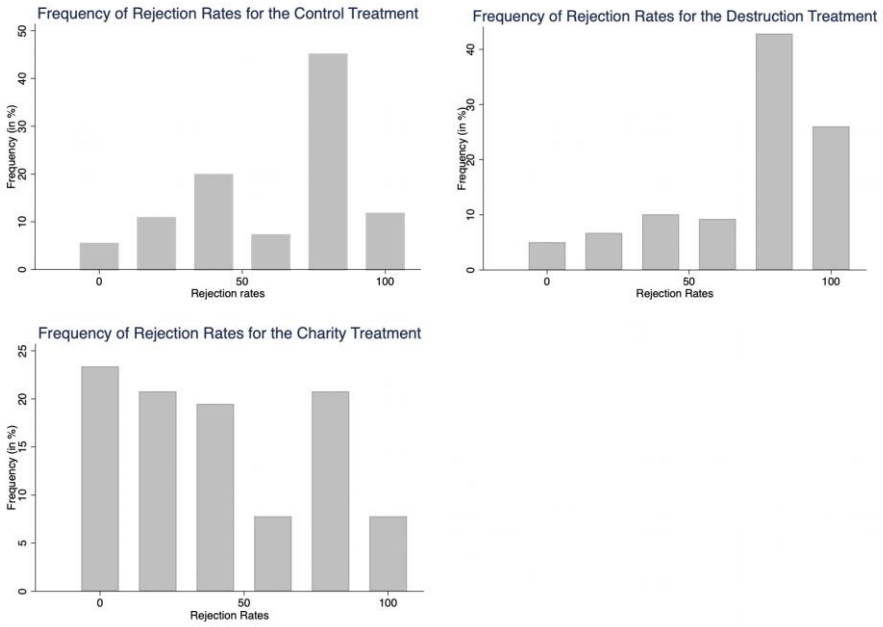
Obs	307
Sum of Wgt.	307
Mean	1556.961
Std. Dev.	12312.45
Variance	1.52e+08
Skewness	14.20155
Kurtosis	218.1789

Table A.1.4. Summary statistics: Sex of participants

	Freq.	Percent	Cum.
Male	184	59.93	59.93
Female	123	40.07	100.00
Total	307	100.00	

NB: approximate, used names to determine sex

Figure A.1.1. Distribution of rejection rates for unfair proposals across treatments



Frequency of each rejection rate of unfair proposals for each treatment group. 0 implies that that 0 out of 5 unfair proposals were accepted. 40% implies that 2 out of 5 unfair proposals were rejected.

9.2.2. Statistical tests

Table A.2.1. Kruskal-Wallis rank test for proposals and rejection rates

	Proposals	Rejection rates
Chi-squared with ties	3.833	39.943
Probability	0.1471	0.0001

Fail to reject null hypothesis for proposals at the 5% level. Reject the null hypothesis at the 1% level for rejection rates of unfair proposals.

Table A.2.2. Mann-Whitney U-tests for rejection rates across treatments

	Treatment 1 = 2	Treatment 2 = 3	Treatment 1 = 3
z	-2.720	5.983	4.306
Prob > |z|	0.0065	0.0000	0.0000

Reject the null hypothesis at the 1% level for all tests.

Table A.2.3. Exploratory Analysis: Chi-square pairwise comparisons of rejection rates for each proposal

P-values	0/1000	100/900	200/800	300/700	400/600	500/500	600/400	700/300	800/200	900/100	1000/0
Control + Destruction	0,508	0,348	0,170	0,170	0,333	1,000	0,901	0,330	0,026	0,056	0,004
Destruction + Charity	0,076	0,041	0,015	0,163	0,025	0,012	0,000	0,000	0,000	0,000	0,005
Control + Charity	0,020	0,006	0,001	0,015	0,007	0,015	0,000	0,000	0,000	0,000	0,777
Criteria Value	0,05										

Cells in red are cells whose p-values are under 0.05

Table A.2.4. Chi-square test: rejection rates of control vs. charity for the 500/500 proposal

Pearson chi2 (1)	5.8916
Prob	0.015

Reject the null hypothesis at the 5% level, but inappropriate since frequencies below 5.

9.2.3. Robustness Tests

Table A.3.1. Kruskal-Wallis H-tests: Control variables across treatments

	Time to complete	Sex	Age
Chi-squared with ties	4.867	0.989	2.896
Probability	0.0877	0.6098	0.2351

Fail to reject all at the 5% level.

Exploratory Analysis

In this subsection of the Appendix I show the results of the robustness checks I performed. Treatment_01 is the Control treatment, Treatment_02 is the Destruction treatment, and Treatment_03 is the Charity treatment.

Table A.3.2. Kruskal-Wallis H-tests: Proposals and Rejection rates

	Proposals	Rejection rates
Chi-squared with ties	4.665	40.784
Probability	0.0971	0.0001

Fail to reject at the 5% level for the proposals. Reject at the 5% level for the Rejection rates of unfair proposals.

Table A.3.3. Mann-Whitney U-tests: Proposals

	Treatment 1 = 2	Treatment 2 = 3	Treatment 1 = 3
z	0.118	1.932	1.957
Prob > |z|	0.9061	0.0534	0.0504

Fail to reject all at the 5% level.

Table A.3.4. Mann-Whitney U-tests: Rejection rates of unfair proposals

	Treatment 1 = 2	Treatment 2 = 3	Treatment 1 = 3
z	-2.514	6.057	4.488
Prob > |z|	0.0119	0.0000	0.0000

Reject all at the 5% level.

Table A.3.5. Chi-square test: Pairwise comparison of rejection rates for each proposal

ROBUSTNESS CHECK Chi-Squared Pairwise comparisons of Rejection Rates for each Proposal											
P-values	0/1000	100/900	200/800	300/700	400/600	500/500	600/400	700/300	800/200	900/100	1000/0
Control + Destruction	0,425	0,207	0,095	0,173	0,336	1,000	0,872	0,414	0,048	0,089	0,004
Control + Charity	0,039	0,006	0,001	0,015	0,007	0,015	0,001	0,000	0,000	0,000	0,637
Destruction + Charity	0,170	0,085	0,040	0,159	0,024	0,012	0,000	0,000	0,000	0,000	0,003
	0,050										
Chi-Squared Pairwise comparisons of Rejection Rates for each Proposal											
P-values	0/1000	100/900	200/800	300/700	400/600	500/500	600/400	700/300	800/200	900/100	1000/0
Control + Destruction	0,508	0,348	0,170	0,170	0,333	1,000	0,901	0,330	0,026	0,056	0,004
Control + Charity	0,020	0,006	0,001	0,015	0,007	0,015	0,000	0,000	0,000	0,000	0,777
Destruction + Charity	0,076	0,041	0,015	0,163	0,025	0,012	0,000	0,000	0,000	0,000	0,005

Table A.3.6. ANOVA tests: Sex and Age of participants

	Sex	Age
chi2(2)	0.0850	0.8590
Prob>chi2	0.958	0.651

9.3. Pre-analysis plan

This section consists of the full pre-registration as presented on the OSF. It was published on February 22nd and is publicly verifiable and available on the OSF website. I have also added the raw data from the experiment in order for future researchers to be able to replicate my results. It is important to note that the names of treatments might be different from the above design plan, but I have not made any changes to the actual essence of the experiment. The changes that I have made, for example with the exclusion criteria, are explicitly noted, and a robustness test has been included for these changes.

9.3.1. Introduction / Description

In this paper we investigate whether changing the context of rejected offers could influence decisions in an ultimatum game. Whereas a large strand of literature in the field of behavioural economics has researched the effects of fairness, altruism, and the joy-of-destruction, these studies have performed experiments using nonzero positive monetary outside options. Our hypothesis is that players' efficiency and fairness concerns can be highlighted by studying the effects of a different type of outside option which does not involve monetary payoffs to the players. We consider one control group, which plays the standard ultimatum game, and two treatments. In the first treatment, participants are informed that the rejection of the offer by the proposer leads to the destruction of the money. For the second treatment, we inform players that in case the responder rejects the offer, the money will be given to charity. Conditional on the results, we can shed light on the effects that nonmonetary outside options may have on proposers and/or responders in the ultimatum game.

9.3.2. Hypotheses

In this study, we are testing two hypotheses. Firstly, we expect that the offers made by proposers differ across the three treatment groups, which are (0) control i.e. the classic ultimatum game, (1) destroying the money following a rejection, (2) giving the money to charity following a rejection. Secondly, we predict that the rejection rates will not be the same across these three treatment groups. We predict that the Proposals will be higher (More money to the Proposer) in the (1) treatment where the money is destroyed in case of rejection. We predict that the Rejection rates will be lower in the (1) treatment where the money is destroyed in case of rejection. We predict that the Proposals will be lower (less money to the Proposer) in the (2) treatment group where the money is given to charity in case of rejection. We predict that the Rejection rates will be higher in the (2) treatment group where the money is given to charity in case of rejection.

9.3.3. Design plan

Study type

Experiment - A researcher randomly assigns treatments to study subjects; this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Design Plan

We investigate whether changing the context of rejected offers in a fashion that does not affect monetary payoffs could influence decisions made by either party in an ultimatum game. Previous research has assumed that players do not care much about efficiency in the context of ultimatum games and are happy to 'burn money' (Güth & Kocher, 2014). However, this does not account for the fact that if a responder rejects a proposal, the money is not truly 'burnt' - it generally is simply not spent and stays in the experimental budget. Our experiment seeks to answer two questions: first, does changing the default rejection scenario from a neutral to a strictly efficiency-lowering (i.e. actually destroying money) or to a positive non-monetary outside option (i.e. giving money to charity) scenario lead to different offers from proposers? Secondly, do responders significantly change their rejection rates in response to low offers, depending on those changes to the default?

Previous research suggests that there is a gap in the literature with regards to understanding the relationship between efficiency, fairness, and altruism concerns in the ultimatum game. We hope that our two treatments will help illuminate these competing

concerns and lead to a better understanding of the ultimatum game and economic behaviour more generally.

In the following section we go through the design of the experiment including a description of our treatments, subjects, procedure, and statistical tests, as well as a discussion on why we made this particular choice. We begin with our treatments. Treatments

Our study involves three treatments:

- Treatment 0: The control group

- Treatment 1: The "Take it or destroy it" group

- Treatment 2: The "Take it or donate it" group

Treatment 0 corresponds to the control group. The control group will play a standard ultimatum game: the proposer makes an offer on how to split the given sum, and the responder will either accept or reject. Acceptance leads to a split of the money as suggested by the proposer. In case of rejection, the participants will not receive the money. It is important to note that in this case the instructions do not specify what will happen to the money in case of a rejection.

The Participants in this experiment get to play the Strategy Method of the Ultimatum game, where they act as both Proposer and Responder. First the player will submit the proposal they would make if they were the Proposer, and then they answer which proposals they would accept if they were the Responder (11 different proposals, intervals of 100 SEK).

Treatment 1 is similar to Treatment 0. The only difference is that participants are now informed that, in case of a rejection by the responder, the money will be destroyed, i.e. through the use of a lab furnace or similarly non-entertaining method.

Treatment 2 is another variation along the lines of Treatment 1. In this treatment participants are informed that if the responder rejects the offer, the whole pot will be given to a charity, Save the Children Sweden (Rädda Barnen).

Subjects / Participants

The participants in this study will mostly be students given that the survey will be distributed by email to the students and alumni from Stockholm School of Economics (SSE). The survey will also be distributed on social media in order to get as many participants as possible. The experiment is written in English and is based on the assumption that everybody who studies at SSE should have a fluent level of English and should thus be able to perfectly understand the questions being asked. Procedure A copy of the survey is attached in the Appendix. An email has been sent out to all students at SSE asking them to participate in the experiment. Students are also asked not to discuss

the experiment with others once they have finished. People will be able to take the survey up until the 31st of March when the survey will be put made inaccessible. Six participants in this experiment will be randomly selected and then randomly assigned to one of the two roles. Thus, if somebody is selected, their actions in this experiment will be matched with the actions of another player. If a pair agrees on the proposal, the money will be awarded to the participants. If the Responder rejects the proposal then the money will either be destroyed, donated to charity, or given back to the experimenters depending on the treatment.

Randomization

The Treatments will be randomized automatically through the chosen survey developer (Qualtrics). When somebody clicks on the link to the survey, they will be automatically given only one of the three treatments. This randomization software is built in to Qualtrics.

9.3.4. Sampling plan

Existing Data

Registration prior to creation of data

Data collection procedures

Data Collection The data will be collected through Qualtrics, the developer of the survey software. Participants will be able to access the survey through a link sent to them via email or accessible on social media channels.

Population

Since randomization occurs through the link, all participants will be randomly placed in each of the treatment groups regardless of how the found the link. This implies that participants can technically be any adults in Sweden. The participants will however most likely be students in the Stockholm region, and more specifically students from Stockholm School of Economics. I will be tracking the number of participants who have taken the survey through Qualtrics without looking at the data. In order to have significant results I am aiming to have at least 100 participants per treatment group, that is at least 300 participants in total. I will start out by sending out a link to SSE students. Depending on the number of responses I get from this group of students, I might have to try to access other groups of students such as students at Stockholm University or KTH.

Payment for Participation

Participants will not have a guaranteed payment for their participation. Instead, a lottery will take place once the survey has been closed whereby 6 participants (2 in each treatment group) will be randomly chosen and randomly assigned one of the roles. Thus,

in each treatment group we will have a randomly chosen Proposer and Responder. We then match their answers from the survey with each other in the respective roles. If the Responder agrees with the Proposal, then the participants will share the 1,000 SEK according to the proposal. The money will be sent to their bank accounts. If the Responder rejects the proposal, then the money will either be destroyed or given to the charity or handed back to the experimenters depending on the treatment group. This implies that the expected value of payment per participant will only be approximately 10 SEK given a minimum of 300 participants.

Exclusion Criteria

There will not be Exclusion Criteria for this experiment. Research shows that both monotonic and non-monotonic Responder answers make sense for an Ultimatum Game (Strategy Method).

Study Timeline

The Experiment / Survey will be launched on the 22nd February and will be available and marketed up until the 31st of March. I will only look at the number of participants having partaken in the experiment in order to better manage my efforts in marketing the survey. No detailed responses or data will be analysed during this period.

Sample Size

My target sample size is 300 participants. With 300 total participants and a well-functioning randomization software through Qualtrics, this should imply approximately 100 participants in each treatment group. 100 participants in each treatment implies 100 Proposals and 100 sets of Responder answers (Responders have to specify if they would accept or reject 11 different proposals, ranging from a 0/1000 division in favour of the Proposer to a 1000/0 division in favour of the Responder.

I will make an effort to try and achieve more than 300 participants in total, but this depends on how successful I am at getting people to do the experiment.

Sample size rationale

A final possible important internal limitation to note is that of experimental power. To make sure that our tests can correctly reject the null hypotheses, we have chosen sufficiently large sample sizes (>50) per treatment group. We hope that this will help us avoid type M and type S errors (Gelman & Carlin, 2014). It may still be the case that the effect sizes of our treatments are not large enough to give us significant results. Considering our treatments add relevant outside options we do not anticipate low power due to a low effect size being a fundamental issue in our experimental design, excluding the limitations discussed above. Amir et al. (2012) find some significant effects in the basic ultimatum game when dealing with 1 USD stakes online, and with our stakes an order of magnitude higher we anticipate participants' reactions will be statistically

noticeable. Although the absolute sum of the winner is higher than most Ultimatum Game experiments it is important to consider that when considering expected payoffs, the ex-post expected payoff for the participants was at best 10 SEK (about 0,94 EUR), which is not particularly high. Nonetheless, the survey is about how to deal with a sum of 1,000 SEK and studies have shown that people tend to overestimate their chances of winning. Both of these facts will be taken into account when analysing whether the impact of the treatments is due to higher stakes.

Stopping rule

If by some chance, the survey ends up being taken by people outside of Sweden for example then this will have an unexpected impact on the survey by including other unexpected populations. If by some small chance this ends up happening, then the observations will probably have to be removed because the survey has been intended primarily to be distributed to 1) students at SSE 2) students in Stockholm 3) adults in Stockholm 4) students and adults in Sweden. The study has been designed for these groups in particular, by choosing for example a survey in English (accessible for both Swedish students and International students in Sweden), by choosing a charity that Swedish people would recognise more easily (Save the Children Sweden / Rädda Barnen), and also by the fact that the currency is in SEK and that the destroying money treatment might not be possible in other countries (it is legal in Sweden to destroy money). All of these factors and more are intended for Swedish participants. I would be afraid that including a wider population could diminish the validity of the survey.

Although the link could technically be shared by some people into other populations, I think the chances of this happening are low. I will only find out if a group of unexpected have taken the survey through the geographical locations of the IP addresses that have been used to take the survey. Technically the names and email addresses and age variables could also be indicators of something having gone wrong.

Given that I will not have access to the survey data until the end of March when I close the experiment and stop gathering data, I will not be able to notice whether anything happens until the very end. This means that a stopping rule will not be implemented, and I will only be able to clean the data later on.

Measured variables

One outcome variable will be the Proposal made by the Proposer. We will measure this by asking participants "Given that I would be selected as the Proposer, I would like to submit the following proposal:" Participants will be asked to choose one of the 11 alternatives going from "The Proposer receives 0 (1000 for the Responder)", "The Proposer receives 100 (900 for the Responder)", step-wise on a 100 SEK scale, all the way to "The Proposer receives 1000 (0 for the Responder)".

The second set of outcome variables will be the answers to whether the participant would accept or reject the above 11 proposals. We will measure this by asking participants "Given that I would be selected as the Responder, I would accept the following offer(s):" Participants will then have to pick either "Accept" or "Reject" for each of the 11 possible proposals. In the Charity Treatment, or Treatment 2, the first proposal would be "Proposer receives 0, Responder receives 1000" which the participant could "Accept", or the participant could Reject which would imply that "Reject (Proposer receives 0, Responder receives 0, Save the Children Sweden (Rädda Barnen) receives 1,000)".

In the end we will have in total 12 outcome variables for each Treatment Group.

Indices

The outcome variables for the Responder will first be in binary form, where we will recode Accept and Reject into 1 and 0 respectively. In order to be able to test the difference between all the 'unfair proposals' (i.e. all proposals where the Responder gets less than the Proposer, that is, all proposals where the Responder gets less than or equal to 400 SEK), we will construct a new variable which will be the mean acceptance rate of 'unfair proposals'. We will take the average of the outcome variables for the proposals whereby the Responder would get less than or equal to 400 SEK. This new variable will allow us to compare the mean acceptance rate of unfair proposals between treatment groups.

9.3.5. Analysis plan

Statistical models

In this study, we are testing two hypotheses. Firstly, we expect that the offers made by proposers differ across the three treatment groups, which are (0) control i.e. the classic ultimatum game, (1) destroying the money following a rejection, (2) giving the money to charity following a rejection. Secondly, we predict that the rejection rate for unfair offers will not be the same across these three treatment groups.

We perform the Kruskal-Wallis H test for each of the two hypotheses. This test is an extension of the Mann-Whitney U test to more than two groups. It tests the hypothesis that the medians of different treatment groups are equal (Siegel & Castellan, 1988). The Kruskal-Wallis H test is a nonparametric method, and thus does not require an underlying normal distribution of the residuals (unlike the one-way ANOVA test).

However, this test does not provide information on which treatment group differs from the others. Therefore, when we find that we can reject the null hypothesis that the medians of all treatment groups are equal, we will perform pairwise Mann-Whitney U tests to identify which group behaves differently from the control group.

Transformations

Many variables will have to be transformed.

The output variables from Qualtrics will give us independent outcome variables for each of the treatments. In order to collect all of the answers for the Proposal by the Proposer, we will have to create a new variable called "Proposal" where we will collect all of the answers from the different treatment groups.

We will have to create the same sort of variables for the Responder answers to each of the Proposals which we will call "Responder 0/1000", "Responder 100/900", "Responder 200/800", "Responder 300/700", "Responder 400/600", "Responder 500/500", "Responder 600/400", "Responder 700/300", "Responder 800/200", "Responder 900/100", and "Responder 1000/0". The answer for each of these variables will be Accept or Reject which will be recoded to 1 or 0.

We will have to create a new variable called "Treatment" for each observation so that we can track which participant had which treatment. We will call the treatments 0, 1, and 2 for the Control, Destruction, and Charity treatments respectively.

Inference criteria

If we are able to reject the null at the 5%-significance level for the Kruskal-Wallis H test, we can conclude that the rejection rate differs among the different treatment groups. To identify which treatment group differs from the control group, we again perform pairwise Mann-Whitney U tests.

Exploratory analysis

Chi-squared test for each Responder Accept/Reject ratio. I expect that the treatments will have an effect on some of the rejection rates for every proposal.