ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Master Thesis [Economics and Business – Behavioural Economics]

Pro-social Behaviour in Dictator Games: Just an Artefact?

An Experimental Analysis

Name student: Lonneke (L.J.) de Jong
Student ID number: 402957

Supervisor: Jan (J.P.M.) Heufer
Second assessor: Cem (A.C.) Peker

Date final version: 01/03/2019

Abstract

This thesis seeks to understand how informative behaviour observed in dictator game experiments is in terms of reflecting preferences. In particular, it investigates whether a less artificial situation within dictator game experiments leads to behaviour that better reflects preferences. For this purpose, this thesis partially replicates the study by Brosig-Koch, Riechmann, and Weimann (2017), who investigated dynamics of behaviour in dictator games with different relative prices by conducting the same games on the same subjects three times with several weeks in between. In these dictator games, the recipients are students, as is typically the case in dictator games. This thesis replicates Brosig-Koch et al.'s findings in the sense that selfish behaviour increases over repetitions of the games, which Brosig-Koch et al. partly explain by a diminishing 'experimenter demand effect'. This thesis extends their study by setting up a new treatment condition in which a less artificial situation is created by replacing the recipients by well-known charities. It is found that selfish behaviour increases less over repetitions of the games in the charity treatment as compared to the treatment with student recipients. How much more stable behaviour in the charity treatment tends to be depends on the relative price of exhibiting selfish behaviour. It is speculated that a smaller decrease in the experimenter demand effect in the charity treatment at least partly explains that more stable behaviour is observed. As this would imply that the experimenter demand effect must have been less present in this treatment, due to the experimental designs, it appears that a less artificial situation within dictator game experiments leads to behaviour that better reflect preferences. Thus, how informative behaviour observed in dictator game experiments is in terms of reflecting preferences appears to depend on how artificial the situation within the dictator game is.

Keywords: social preferences, pro-social behaviour, selfish behaviour, artificiality, experiments, methodology

Acknowledgements

This thesis completes my time at the Erasmus University. I look back to fours years in which I learned a lot and developed myself as a person. I especially enjoyed the last year of it: the period of the master 'Behavioural Economics'. I found the subjects that were taught in this program very interesting and, though sometimes stressful, I learned a lot from the seminar, in which we could put into practice what we had learned from the courses. Besides that, I met a lot of great people from all different kinds of nationalities and we had a good time together. I am glad to finish this period with this thesis and I would like to express my appreciation to a number of people who were involved in the process of it.

First of all, I would like to thank my supervisor, Jan Heufer. After he had supervised my bachelor thesis, I very much hoped that he would be assigned as my supervisor again for my master thesis. I am grateful that this happened for a number of reasons. At the very beginning of the thesis process, he really took his time to brainstorm with me about the ideas I had in mind for this project. This resulted in a topic that I found truly interesting. Throughout the entire process, he was always willing to answer any questions, either during meetings or online. Overall, his quick replies and profound answers made sure I could move on quickly. All in all, Jan's expertise in the field of experimental economics and social preferences, together with his insightful feedback make him the best supervisor I could wish for. I am also thankful to the teachers Sander Opschoor and Leon de Gier from the secondary school 'Dalton Lyceum Barendrecht' for letting me conduct questionnaires in their classes; and to the students who took the questionnaires. I am grateful to my parents for always believing in me and supporting me. My younger sister, Daniëlle de Jong, should be thanked as well since I could always count on her for moral support; she also took the time to brainstorm together, which provided me with useful insights. Finally yet importantly, I would like to thank my friend Merel Vosselman who has been a great help by proofreading this thesis for any English mistakes and by providing moral support.

Table of contents

1. Introduction

1.1 Theoretical Framework

According to the revealed preference theory, pioneered by Samuelson (1938, 1948), preferences of individuals can be inferred from the choices that they make. This theory has been applied broadly in economic lab experiments in order to elicit preferences; for example risk preferences, time preferences, and social preferences. A game that is widely used in lab experiments with the purpose of eliciting pro-social preferences is the dictator game, developed by Forsythe, Horowitz, Savin, and Sefton (1994). In this game, a subject, called the dictator, is endowed with a certain amount of money and is asked to divide this amount between themselves and another subject, called the recipient. The typical finding is that dictators do not take the whole amount for themselves, but give part of it to the recipient. Specifically, on average, they give away 28.35% of the amount to be shared, according to Engel's (2011) meta-study on dictator games covering 129 papers. Dictators that give part of the money away, or even all, 'revealed preferred' giving to keeping everything. The conclusion that experimental economists draw from the observed pro-social *behaviour* is that people have pro-social *preferences*. This is a rather impactful conclusion as it refutes the classical assumption of selfish preferences, an assumption made by almost all economic models about 20 years ago (Fehr & Schmidt, 1999). Inspired by the departures from selfish behaviour in economic lab experiments, models have been developed that incorporate forms of social preferences. For example, the models developed by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000) incorporate inequality aversion: They take into account that people may dislike being worse off as well as being better off compared to others. Other models have been developed that not only take into account the final allocation of payoffs, but also fellow players' intensions, i.e. models that incorporate forms of reciprocity (Rabin, 1993; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006). Charness and Rabin (2002) developed a model that incorporates inequality aversion, reciprocity, and social welfare preferences. There has thus been a shift from models with the unrealistic assumption of selfish preferences to seemingly more realistic models, based on experimental evidence. However, recent research calls into question whether economic lab experiments, in particular dictator game experiments, do in fact reveal preferences, raising doubt about the doubt about the 'unrealistic' economic models of about 20 years ago. The doubt raised by this recent research is, in turn, challenged by the present study. Is pro-social behaviour in dictator games just an artefact? Or does pro-social behaviour in dictator games reflect pro-social preferences?

An example of recent research calling into question whether dictator game experiments reveal preferences well is the research by Bardsley (2008). He compared behaviour observed in three types of dictator games. In each type, there was an initial endowment that provided money to both the dictator and the recipient. In the first type of game, dictators were provided with 'give options', i.e. options to increase the recipients' payoff by decreasing their own. In the second type of game, dictators were provided with the exact same options as in the first type, but 'take options' were added, i.e. options to increase the dictator's own payoff by decreasing the recipient's payoff.[1] In the third type of game, dictators were provided with only take options. The budget lines of the second and third types of games had the same gradient and endowment points as the budget lines of the first type. Please note that an "indifference curve can be tangential with a budget line, of arbitrary length, at one point only" (Bardsley, 2008), assuming strict convex indifference curves. Therefore, if pro-social *preferences* would be the motive behind pro-social *behaviour* in dictator games, the proportion of givers in the first type of game should be equal to that in the second type of game; and the proportion of givers in the first type of game should not exceed the proportion of dictators choosing the endowment point in third type of game. However, this is not what Bardsley found: The proportion of givers in the first type of game was greater than both the proportion of givers in the second type of game and the proportion of dictators choosing the endowment point in third type of game. According to Bardsley, the data provide strong support for the interpretation that subjects react to experimental "demand characteristics", i.e. "the cues the protocol supplies about appropriate behaviour", "with the choice set contributing to the cues which indicate appropriate behaviour". According to this explanation, subjects in the first type of game infer that it is tested how much they give away "since they can either do nothing or give", inducing them to give. When take options are included, as in the second type of game, subjects cannot as easily infer that it is tested how much they give away, since options to take could introduce a "conflicting message". In the third type of games it is even harder to infer this, as there are only take options. Bardsley concludes that, rather than representing pro-social preferences, pro-social behaviour in dictator games may be an artefact of experimentation.

Another example of recent research calling into question whether dictator game experiments reveal preferences well is that by Brosig-Koch, Riechmann, and Weimann (2017). As stated by Brosig-Koch

[1] The amount of money for the recipient when the dictator would choose the initial endowment is meant with the 'recipient's payoff'.

et al., inherent to the models that incorporate forms of social preferences is their assumption that social behaviour is stable over time. In other words, it is expected that "people make identical decisions in identical repetitions of the decision task". If unstable behaviour in identical repetitions is, however, observed, this is generally ascribed to "strategic considerations such as reputation building or to some form of learning from information feedback" (Brosig-Koch et al., 2017). Brosig-Koch et al. examined the dynamics of people's pro-social behaviour without the presence of these considerations. They used modified dictator games for this, introduced by Andreoni and Miller (2002). This version differs from the original dictator game in that different relative prices are used: In some games the amount that recipients receive is less than the amount dictators actually gave away, while in other games recipients receive more than actually given away by dictators. The dictator games used by Brosig-Koch et al. consisted of four take games and four give games; their sample comprised of German university students. In these games, the subjects were asked to choose between 11 different allocations. In both the take games and the give games, one of the allocation options resulted in exactly the same payoff for both the dictator and the recipient, namely, 500 eurocents. The difference between the take games and the give games is that in the take games the dictator was given the option to increase their own payoff by decreasing the recipient's payoff; whereas in the give games the dictator was given the option to increase the recipient's payoff by decreasing their own. The idea of the differing relative prices, as introduced by Andreoni and Miller (2002), can be applied analogously in the give games and the take games. Apart from having their subjects participate in take and give games, Brosig-Koch et al. (2017) also had them participate in sequential prisoner's-dilemma games. They created a time dimension by repeating the take games, give games, and prisoner's-dilemma games twice, on the same subjects, each with a time period of four weeks in between. The first experiment is referred to as "wave 1", the second "wave 2", and third "wave 3". These terms will be adopted in the present study as well.

Their results show that pro-social behaviour is not stable over time: Selfish behaviour increased over the waves in all types of games, with behaviour in take games affected the most. In the take games, the dictators did on average not take the whole amount in wave 1, but left part of it for the recipient. More specifically, dictators on average took 301.80 eurocents of a maximum of 500 eurocents. However, the amount taken by the dictators increased over the waves. The average amount taken increased to 425.40 eurocents in wave 2 and increased further to 496.40 eurocents in wave 3. That

is, while playing the games for the first time led dictators to exhibit pro-social behaviour to a great extent, repeating the games twice made them behave almost entirely selfishly.

Brosig-Koch et al. (2017) offer two possible explanations for the increase in selfish behaviour over the waves. One of them is "moral licensing", implying that people have the feeling that they are allowed to behave more selfishly if they have behaved pro-socially before, i.e. in a previous wave. The alternative explanation they discuss is that the "experimenter demand effect" [2] ("EDE", hereafter) diminishes over time. They argue that subjects are uncertain about the decision environment when they participate in an experiment for the first time. As a result, "everything implicitly or explicitly communicated by the experimenter will be used as a signal from which subjects try to infer what the experiment is about and how they should behave". In other words, due to their uncertainty, subjects search for hints that tell them what is the "right" thing to do. Thus, in the light of the experiments conducted by Brosig-Koch et al. (2017), if participants correctly infer that the experimenter is interested in their pro-social behaviour in dictator games, subjects might act accordingly, by taking (giving) less (more) than they would in the absence of the EDE. In repetitions of the games, subjects might feel less uncertain about the decision environment, causing the EDE to diminish and therewith selfish behaviour to increase.

A diminishing EDE driving the increase in selfish behaviour over the waves suggests that behaviour observed in dictator games is an artefact of experimentation: Rather than caring about the recipient, uncertainty about the decision environment would have caused people to exhibit pro-social behaviour. Accordingly, pro-social *behaviour* in dictator games does not necessarily imply pro-social *preferences*. If it is moral licensing instead which drives the instability of behaviour, the lesson to learn is that the preferences change over time, if people have acted pro-socially before. However, behaviour would reveal preferences in dictator games. While Brosig-Koch et al. cannot judge their relative strength, they expect both factors to be at play. A diminishing EDE at least partly driving the results implies that pro-social preferences are overestimated by the observed pro-social behaviour, suggesting that dictator games may not be as useful as its widespread use suggests.

[2] Note that the experimenter demand effect is similar to the phenomenon of subjects reacting to experimental "demand characteristics", as described by Bardsley (2008).

The research by Brosig-Koch et al. and Bardsley thus suggest that observed pro-social behaviour in dictator game experiments may be an artefact of experimentation rather than a representation of preferences. This thesis further explores this. A broad research question is explored in this thesis, but a closer look will be taken to a specific case to contribute a bit to the answer. The research question reads:

Research Question: *How informative is behaviour observed in dictator game experiments in terms of reflecting preferences?*

In the typical dictator game, also applied by Bardsley (2008) and Brosig-Koch et al. (2017), the donation recipient is an anonymous student, randomly drawn from a large subject pool. Dictators do not know whether the recipient they will be matched with is in need of the money (some students may be while others may not be). The dictators also do not know what the recipients will do with the donated money. Dictators thus do not have much information to base their donation decision on, whereas in real life, donations are typically made in an informative setting. The typical dictator game thus creates a situation that is rather artificial. The artificiality of the situation likely causes uncertainty about the decision environment among subjects, which, according to Brosig-Koch et al., in turn causes an EDE. Pro-social behaviour would then indeed overestimate pro-social preferences. In a less artificial situation, subjects would likely feel less uncertain, leading to less of an EDE. Behaviour would then better reflect preferences. It is thus expected that the answer to the research question depends on how artificial the situation within the dictator game is. A sub-research question is therefore formulated:

Sub-Research Question: *Does a less artificial situation within dictator game experiments lead to behaviour that better reflects preferences?*

A less artificial situation will be created by replacing the typically applied students as recipients by well-known charities. It is less artificial as it creates a more informative setting. It does so in two ways. Frist, while dictators do not know whether the particular student they will be matched with is in need of the money, it is common knowledge that charities do have a serious need of money. Second, while dictators do not know what both students and charities will exactly do with the donated money, they at least know the goal of charities: to do good. It is common for people to

donate to charities[3] and there are indications that it is also not uncommon for winners of lotteries to donate part of their windfall to charities[4]. This indicates that people are generally familiar with donating to charities, either by means of their own income or by means of a windfall (which is the case in the dictator games). This supports the claim that replacing students as recipients by well-known charities creates a less artificial situation. The type of dictator games with students as recipients will from now on be referred to as 'standard dictator games'; the type of dictator games with charities as recipients as 'charity dictator games'.

The remainder of the paper is organised as follows: The rest of the introduction discusses the hypothesis and provides the definition of (pro-) social preferences and the measure of pro-social behaviour used in this study. Section 2 discusses the data used and methodology applied in this study. Section 3 deals with the results obtained, and thereby provides an answer to the sub-research question and the research question. Section 4 compares findings of this study with those from related literature, discusses limitations, and proposes suggestions for future research. Section 5 concludes.

1.2 Hypothesis

Brosig-Koch et al. speculate that both moral licensing and a diminishing EDE drove the increase in selfish behaviour in their study. In the present study, moral licensing is also likely to be at play, both in the standard dictator games and in the charity dictator games: Whether the recipient is a student or a charity, having behaved pro-socially in a previous wave probably gives people the feeling that they are allowed to behave more selfishly in future waves. The simplifying assumption is made that moral licensing is *equally* present in both types of dictator games.

The simplifying assumption implies that the EDE is isolated: A lower increase in selfish behaviour in charity dictator games compared to standard dictator games must then be due to a lower decrease in EDE. As the experimental designs of the present study are such that the EDE would decrease approximately equally if the EDE would have been present equally in the first place (in wave 1), a lower decrease in EDE implies that it the EDE was less present in the first place. More stable behaviour thus implies that behaviour better reflects preferences. Accordingly, stability of behaviour will be used to measure how well behaviour reflects preferences.

[3] According to the most recent annual report of Goede Doelen Nederland (2018), the private sector in the Netherlands donated in total 982 million euros to charities in 2017, based on the 146 charities that participated in the investigation.

[4] According to a financial advisor/wealth manager who provides financial advice to lottery winners, lottery winners like to donate part of their winnings to charities (de Ronde, 2008). There are some noticeable examples in which lottery winners donated millions of euros of their winnings to charities (Bacon, 2018; Drewett, 2018; Ensor, 2014).

To sum up, it is expected that a less artificial situation within dictator game experiments leads to behaviour that better reflects preferences, charity dictator games create a less artificial situation than standard dictator games, and behaviour that better reflect preferences is more stable. This results in the following hypothesis:

Hypothesis: *People exhibit more stable behaviour in charity dictator games compared to standard dictator games*

1.3 Definition of (pro-) social preferences & measure of pro-social behaviour

In terms of payoffs, an individual with social preferences does not only care about their own payoff, but also about that of other people. This results in a utility function that includes not only their own payoff, but also that of others. In case of the dictator game, where only the dictator and the recipient participate, the utility function of a dictator with social preferences can be illustrated as $U_d = f(\pi_d, \pi_r)$, where π_d and π_r represent the dictator's and recipient's payoffs respectively. Having social preferences does not necessarily mean that the dictator's utility increases in the recipient's payoff ($\frac{\partial U_d}{\partial \pi_r} > 0$); it can as well be the case that it decreases in the recipient's payoff ($\frac{\partial U_d}{\partial \pi_r} < 0$). Accordingly, the term 'social preferences' only tells that the dictator's utility is influenced by the recipient's payoff in *some* way. To indicate that the dictator's utility *increases* in the recipient's payoff, the term 'pro-social preferences' will be used in this study.

As will be explained in Section 2.2.1, the take games of Brosig-Koch et al. will be applied in this study. Pro-social behaviour is measured by the amount that is not taken in these games: The smaller the amount taken, the greater the extent of pro-social behaviour. Note that pro-social behaviour is a specific case of social behaviour, where an individual's behaviour is influenced by others in *some* way, but not necessarily *beneficial* for the others.

2. Data & Methodology

2.1 Sample

The research method employed in this study is quantitative empirical research. Data was obtained experimentally by conducting questionnaires on students from four classes in a Dutch secondary school, Dalton Lyceum Barendrecht. This secondary school has three educational attainment levels: 1) *vmbo* (in English 'preparatory secondary vocational education'), 2) *havo* (in English 'senior general secondary education'), and 3) *vwo* (in English 'pre-university education'). *Vmbo* is the lowest level,

havo is the second highest level, and *vwo* is the highest level of secondary education in the Netherlands. In order to increase the likelihood of subjects understanding the questionnaires, the classes highest in age and educational attainment level would ideally be selected. The third, fifth, and sixth grade of *vwo* comprise some students that have participated in the experimenter's previous dictator games (de Jong, 2017; de Jong, Mol, Phillips, Straka, & de Wilde, 2017). As this study particularly aims to examine the way in which behaviour changes from the first time one participates in such a game to repetitions over the games, these classes could not be selected for the present study. The best remaining option was to use four fourth grade *vwo* classes. Two of those classes could be selected to conduct the questionnaires. The other two selected classes were fourth grade *havo* classes (fifth grade *havo* classes were not available due to near exams). In all four classes, the questionnaires were conducted during the course 'economics'.

This set of participants is well suited for this research for at least two reasons. First, as will be explained in Section 2.2, two similar groups could be constructed from this set of participants, which implies that any observed difference in behaviour between the two groups is likely to be due to the different treatments they received rather than different characteristics of participants between the groups. Second, as this study is particularly interested in dynamics of behaviour over time, it is important that the participants were present every time a questionnaire was conducted. As students are legally obliged to go to class, there was a good likelihood that most students were indeed present every time. In addition to the set of participants, the physical environment in which the experiments took place, classrooms, was also well suited: Besides the fact that it is a highly controlled setting (the participants could, for example, not communicate with each other without the experimenter noticing), Brosig Koch et al. (2017) had their subjects participate in a similar setting.

2.2 Method

In order to be able to compare behaviour in standard dictator games to behaviour in charity dictator games, this study applied a between-subjects design. More specifically, one *havo* and one *vwo* class participated in standard dictator games. This group will be referred to as the 'standard group' from now on. The remaining *havo* and *vwo* classes participated in charity dictator games; that is, they received the 'charity treatment'. This group will be referred to as the 'charity group' from now on. As the composition of classes with different educational attainment levels is the same in both the standard group and the charity group; and as it concerns secondary school classes following the same course (economics), the standard group and charity group are similar to each other. The next two

sections, namely 2.2.1 and 2.2.2, describe both treatments. Which *havo* and *vwo* classes received which treatment was determined randomly by flipping a coin, as explained in Part 1 of Appendix E. This procedure resulted in the *havo* class and *vwo* class in which the experiment was conducted first to receive the standard treatment and the classes in which the experiment was conducted later to receive the charity treatment. Besides the between-subjects design that this study applied, it also applied a within-subjects design. The purpose of the latter type of design is to observe the participants' behaviour over time. Each subject participated in the same experiment three times; i.e. in waves 1, 2, and 3, each time with 13 days in between.

2.2.1 Standard treatment

The games in the standard treatment are identical to Brosig-Koch et al's. (2017) four take games. The reason that take games rather than give games were employed is that Brosig-Koch et al. found that the former games provided the strongest results in terms of the increase in selfish behaviour over the waves; in the give games the average amount given away in the first wave was already close to zero. Table 1, adopted from Brosig-Koch et al., presents the structure of these games[5]:

Table 1. Payoffs in the four take games (Brosig-Koch et al., 2017)

Game	π	1	2	3	4	5	6	7	8	9	10	11
T1	π_d, π_r	500, 500	600, 450	700, 400	800, 350	900, 300	1000, 250	1100, 200	1200, 150	1300, 100	1400, 50	1500, 0
T2	π_d, π_r	500, 500	575, 450	650, 400	725, 350	800, 300	875, 250	950, 200	1025, 150	1100, 100	1175, 50	1250, 0
T3	π_d, π_r	500, 500	550, 450	600, 400	650, 350	700, 300	750, 250	800, 200	850, 150	900, 100	950, 50	1000, 0
T4	π_d, π_r	500, 500	525, 450	550, 400	575, 350	600, 300	625, 250	650, 200	675, 150	700, 100	725, 50	750, 0

As illustrated in Table 1, the subjects were provided with 11 different allocations in each game. In each game, one of the allocation options resulted in exactly the same payoff for both the dictator and the recipient, namely, 500 eurocents. Adhering to Brosig-Koch et al.'s terminology, this can be expressed as follows: $(\pi_d, \pi_r) = (500{,}500)$, in which π_d and π_r represent the payoff to the dictator and to the recipient respectively. By choosing another option, the dictator increased their own payoff by decreasing the recipient's payoff at the constant relative price of taking of $p_d = |d\pi_r/d\pi_d|$. This implies $\pi_d = 500 + \frac{1}{p_d}(500 - \pi_r)$. Equivalently, $\pi_r = 500 + p_d(500 - \pi_d)$, which represents the budget line of the take games. The only aspect in which the take games differ concerns the slope of the

[5] It should be noted that the original terms for 'dictator' and 'recipient' used by Brosig-Koch et al., namely 'a' and 'b', are replaced by 'd' and 'r' for the purpose of this thesis.

budget line; that is, they only differ in their relative prices of taking: Games 1, 2, 3, and 4 have relative prices of taking of ½, ⅔, 1, and 2 respectively. This implies that compared to game 3 (in which it is neither expensive nor cheap to take money due to the relative price of 1), it is relatively cheap to take money from the recipient in games 1 and 2, whereas it is relatively expensive to do so in game 4.

In wave 1, the subjects were informed that they would be randomly matched with one of the other students in the class. They would, however, not find out who their partners would be. In waves 2 and 3, the subjects were additionally told that they would not be matched with a student whom they had been matched with previously. An English translation of the standard group questionnaire can be found in Appendix A. The original Dutch version of the questionnaire can be found in Appendix B.

2.2.2 Charity treatment

The games in the charity treatment are identical to those in the standard treatment with the only difference being the recipients. More specifically, the recipients in the charity treatment are well-known charities instead of students: "KWF Kankerbestrijding", "UNICEF", and "De Dierenbescherming". In addition, the recipient group consists of only three recipients rather than a whole secondary school class of recipients. In wave 1, the subjects were informed that they would be randomly matched with one of the three charities, but were left ignorant about which one. In waves 2 and 3, the subjects were additionally told that they would not be matched with a charity they had been matched with previously. An English translation of the charity group questionnaire can be found in Appendix C. The original Dutch version of the questionnaire can be found in Appendix D.

The experimental procedure followed in the classrooms for both treatments is explained in Part 2 of Appendix E.

2.2.3 Differences from Brosig Koch et al.'s design

In order to increase the likelihood of replicating Brosig-Koch et al.'s findings in the standard dictator games, the experimental design of the standard treatment was made as similar to Brosig-Koch et al.'s experimental design as possible. Moreover, in order to be able to compare the stability of behaviour in standard dictator games to the stability of behaviour in charity dictator games, the experimental design of the charity treatment was made as similar as possible to that of the standard treatment (and

therewith to Brosig-Koch et al's experimental design); apart from, of course, the variable of interest; i.e. the recipient group. The designs of the standard and charity treatments are, however, not completely equal to Brosig-Koch et al.'s design. The main differences are explained below.

1. Volunteers vs. pseudo-volunteers

In terms of Eckel and Grossman (2000), the kind of subjects Brosig-Koch et al. (2017) used in their experimenters are "volunteers", meaning that subjects are "recruited to turn up at a particular time and place, and [are] promised a chance to earn a significant amount of money by participating". The kind of subjects in the present study are "pseudo-volunteers": subjects on whom the experiments are immediately conducted during class time. Eckel and Grossman (2000) found a higher level of pro-social behaviour among "pseudo-volunteers" compared to "volunteers". Therefore, a higher level of pro-social behaviour may be expected in the present study compared to Brosig-Koch et al.'s study.

2. Games to be played

Apart from having their subjects play four take games, Brosig-Koch et al. also had their subjects participate in four give games and two prisoner's dilemma games. In each wave, there were two sessions that took place during one week. In session 1, the take games as well as one prisoner's dilemma game were played; in session 2, the give games and the remaining prisoner's dilemma game were played. As the increase in selfish behaviour was largest in the take games in Brosig-Koch et al.'s study, those games are most relevant for the present study and are therefore the games that were used. There was no need for two sessions in each wave, as only four games were to be played in both the standard and the charity treatment.

3. Time period in between the waves

Brosig-Koch et al. adhered to a time period of four weeks in between the waves. Their reason to conduct the questionnaires with a time period in between rather than conducting them sequentially is that they were "particularly interested in the dynamics of social behaviour when repetitions can be considered as more disjunctive events (as is often the case in the field)". Due to practical limitations with respect to the possibilities at the secondary school and limitations on the amount of time, it was not feasible to adopt this four-week time period. Given these constraints, maximising the time period in between the waves resulted in 13 days between each wave. The dates on which the questionnaires were conducted are mentioned under the heading 'dates of the experiments' in Section 2.2.4.

4. Possibility of participating in future experiments

In wave 1, Brosig-Koch et al. communicated to their subjects that there was a possibility of them having to participate in more than one experiment, without revealing how many experiments they would have to participate in. In the present study, the possibility of being asked to join future experiments was, apart from in wave 1, also mentioned in wave 2. The reason for this is the conjecture that subjects would deem it illogical that there was yet another experiment after the second one, when they had only been informed about the possibility of participating in future experiments in wave 1. Another reason for this is that the teachers wanted the experimenter to be as transparent to the students as possible.

5. Informing the subjects about the games in wave 3

In wave 3, Brosig-Koch et al. (2017) informed their subjects about the games prior to the start of the experiment, which they did not do in waves 1 and 2, by providing a list of the decision situations in the Appendix. They did this to "mimic the experience-enhancing effect of repeating the experiment". However, informing the subjects only in wave 3 about the games creates different situations over the waves. This is not desirable as potential instability of behaviour could then possibly be due to the different situations created over the waves rather than due to the repetition of the games. Therefore, the subjects in the present study were not informed about the games in wave 3 prior to the experiment.

6. Conducting the questionnaires online vs. on paper

Brosig-Koch et al. conducted their questionnaires online using the 'z-Tree' software package. This allowed them to have the subjects submit their choices one after the other. Contrastively, this study employed a paper-based format, due to some practical reasons: 1) the secondary school where the experiments took place sometimes deals with faulty wireless internet connections, 2) students may not have their mobile phone with them, or 3) they may have an out-dated mobile phone on which the online questionnaire could not be completed. A potential disadvantage of conducting the questionnaires on paper is that it cannot be ascertained that subjects would not read ahead.

7. Forbidding subjects to talk to others

Brosig-Koch et al. added the following sentence to the instruction of their questionnaire: "Moreover, we would ask you not to talk about the experiment to others in order to avoid influencing the

behaviour of potential future participants". It was decided to exclude this from the questionnaire of the present research. The reason for this is that it may induce some students, probably the ones who typically do not adhere to the rules, to do talk to others. In Brosig-Koch et al.'s experiment, this threat may not be as present as it is in this experiment, since the subjects in their experiment had fewer opportunities to talk to each other: Their subjects come from "different fields and levels of study", whereas the subjects in the present study come from the same school and some even come from the same class.

8.1 Striving for identical decision situations over the waves in the standard treatment

Brosig-Koch et al. "randomly recruited subjects from different fields and levels of study via ORSEE". Furthermore, their experiment was designed as such that in each wave there was one group of dictators who were matched with newly recruited recipients. They aimed to create identical decision situations throughout the waves by doing this.

Having one group of dictators and a new group of recipients in each wave was not feasible in the present research due to a limited amount of classes available at the selected secondary school. Rather than having two groups in which one group consisted of only dictators and the other of only recipients, this study only used one group comprising both dictators and recipients. Each subject made decisions as if they were dictators. Only after this, it was randomly determined whether subjects were dictators or recipients. Following Brosig-Koch et al.'s approach, dictators were matched with recipients they had not been matched with before.[6] If payments of wave 1 would already be made after wave 1, before the next wave(s), a recipient could infer how much a dictator gave away to them, which could influence the recipient's decision in the next wave. In order to overcome this strategic interaction problem, the moment of payment occurred only at the end of wave 3 (more information about the payments can be found under the heading 'payment structure' in the current section). This method had as a result that the three situations were made as similar as possible, given the limitation of not being able to have one group of dictators being matched with newly recruited recipients each wave.

[6] Note that otherwise identical situations throughout the waves would not be ensured: The situation of someone making decisions regarding a recipient he or she has not previously made decisions for (which would be the case in wave 1) differs from the situation in which someone makes decisions regarding a recipient he or she has previously made decisions for (which would be the case in waves 2 and 3).

It appears that Brosig-Koch et al.'s reason to recruit *new* recipients in each wave is to prevent subjects from getting better acquainted with each other over the course of the waves, thus creating decision situations that are not identical throughout the waves. However, the subjects in the present study already knew each other and it is expected that a couple of weeks will not greatly affect their relationship. Thus, 'learning' in this sense is not expected to be a problem.

8.2 Striving for identical decision situations over the waves in the charity treatment

As already mentioned in Section 2.2.2, in wave 1, the subjects were informed that they would be randomly matched with one of the three charities, but were not told which one. In waves 2 and 3, they were additionally told that they would not be matched with a charity they had previously been matched with. This created identical decision situations throughout the waves. Note that this is necessary in order to be able to compare the stability of behaviour in charity dictator games to the stability of behaviour in standard dictator games, in which it was also made sure that the decision situations over the waves were identical.

Selecting three charities as recipients and stating which charity is the recipient in each wave before subjects fill out the questionnaires would create different situations depending on the wave. Selecting one and the same charity that is the recipient in all three waves would also create different situations over the waves for the reason explained in footnote 6. Making sure that a dictator would never be matched with the same charity is of particular importance in the present study due to the simplifying assumption of moral licensing, which states that moral licensing is equally present in standard dictator games and charity dictator games. This assumption is less likely to hold, and therefore less justified, when dictators in the standard treatment are ensured of being never matched with the same recipient more than once while dictators in the charity treatment would be matched with the same recipient more than once. This is because the effect of moral licensing is probably stronger in charity dictator games then: Dictators could tell themselves they had already donated to the *same* recipient.

9. Payment structure

In Brosig-Koch et al.'s design, all subjects were paid according to the dictators' decisions in each game, maintaining the exchange rate of '150 laboratory cents = 100 eurocents'. The moment of payments was at the end of each session, i.e. subjects were paid each wave after playing the games.

It was not possible to pay all subjects for all games in the present study due to capital constraints. Instead, the following applied for each secondary school class in both treatments: In wave 1, one randomly selected game was paid out to one randomly selected dictator and one randomly selected recipient. The dictator that had been selected in wave 1 remained the dictator in waves 2 and 3 as well. The recipients and games were randomly selected again in the latter waves, while making sure that a dictator was never matched with the same recipient more than once. The reason for having a single dictator throughout all waves rather than a different dictator in each wave is to treat the dictators in the present study that were paid out the same as the dictators in Brosig-Koch et al.'s study that were paid out, given the limitation that not all dictators could be paid out in the present study. The payment procedure can be found in Part 3 of Appendix E. The exchange rate used in this experiment was '200 laboratory cents = 100 eurocents', again due to capital constraints.

Rather than paying them after each wave, subjects in both treatments were paid only at the end of wave 3. The reason why this is the case for the standard group has already been explained by the text under the heading 'striving for identical decision situations over the waves in the standard treatment' in the current section. It is noteworthy that subjects in the charity group could have been paid after each wave; there would still be identical decision situations throughout the waves. However, the moment that subjects get paid, they learn that they actually get the money they 'won' and that their decisions remain anonymous. According to Brosig-Koch et al., assuming that the EDE is rooted in subjects' uncertainty, learning about the decision environment causes the EDE to reduce. Paying out subjects in the charity treatment after each wave and subjects in the standard treatment only after wave 3 would thus imply that EDE would decrease more in the charity treatment than in the standard treatment over the waves. This would suggest that the EDE was more present in the charity treatment in the first place, while that does not have to be the case at all. In order to make the decrease in the EDE equal in both treatments if the EDE was equally present in wave 1, subjects in both treatments should learn about the decision environment equally. Subjects in the charity treatment were therefore also only paid in wave 3.

2.2.4 Explanations

Selection of subjects for the analysis

It is worth noting that, as mentioned in the experimental procedure followed in the classrooms, which can be found in Part 2 of Appendix E, students were informed that if they had not participated in a previous wave, they could not participate in the current wave. In order to find out if

students were present who had not participated in (a) previous wave(s), they were simply asked. However, by observing the student numbers after conducting the experiments, it appeared that some students had not honestly reported this. The data from these students will not be used in this study and these students were also excluded from having the possibility to win money in the waves in which they were not supposed to participate. Table A in Appendix F shows the number of subjects in each wave per treatment and secondary school class. The table distinguishes between subjects who filled out the questionnaire and should have done so and subjects who filled it out while they should not have done so. As can be observed in the table, the numbers of usable subjects are 14, 20, 19, and 19 in the *havo* standard treatment, *havo* charity treatment, *vwo* standard treatment, and *vwo* charity treatment respectively. This results in data from 33 subjects in the standard group and data from 39 subjects in the charity group. Please note that using the data of only those subjects who were present in all waves provided a balanced panel; i.e. all subjects were observed in all waves. The statistical analyses could, however, have already been performed with those subjects who were present in at least two waves. But to stay as close as possible to Brosig-Koch et al's design, it was decided to only use the subjects who were present in all waves.

Dates of the experiments

In the *havo* classes, wave 1 took place on 17 May for the standard group and on 18 May for the charity group. Wave 2 took place on 31 May and 1 June for the standard group and charity group respectively; and wave 3 on 14 and 15 June for the standard group and charity group respectively. In the *vwo* classes, wave 1 took place on 22 and 23 May for the standard and charity group respectively. Wave 2 took place on 5 June for the standard group and for the charity group on 6 June; and wave 3 took place on 19 and 20 June for the standard group and charity group respectively.

Anonymity

This study applied a double-blind procedure, i.e. anonymity between the subjects as well as between the subjects and the experimenter was maintained. Anonymity between the subjects was established by not informing the subjects whom they were matched with. Anonymity between the subjects and the experimenter was established by the experimental procedure followed in the classrooms and the payment procedure. As can be seen in the questionnaires, subjects were asked to indicate their student e-mail address (containing student numbers rather than names), class, and gender on their questionnaire. Once the subjects had finished their questionnaire, they turned it over with the text

facing the table so that subjects' personal information was not visible. In this way, the experimenter could not link student e-mail addresses to the students when collecting the questionnaires. Also, the questionnaires were shuffled after all questionnaires had been collected so that the experimenter could also not link questionnaires to students by remembering the order in which the students were seated in class. Concerning the payment procedure: Winning subjects did not have to reveal their names in order to get paid, as it is possible to transfer money entering a random letter for the name of the receiver of the money. The only additional personal information the winning subjects would be asked for was their bank account number – and the subjects were informed about this before they filled out the questionnaires. The experimenter could thus not link decisions to subjects. Anonymity between the subjects and the teacher was also guaranteed as the teacher's only role was to verify that the experimenter had transferred the money to the subjects by observing anonymised screenshots of the money transfers. The teacher was in no way informed about the decisions of the subjects.

The reason to apply a double-blind procedure is to replicate Brosig-Koch et al.'s approach. It is important to do so as it has been found that the degree to which dictators are anonymous with respect to each other and with respect to the experimenter influences dictators' degree of pro-social behaviour (Bohnet & Frey, 1999; Hoffman, McCabe, & Smith, 1996).

Verification

As illustrated in the experimental procedure followed in the classrooms, which can be found in Part 2 of Appendix E, the experimenter was slightly unclear about when the 'winning' subjects would be contacted and paid out. The reason for this is that the subjects would be contacted only after wave 3, but the subjects were not allowed to know that there were going to be three waves. This obscurity may have caused a certain lack of credibility regarding whether the payments would be carried out, which could result into subjects not putting in their utmost effort with respect to the decision situations. Another factor that may have caused a certain lack of credibility is the double-blind procedure: "Bolton, Katok and Zwick (1998) argue that greater anonymity makes the participants sceptical about whether the transfers will be carried out" (as qtd. in Andreoni, Harbaugh, & Vesterlund, 2008). In an attempt to increase the credibility that the payments would actually be carried out, the experimenter assured the subjects to send anonymised screenshots of the transfers to their teacher so that the payments would be verified.

All questionnaires conducted by experimenter

The extent and direction of the EDE in experiments can depend on who the experimenter is. As Kantowitz, Roediger, and Elmes (2008) state in their textbook in experimental psychology: "The gender, race, and ethnicity of the experimenter are also potential experimenter effects". Therefore, it was deemed necessary that the same person conducted all experiments. In the first wave, the experimenter needed to verbally explain some things to the subjects, which could be best done by the experimenter herself. It was therefore decided that the experimenter was the one conducting all the experiments.

Tracking the subjects

The subjects were tracked throughout the waves by observing the student e-mail addresses on the questionnaires.

Selecting the number and nature of charities in the charity treatment

In order to create a situation as inartificial as possible, the pool of charities was made as small as possible (in real life, donations are typically made to a particular recipient). Given that at least three charities needed to be selected in order to create identical situations over the waves, as explained under the heading 'striving for identical decision situations over the waves in charity dictator treatment' in Section 2.2.3, the pool in the charity treatment consisted of three charities. The charities were carefully selected, making sure that they were both well-known (contributing to creating a less artificial situation) and uncontroversial.

While subjects in the standard treatment probably like part of the recipients and also dislike part of the recipients, the *uncontroversial* charities should have the result that subjects in the charity treatment like (or at least not dislike) all of the recipients. This may cause that the absolute level of pro-social behaviour is greater in the charity dictator games than in the standard dictator games, which makes the study conservative in the sense that the design was made as such that in the charity dictator games there could be more room for selfish behaviour to increase.

2.3 Hypothesis testing

The statistical software package 'Stata' will be used in order to analyse the results. Differences will be termed 'significant' for p-values lower than 0.050 and termed 'marginally significant' for p-values lower than 0.100 and higher than or equal to 0.050. The results section is split into four parts. Part 1

investigates whether behaviour is stable in standard dictator games or whether selfish behaviour increases over waves, as found by Brosig-Koch et al. Part 2 investigates this for charity dictator games. Part 3 investigates whether behaviour in charity dictator games is more stable than in standard dictator games, and will thereby test the hypothesis of this study: *People exhibit more stable behaviour in charity dictator games compared to standard dictator games*. Using the information obtained in the previous parts, part 4 provides an answer to the sub-research question and the research question.

2.3.1 Parts 1 & 2

The statistical tests that will be performed in the first two parts are a subset of the statistical tests used by Brosig-Koch et al. For both parts, this concerns the following tests: the Wilcoxon signed-rank test, Jonckheere-Terpstra test, an ordered probit regression, and an ordered logit regression. It should be noted that in the present study there is a specified order of ranks a priori, based on Brosig-Koch et al.'s findings: It is expected that the amounts taken increase over the waves. This implies that one-tailed tests will be applied in the present study, as opposed to the two-tailed Wilcoxon signed-rank tests and two-tailed ordered probit and ordered logit regressions applied by Brosig-Koch et al, who did not have a prior expectation as to whether selfish behaviour would increase or decrease. The tests in the present study will be applied on the aggregates of the games rather than on each game separately. Parts 1 and 2 will use the data concerning the usable subjects in the standard group (33 subjects) and in the charity group (39 subjects) respectively.

The Wilcoxon signed-rank test is a rank-based non-parametric test that detects if the distributions of two paired samples are the same. Both in part 1 and in part 2, the test will be performed three times: once comparing the amounts taken in wave 1 with those taken in wave 2; once comparing the amounts taken in wave 2 with those taken in wave 3; and once comparing the amounts taken in wave 1 with those taken in wave 3. The null-hypothesis in each case is that the average amount taken in the earlier wave is higher than or equal to the average amount taken in the later wave. The alternative hypothesis in each case is that the average amount taken in the earlier wave is lower than the average amount taken in the later wave. That is, if the null hypothesis will be rejected, the amounts taken significantly increased from one specific wave to another. It should be noted that there are four observations per subject in each wave since each subject played four dictator games in each wave. To perform the Wilcoxon signed-rank tests, however, there should be just one observation per subject in each wave. Therefore, the average over the four observations will be calculated for each subject in each wave - and the tests will be applied on these averages. Note that all observations are taken into

account in this way. Although Brosig-Koch et al. do not explicitly state whether they followed a similar approach or performed Wilcoxon signed-rank tests on each game separately, their paper suggests the former.

For the same reason that Brosig-Koch et al. performed the Jonckheere-Terpstra test, the purpose of the test in the present study is to check the results obtained from the Wilcoxon signed-rank tests. It is a rank-based non-parametric test that detects if the distributions of k different samples are the same. It is appropriate to use when there is a specified order of ranks a priori, which is the case in this study, as already discussed above. The null hypothesis is that the median amount taken in wave 1 is higher than or equal to the median amount taken in wave 2; and the median amount taken in wave 2 is higher than or equal to the median amount taken in wave 3. The alternative hypothesis is that the median amount taken in wave 1 is lower than or equal to the median amount taken in wave 2; and the median amount taken in wave 2 is lower than or equal to the median amount taken in wave 3, with at least one strict inequality. That is, if the null-hypothesis will be rejected, there is a significant overall upward trend in the amounts taken over the waves. It should be noted that the Jonckheere-Terpstra test assumes that the observations are independent, both within and between groups, in which 'groups' indicate the waves in this case. In order to meet the assumption of independent observations within groups, the test will be applied on the averages over the games per subject and per wave. The assumption of independent observations between groups is not met, however, as the same subjects are observed in each wave. Not meeting this assumption implies that the data are not optimally used: Whereas finding a significant effect does mean that there is an effect, finding no significant effect does not necessarily mean that there is no effect.

The ordered probit regression and ordered logit regression are choice data regression models for ordered multinomial outcomes. That is, these regressions are appropriate to use when the dependent variable is measured in multiple categories that are ordered.[7] They can be applied on cross-sectional data as well as on panel data. The latter is the case in this study. As the observations should be treated "as independent across subjects, but not within" (Brosig-Koch et al., 2017), a clustering specification will be used. The clustering will be done at the subjects level, replicating Brosig-Koch et al.'s approach. Both the ordered probit regression and the ordered logit regression will regress the amount taken, in categories, on dummies for waves 2 and 3 and games 2, 3, and 4 as explanatory

[7] Note that the dependent variable in this study, the amount taken, is measured in 11 categories that are clearly ordered.

variables, which replicates Brosig-Koch et al.'s regressions. Positive wave and game coefficients indicate that the probability to be in a higher category increases with (a) repetition(s) of a particular game and with a higher relative price respectively; for negative coefficients it is the other way around. As higher categories correspond with higher amounts taken in this study, a positive coefficient implies that the probability to take more increases; i.e. that the amount taken increases. While the exact magnitude of the coefficients cannot be directly interpreted from the regression output, what it roughly tells is the greater the magnitude of the coefficient, the higher the probability to be in a higher category; i.e. the greater the increase in the amount taken. The sign and significance of the two wave coefficients will be inspected to determine whether playing a specific game a second/third time, compared to playing it the first time, has a significantly positive effect on the amount taken. The effect on the amount taken of playing a specific game a third time, compared to playing it the second time, cannot be directly inferred from the regression output, however. In order to obtain this information, the lincom command will be used.[8] Additionally, the effect of relative prices on taking behaviour will be investigated. The sign and significance of the three game coefficients will be inspected to determine whether the amount taken significantly decreases from game 1 to games 2, 3, and 4; i.e. when the relative price of taking increases.

2.3.2 Part 3

In part 3, difference in difference analyses will be performed, using one-tailed tests. The data concerning both the usable subjects in the standard group and in the charity group will be used for this, i.e. data from 72 subjects. First, a difference in difference analysis will be performed, on the aggregates of the games, with the ordered probit and ordered logit regressions from the first two parts, adding the following independent variables: interaction dummy variables between the charity treatment and waves; and a dummy variable for the charity treatment. These regressions are performed to stay close to Brosig-Koch et al.'s approach. However, as a dependent variable consisting of 11 categories is in fact enough for it to be considered as continuous (Johnson & Creech, 1983), the difference in difference analysis will be repeated by a linear panel data regression. The independent variables will remain the same as in the ordered probit and ordered logit regressions, but the dependent variable, the amount taken, will be measured in amounts rather than in categories.

[8] Command 'lincom' can provide information on linear combinations of coefficients after any estimation command.

Either a random effects model or a fixed effects model will be performed. The random effects model exploits both within and between variations while the fixed effects model exploits only within variation. As a result, the former model is more efficient. However, for the random effects model to be consistent and thus the appropriate model to use, the fixed part of the error term should be uncorrelated with the variables of interest, which is not necessary for the fixed effects model to be consistent because all components that stay fixed over time cancel out in this type of model. Whether the random effects model or the fixed effects model is the appropriate model to use will be tested with the Hausman test. The null-hypothesis of the Hausman test is that the difference in coefficients is not systematic. Rejecting this hypothesis indicates that there is a significant systematic difference between the coefficients obtained by the random effects model and the fixed effects model. This suggests that the fixed part of the error term is correlated with the variables of interest, implying the random effects model is inconsistent. Although the fixed effects model is less efficient, this is the appropriate model to use in this case. Failing to reject the null-hypothesis implies that there is no evidence for a significant systematic difference between the coefficients obtained by the random effects model and the fixed effects model. This suggests that the fixed part of the error term is uncorrelated with the variables of interest, implying that the random effects model is consistent. The random effects model is the appropriate model to use in this case as it is likely to be consistent and is more efficient than the fixed effects model as well.

Besides investigating whether behaviour in charity dictator games is more stable than in standard dictator games on the aggregates of the games, this will also be investigated for the games separately. By doing so, it is investigated whether a potential difference in stability of behaviour in standard dictator games and charity dictator games depends on the relative price of taking. Linear panel data regression models will be used for this purpose with the same dependent and independent variables as used in the linear panel data regression of the aggregates of the games (except that no dummy variables for the games are included as independent variables). Whether random effects models or fixed effects models are the appropriate models to use will again be tested with the Hausman test.

Table 2 provides an overview of the linear combinations of variables that should be inspected to investigate whether the amount taken increases moving from one wave to another. It does so for the standard treatment and the charity treatment. More importantly for testing the hypothesis, though,

the table also shows which variables provide the difference in difference coefficients; i.e. the linear combinations of variables that should be inspected to investigate whether the amount taken increases less in the charity treatment compared to the standard treatment moving from one wave to another.

Table 2: Linear combinations of variables that should be inspected

		Variables
ST	W12	Wave 2
	W23	Wave 3 – Wave 2
	W13	Wave 3
CT	W12	Wave 2 + Charity * Wave 2
	W23	Wave 3 + Charity * Wave 3 – Wave 2 – Charity * Wave 2
	W13	Wave 3 + Charity * Wave 3
Diff	W12	Charity * Wave 2
	W23	Charity * Wave 3 – Charity * Wave 2
	W13	Charity * Wave 3

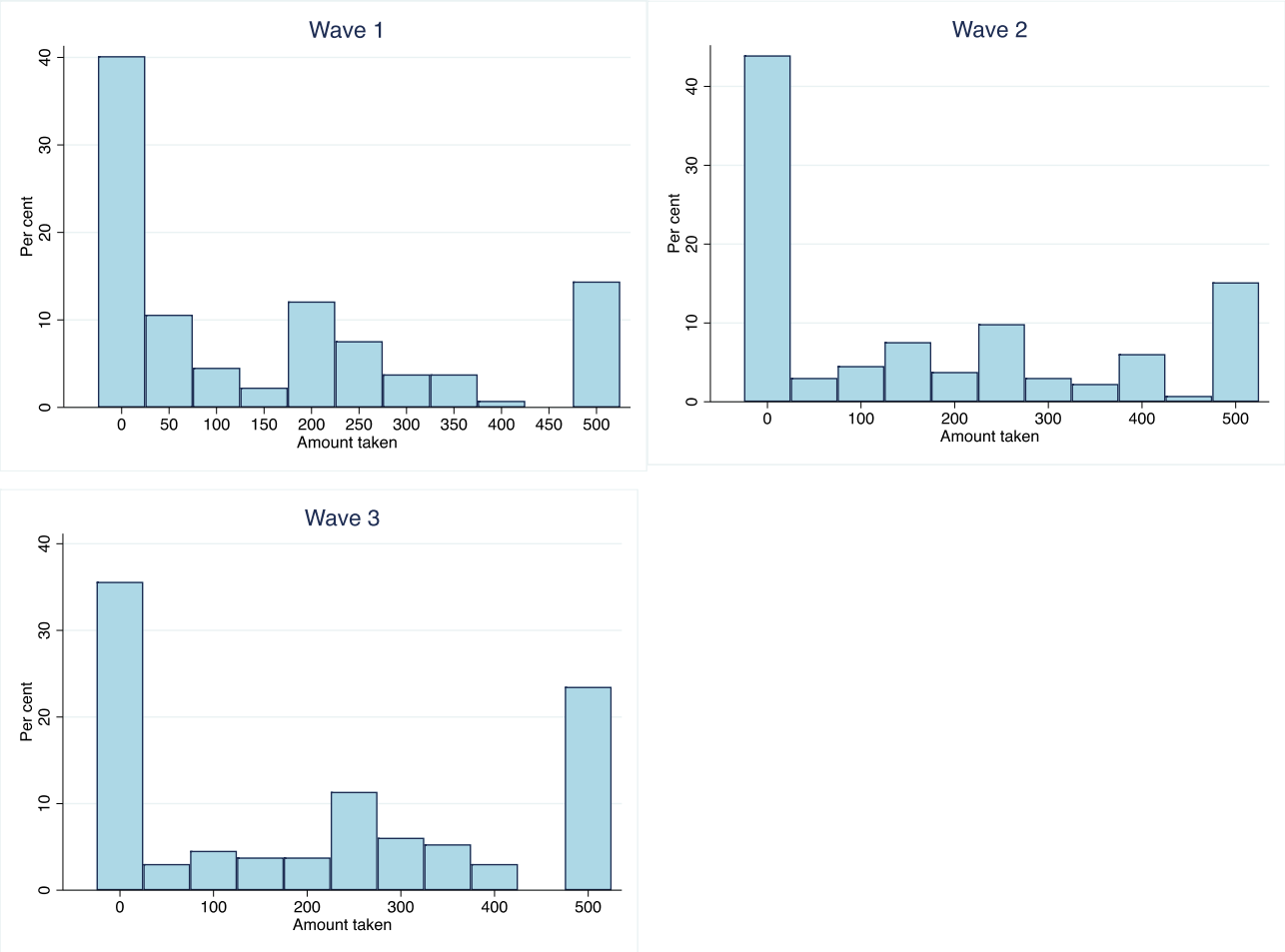
Notes: 'ST', 'CT', and 'Diff' represent 'standard treatment', 'charity treatment', and 'difference between the treatments' respectively. 'W12, 'W23', and 'W13' stand for 'the increase from wave 1 to wave 2', 'the increase from wave 2 to wave 3', and 'the increase from wave 1 to wave 3' respectively. 'Charity * Wave 2' and 'Charity * Wave 3' represent interaction terms between the charity treatment and wave 2; and the charity treatment and wave 3 respectively.

3. Results

3.1 Standard dictator games

First, some descriptive statistics will be provided in order to raise a general understanding of the results derived from the standard dictator games. Figure 1 displays histograms of the amounts taken per wave, aggregated over the games; i.e., each wave contains the data gathered from all four games. Please note that this implies that the observations are in terms of games rather than in terms of dictators and that the number of observations is 132 in each wave (33 usable subjects playing four games per wave).

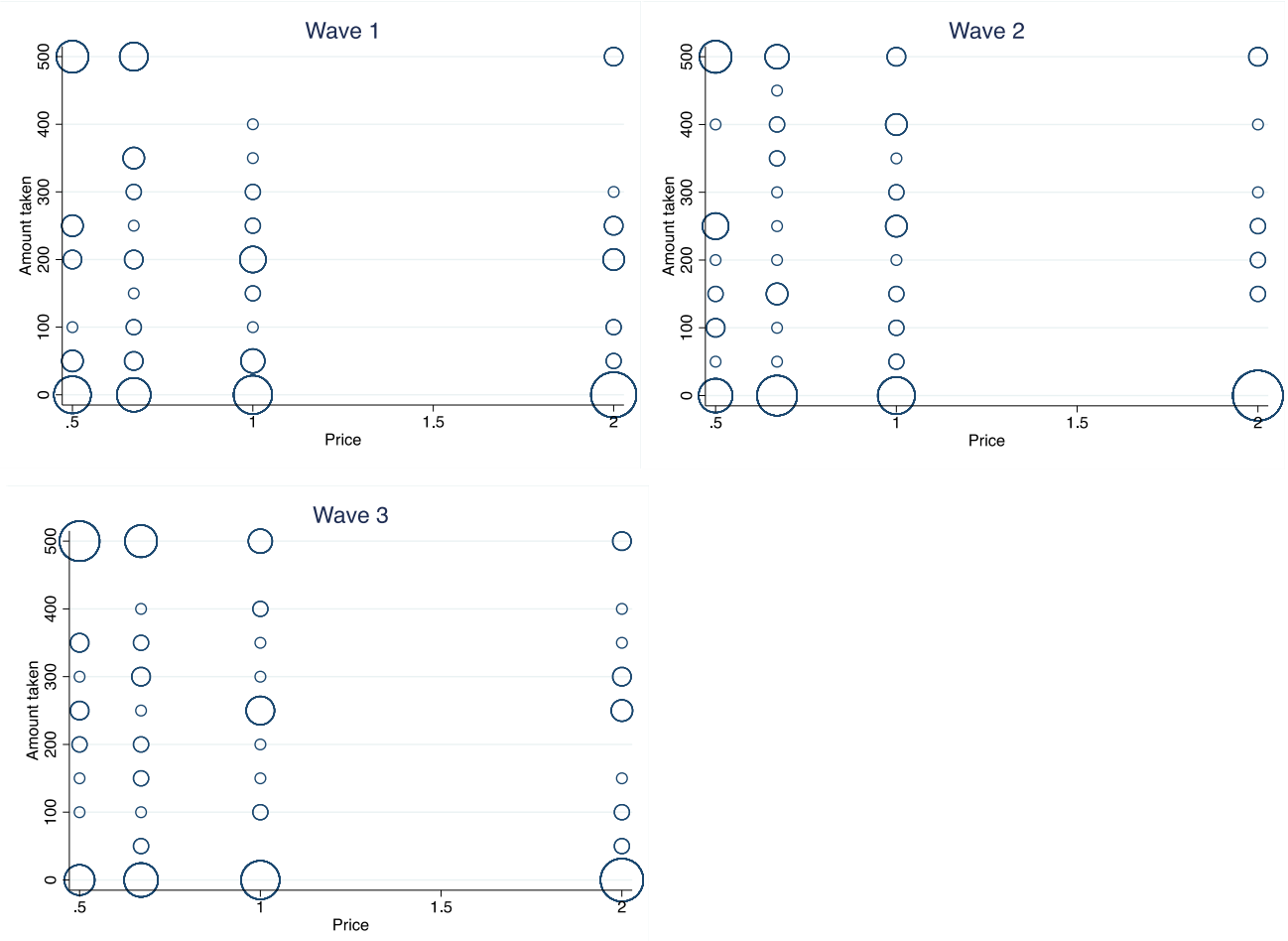
Figure 1. Histograms of the amounts taken per wave in the standard group



The amount taken most frequently is 0 eurocents in each wave, the minimum amount to be taken, followed by the amount taken of 500 eurocents, the maximum amount to be taken. In wave 1, dictators took nothing in 53 games and everything in 19 games; in the remaining 60 games they took something but less than the maximum amount. The average amount taken is 156.06 eurocents, with a standard deviation of 178.74 eurocents. In wave 2, both the frequency of games in which dictators took nothing and the frequency of games in which they took everything increased slightly: Dictators took nothing in 58 games and everything in 20 games. The average amount taken increased to 170.08 eurocents, with a standard deviation of 189.40 eurocents. In wave 3, the frequency of games in which dictators took nothing decreased to 47 while the frequency of games in which they took everything further increased to 31. The average amount taken further increased to 214.02 eurocents, with a standard deviation of 200.03 eurocents.

As the analysis above does not consider each game separately but aggregates over the games per wave, it ignores the effect that different relative prices may have on taking behaviour. The scatterplots in Figure 2 account for this. Matching the approach of Brosig-Koch et al., the size of the dots is proportional to the number of observations.

Figure 2. Scatterplots of the amounts taken per wave in the standard group



The descriptive statistics reveal that dictators, on average, did not take the maximum amount of money, but left a considerable part for the recipients. They also suggest that selfish behaviour increased over the waves: The average amount taken increases, the frequency of games in which dictators took the maximum (minimum) amount increases (decrease), and the distributions of the amounts taken shift to the right over the waves. Whether selfish behaviour *significantly* increased over the waves will be statistically tested with the Wilcoxon signed-rank test, the Jonckheere-Terpstra test, an ordered probit regression, and an ordered logit regression.

Three Wilcoxon signed-rank tests are performed: one comparing the amounts taken in wave 1 with those taken in wave 2; one comparing the amounts taken in wave 2 with those taken in wave 3; and one comparing the amounts taken in wave 1 with those taken in wave 3. By doing so, p-values of respectively 0.326, 0.008, and 0.053 are obtained. That is, the null-hypothesis that the average amount taken in the earlier wave is higher than or equal to the average amount taken in the later wave is rejected when comparing wave 2 and wave 3 to each other and marginally rejected when comparing wave 1 and wave 3. This implies that the amounts taken significantly increased from wave 2 to wave 3. Additionally, there is some evidence that the amounts taken increased from wave 1 to wave 3 as well.

Performing the Jonckheere-Terpstra test results in a p-value of 0.078, which leads to marginally rejecting the null-hypothesis that the median amount taken in wave 1 is higher than or equal to the median amount taken in wave 2; and the median amount taken in wave 2 is higher than or equal to the median amount taken in wave 3. That is, there is some evidence for an overall upward trend in the amounts taken over the waves.

The results obtained from applying an ordered probit regression and an ordered logit regression on the amount taken with dummies for waves 2 and 3 and games 2, 3, and 4 as explanatory variables can be found in Table 3.

Table 3: Ordered probit and logit models on the amount taken in the standard group

	Ordered probit			**Ordered logit**		
	Coefficient	z	p-value	Coefficient	z	p-value
Wave 2	0.021 (0.166)	0.13	0.448	0.077 (0.270)	0.28	0.390
Wave 3	0.398** (0.184)	2.16	0.015	0.730** (0.317)	2.30	0.011
Game 2	-0.236* (0.145)	-1.62	0.053	-0.329* (0.243)	-1.36	0.087
Game 3	-0.552*** (0.224)	-2.46	0.007	-0.829** (0.374)	-2.22	0.013
Game 4	-0.960*** (0.248)	-3.87	0.000	-1.567*** (0.429)	-3.65	0.000
Observations	396			396		

Notes: Standard errors are in parentheses, clustered at subjects level.
*** Significant at the 1% level.
** Significant at the 5% level.
* Significant at the 10% level.

The positive coefficients of 'wave 2' suggest that playing a specific game a second time, compared to playing it the first time, has a positive effect on the amount taken. However, this effect is not significant. Playing a specific game a third time, compared to playing it the second time, does have a significantly positive effect on the amount taken.[9] Playing a specific game a third time, compared to playing it the first time, also has a significantly positive effect on the amount taken. Increasing the relative price of taking from game 1 to game 2 has a marginally significantly negative effect on the amount taken; increasing the relative price from game 1 to game 3 or to game 4 has a significantly negative effect.

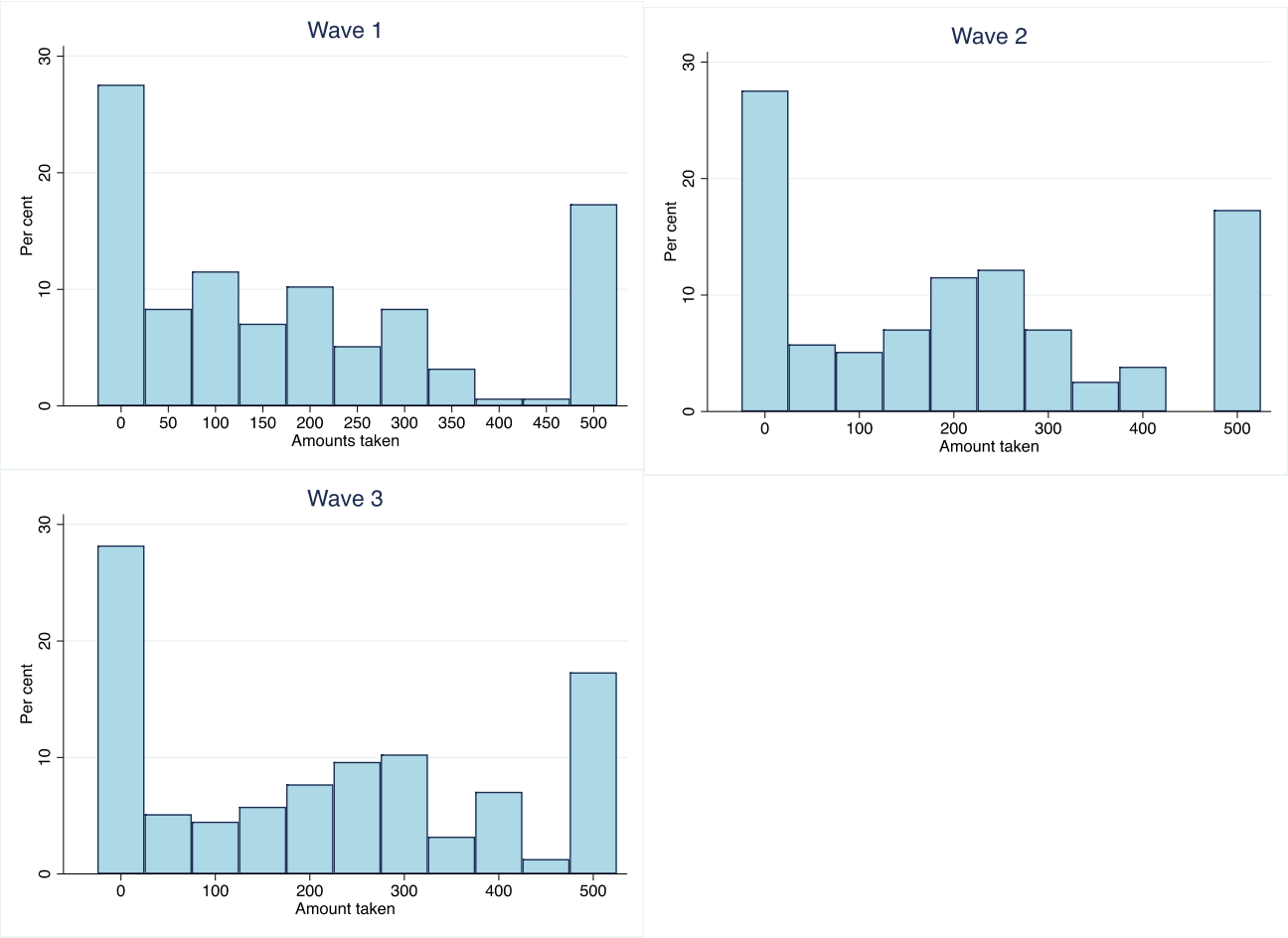
It can thus be concluded from the statistical tests performed in the current section that selfish behaviour increases over the waves in standard dictator games. By and large, the findings in the standard dictator games of the present study replicate the findings of Brosig-Koch et al.'s original study. Please refer to Section 4.1.1 for a detailed comparison.

[9] Coefficients of 0.377 and 0.654 with corresponding p-values of 0.002 and 0.001 are obtained by the ordered probit regression and orderd logit regression respectively.

3.2 Charity dictator games

In order to raise a general understanding of the results derived from charity dictator games, some descriptive statistics will be provided first. Figure 3 presents histograms of the amounts taken per wave, aggregated over the games; i.e., each wave contains the data gathered from all four games. As in the histograms in Section 3.1, this implies that the observations are in terms of games. The number of observations is 156 in each wave (39 usable subjects playing four games per wave).

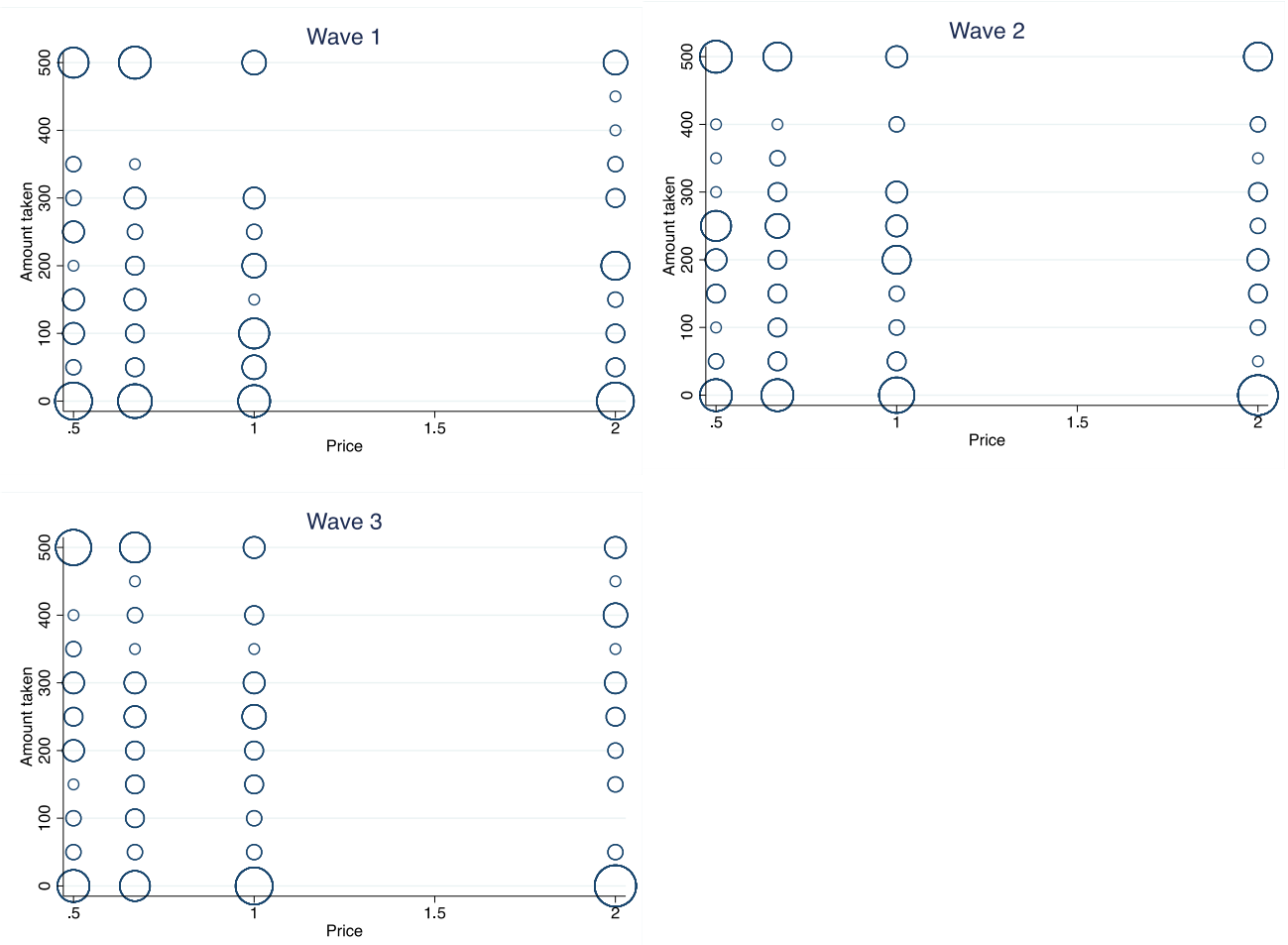
Figure 3. Histograms of the amounts taken per wave in the charity group



Matching the findings in Section 3.1, the amount taken most frequently is 0 eurocents in each wave, followed by the amount taken of 500 eurocents. The frequency of games in which dictators took nothing is 43 in waves 1 and 2 and 44 in wave 3; the frequency of games in which they took everything is 27 in each wave. The average amount taken is 187.82 eurocents in wave 1. This amount increased to 204.17 eurocents in wave 2, and further increased to 217.63 eurocents in wave 3, with standard deviations of 179.46, 178.61, and 184.72 eurocents respectively.

While the analysis above aggregates over the games, the scatterplots in Figure 4 contain information on each game separately. As in the previous scatterplots, the size of the dots is proportional to the number of observations

Figure 4. Scatterplots of the amounts taken per wave in the charity group



As in the previous section, the descriptive statistics reveal that dictators, on average, did not take the maximum amount of money, but left a considerable part for the recipients. They do not provide an unambiguous judgement as to whether selfish behaviour increased over the waves, however: The average amount taken increases slightly and the distributions of the amounts taken shift to the right to some extent over the waves. However, the frequencies of games in which dictators took the minimum amount and the maximum amount hardly change. Whether or not selfish behaviour significantly increased over the waves will be statistically tested with the Wilcoxon signed-rank test, the Jonckheere-Terpstra test, an ordered probit regression, and an ordered logit regression.

Three Wilcoxon signed-rank tests are performed. They result in the following p-values: '0.140' when comparing the amounts taken in wave 1 with those taken in wave 2; '0.305' when comparing the amounts taken in wave 2 with those taken in wave 3; and '0.117' when comparing the amounts taken in wave 1 with those taken in wave 3. The null-hypothesis that the average amount taken in the earlier wave is higher than or equal to the average amount taken in the later wave is thus not rejected comparing any of the waves. There is thus no evidence for an increase in the amounts taken over the waves.

Performing the Jonckheere-Terpstra test results in a p-value of 0.189. The null-hypothesis that the median amount taken in wave 1 is higher than or equal to the median amount taken in wave 2; and the median amount taken in wave 2 is higher than or equal to the median amount taken in wave 3 is thus not rejected. That is, there is no evidence for an overall upward trend in the amounts taken over the waves.

The results obtained from applying an ordered probit regression and an ordered logit regression on the amount taken with dummies for waves 2 and 3 and games 2, 3, and 4 as explanatory variables can be found in Table 4.

Table 4: Ordered probit and logit models on the amount taken in the charity group

	Ordered probit			**Ordered logit**		
	Coefficient	z	p-value	Coefficient	z	p-value
Wave 2	0.145 (0.154)	0.94	0.174	0.246 (0.275)	0.89	0.187
Wave 3	0.276* (0.189)	1.46	0.072	0.486* (0.334)	1.46	0.072
Game 2	-0.042 (0.122)	-0.35	0.363	0.038 (0.178)	0.21	0.417
Game 3	-0.459** (0.208)	-2.20	0.014	-0.553* (0.357)	-1.55	0.061
Game 4	-0.419** (0.231)	-1.81	0.035	-0.431 (0.393)	-1.10	0.136
Observations	468			468		

Notes: Standard errors are in parentheses, clustered at subjects level.
*** Significant at the 1% level.
** Significant at the 5% level.
* Significant at the 10% level.

The positive coefficients of 'wave 2' suggest that playing a specific game a second time, compared to playing it the first time, has a positive effect on the amount taken. However, this effect is not significant. Comparing playing a specific game a third time to playing it the second time generates positive coefficients, suggesting a positive effect on the amount taken. However, also this effect is not significant.[10] Playing a specific game a third time, compared to playing it the first time, has a marginally significantly positive effect on the amount taken. Increasing the relative price of taking from game 1 to game 3 has significantly negative effect on the amount taken according to the ordered probit model and a marginally significantly negative effect according to the ordered logit model; increasing the relative price of taking from game 1 to game 4 has a significantly negative effect according to the ordered probit model, but according to the ordered logit model this effect is not significant.

[10] Coefficients of 0.131 and 0.241 with corresponding p-values of 0.164 and 0.149 are obtained by the ordered probit regression and orderd logit regression respectively.

Based on the statistical tests performed in the current section, it can thus not be concluded that selfish behaviour increases over the waves in charity dictator games.

3.3 Comparison of standard and charity dictator games

The previous two sections seem to indicate that behaviour is more stable in charity dictator games than in standard dictator games: the Wilcoxon test found no significant increase in the amounts taken comparing any two waves in the charity dictator games while it found a significant increase comparing waves 2 and 3 and a marginally significant increase comparing waves 1 and 3 in the standard dictator games; the Jonckheere test found no significant overall upward trend in the amounts taken over the waves in the charity dictator games and a marginally significant overall upward trend in the standard dictator games; regarding the ordered probit and ordered logit regressions, the greater magnitude of the coefficients in the standard dictator games (except for the comparison of playing a specific game a second time to playing it for the first time) plus the fact that they were more often significant also provides an indication that behaviour is more stable in charity dictator games.

In order to statistically test whether behaviour in charity dictator games is indeed more stable than in standard dictator games, difference in difference analyses will be performed. First, a difference in difference analysis will be performed, on the aggregates of the games, with the ordered probit and ordered logit regressions from Sections 3.1 and 3.2, adding the following independent variables: interaction dummy variables between the charity treatment and waves; and a dummy variable for the charity treatment. This generates the results presented in the columns 'ordered probit' and 'ordered logit' in Table B in Appendix G. The difference in difference analyses obtained from these regressions are shown in the columns 'ordered probit' and 'ordered logit' in Table 5 below.

Table 5: Difference in difference analyses on the aggregates of the games

		Ordered probit			**Ordered logit**			**Random effects**		
		Coefficient	z	p-value	Coefficient	z	p-value	Coefficient	z	p-value
ST	W12	0.046 (0.188)	0.25	0.401	0.148 (0.319)	0.46	0.323	14.015 (16.055)	0.87	0.192
	W23	0.423*** (0.153)	2.77	0.003	0.789*** (0.279)	2.83	0.002	43.939*** (16.055)	2.74	0.003
	W13	0.470** (0.213)	2.21	0.014	0.936*** (0.395)	2.37	0.009	57.955*** (16.055)	3.61	0.000
CT	W12	0.101 (0.126)	0.80	0.212	0.157 (0.210)	0.75	0.227	16.346 (14.769)	1.11	0.134
	W23	0.104 (0.111)	0.93	0.176	0.184 (0.183)	1.01	0.156	13.462 (14.769)	0.91	0.181
	W13	0.205* (0.155)	1.32	0.093	0.342* (0.257)	1.33	0.092	29.808** (14.769)	2.02	0.022
Diff	W12	0.054 (0.226)	0.24	0.405	0.010 (0.380)	0.03	0.488	2.331 (21.815)	0.11	0.456
	W23	-0.320** (0.190)	-1.68	0.047	-0.604** (0.334)	-1.81	0.035	-30.478* (21.815)	-1.40	0.081
	W13	-0.265 (0.263)	-1.01	0.156	-0.595 (0.468)	-1.27	0.102	-28.147* (21.815)	-1.29	0.099

Notes: 'ST', 'CT', and 'Diff' represent 'standard treatment', 'charity treatment', and 'difference between the treatments' respectively. 'W12', 'W23', and 'W13' stand for 'the increase from wave 1 to wave 2', 'the increase from wave 2 to wave 3', and 'the increase from wave 1 to wave 3' respectively. Standard errors are in parentheses, clustered at subjects level for ordered probit and ordered logit regressions.

*** Significant at the 1% level.

** Significant at the 5% level.

* Significant at the 10% level.

It is found that the increase in the amount taken when playing a specific game a second time, compared to playing it the first time, is greater in the charity dictator games than in the standard dictator games. However, this effect is not significant. The increase in the amount taken when playing a specific game a third time, compared to playing it the second time, is significantly greater in the standard dictator games than in the charity dictator games. The increase in the amount taken

when playing a specific game a third time, compared to playing it the first time, is also greater in the standard dictator games than in the charity dictator games. However, this effect is not significant. Thus, based on the ordered probit and ordered logit models, instability of pro-social behaviour between waves 1 and 2 and between waves 1 and 3 does not significantly differ between standard dictator games and charity dictator games, but between waves 2 and 3, pro-social behaviour is significantly more stable in charity dictator games than in standard dictator games.

The difference in difference analysis will be repeated by a linear panel data regression. First, it is tested with the Hausman test whether the random effects model or the fixed effects model is the appropriate model to use. This test results in a p-value of 0.997, which means that the null-hypothesis that the difference in coefficients is not systematic is not rejected. There is thus no evidence for a significant systematic difference between the coefficients obtained by the random effects model and the fixed effects model. This suggests that the fixed part of the error term is uncorrelated with the variables of interest, implying that the random effects model is consistent. As the random effects model is likely to be consistent and is more efficient then the fixed effects model as well, this model will be applied. The results of the random effects regression can be found in Table B in Appendix G; the difference in difference analysis obtained from this regression in Table 5 above.

The increase in the amount taken when playing a specific game a second time, compared to playing it the first time, is greater in the charity dictator games than in the standard dictator games. However, this effect is not significant. The increase in the amount taken when playing a specific game a third time, both compared to playing it the first time and the second time, is marginally significantly greater in the standard dictator games than in the charity dictator games. Thus, based on the fixed effects model, instability of pro-social behaviour between waves 1 and 2 does not significantly differ between standard dictator games and charity dictator games, but between waves 2 and 3 and between waves 1 and 3, pro-social behaviour is marginally significantly more stable in charity dictator games than in standard dictator games.

To sum up, in each model the sign of the difference in difference coefficients that compare wave 2 and wave 3; and wave 1 and wave 3 is in favour of pro-social behaviour being more stable in charity dictator games compared to standard dictator games. The former coefficient is significant or

marginally significant, depending on the model; the latter coefficient is marginally significant according to the random effects model, and although not significant according to the ordered probit and ordered logit models, it is not far from marginally significant. The sign of the difference in difference coefficient that compares wave 1 and wave 2 goes in the opposite direction in each model. However, the magnitude of this coefficient is small compared to the magnitude of the other difference in difference coefficients. Also, the p-value is large compared to the p-values of the other difference in difference coefficients; accordingly, this coefficient is highly insignificant in each model. Thus, while instability of pro-social behaviour between waves 1 and 2 does not significantly differ between standard dictator games and charity dictator games, behaviour is more stable in charity dictator games than in standard dictator games between waves 2 and 3. Additionally, there is some evidence that behaviour is more stable in the charity dictator games between waves 1 and 3 as well. All in all, it can be concluded that *people exhibit more stable behaviour in charity dictator games compared to standard dictator games*, confirming the hypothesis of this study.

Linear panel data regressions are also used to investigate whether the difference in stability of behaviour observed in standard dictator games and charity dictator games depends on the relative price of taking. Whether random effects models or the fixed effects models are the appropriate models to use is tested with the Hausman test. This test results in p-values of 1.000, 1.000, 0.774, and 0.564 for games 1, 2, 3, and 4 respectively, which means that the null-hypothesis that the difference in coefficients is not systematic is not rejected for any game. There is thus no evidence for a significant systematic difference between the coefficients obtained by the random effects model and the fixed effects model. This implies that the random effects model will be applied for all games. The results of the random effects regressions can be found in Table C in Appendix G; the difference in difference analyses obtained from these regressions in Table D in Appendix G.

The only differences in stability of behaviour observed in the standard treatment and charity treatment that are significant (marginally significant, strictly speaking) are differences found in games 1 and 3 and concern differences that indicate that behaviour is more stable in charity dictator games than in standard dictator games. That is, in particular when it is neither expensive nor cheap to take money; and when it is particularly cheap, behaviour tends to be more stable in charity dictator games than in standard dictator games.

As a side note, as mentioned in Sections 3.1 and 3.2, the average amounts taken in the standard group are 156.06 eurocents, 170.08 eurocents, and 214.02 eurocents in waves 1, 2, and 3 respectively; the average amounts taken in the charity group are 187.82 eurocents, 204.17 eurocents, and 217.63 eurocents in waves 1, 2, and 3 respectively. That is, in all waves, the average amount taken is higher in charity dictator games than in standard dictator games. These raw findings contradict the intuition that people are more likely to donate to charities. However, according to the regressions that were performed on the aggregates of the games in the current section, in none of the waves the amounts taken in the charity dictator games are significantly higher than the amounts taken in the standard dictator games (all p-values≥0.129). The same can be concluded when considering the games separately, based on the performed random effects regressions of the current section, with wave 1 in game 3 and waves 1, 2, and 3 in game 4 as exceptions (the p-value of comparing the first waves in game 3 is '0.092'; the p-values of comparing the first waves, the second waves, and the third waves in game 4 are '0.044', '0.015', and '0.099' respectively). That is, whether or not the average amounts taken are significantly higher in charity dictator games depends somewhat on the relative price of taking. But overall, the differences between the two treatments are not significant.

3.4 Answer to (sub-) research question

It is thus found that people exhibit more stable behaviour in charity dictator games compared to standard dictator games. There must thus have been a lower decrease in EDE in the charity dictator games, under the simplifying assumption of moral licensing being equally present in both types of dictator games. The experimental designs ensured that subjects in both treatments learned about the decision environment equally, which should lead to an equal reduction in uncertainty and therewith an equal reduction in EDE if the EDE would have been present equally in wave 1. A lower decrease in EDE therefore implies that the EDE was less present in the first place. More stable behaviour being observed in charity dictator games thus implies that the EDE was less present in charity dictator games from the first time on. This, in turn, implies that behaviour better reflects preferences in the charity dictator games.

As charity dictator games create a less artificial situation than standard dictator games, it appears that *a less artificial situation within dictator game experiments leads to behaviour that better reflects preferences*, answering the sub-research question. Getting back to the research question, *how informative behaviour observed in dictator game experiments is in terms of reflecting preferences, thus appears to depend on how artificial the situation within the dictator game is.*

4. Discussion

4.1 Comparison to related literature

4.1.1 Findings Brosig-Koch et al. vs. findings standard dictator games present study

The present study replicated Brosig Koch et al.'s (2017) experimental design as much as possible in the standard group and, therefore, findings similar to theirs were expected. Brosig-Koch et al. found that "dictators, on average, do not take the maximum amount in wave 1, but leave a considerable amount of money for recipients". More specifically, dictators took on average 301.80 eurocents in wave 1 of a maximum of 500 eurocents. However, the Wilcoxon signed-rank test, Jonckheere-Terpstra test, ordered probit regression, and ordered logit regression all found that the amount taken increased over the waves. According to the Wilcoxon signed-rank test, the average amount taken significantly increased to 425.40 eurocents in wave 2 and significantly increased further to 496.40 eurocents in wave 3. The Jonckheere-Terpstra test indicated that there was a significant overall upward trend in the amounts taken over the waves. Ordered probit and ordered logit regressions indicated that playing a specific game a second or a third time, both compared to playing it the first time, has a significantly positive effect on the amount taken; increasing the relative price from game 1 to game 3 or to game 4 has a significantly negative effect. The results from these tests thus indicate that people exhibit pro-social behaviour, but selfish behaviour increases over the waves; and that dictators tend to take more when it is relatively inexpensive to do so.

Similar to the dictators in Brosig-Koch et al's study, dictators in the standard group in the present study also left a considerable amount of money for recipients: they took on average 156.06 eurocents in wave 1 of a maximum of 500 eurocents. In waves 2 and 3, dictators took on average 170.08 eurocents and 214.02 eurocents respectively. These amount are, however, noticeably smaller than the average amounts taken by dictators in Brosig-Koch et al.'s study in corresponding waves. As discussed in Section 2.2.3, this may have been expected because of the different recruitment method applied in the two studies. Brosig-Koch et al. had "volunteers" participating while the present study had "pseudo-volunteers" participating; and according to the results found by Eckel and Grossman (2000), pseudo-volunteers tend to exert a higher level of generosity compared to volunteers. An alternative explanation is that the subjects in the standard group in the present study knew each other while the subjects in Brosig-Koch et al.'s study did not. It is quite plausible that knowing the

recipients, maybe even be friends, leads to behaviour that is more pro-social compared to not knowing the recipients.

Similar to Brosig-Koch et al.'s findings, the Wilcoxon signed-rank test, Jonckheere-Terpstra test, ordered probit regression, and ordered logit regression all found that the amount taken increased over the waves in the present study. However, it seems to take a bit longer in the present study for selfish behaviour to increase and the increase in selfish behaviour seems to be smaller. Wilcoxon signed-rank tests indicated that while the increase in average amount taken is significant between waves 2 and 3 and marginally significant between waves 1 and 3, it is not significant between waves 1 and 2. The positive coefficients in the ordered probit and ordered logit regressions of this study indicated that repeating a specific game has a positive effect on the amount taken. While this effect is significant comparing playing a specific game a third time to playing it the first time and the second time, it is not significant comparing playing a specific game a second time to playing it the first time. As opposed to these findings, Brosig-Koch et al. did find the increase in amount taken to be significant between waves 1 and 2, based on both the Wilcoxon signed-rank test and the ordered probit and ordered logit regressions, suggesting that it takes a bit longer in the present study for selfish behaviour to increase. Additionally, the increase in selfish behaviour seems to be smaller in this study compared to that of Brosig-Koch et al. as the increases in average amounts taken over the waves are smaller and the magnitude of the wave coefficients in the ordered probit and ordered logit regressions are smaller in the present study.

Three potential explanations are offered for the increase in selfish behaviour that occurs a bit later and to a lesser degree in the present study. First, as opposed to Brosig-Koch et al., who paid all subjects; and after each experimental session, the present study only paid a small part of their subjects; and only after all experiments had been conducted. The moment that subjects get paid, they learn that they actually get the money they 'won' and that their decisions remain anonymous. In the present study, the subjects that had won money learned this only after all experiments were conducted, while all subjects in Brosig-Koch et al.'s study learned this already after wave 1. According to Brosig-Koch et al., assuming that the EDE is rooted in subjects' uncertainty, learning about the decision environment causes the EDE to reduce. Paying only at the end of all experiments, and to only a small fraction of subjects, may thus have caused the EDE to diminish later and less, and therewith selfish behaviour to increase later and less. Second, Brosig-Koch et al. had their

subjects not only participate in take games, but also in give games and prisoner's dilemma games in each wave. Having subjects participate in multiple types of dictator games within a wave rather than just take games may cause uncertainty about the decision environment to reduce more within a wave, plausibly leading to a greater decrease in EDE over waves as compared to the present study. Another consequence of having subjects participate in multiple types of dictator games within a wave is that subjects might have behaved pro-socially more often within a wave, simply because they were provided with more opportunities to do so. As a result, moral licensing could be more present in future waves in Brosig-Koch et al.'s study as compared to the present study. Both a greater decrease in EDE over waves and moral licensing being more present in Brosig-Koch et al.'s study can potentially explain the greater increase in selfish behaviour observed in their study. Third, some subjects in the present study may not have been triggered enough to think well about the decision problems due to the rather low possibility to actually be paid out their decisions. If these students would have increased the amount taken if they would have thought well about the decision problem but now took the same amount as they did in the previous wave(s) just to be done with it quickly, this would result in behaviour that is more stable compared to the situation with higher monetary incentives, as in Brosig-Koch et al.'s study.[11]

The ordered probit and ordered logit regression of the present study indicated that dictators tend to take more when it is relatively inexpensive to do so, as found by Brosig-Koch et al. as well. The effects in the present study seem to be a bit stronger, though: Increasing the relative price from game 1 to game 2 had no significant effect on the amount taken in Brosig-Koch et al.'s study, while it had a marginally significant negative effect on the amount taken in the study at hand. Furthermore, the magnitude of the remaining game coefficients (those of game 3 and game 4), which were significant in both studies, were greater in the present study as compared to Brosig-Koch et al.'s study.

All in all, even though there are differences between the results of Brosig-Koch et al. and of this study that should not go unmentioned, the findings of the standard group in the present study replicate those of Brosig-Koch et al. pretty well in the sense that people exhibit pro-social behaviour

[11] Note that all explanations are not problematic for the comparison of behaviour in standard dictator games with behaviour in charity dictator games as the experimental designs are as such that if it occurs in one of the treatment groups, it will also in the other treatment group.

with selfish behaviour increasing over the waves; and that dictators tend to take more when it is relatively inexpensive to do so.

4.1.2 Typical finding pro-social behaviour vs. finding present study

As mentioned in the introduction of this thesis, a meta-study on dictator games revealed that dictators give away on average 28.35% of the amount to be shared to the recipient (Engel, 2011). Calculated to percentages and in terms of giving rather than taking, dictators in the standard group in the present study gave away on average 68.79%, 65.98%, and 57.20% in waves 1, 2, and 3 respectively; dictators in the charity group gave away on average 62.44%, 59.17%, and 56.47% in waves 1, 2, and 3 respectively. These percentages are considerably higher than the percentage obtained by the meta-study. A potential explanation for this is that the context of the dictator game may influence the amounts dictators give away. For example, in the present study dictators were provided with 11 division options to choose from, with the first option being the most generous one: an equal division of money between themselves and the recipient. People could have 'anchored' on this option, and insufficiently adjusted afterwards (Tversky & Kahneman, 1974), causing them to behave more pro-socially than they would have if they had been presented with a less generous first option or if they would not have been presented with options at all but were free to choose a division themselves, which may be the case in most dictator game studies that are used for Engel's meta-study. An alternative explanation is that dictators' action space in the present study is limited such that it does not include options to receive a lower payoff than recipients. More specifically, dictators receive a minimum of 500 eurocents themselves and got to divide 500 eurocents on top of that. Donating (part of) the amount to be shared to the recipient reduces the payoff to the dictator, but at the same time it also reduces the inequality between the dictator and the recipient. The latter is favourable for the dictator, assuming inequality aversion (Fehr & Schmidt, 1999). This may be different from most dictator games that are used for Engel's meta-study in which dictators' action space may not have been limited in the sense that it does not include options to receive a lower payoff than recipients. In these games, donating generates a lower payoff to the dictator, similar to the take games, but only reduces the inequality between the dictator and the recipient up to an equal division. After the equal division, the dictator would receive less than the recipient. Donating after the equal division thus reduces the payoff to the dictator and, at the same time, it increases the inequality between the dictator and the recipient. Assuming inequality aversion, dictators should not donate more than half of the amount to be shared in these dictator games, whereas in the dictator games of the present study donating everything can be in line with inequality aversion.

4.2 Limitations of this study

4.2.1 Experimental design

Communication

A limitation of this study is the possibility of subjects to communicate, both in their own and between other secondary school classes. Concerning communication between classes, in each wave, the questionnaires would ideally have been conducted on the same day, right after each other in order to reduce the possibility of part of the subjects obtaining information about the questionnaire before conducting them (which potentially results in a lower EDE for this part of the subjects, assuming the EDE is rooted in one's uncertainty). Due to practical limitations of the secondary school, however, this was not feasible. Instead, in both the *havo* and *vwo* classes the questionnaires were conducted on the charity group one day after they had been conducted on the standard group; and for each wave, the questionnaires in the *vwo* classes were conducted one week after the *havo* classes. The standard groups could inform the charity groups about the questionnaire they had received. A factor that limits the transition of communication about the questionnaires before they were conducted in the charity groups is that, at least in the first wave, the students did not know that the charity groups would also receive questionnaires. Informing the charity groups about what to expect could thus not be a reason to communicate about the questionnaires. Also, the *havo* classes could inform the *vwo* classes about the questionnaires. However, according to a current student from the secondary school, there is not much communication between the two educational attainment levels. Even if there would be communication between the *havo* and *vwo* classes about the questionnaires before the *vwo* classes received them, it is unlikely that it threatened the results. This is because the *havo* standard group was pooled with the *vwo* standard group, and the *havo* charity group was pooled with the *vwo* charity group - and there is no reason to believe that only one of both vwo groups will obtain information from the *havo* classes. Concerning communication within the secondary school classes, after wave 1 communication between subjects was possible, potentially influencing decisions that are made in future waves. This threatens the attempt to create identical decision situations throughout the waves. Communication within subject groups is a limitation to all studies that have a group of the same subjects participating in experiments multiple times. But in this study the subjects within a subject group could communicate relatively easy as part of them came from the same secondary school class.

Payment structure

As mentioned in Part 3 of Appendix E, the 'winning' subjects have thus far not replied to the experimenter's e-mail in which they are asked for their bank account number. This might be due to students not having checked their student e-mail. One could argue that this illustrates that the subjects considered the chance that their decisions were the ones that were going to be paid out small, which could have resulted into subjects not putting in enough effort to truthfully answer the questions in the questionnaires. However, it may as well be the case that subjects did consider the chance that their decisions would be paid out high enough to put in enough effort to truthfully answer the questions, but that the monetary incentive was simply not high enough to make them exert effort to check their student mail and send their bank account number after the questionnaires had been performed. At the moment they filled out the questionnaires, they might not have anticipated that they would be reluctant to exert effort in order to collect their payments, which would imply that it does not threaten the results that the winning subjects have not yet replied.

One-shot vs. repeated dictator games

It should also be noted that this study draws a conclusion regarding 'one-shot' dictator games, while the games played in the experiment were repeated. Knowing that the same games will be played in the future at the moment of playing the games could influence behaviour in the present games by taking into account decisions for future games. That would generate behaviour in the present games that differs from behaviour had it been one-shot dictator games. In an attempt to prevent this, subjects had only been informed that there would be a possibility to be asked to participate in future experiments rather than also about the fact that it concerns the exact same games, following Brosig-Koch et al.'s approach. Because of this, it is deemed acceptable to draw a conclusion about one-shot dictator games by having performed repeated dictator games.

Artificiality

One might ask whether charity dictator games really create a less artificial situation than standard dictator games. The fact that the dictators in this study are students, i.e. dictators who may be more familiar with 'donating' money to fellow students than donating to charities, suggests that, for these particular dictators, standard dictator games may be less artificial. However, the lack of information dictators have to base their donation decision on in standard dictator games makes this type of games artificial, even though the dictators may be familiar with donating to fellow students. A more informative setting, whether dictators are or are not familiar with donating to the particular recipient,

should create a less artificial situation. This is because, in real life, donations are typically made in informative settings.

4.2.2 Simplifying assumption moral licensing

The simplifying assumption was made that moral licensing is equally present in both types of dictator games. This implied that the EDE was isolated: A lower increase in selfish behaviour in charity dictator games must have been due to a lower decrease in EDE. Due to the experimental designs, this meant that more stable behaviour implied that behaviour better reflects preferences. It is, however, possible that moral licensing was less present in charity dictator games than in standard dictator games: A charity recipient is probably in more serious need of money than a student recipient. This may imply that not donating, just because having already done so before, cannot as easily be justified in case of the recipient being a charity compared to it being a student. A lower increase in selfish behaviour in charity dictator games can thus also possibly be due to moral licensing being less present in charity dictator games than in standard dictator games. This study cannot judge whether moral licensing had a role to play in the lower increase in selfish behaviour in charity dictator games or whether it was only a lower decrease in EDE. However, a diminishing EDE is expected to at least partly drive the results of more stable behaviour in charity dictator games – and this is enough to conclude that a less artificial situation within dictator game experiments leads to behaviour that better reflects preferences.

4.2.3 Degree to which behaviour reflects preferences in charity dictator games

The conclusion that behaviour reflects preferences better in charity dictator games than in standard dictator games relies on the fact that the EDE as discussed in this thesis could have played less of a role in the former type of games, assuming that moral licensing is equally present in both types of games. It should be noted, though, that this does not necessarily mean that behaviour in charity dictator games perfectly reflect preferences. This is because the EDE could still play some role in this type of games, as behaviour was not perfectly stable. But even if behaviour would have been perfectly stable, behaviour might not perfectly reflect preferences: Besides *uncertainty* (as discussed in this thesis), *being observed* by the experimenter may also cause an EDE[12], implying that preferences may be less pro-social than the pro-social behaviour suggests. The reason that this effect may even

[12] Research by Haley and Fessler (2005) shows that, in double-blind dictator game experiments, even subtle cues that give the impression of being observed leads to a change in behaviour. More specifically, they found that the presence of "stylized eye-like shapes on a computer screen" leads to an increase in pro-social behaviour, despite no differences in actual anonymity.

occur when behaviour is perfectly stable is because the effect is probably equally present in each wave as subjects were being observed to the same extent in each wave (subjects filled out the questionnaire in the presence of other subjects, their teacher, and the experimenter in each wave). Note that also in both treatments the subjects were being observed to the same extent, which means that the effect of being observed should be about equally present in both types of dictator games. Therefore, it can still be concluded that behaviour reflects preferences better in charity dictator games than in standard dictator games; it should just be kept in mind that behaviour may still not perfectly reflect preferences in charity dictator games. Only if subjects in the charity treatment are more sensitive to being observed compared to subjects in the standard treatment[13], it cannot be concluded that behaviour reflects preferences better in charity dictator games, because even though the EDE caused by *uncertainty* would be present less, the EDE caused by *being observed* would be present more – and it is unsure which effect dominates.

4.3 Suggestions for future research

Despite its limitations, this study provided some evidence that charity dictator games result in more stable behaviour than standard dictator games: Behaviour between waves 2 and 3 was found to be significantly or marginally significantly, depending on the model, more stable in charity dictator games compared to standard dictator games. Additionally, behaviour between waves 1 and 3 was found to be marginally significantly or almost marginally significantly, depending on the model, more stable in charity dictator games. This indicates that more research is required to find out if more evidence will be found when the experimental design is improved. For example, the payment structure could be improved by providing stronger monetary incentives and by paying out subjects after each wave rather than only at the end of wave 3. The latter implies that, in the standard group, a group of subjects should not consist of both dictators and recipients. If it would, a recipient could infer how much a dictator gave away to them, which could influence the recipient's decision in the future waves. That is, the group of dictators should be separated from the group of recipients. In an attempt to limit communication both within a group of dictators and between groups of dictators that obtain a different treatment, the dictator groups should not only be separated from recipient groups, but the dictator groups should also be composed of subjects that do not know each other.

[13] Note that it is possible that this is the case. As the recipients in the charity dictator games are probably in a more serious need of money than most of the recipients in the standard dictator games, not donating in the former type of game can be seen as more 'unkind'. Therefore, decreasing subjects' perception of privacy by having them being observed possibly affects behaviour in charity dictator games more than in standard dictator games.

Another suggestion for future research concerns investigating a secondary finding of this study: a potential difference in responsiveness to relative prices in standard dictator games and charity dictator games. Andreoni and Vesterlund (2001) noticed seemingly contradictory results in previous studies concerning how altruism depends on gender: Some studies found men to be kinder, some found women to be kinder, and some found no differences between men and women at all. In order to further investigate this, Andreoni and Vesterlund applied modified dictator games, using different budgets and different relative prices, on both men and women. They found that whether males or females are more altruistic depends on the relative price of giving: When it is relatively expensive to give, women are more altruistic; when it is relatively cheap to give, men are more altruistic. This can be explained by men being more responsive to price changes than women.

A difference in responsiveness to relative prices seems to be found in the present study as well: Regressions from Section 3.1 indicated that, in the standard group, increasing the relative price of taking from game 1 to game 2 has a marginally significantly negative effect on the amount taken; increasing the relative price of taking from game 1 to game 3 or to game 4 has a significantly negative effect. Regressions from Section 3.2 indicate that, in the charity group, increasing the relative price of taking from game 1 to game 2 has no significant effect on the amount taken. Increasing the relative price of taking from game 1 to game 3 has significantly negative effect on the amount taken according to the ordered probit model and a marginally significantly negative effect according to the ordered logit model. Increasing the relative price of taking from game 1 to game 4 has a significantly negative effect according to the ordered probit model and no significant effect according to the ordered logit model. The fact that more differences between game 1 and the other games are (marginally) significant in the standard group than in the charity group, plus that the magnitude of the game coefficients are greater in the standard group, suggest that dictators in charity dictator games seem to respond less to relative prices than dictators in standard dictator games.

Andreoni and Vesterlund (2001) consider what their results could mean in the realm of charitable giving. Quoted from them, "for years, public economists have been studying the effect of the charitable deduction for taxpayers, which acts as a subsidy to lower the price of giving". They explain that if men and women systematically respond differently to price changes, tax policy could affect behaviour of males and females differently, which has its impact on charities "whose donors may vary by sex". In light of Andreoni and Vesterlund's findings, this would imply that charitable

deduction may lead men to increase their donations to charities more than women. However, both men and women may in fact not increase their donations at all when it gets cheaper to give to charities. This is because Andreoni and Vesterlund's finding that men are more responsive to price changes than women is based on dictator games in which the recipients are students; and the present study suggests that people are less responsive to price changes when the recipients concern actual charities.

Future research could further investigate people's responsiveness to price changes in charity dictator games; and whether this responsiveness depends on certain individual characteristics, in particular potential differences between men and women is something to look out for. The fact that there is a tax deduction in the first place seems to indicate that politicians try to stimulate donations to charities. Knowing more about people's responsiveness to price changes when donating to charities may help lawmakers to determine the optimal charitable deduction. Once the optimal charitable deduction has been established, charities could make use of people's different responsiveness to price changes by emphasising the charitable deduction especially to the most responsive ones. For charities it would thus be useful to know how responsiveness of price changes depends on individual characteristics.

5. Conclusion

Nowadays, the dictator game is widely used in economic lab experiments in order to elicit pro-social preferences. The results obtained by these games function as experimental evidence that people indeed have pro-social preferences. However, research by Brosig-Koch et al. (2017) and Bardsley (2008) suggest that pro-social behaviour observed in dictator games may be an artefact of experimentation rather than a representation of pro-social preferences: Rather than caring about other people's payoff, dictators may exhibit pro-social behaviour because they have inferred that this is what the experiment is about and is thus the 'right' thing to do. Accordingly, pro-social *behaviour* in dictator games does not necessarily imply pro-social *preferences*. Their research thus calls into question whether economic lab experiments, in particular dictator game experiments, do in fact reveal preferences.

This thesis investigated how informative behaviour observed in dictator game experiments is in terms of reflecting preferences. Accordingly, the research question was:

How informative is behaviour observed in dictator game experiments in terms of reflecting preferences?

It was expected that the answer to the research question depends on how artificial the situation is within the dictator game. A sub-research question was therefore formulated:

Does a less artificial situation within dictator game experiments lead to behaviour that better reflects preferences?

A less artificial situation was created by replacing the typically applied students as recipients by well-known charities. The present study has demonstrated that in typical dictator games, pro-social behaviour drops when repeating the same games on the same subjects, which replicated the finding by Brosig-Koch et al. This can possibly be due to pro-social behaviour being an artefact, implying that pro-social preferences are overestimated by the observed behaviour. This is, however, debatable. An alternative explanation of pro-social behaviour that drops is moral licensing, implying that the observed behaviour does represent pro-social preferences accurately; the preferences just change after having been generous before. In charity dictator games, pro-social behaviour tends to be more stable over repetitions of the games. How much more stable depends on the relative price of exhibiting selfish behaviour: In particular when it is neither expensive nor cheap to take money; and when it is particularly cheap, behaviour tends to be more stable in charity dictator games than in standard dictator games. But overall, behaviour in charity dictator games tends to be more stable compared to behaviour in standard dictator games. This confirms the hypothesis of this study: *People exhibit more stable behaviour in charity dictator games compared to standard dictator games*.

As stability of behaviour functioned as a measure of how well behaviour reflects preferences, it appears that *a less artificial situation within dictator game experiments leads to behaviour that better reflects preferences*, answering the sub-research question. Getting back to the research question, *how informative behaviour observed in dictator game experiments is in terms of reflecting preferences, thus appears to depend on how artificial the situation within the dictator game is.*

Is pro-social behaviour in dictator games just an artefact? Perhaps this is the case in the typical dictator game, although this is debatable. In less artificial dictator games, however, pro-social behaviour appears to be less of an artefact; it seems to reflect preferences instead. In general, dictator game experiments thus seem to reveal preferences well, under the condition that the situation created in the experiment is not too artificial.

6. Appendices

Appendix A: English translation of the standard group questionnaire

Welcome to the experiment! First of all, please fill out the following:

Student e-mail address:[1]
Class:
Gender: Male/Female

It is very important that you do not communicate with each other from now on. If you do so, you are unfortunately not allowed to finish the questionnaire, which means that your chance of winning money disappears.

In this experiment, you will be presented with several decision situations. In the decision situations, you are asked to divide money between yourself and a randomly chosen other student in this class, called "Partner" in the decision situations.

The choice of one randomly chosen student in this class of one randomly chosen decision situation will be paid out with real money. If you are this student, you will receive the money you chose to keep for yourself and your partner will receive the money you chose to give away. It will be randomly determined who your partner will be. Note that this implies that you don't know between which student and yourself you're dividing money.

[*Addition to wave 1 in wave 2*: But it won't be the same student as in the previous experiment.]

[*Addition to wave 1 in wave 3*: But it won't be the same student as in the previous experiments.]

It should be noted that the ones who win money in this experiment will be paid out half of the amount mentioned in the experiment.

Remember that no one will get to know your decision. If anything is unclear, please raise your hand and wait with asking your question until I am there to help. Please think carefully about every decision situation. You can now go to decision situation 1.

[1] i.e. 'studentnumber@edu.ozhw.nl'

Decision situation 1

Here you can determine the payoff to yourself and the payoff to your partner. How would you divide the money? Please choose **one** of the following combinations of payoffs:

Choice:

- You: 500 eurocent Partner: 500 eurocent
- You: 600 eurocent Partner: 450 eurocent
- You: 700 eurocent Partner: 400 eurocent
- You: 800 eurocent Partner: 350 eurocent
- You: 900 eurocent Partner: 300 eurocent
- You: 1000 eurocent Partner: 250 eurocent
- You: 1100 eurocent Partner: 200 eurocent
- You: 1200 eurocent Partner: 150 eurocent
- You: 1300 eurocent Partner: 100 eurocent
- You: 1400 eurocent Partner: 50 eurocent
- You: 1500 eurocent Partner: 0 eurocent

Decision situation 2

Here you can determine the payoff to yourself and the payoff to your partner. How would you divide the money? Please choose **one** of the following combinations of payoffs:

Choice:

- You: 500 eurocent Partner: 500 eurocent
- You: 575 eurocent Partner: 450 eurocent
- You: 650 eurocent Partner: 400 eurocent
- You: 725 eurocent Partner: 350 eurocent
- You: 800 eurocent Partner: 300 eurocent
- You: 875 eurocent Partner: 250 eurocent
- You: 950 eurocent Partner: 200 eurocent
- You: 1025 eurocent Partner: 150 eurocent
- You: 1100 eurocent Partner: 100 eurocent
- You: 1175 eurocent Partner: 50 eurocent
- You: 1250 eurocent Partner: 0 eurocent

Decision situation 3

Here you can determine the payoff to yourself and the payoff to your partner. How would you divide the money? Please choose **one** of the following combinations of payoffs:

Choice:

- You: 500 eurocent Partner: 500 eurocent
- You: 550 eurocent Partner: 450 eurocent
- You: 600 eurocent Partner: 400 eurocent
- You: 650 eurocent Partner: 350 eurocent
- You: 700 eurocent Partner: 300 eurocent
- You: 750 eurocent Partner: 250 eurocent
- You: 800 eurocent Partner: 200 eurocent
- You: 850 eurocent Partner: 150 eurocent
- You: 900 eurocent Partner: 100 eurocent
- You: 950 eurocent Partner: 50 eurocent
- You: 1000 eurocent Partner: 0 eurocent

Decision situation 4

Here you can determine the payoff to yourself and the payoff to your partner. How would you divide the money? Please choose **one** of the following combinations of payoffs:

Choice:

- You: 500 eurocent Partner: 500 eurocent
- You: 525 eurocent Partner: 450 eurocent
- You: 550 eurocent Partner: 400 eurocent
- You: 575 eurocent Partner: 350 eurocent
- You: 600 eurocent Partner: 300 eurocent
- You: 625 eurocent Partner: 250 eurocent
- You: 650 eurocent Partner: 200 eurocent
- You: 675 eurocent Partner: 150 eurocent
- You: 700 eurocent Partner: 100 eurocent
- You: 725 eurocent Partner: 50 eurocent
- You: 750 eurocent Partner: 0 eurocent

This is the end of the questionnaire. Please check if you filled out your student e-mail address, class, and gender at the top of the questionnaire and if you ticked one box in every decision situation. Once you have finished this, please turn the questionnaire over with the text facing the table so that your student e-mail address, class, and gender are not visible. Then raise your hand so that I can pick up your questionnaire. Please remain quiet until I have picked up all questionnaires. Thank you for your participation!

Appendix B: Original standard group questionnaire

Welkom bij het experiment! Vul om te beginnen alsjeblieft de volgende gegevens in:

Leerlingen e-mailadres:[1]
Klas:
Geslacht: Man/Vrouw

Het is heel belangrijk dat je niet communiceert met elkaar vanaf nu. Als je dit toch doet, mag je de vragenlijst helaas niet afmaken, wat betekent dat je geen kans meer hebt op het winnen van geld.

In dit experiment zal je verschillende beslissing-situaties te zien krijgen. In de beslissing-situaties wordt je gevraagd geld te verdelen tussen jezelf en een willekeurig gekozen andere leerling in deze klas, genoemd "Partner" in de beslissing-situaties.

De keuze van één willekeurig gekozen leerling in deze klas van één willekeurig gekozen beslissing-situatie zal uitbetaald worden met echt geld. Als jij deze leerling bent, zal je het geld ontvangen dat je gekozen hebt voor jezelf te houden en jouw partner zal het geld ontvangen dat je gekozen hebt weg te geven. Het zal willekeurig bepaald worden wie jouw partner zal zijn. Merk op dat dit betekent dat je niet weet tussen welke leerling en jezelf je geld verdeelt.

[*Toevoeging aan wave 1 in wave 2*: Maar het zal niet dezelfde leerling zijn als in het vorige experiment.]

[*Toevoeging aan wave 1 in wave* 3: Maar het zal niet dezelfde leerling zijn als in de vorige experimenten.]

Let op: Degenen die geld winnen in dit experiment krijgen de helft uitbetaald van het bedrag dat genoemd wordt in het experiment.

Onthoud dat niemand achter je beslissing zal komen. Als er iets niet duidelijk is, steek dan alsjeblieft je hand op en wacht met het stellen van je vraag totdat ik bij je ben om te helpen. Denk alsjeblieft goed na over elke beslissing-situatie. Je kunt nu naar beslissing-situatie 1 gaan.

[1] Het leerlingen e-mailadres is als volgt opgebouwd: 'leerlingnummer@edu.ozhw.nl'

Beslissing-situatie 1

Hier kan je de uitbetaling voor jezelf en die voor je partner bepalen. Hoe zou jij het geld verdelen? Kies alsjeblieft **één** van de volgende combinaties van uitbetalingen:

Keuze:

- Jij: 500 eurocent Partner: 500 eurocent
- Jij: 600 eurocent Partner: 450 eurocent
- Jij: 700 eurocent Partner: 400 eurocent
- Jij: 800 eurocent Partner: 350 eurocent
- Jij: 900 eurocent Partner: 300 eurocent
- Jij: 1000 eurocent Partner: 250 eurocent
- Jij: 1100 eurocent Partner: 200 eurocent
- Jij: 1200 eurocent Partner: 150 eurocent
- Jij: 1300 eurocent Partner: 100 eurocent
- Jij: 1400 eurocent Partner: 50 eurocent
- Jij: 1500 eurocent Partner: 0 eurocent

Beslissing-situatie 2

Hier kan je de uitbetaling voor jezelf en die voor je partner bepalen. Hoe zou jij het geld verdelen? Kies alsjeblieft **één** van de volgende combinaties van uitbetalingen:

Keuze:

- Jij: 500 eurocent Partner: 500 eurocent
- Jij: 575 eurocent Partner: 450 eurocent
- Jij: 650 eurocent Partner: 400 eurocent
- Jij: 725 eurocent Partner: 350 eurocent
- Jij: 800 eurocent Partner: 300 eurocent
- Jij: 875 eurocent Partner: 250 eurocent
- Jij: 950 eurocent Partner: 200 eurocent
- Jij: 1025 eurocent Partner: 150 eurocent
- Jij: 1100 eurocent Partner: 100 eurocent
- Jij: 1175 eurocent Partner: 50 eurocent
- Jij: 1250 eurocent Partner: 0 eurocent

Beslissing-situatie 3

Hier kan je de uitbetaling voor jezelf en die voor je partner bepalen. Hoe zou jij het geld verdelen? Kies alsjeblieft **één** van de volgende combinaties van uitbetalingen:

Keuze:

- Jij: 500 eurocent Partner: 500 eurocent
- Jij: 550 eurocent Partner: 450 eurocent
- Jij: 600 eurocent Partner: 400 eurocent
- Jij: 650 eurocent Partner: 350 eurocent
- Jij: 700 eurocent Partner: 300 eurocent
- Jij: 750 eurocent Partner: 250 eurocent
- Jij: 800 eurocent Partner: 200 eurocent
- Jij: 850 eurocent Partner: 150 eurocent
- Jij: 900 eurocent Partner: 100 eurocent
- Jij: 950 eurocent Partner: 50 eurocent
- Jij: 1000 eurocent Partner: 0 eurocent

Beslissing-situatie 4

Hier kan je de uitbetaling voor jezelf en die voor je partner bepalen. Hoe zou jij het geld verdelen? Kies alsjeblieft **één** van de volgende combinaties van uitbetalingen:

Keuze:

- Jij: 500 eurocent Partner: 500 eurocent
- Jij: 525 eurocent Partner: 450 eurocent
- Jij: 550 eurocent Partner: 400 eurocent
- Jij: 575 eurocent Partner: 350 eurocent
- Jij: 600 eurocent Partner: 300 eurocent
- Jij: 625 eurocent Partner: 250 eurocent
- Jij: 650 eurocent Partner: 200 eurocent
- Jij: 675 eurocent Partner: 150 eurocent
- Jij: 700 eurocent Partner: 100 eurocent
- Jij: 725 eurocent Partner: 50 eurocent
- Jij: 750 eurocent Partner: 0 eurocent

Dit is het einde van de vragenlijst. Check alsjeblieft of je je leerlingen e-mailadres, klas en geslacht hebt ingevuld bovenaan de vragenlijst en of je één vakje hebt aangekruist in elke beslissing-situatie. Wanneer je hiermee klaar bent, leg je vragenlijst dan alsjeblieft met de tekst naar beneden neer zodat je leerlingen e-mailadres, klas en geslacht niet zichtbaar zijn. Steek dan je hand op zodat ik je vragenlijst kan komen ophalen. Blijf alsjeblieft stil totdat ik alle vragenlijsten heb opgehaald. Bedankt voor je deelname!

Appendix C: English translation of the charity group questionnaire

Welcome to the experiment! First of all, please fill out the following:

Student e-mail address:[1]
Class:
Gender: Male/Female

It is very important that you do not communicate with each other from now on. If you do so, you are unfortunately not allowed to finish the questionnaire, which means that your chance of winning money disappears.

In this experiment, you will be presented with several decision situations. In the decision situations, you are asked to divide money between yourself and a charity. The possible charities are "KWF Kankerbestrijding", "UNICEF", and "De Dierenbescherming".

The choice of one randomly chosen student in this class of one randomly chosen decision situation will be paid out with real money. If you are this student, you will receive the money you chose to keep for yourself and one of the above mentioned charities will receive the money you chose to give away. It will be randomly determined which charity that is. Note that this implies that you don't know between which of the above-mentioned charities and yourself you're dividing money.

[*Addition to wave 1 in wave 2*: But it won't be the same charity as in the previous experiment.]

[*Addition to wave 1 in wave 3*: But it won't be the same charity as in the previous experiments.]

It should be noted that the ones who win money in this experiment will be paid out half of the amount mentioned in the experiment.

Remember that no one will get to know your decision. If anything is unclear, please raise your hand and wait with asking your question until I am there to help. Please think carefully about every decision situation. You can now go to decision situation 1.

[1] i.e. 'studentnumber@edu.ozhw.nl'

Decision situation 1

Here you can determine the payoff to yourself and the payoff to one of the three charities mentioned above. How would you divide the money? Please choose **one** of the following combinations of payoffs:

Choice:

- You: 500 eurocent Charity: 500 eurocent
- You: 600 eurocent Charity: 450 eurocent
- You: 700 eurocent Charity: 400 eurocent
- You: 800 eurocent Charity: 350 eurocent
- You: 900 eurocent Charity: 300 eurocent
- You: 1000 eurocent Charity: 250 eurocent
- You: 1100 eurocent Charity: 200 eurocent
- You: 1200 eurocent Charity: 150 eurocent
- You: 1300 eurocent Charity: 100 eurocent
- You: 1400 eurocent Charity: 50 eurocent
- You: 1500 eurocent Charity: 0 eurocent

Decision situation 2

Here you can determine the payoff to yourself and the payoff to one of the three charities mentioned above. How would you divide the money? Please choose **one** of the following combinations of payoffs:

Choice:

- You: 500 eurocent Charity: 500 eurocent
- You: 575 eurocent Charity: 450 eurocent
- You: 650 eurocent Charity: 400 eurocent
- You: 725 eurocent Charity: 350 eurocent
- You: 800 eurocent Charity: 300 eurocent
- You: 875 eurocent Charity: 250 eurocent
- You: 950 eurocent Charity: 200 eurocent
- You: 1025 eurocent Charity: 150 eurocent
- You: 1100 eurocent Charity: 100 eurocent
- You: 1175 eurocent Charity: 50 eurocent
- You: 1250 eurocent Charity: 0 eurocent

Decision situation 3

Here you can determine the payoff to yourself and the payoff to one of the three charities mentioned above. How would you divide the money? Please choose **one** of the following combinations of payoffs:

Choice:

- You: 500 eurocent Charity: 500 eurocent
- You: 550 eurocent Charity: 450 eurocent
- You: 600 eurocent Charity: 400 eurocent
- You: 650 eurocent Charity: 350 eurocent
- You: 700 eurocent Charity: 300 eurocent
- You: 750 eurocent Charity: 250 eurocent
- You: 800 eurocent Charity: 200 eurocent
- You: 850 eurocent Charity: 150 eurocent
- You: 900 eurocent Charity: 100 eurocent
- You: 950 eurocent Charity: 50 eurocent
- You: 1000 eurocent Charity: 0 eurocent

Decision situation 4

Here you can determine the payoff to yourself and the payoff to one of the three charities mentioned above. How would you divide the money? Please choose **one** of the following combinations of payoffs:

Choice:

- You: 500 eurocent Charity: 500 eurocent
- You: 525 eurocent Charity: 450 eurocent
- You: 550 eurocent Charity: 400 eurocent
- You: 575 eurocent Charity: 350 eurocent
- You: 600 eurocent Charity: 300 eurocent
- You: 625 eurocent Charity: 250 eurocent
- You: 650 eurocent Charity: 200 eurocent
- You: 675 eurocent Charity: 150 eurocent
- You: 700 eurocent Charity: 100 eurocent
- You: 725 eurocent Charity: 50 eurocent
- You: 750 eurocent Charity: 0 eurocent

This is the end of the questionnaire. Please check if you filled out your student e-mail address, class, and gender at the top of the questionnaire and if you ticked one box in every decision situation. Once you have finished this, please turn the questionnaire over with the text facing the table so that your student e-mail address, class, and gender are not visible. Then raise your hand so that I can pick

up your questionnaire. Please remain quiet until I have picked up all questionnaires. Thank you for your participation!

Appendix D: Original charity group questionnaire

Welkom bij het experiment! Vul om te beginnen alsjeblieft de volgende gegevens in:

Leerlingen e-mailadres:[1]
Klas:
Geslacht: Man/Vrouw

Het is heel belangrijk dat je niet communiceert met elkaar vanaf nu. Als je dit toch doet, mag je de vragenlijst helaas niet afmaken, wat betekent dat je geen kans meer hebt op het winnen van geld.

In dit experiment zal je verschillende beslissing-situaties te zien krijgen. In de beslissing-situaties wordt je gevraagd geld te verdelen tussen jezelf en een goed doel. De mogelijke goede doelen zijn "KWF Kankerbestrijding", "UNICEF" en "De Dierenbescherming".

De keuze van één willekeurig gekozen leerling in deze klas van één willekeurig gekozen beslissing-situatie zal uitbetaald worden met echt geld. Als jij deze leerling bent, zal je het geld ontvangen dat je gekozen hebt voor jezelf te houden en één van de bovengenoemde goede doelen zal het geld ontvangen dat je gekozen hebt weg te geven. Het zal willekeurig bepaald worden welk goed doel dat is. Merk op dat dit betekent dat je niet weet tussen welk van de bovengenoemde goede doelen en jezelf je geld verdeelt.

[*Toevoeging aan wave 1 in wave 2*: Maar het zal niet hetzelfde goede doel zijn als in het vorige experiment.]

[*Toevoeging aan wave 1 in wave 3:* Maar het zal niet hetzelfde goede doel zijn als in de vorige experimenten.]

Let op: Degenen die geld winnen in dit experiment krijgen de helft uitbetaald van het bedrag dat genoemd wordt in het experiment.

Onthoud dat niemand achter je beslissing zal komen. Als er iets niet duidelijk is, steek dan alsjeblieft je hand op en wacht met het stellen van je vraag totdat ik bij je ben om te helpen. Denk alsjeblieft goed na over elke beslissing-situatie. Je kunt nu naar beslissing-situatie 1 gaan.

[1] Het leerlingen e-mailadres is als volgt opgebouwd: 'leerlingnummer@edu.ozhw.nl'

Beslissing-situatie 1

Hier kan je de uitbetaling voor jezelf en die voor één van de drie bovengenoemde goede doelen bepalen. Hoe zou jij het geld verdelen? Kies alsjeblieft **één** van de volgende combinaties van uitbetalingen:

Keuze:

- Jij: 500 eurocent Goed doel: 500 eurocent
- Jij: 600 eurocent Goed doel: 450 eurocent
- Jij: 700 eurocent Goed doel: 400 eurocent
- Jij: 800 eurocent Goed doel: 350 eurocent
- Jij: 900 eurocent Goed doel: 300 eurocent
- Jij: 1000 eurocent Goed doel: 250 eurocent
- Jij: 1100 eurocent Goed doel: 200 eurocent
- Jij: 1200 eurocent Goed doel: 150 eurocent
- Jij: 1300 eurocent Goed doel: 100 eurocent
- Jij: 1400 eurocent Goed doel: 50 eurocent
- Jij: 1500 eurocent Goed doel: 0 eurocent

Beslissing-situatie 2

Hier kan je de uitbetaling voor jezelf en die voor één van de drie bovengenoemde goede doelen bepalen. Hoe zou jij het geld verdelen? Kies alsjeblieft **één** van de volgende combinaties van uitbetalingen:

Keuze:

- Jij: 500 eurocent Goed doel: 500 eurocent
- Jij: 575 eurocent Goed doel: 450 eurocent
- Jij: 650 eurocent Goed doel: 400 eurocent
- Jij: 725 eurocent Goed doel: 350 eurocent
- Jij: 800 eurocent Goed doel: 300 eurocent
- Jij: 875 eurocent Goed doel: 250 eurocent
- Jij: 950 eurocent Goed doel: 200 eurocent
- Jij: 1025 eurocent Goed doel: 150 eurocent
- Jij: 1100 eurocent Goed doel: 100 eurocent
- Jij: 1175 eurocent Goed doel: 50 eurocent
- Jij: 1250 eurocent Goed doel: 0 eurocent

Beslissing-situatie 3

Hier kan je de uitbetaling voor jezelf en die voor één van de drie bovengenoemde goede doelen bepalen. Hoe zou jij het geld verdelen? Kies alsjeblieft **één** van de volgende combinaties van uitbetalingen:

Keuze:

- Jij: 500 eurocent Goed doel: 500 eurocent
- Jij: 550 eurocent Goed doel: 450 eurocent
- Jij: 600 eurocent Goed doel: 400 eurocent
- Jij: 650 eurocent Goed doel: 350 eurocent
- Jij: 700 eurocent Goed doel: 300 eurocent
- Jij: 750 eurocent Goed doel: 250 eurocent
- Jij: 800 eurocent Goed doel: 200 eurocent
- Jij: 850 eurocent Goed doel: 150 eurocent
- Jij: 900 eurocent Goed doel: 100 eurocent
- Jij: 950 eurocent Goed doel: 50 eurocent
- Jij: 1000 eurocent Goed doel: 0 eurocent

Beslissing-situatie 4

Hier kan je de uitbetaling voor jezelf en die voor één van de drie bovengenoemde goede doelen bepalen. Hoe zou jij het geld verdelen? Kies alsjeblieft **één** van de volgende combinaties van uitbetalingen:

Keuze:

- Jij: 500 eurocent Goed doel: 500 eurocent
- Jij: 525 eurocent Goed doel: 450 eurocent
- Jij: 550 eurocent Goed doel: 400 eurocent
- Jij: 575 eurocent Goed doel: 350 eurocent
- Jij: 600 eurocent Goed doel: 300 eurocent
- Jij: 625 eurocent Goed doel: 250 eurocent
- Jij: 650 eurocent Goed doel: 200 eurocent
- Jij: 675 eurocent Goed doel: 150 eurocent
- Jij: 700 eurocent Goed doel: 100 eurocent
- Jij: 725 eurocent Goed doel: 50 eurocent
- Jij: 750 eurocent Goed doel: 0 eurocent

Dit is het einde van de vragenlijst. Check alsjeblieft of je je leerlingen e-mailadres, klas en geslacht hebt ingevuld bovenaan de vragenlijst en of je één vakje hebt aangekruist in elke beslissing-situatie. Wanneer je hiermee klaar bent, leg je vragenlijst dan alsjeblieft met de tekst naar beneden neer zodat je leerlingen e-mailadres, klas en geslacht niet zichtbaar zijn. Steek dan je hand op zodat ik je

vragenlijst kan komen ophalen. Blijf alsjeblieft stil totdat ik alle vragenlijsten heb opgehaald. Bedankt voor je deelname!

Appendix E: Experimental procedure

Part 1 – Procedure followed to assign treatments

First, it needed to be randomly determined which havo and vwo classes would receive which treatment. The experimenter did this by flipping a coin. For both the havo and vwo classes throwing head would have the result that the class in which the experiment was conducted first would receive the standard treatment and the class in which the experiment was conducted later the charity treatment. Throwing tail, in contrast, would have the result that the class in which the experiment was conducted first would receive the charity treatment and the class in which the experiment was conducted later the standard treatment.

Part 2 - Procedure followed in the classrooms

The procedure followed in the classrooms in waves 1, 2 and 3 are explained below. There was a time period of 13 days between waves 1 and 2 as well as between waves 2 and 3.

Wave 1

The experimenter explained the following verbally to the subjects:

Hello everyone. I am Lonneke, a former student at Dalton Lyceum Barendrecht. I do the master "Behavioural Economics" at the Erasmus University Rotterdam. This school year is coming to an end, which means that we have to write our graduation project, called "master's thesis". For this thesis, I gather my own data by means of questionnaires. That is why I am here: You would be of great help by filling out a questionnaire.

The questionnaire consists of several decision situations and you have the chance to win money. It should be noted that this money is neither from Dalton Lyceum nor from the Erasmus University, but is provided by an external source. In the questionnaire, you will be asked for your student e-mail address so that if you won money in the experiment, I can contact you to ask for your bank account number and transfer the money. However, I won't contact you immediately after conducting this questionnaire. It will take some time due to some organisational reasons. Once I have transferred the money, I will send anonymised screenshots of the transfers to your teacher so that he can verify it.

I will not be able to link your student e-mail address to your face, and the rest of the experiment is also designed in such a way that neither your teacher and fellow students nor I will get to know your decisions.

If you have any questions, please raise your hand and wait with asking your question until I am there to help. Also, please ask your question in a low voice so that no other student can hear it. If you participate in this experiment, it is possible that I will ask you to participate in future experiments. Is there anyone who does not want to participate?

I will now hand out the questionnaires.

The experimenter then handed out the questionnaires and the subjects started filling them out when everyone had received their questionnaire. If the subjects had any questions, they raised their hands and asked their question in a low voice, so that no other subject could hear the question. Once the subjects were done filling out the questionnaire, they turned them over with the text facing the table and raised their hands (as they were instructed in the questionnaire to do so). The experimenter picked up the questionnaires and shuffled them so that the questionnaires could not be linked to the subjects because of the order in which the questionnaires were picked up. The experimenter then told the subjects the following: *As I already explained, if you are one of the winners, you will be sent an e-mail to ask for you bank account number, but this may take a while*. The experimenter thanked the students and then left the classroom.

Wave 2

The procedure in wave 2 was the same as in wave 1, except in the following aspects:

Start of verbal explanations in wave 1

Hello everyone. I am Lonneke, a former student at Dalton Lyceum Barendrecht. I am doing the master "Behavioural Economics" at the Erasmus University Rotterdam. This school year is coming to an end, which means we have to write our graduation project, named a master's thesis. For this thesis, I gather my own data by means of questionnaires. And that is why I am here: you would be of great help by filling out a questionnaire.

Start of verbal explanations in wave 2

Hello everyone. You probably remember me from approximately two weeks ago, when you filled out a questionnaire for my master's thesis. Now again, I would like to ask you to fill out a questionnaire. If you did not participate last time, you, unfortunately, cannot participate. As I discussed with your teacher, you can work on something on your own, quietly. Are people present who did not participate last time?

Wave 3

The procedure in wave 3 was nearly the same as in wave 2. A difference was that the subjects were not informed that it is possible that the experimenter would ask them to participate in future experiments. And instead of informing the students that if they did not participate last time, they cannot participate this time; they were informed that if they did not participated last time or the time before, they could not participate this time. Another difference was that instead that the subjects were told that it takes a while before the winners will receive an e-mail, they were told: *This was the last time I needed you to fill out a questionnaire. This implies that all winners of the three experiments will be sent an e-mail tonight to ask for your bank account number.*

Part 3 - Payment procedure

Standard group

After wave 3, for each class a subject needed to be randomly picked to be the dictator for all three waves. In addition, for each class and each wave a recipient and a game needed to be randomly picked. This was done in the following way:

In order to determine the dictator, the experimenter shuffled the questionnaires from wave 1 and picked one questionnaire randomly (with the text facing the table so that no text was visible). The student who had filled out this questionnaire was supposed to be the dictator for all three waves. There was, however, a possibility that this student had not been present in all waves. If this was the case, the experimenter randomly picked another questionnaire as to determine the dictator for the wave(s) the former dictator had not been present. After this, the experimenter randomly picked a questionnaire, without replacement, to determine the recipient in wave 1. After this, the experimenter removed the questionnaire of the recipient from wave 1 out of the questionnaires of wave 2. The experimenter then shuffled the questionnaires from wave 2 and randomly picked one. A recipient for wave 3 was picked in similar fashion.

Then four tickets with the numbers 'one', 'two', 'three', and 'four' were put in a hat and the experimenter picked one ticket. The number on this ticket is the number of the game that was going to be paid out for wave 1. This number was put back in the hat and the procedure was repeated twice in order to determine the games that were going to be paid out for wave 2 and 3 as well.

Now that the dictator(s) and the recipients were known as well as the games that had to be paid out, it was determined how much the dictator(s) had kept and how much s/he had given away for the games that were picked. The dictator(s) and recipients were sent an e-mail to ask for their bank account number. Upon receiving e-mail with their bank account number, the money would be transferred and anonymised screenshots of it would be sent to the teacher. However, thus far the experimenter has not received their bank account numbers. Therefore, the money could not have been transferred yet.

Charity group

After wave 3, for each class a subject needed to be randomly picked to be the dictator for all three waves. In addition, for each class and each wave a recipient and a game needed to be randomly picked. This was done in the following way:

In order to determine the dictator and the games that were going to be paid out, the same procedure was followed as in the standard group. To determine which charity was the recipient in what wave, tickets with numbers 'one', 'two' and 'three' were put in a hat. 'One', 'two', and 'three' stand for the charities "KWF Kankerbestrijding", "UNICEF", and "De Dierenbescherming" respectively. The experimenter picked three tickets out of the hat without replacement. The number on the first drawn ticket corresponds to the charity to which the dictator is matched in the first wave and the numbers on the second and third drawn tickets to the charity to which the dictator is matched in the second and third wave respectively.

Now that the dictator(s) and the order of charities were known as well as the games that had to be paid out, it was determined how much the dictator(s) had kept and how much s/he had given away for the games that were picked. The dictator(s) was/were sent an e-mail to ask for their bank account number. The money was transferred to the charities. And upon receiving e-mail with the dictator(s)' bank account number, the money would be transferred to the dictator(s) as well and anonymised screenshots of it would sent to the teacher. However, thus far the experimenter has not received their bank account numbers. Therefore, the money could not have been transferred yet.

Appendix F: Number of subjects

In table A, the abbreviations 'HST', 'HCT', 'VST', and 'VCT' stand for '*havo* standard treatment', '*havo* charity treatment', '*vwo* standard treatment', and '*vwo* charity treatment' respectively. '# correct' represents the number of subjects who filled out the questionnaire and should have done so and '# incorrect' represents the number of subjects who filled it out while they should not have done so.

Table A. Number of subjects

	Wave 1	**Wave 2**		**Wave 3**	
	# Subjects	*#Subjects*		*# Subjects*	
		# Correct	*# Incorrect*	*# Correct*	*# Incorrect*
HST	27	22	0	14	3
HCT	25	22	1	20	3
VST	22	21	0	19	0
VCT	21	20	0	19	0

Appendix G: Test results

Table B: Regression models on the amount taken on the aggregates of the games

	Ordered probit			**Ordered logit**			**Random effects**		
	Coefficient	z	p-value	Coefficient	z	p-value	Coefficient	z	p-value
Constant							194.02*** (26.328)	7.37	0.000
Wave 2	0.046 (0.188)	0.25	0.401	0.148 (0.319)	0.46	0.323	14.015 (16.055)	0.87	0.192
Wave 3	0.470** (0.213)	2.21	0.014	0.936*** (0.395)	2.37	0.009	57.955*** (16.055)	3.61	0.000
Charity * Wave 2	0.054 (0.226)	0.24	0.405	0.010 (0.380)	0.03	0.488	2.331 (21.815)	0.11	0.456
Charity * Wave 3	-0.265 (0.263)	-1.01	0.156	-0.595 (0.468)	-1.27	0.102	-28.147* (21.815)	-1.29	0.099
Charity	0.367 (0.367)	1.00	0.159	0.730 (0.648)	1.13	0.129	31.760 (34.215)	0.93	0.176
Game 2	-0.139* (0.093)	-1.49	0.068	-0.149 (0.148)	-1.01	0.156	-16.898* (12.551)	-1.35	0.089
Game 3	-0.493*** (0.150)	-3.29	0.001	-0.684*** (0.251)	-2.73	0.003	-61.111*** (12.551)	-4.87	0.000
Game 4	-0.697*** (0.166)	-4.20	0.000	-1.006*** (0.291)	-3.45	0.000	-73.843*** (12.551)	-5.88	0.000
Observations	864			864			864		

Notes: 'Charity * Wave 2' and 'Charity * Wave 3' represent interaction terms between the charity treatment and wave 2; and the charity treatment and wave 3 respectively. Standard errors are in parentheses, clustered at subjects level for ordered probit and ordered logit regressions.

*** Significant at the 1% level.

** Significant at the 5% level.

* Significant at the 10% level.

Table C: Regression model on the amount taken on the games separately

	Random effects											
	Game 1			*Game 2*			*Game 3*			*Game 4*		
	Coefficient	z	p-value	Coefficient	z	p-value	Coefficient	z	p-value	Coefficient	z	p-value
Constant	193.939*** (34.377)	5.64	0.000	207.576*** (33.347)	6.22	0.000	112.121*** (28.741)	3.90	0.000	110.606*** (30.817)	3.59	0.000
Wave 2	25.758 (27.805)	0.93	0.176	-27.273 (27.286)	-1.00	0.159	65.152*** (23.652)	2.75	0.003	-7.576 (30.783)	-0.25	0.401
Wave 3	101.515*** (27.805)	3.65	0.000	24.242 (27.286)	0.89	0.187	77.273*** (23.652)	3.27	0.001	28.788 (30.783)	0.94	0.174
Charity * Wave 2	10.140 (37.780)	0.27	0.394	28.555 (37.074)	0.77	0.221	-48.485* (32.137)	-1.51	0.066	19.114 (41.826)	0.46	0.323
Charity * Wave 3	-45.105 (37.780)	-1.19	0.117	2.681 (37.074)	0.07	0.472	-52.914* (32.137)	-1.65	0.050	-17.249 (41.826)	-0.41	0.341
Charity	0.932 (46.709)	0.02	0.492	2.681 (45.309)	0.06	0.476	51.981* (39.051)	1.33	0.092	71.445** (41.873)	1.71	0.044
Observations	216			216			216			216		

Notes: 'Charity * Wave 2' and 'Charity * Wave 3' represent interaction terms between the charity treatment and wave 2; and the charity treatment and wave 3 respectively.
Standard errors are in parentheses.
*** Significant at the 1% level.
** Significant at the 5% level.
* Significant at the 10% level.

Table D: Difference in difference analyses on the games separately

		Random effects											
		Game 1			*Game 2*			*Game 3*			*Game 4*		
		Coefficient	z	p-value	Coefficient	z	p-value	Coefficient	z	p-value	Coefficient	z	p-value
ST	W12	25.758 (27.805)	0.93	0.176	-27.273 (27.286)	-1.00	0.159	65.152*** (23.652)	2.75	0.003	-7.576 (30.783)	-0.25	0.401
	W23	75.758*** (27.805)	2.72	0.003	51.515** (27.286)	1.89	0.029	12.121 (23.652)	0.51	0.305	36.364 (30.783)	1.18	0.119
	W13	101.515*** (27.805)	3.65	0.000	24.242 (27.286)	0.89	0.187	77.273*** (23.652)	3.27	0.001	28.788 (30.783)	0.94	0.174
CT	W12	35.897* (25.577)	1.40	0.081	1.282 (25.099)	0.05	0.480	16.667 (21.757)	0.77	0.221	11.538 (28.316)	0.41	0.341
	W23	20.513 (25.577)	0.80	0.212	25.641 (25.099)	1.02	0.154	7.692 (21.757)	0.35	0.363	0.000 (28.316)	0.00	0.500
	W13	56.410** (25.577)	2.21	0.014	26.923 (25.099)	1.07	0.142	24.359 (21.757)	1.12	0.131	11.538 (28.316)	0.41	0.341
Diff	W12	10.140 (37.780)	0.27	0.394	28.555 (37.074)	0.77	0.221	-48.485* (32.137)	-1.51	0.066	19.114 (41.826)	0.46	0.323
	W23	-55.245* (37.780)	-1.46	0.072	-25.874 (37.074)	-0.70	0.242	-4.429 (32.137)	-0.14	0.444	-36.364 (41.826)	-0.87	0.192
	W13	-45.105 (37.780)	-1.19	0.117	2.681 (37.074)	0.07	0.472	-52.914* (32.137)	-1.65	0.050	-17.249 (41.826)	-0.41	0.341

Notes: 'ST', 'CT', and 'Diff' represent 'standard treatment', 'charity treatment', and 'difference between the treatments' respectively. 'W12', 'W23', and 'W13' stand for 'the increase from wave 1 to wave 2', 'the increase from wave 2 to wave 3', and 'the increase from wave 1 to wave 3' respectively. Standard errors are in parentheses.

*** Significant at the 1% level.

** Significant at the 5% level.

* Significant at the 10% level.

7. Bibliography

Andreoni, J., Harbaugh, W. T., & Vesterlund, L. (2008). Altruism in experiments. *The New Palgrav Dictionary of Economics: Volume 1–8*, 134-138.

Andreoni, J., & Miller, J. (2002). Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737-753.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly Journal of Economics*, *116*(1), 293-312.

Bacon, J. (2018). These lucky charities got cash from mystery $560M Powerball winner. *USA Today*. Retrieved from https://eu.usatoday.com/story/news/nation/2018/03/08/mystery-560-m-powerball-winner-gave-cash-these-2-lucky-charities/406231002/

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

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

Bolton, G. E., Katok, E., & Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International journal of game theory*, *27*(2), 269-299.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review*, *90*(1), 166-193.

Brosig-Koch, J., Riechmann, T., & Weimann, J. (2017). The dynamics of behavior in modified dictator games. *PloS one*, *12*(4), e0176199.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817-869.

Drewett, Z. (2018). Britain's most charitable lottery winner gave away £5,500,000 of her winnings. *Metro*. Retrieved from https://metro.co.uk/2018/05/22/britains-charitable-lottery-winner-gave-away-5500000-winnings-7569816/

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, *47*(2), 268-298.

Eckel, C. C., & Grossman, P. J. (2000). Volunteers and pseudo-volunteers: The effect of recruitment method in dictator experiments. *Experimental Economics*, *3*(2), 107-120.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583-610.

Ensor, J. (2014). EuroMillions lottery winner to give away €50 million to charity. *The Telegraph*. Retrieved from https://www.telegraph.co.uk/news/worldnews/europe/france/10864431/EuroMillions-lottery-winner-to-give-away-50-million-to-charity.html

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and economic behavior*, *54*(2), 293-315.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817-868.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic behavior*, *6*(3), 347-369.

Goede Doelen Nederland. (2018). *Feiten & Cijfers Goede Doelen 2017*. Retrieved from https://goededoelen.nl/images/documenten/2017-Feitenencijfers.pdf

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

Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American economic review*, *86*(3), 653-660.

Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, 398-407.

de Jong, L. J. (2017). Differences in Social Preferences Towards Friends and Strangers. Economics. Retrieved from http://hdl.handle.net/2105/38327

de Jong, L. J., Mol, W. M., Phillips, J. O., Straka, E., & de Wilde, K. (2017). Are people truly altruistic? (research report for the seminar behavioural economics). Economics.

Kantowitz, B. H., Roediger, H. L., & Elmes, D. G. (2008). *Experimental Psychology* (9th edn.). Belmont: Wadsworth Cengage Learning.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American economic review*, 1281-1302.

de Ronde, K. (2008). Financiële hulp voor loterijwinnaars. *de Volkskrant*. Retrieved from https://www.volkskrant.nl/economie/financiele-hulp-voor-loterijwinnaars~b6ab7f82/

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, *5*(17), 61-71.

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, *15*(60), 243-253.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124-1131.