



Master Thesis
Behavioral Economics

**The Dictator Game: Gender Differences
in Behavior under Social Pressure**

Author: G.A.M van de Pol

Supervisor: Prof. dr. K.I.M. Rohde

Rotterdam, June 24th 2014

Abstract

In this study my main purpose is to examine gender differences in behavior under social pressure in the dictator game. In addition, I also try to visualize the underlying decision process by thoroughly reviewing existing literature and by means of my own experiment. In different treatments I check for a pure gender difference, a gender difference in behavior under social pressure, and a possible differing social experimenter demand effect in the case of either a male or a female experimenter. Results show that (1) it is likely that a pure gender difference exists with women behaving more altruistically and that (2) a female experimenter positively influences female giving behavior. In addition, participants state that the importance of fairness as a motive for giving is significantly higher than the importance of any other motive.

Content

1 Introduction	4
2 Motive analysis / Literature review	8
2.1 Social pressure	10
2.2.1 Fairness	12
2.2.2 Altruism	13
2.2.3 Warm-glow giving	17
2.2.4 Self-image	18
2.2.5 Social image	19
3 On real and hypothetical rewards	20
4 Treatments and hypotheses	22
5 Experimental design	24
6 Determining sample size	28
7 Methods	31
7.1 Hypothesis testing	31
7.2 Motive analysis	31
8 Results	33
8.1 Hypothesis testing	33
8.2 Motive results	36
8.3 The randomness of the sample	39
9 Conclusions	43
10 Discussion and limitations	45
11 Future research	47
12 References	48
Appendix A	52
A.1 The anonymous treatment answer sheet.....	53
A.2 The non-anonymous treatment main answer sheet	54
A.3 The non-anonymous treatment additional answer sheet	55

1 Introduction

The dictator game is a game in experimental economics where one player, the dictator, decides on how to divide a fixed amount of money between himself and a recipient. The dictator and recipient may or may not know each other's identity. The game is in fact a one-person task of the dictator; the recipient is entirely passive. The recipient simply receives whatever amount the dictator decides on.

Game theory would suggest that individuals care only about their own pay-off, while preferring more money to less, with the result of players keeping all the money for themselves and leaving nothing for the recipients. However, in practice we observe players in this game giving away some percentage of the total endowment. Forsythe et al (1994) find that only 36% of players act in line with game theory predictions in their \$5 dictator game (leaving nothing for the recipients). In addition, 22% of players give away an equal share or more.

Why do people give away this money? This question seems to be a complex one. Many different motives drive this decision; some with thin lines between them. This complex frame of motives includes the following most widely described motives for giving: altruism (Eckel and Grossman (1996) ; Koch and Normann (2008) and Korenok et al (2009)), fairness (Kahneman et al (1986) ; Forsythe et al (1994) ; Bolton et al (1998) and Hoffman et al (1994)), warm-glow (Korenok et al (2009) ; Lucassen and Grossman (2013) ; Andreoni (1990) and Tonin and Vlassopoulos (2011)), self-image (Tonin and Vlassopoulos (2013) ; Grossman and Van der Weele (2013) and Ploner and Regner (2013) and social image (Andreoni and Bernheim (2009) and Bénabou and Tirole (2005)).

To understand behavior better, countless different manipulations of the game have been constructed. To illustrate (although it may seem a little exaggerating), it led

Oechssler (2010) to the words “Please, not another dictator game!”. An interesting example is that when taking was incorporated in the action set in List (2007), people were less likely to give. Also, when dictators were shown physical attractive recipients, they gave more (Rosenblat (2008)).

Gender differences have been checked for as well, with varying results. Andreoni and Vesterlund (2001) and Eckel and Grossman (1998) find women leaving more money for recipients. Gong et al (2009) find a reversed result in their study among the Mosuo, a Matrilineal ethnic minority group in southwestern China. The recipient’s gender is of influence as well. Saad and Gill (2001) find that men as well as women give more money to women. Rosenblat (2008) also concluded that being a women is an advantage for recipients.

In addition (in non-dictator game studies), men and women have also been found to react differently to peer pressure. Brown (1982) finds that men and women assign different strengths to specific pressures. Taylor and Wong (1996) find similar results. For example, the higher girls perceive the value that their friends attach to things associated with popularity, the higher their intrinsic motivation is to go to parties or to have a boyfriend. In a study that concerns a working environment, Bellemare et al (2010) find men significantly reducing their productivity when given a very low signal about the productivity of another worker, while this is not the case for women.

These findings on peer pressure (in non-dictator game studies) and gender-differences in the dictator game raise the question whether men and women act differently under social pressure in the dictator game. This is the main research question of this paper:

Main research question: Do men and women act differently under social pressure in the dictator game?

In this paper we try to answer this question by constructing different experimental treatments where dictators (divided in men and women) feel pressure to give money to the recipient from the experimenter (either a man or a woman). The social pressure that participants feel from the experimenter is called the social experimenter demand effect. It is useful to distinguish between social experimenter demand effects and cognitive experimenter demand effects. Cognitive experimenter demand effects, described by Zizzo (2010), arise when participants, because of information or instructions given about the experiment, behave according to formed beliefs about what seems to be considered appropriate behavior for the task at hand. Evidence has been found by List (2007) and Bardsley (2008) by showing that behavior differs when some form of taking away money from the other player is also incorporated in the action set. In addition, Branas-Garza (2007) found evidence for the related social experimenter demand effect. This effect arises when participants feel pressure from the experimenter by the instructions he or she is giving as an authority. Branas-Garza (2007) showed that the participants (students) were more likely to give away money when the experimenter was a professor.

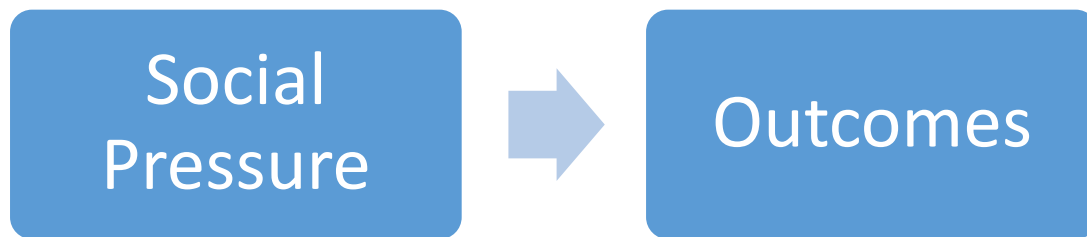
We construct our research such that we do not only answer the question whether men and women act differently under social pressure, we also check whether it matters where the social pressure is coming from in terms of experimenter-gender. We also compare results with control groups where women as well as men participate in a dictator game as anonymously as possible, to feel as little social pressure as possible.

This is to compare outcomes with standard dictator game results. Note that the gender of the recipient is no issue in any treatment.

In the following we will firstly study previous dictator game literature on the (many) main motives for giving. Note that this information is not needed to answer our main research question. However, we think it is important to gain detailed knowledge on the relations in this underlying complex frame of motives, and how social pressure can have a role in this. This will help us with conducting our own experiment and to in the end draw more detailed conclusions. After this we will step by step explain how we constructed our own experiment by talking about real and hypothetical rewards, treatments and hypotheses, our experimental design and our methods. We will then end with results, conclusions, a discussion and future research.

2 Motive analysis / Literature review

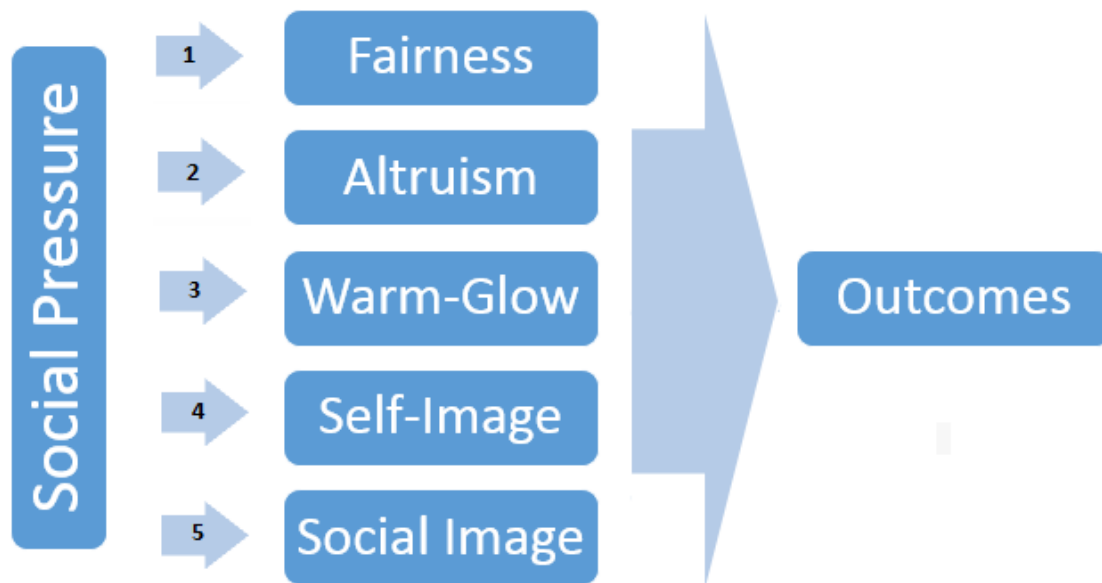
In the following we will, by carefully reviewing previous literature, try to describe the previously mentioned motives for giving away money in the dictator game. Please note that to answer the main research question, we do not need this information on the underlying framework of motives. For the main research question we only need to measure direct effects of social pressure on outcomes, like this:



With outcomes we mean the amounts of money that participants decide to give away. However, as this is far from the whole story, we also seek to obtain knowledge on *how* social pressure influences outcomes. Ideally we would want to point out the main driver behind the decision-making process in dictator games. Understanding the underlying framework will also be useful for constructing our own experiment. Also, as a part of the experiment, we will ask our respondents certain questions to obtain extra knowledge on what *they* perceive to be the main driver for their behavior. In this chapter we will start by showing in what way social pressure might influence motivations (and consequently behavior). The previously mentioned main motives for giving away money in the dictator game are, in no particular order:

- Fairness
- Altruism
- Warm-glow giving
- Self-image
- Social image

The relative importance of every one of these motives might increase or decrease when we increase social pressure during the decision process. We think it is crucial to realize that social pressure might have an effect on one or more of these motives, but is not a motive itself. However, with social pressure present, improving social image becomes a motive. Graphically, the way social pressure might influence outcomes is like this:



The role of social image might need a little more explanation. Social image is a bit different from the other motives because it only arises when social pressure is present. Also, without the presence of social pressure participants are not concerned with what others think is (for example) fair. However, *with* social pressure, participants might start to act more fairly because they think this will improve their social image. This means that in this case motives might influence each other and in this sense the above representation might be a little simplified. Still, we hope it helps to better understand the framework. We hope to have illustrated the relation between social pressure, motives and outcomes.

In the following we will start to talk about previous literature on social pressure. After this we will one by one discuss previous literature on the main motives for giving

away money in the dictator game. But before we can do this we have to describe their exact meaning in detail, as this is not always straightforward. All this will help us gain understanding of the more complex framework of the decision-making process. In our own experiment we will try to gain even more information.

2.1 Social pressure

Social pressure is the pressure someone can feel from another person or group to behave in a certain manner. In many papers it appears to be the case that social pressure is a factor in the dictator game. We start with the following: Zizzo and Fleming (2011) examined this in a unique way when it comes to dictator games, and find evidence for social pressure being an important factor by showing that outcomes can be linked to measures of social pressure sensitivity (Stöber (2001)). The measures are questionnaire scales where people state how much they agree with varying statements about their own behavior in 17 different situations. This way conclusions can be drawn on how sensitive people are to social pressure. Two examples are:

- I always stay friendly and courteous with other people, even when I am stressed out.
- There has been at least one occasion when I failed to return an item that I borrowed.

Stöber (2011) found a way to use these statements to help identifying people that are sensitive to social pressure. The research of Zizzo and Flemming (2011) is unique in the sense that, by using these statements, the importance of social pressure could be measured in a somewhat indirect way. The research showed that the people that are sensitive to social pressure, are also the people giving away money in the dictator game. In Zizzo and Flemming (2011) participants were even split up in two sessions, where

in the first their social pressure sensitivity was checked, and in the second the dictator game was played (with the same participants).

More research on social pressure in the dictator game has been done by examining social experimenter demand effects. As mentioned earlier, these effects arise when participants implicitly or explicitly perceive some kind of social pressure from the experimenter. For example, in Branas-Garza (2007) it is shown that students are more likely to give away money when the experimenter is a professor, compared to when they do not know the experimenter. This means that the experimenter influences outcomes, although we have to be careful with the exact interpretation here. Social pressure having a role seems likely though, also because of the fact that Zizzo and Fleming (2011) already showed that this is a factor. Also, we have found no evidence that people might give more with different experimenters because of other reasons than social pressure. We conclude that giving behavior can be influenced by the way the experimenter acts, or simply by the fact that he is a well-respected person. In this last case behavior is apparently influenced towards more giving.

However, Koch and Normann (2008) find no evidence for social pressure having an influence in their dictator game. There does not seem to be a difference in the one case where recipients are aware of how the money is allocated to them, and in the other where the recipients receive the money without any indication of where it came from.

Above results show that social pressure can be of importance in some researches, but not in all. It can be the case that social pressure does not have an influence on all motives. We will see what the results will be in our own experiment, to show how social pressure influences outcomes. As explained earlier, in our experiment we will, by asking the participants questions, try to more concretely give insights on

the indirect path of how social pressure influences outcomes. We will now discuss the most important motives for giving away money in the dictator game.

2.2.1 Fairness

Fairness is one of the most important concepts when it comes to social interactions. However it is not simply defined. A phrase in Bicchieri (2005) describes this: "We can come to accept the most onerous tasks if we are convinced that the decision procedure was fair and, conversely, we may reject a profitable exchange if we feel treated unfairly. Since the dawn of philosophy, a concern with fairness, what is it, how to define it, has been central to the philosophers quest".

This suggests that the decision to give or keep money may be partly based on what the participant thinks is 'fair'. It might be the case that people do not give away money because they are altruistic (to be better described later) but because they somehow perceive this to be fair, and they want to be fair. This makes sense to us, as many people seem to give away exactly half the money, and not a cent more (Forsythe et al (1994)). We conclude that fairness might well have some influence on the decision, as it seems plausible that an equal division is considered most fair by everyone, and that this is an important factor in the decision. It seems less plausible to us, that for some reason, many people are exactly equally altruistic, all arriving on the same amount (half the total endowment) to give away because of this reason.

We find evidence for the desire to act fairly as a motive in Bolton et al (1998), where it is concluded that dictators are primarily concerned with securing what they consider to be their fair share. When participants had the choice of leaving more or less than the amount they initially decided on, they chose less. We also find another interesting suggestion. Hoffman et al (1994) suggest that fairness might not be an internal, but an external motivation, influenced by judgments of others. This would

mean that people do not act in a fair way because they have an intrinsic desire to do so, but because they feel pressure from other people expecting them to do so. This type of reasoning shows the difficulty of the decision-framework. However, as we only need to measure the direct effect of social pressure on outcomes to answer our main research question, this discussion does not add difficulty here. It does provide insight on the underlying framework. Andreoni and Bernheim (2009) add to this discussion. They find that fairness by itself does not account for certain patterns, but the fact that people like to be perceived as fair does. This accounts for previously unexplained patterns. It would also most likely mean that social pressure would increase giving. People want to act fairly because of the pressure they feel to do so.

We think that fairness has an important role in the decision process. The influence of the desire to act fairly on outcomes seems to be present. Apart from this, it also seems to be the case that social pressure influences fairness. This means that apart from people having an intrinsic desire to act fairly, we also think that social pressure influences the desire to act fairly because people also want to be perceived as fair. By asking our participants additional questions we will try to examine the importance of fairness in the decision process in a more concrete way.

2.2. Altruism

To start explaining how altruism can have its role in the decision process, we have to properly define the term, or at least explain the difficulties with this, as different definitions of altruism exist. The first distinction we see is the one between evolutionary and psychological altruism, and we want to start by making clear that, although this is rarely explicitly stated, dictator game papers are about psychological altruism. This is because the main focus of evolutionary altruism is based only on behavior, not taking into account the reasons behind the behavior, as is described in Wilson (1992). To

illustrate: when an animal increases the fitness level of another animal, only to see his own decrease, then from an evolutionary perspective this is considered altruism. In the case of animals (some of them like insects even assumed to be incapable of thought) it does not seem objectionable to speak of altruism in this case. However, in the case of dictator games, if we observe a human individual donating money to improve his image with the expectation of receiving something back in the future, in general we do not want to consider this altruism. Therefore and for similar cases, we sometimes need a narrower definition of altruism. Psychological altruism does take motives into account, and Batson (1991) defines (psychological) altruism as a motivational state with the ultimate goal of increasing another's welfare. This would imply the following:

1. Donating money because it makes you feel better about yourself is not altruistic.
2. Donating money because it makes you feel better as a result of derived pleasure purely from the act of giving is not altruistic.
3. Donating money because of the fact that increasing another's welfare makes *you* feel happy is not altruistic per se.
4. Donating money because increasing another's welfare is your ultimate goal is altruism.

In the literature, we find examples of other researches where acts similar as number two and three described above seem to be considered as altruism. For example in Eckel and Grossman (1996) it seems to be the case that deriving pleasure from the happiness of others is considered altruism. They point out that it is no surprise that Hoffman et al (1994) find no evidence for altruism, because as they see it, Hoffman et al (1994) have removed virtually every reason for giving money away in their double-blind procedure where all participants are as untraceable as possible. However, using the definition as

stated in Batson (1991), the approach of Hoffman et al (1994) seems correct, as an increased welfare of others is the only consequence of donating money known to the players. The possibility of deriving pleasure from it yourself is arguably fully taken away by leaving no clue whatsoever of who you are helping. Eckel and Grossman (1996), however, suggest replacing the anonymous recipients with charity organizations, and make comparisons with giving money away to family and friends. In these cases it becomes very hard to tell what the ultimate goal is in giving away money. Since Eckel and Grossman (1996) are not concerned by this, this leads us to the conclusion that they use a different definition of altruism than Batson (1991) and Hoffman et al (1994).

In some other papers where altruism has a central role, researchers have put limited effort in defining or describing the term (see Levine (1998) or Andreoni and Miller (2002)). It is clear though that they do not use the definition of Batson (1991).

Kerr et al (2004) also experience difficulties with the term. When it comes to evolutionary altruism, they conclude the following: "Within evolutionary biology, different authors have interpreted the concept of altruism differently, leading to dissimilar predictions about the evolution of altruistic behavior." They go on to describe four different definitions of altruism, and it is concluded that the different interpretations of the term are the reason of authors arriving at contradicting conclusions. Wilson (1992) states a similar problem, also involving psychological altruism: "Ignoring the precise nature of both psychological and evolutionary definitions has obscured many important issues, including the biological roots of psychological altruism".

Barring interpretation difficulties, we can highlight some previous findings on altruism:

- Hoffman et al (1994) et al observe that dictator game offers lower as social isolation increases. The data is generally supportive of the economic assumption of non-altruistic behavior. However, in the highest degree of anonymity tested, including double-blindness, participants still transfer considerable amounts of money. This means results *may* reflect (narrowly defined) altruistic preferences.
- Zizzo (2011) concludes that dictator games do not measure altruism. A lot of the dictator giving is a result of the experimenter demand effects. Players form some kind of belief of what the experimenter thinks they are supposed to do, and actions can solely be the result of the desire to act accordingly.
- Koch and Normann (2008) contradict these findings by concluding that dictators are purely internally motivated. In this research two double-blind treatments are compared. In the one case recipients are aware of how the money is allocated to them, and in the other the recipients receive the money without any indication of where it came from; they do not even know they are part of a game. Results turn out not to be significantly different. This leads Koch (2005) to the conclusion that (1) players are purely internally motivated (this leaves room for altruistic reasons having a role in the decision process) and (2) that social pressure does not have an influence on altruistic reasoning.
- Bardsley (2008) finds that the reversing of generosity between his two experimental treatments is inconsistent with any theory of dictator game giving which regards underlying altruistic motivation as causing a desire to share the endowment. This is also consistent with Andreoni and Miller (2002).

The main conclusion of the previous analysis is that altruism is a difficult subject where a lot of indistinctness is present across papers. We feel that in the field of dictator games where psychological altruism has a crucial role, it can be useful to put effort in

explaining exactly what is considered altruism in the research. Different interpretations of the term seem to be present, but sometimes results are compared.

In our case where we want to provide a clear picture of what can be of importance in the decision process, we obtained enough information with understanding the difficulties of the discussion. We conclude that although giving away money seems like an act with altruistic motivations behind it, it is far from certain that this is the case. We keep in mind that Bardsley (2008) found no evidence of social pressure influencing altruistic reasoning. However, the more general discussion seems to be a never-ending one. Still, we will try to obtain more information on the importance of altruism by directly asking our respondents questions about this.

2.2.3 Warm-glow giving

People sometimes derive pleasure solely from the act of giving. This means that pleasure derived later as a consequence of seeing the recipient happier, or pleasure from any other consequence, is not taken into account. This phenomenon has first been described as warm-glow giving by Andreoni (1989). The existence of such an effect, also in earlier researches, has been acknowledged many times now (Arrow (1972) ; Sen (1977) ; Margolis (1984) ; Andreoni (1990) ; Tonin and Vlassopoulos (2011) ; Lucassen and Grossman (2013)).

Strong evidence of warm-glow giving for the dictator game is found by Korenok et al (2009) who state that (for their specific dictator game) ‘the behavior of 89% of the dictators can be rationalized by the utility function that includes both altruistic and warm glow effects, while only 16% of the dictators made choices that are consistent with the utility function that excludes the warm glow effects.’ Crumpler and Grossman (2008) draw a similar conclusion in their dictator game.

We conclude that a warm-glow effect exists in dictator games. There is evidence of this effect directly influencing outcomes. We do not find information on social pressure possibly influencing warm-glow.

2.2.4 Self-image

We think self-image can be defined as the idea an individual has of his or her capabilities, appearance and personality. Although (self)-image is mentioned as a reason repeatedly throughout papers, not many try to look at the influence of image in isolation in dictator games. First of all, we find a general result in Bénabou and Tirole (2005). They conclude in their non-dictator game research that image-related rewards create doubts about the true motive for which good deeds are performed. Even warm-glow effects appear to be possible to arise because of self-interest. This means as much as that someone can enjoy the act of giving (sometimes considered a part of (impure) altruism) because this act makes this person feel good about himself. This shows the possible thin line between certain motives. Similar reasoning is also present in Tonin and Vlassopoulos (2013).

Ploner and Regner (2013) did conduct an experiment to examine the role of self-image in the dictator game. Before players played the game, they were part of an experiment where their self-image was manipulated. Consequently, players could cheat to receive a high endowment to play the dictator game with. The participants that cheated, proceeded to give away significantly more money than the participants that did not cheat. The formulated reason is that the cheating participants needed moral cleansing to feel good about themselves again. Murnighan et al (2001) also find evidence for a role of self-image. They test the self-impression management model, which predicts that individuals want to show themselves in a positive light, even when nobody is observing their actions. This self-image model correctly predicts giving

behavior for the three differently classified types of dictators they consider: 'rational', 'equal' and 'other' dictators.

We conclude that self-image has a role in the dictator game. By definition, we think it is unlikely that social pressure has an influence on self-image. However, in our experiment we do not want to rule out this possibility.

2.2.5 Social image

Social image is closely related to social pressure. If social pressure influences outcomes, it is likely that this is the case because people want to improve their social image.

Grossman and Van der Weele (2013) found evidence for social image having a role in the dictator game when they tested a Bayesian signaling model driven on social-images. This model partly predicted outcomes and provides evidence for social image being important.

Andreoni and Bernheim (2009) also examine social image in the dictator game. They, as mentioned earlier, found that fairness by itself does not account for certain patterns, but the fact that people like to be perceived as fair does. This would most likely mean that social pressure would increase giving, because people can improve their social image by acting fairly. Social image is then important because of that fact that if participants do not act a certain way they damage their image.

We conclude that social image has an important role in the dictator game. Apart from findings of Grossman and Van der Weele (2013) and Andreoni and Bernheim (2009) on social image we already found indications of an influence of social pressure in the dictator game. We think that if social pressure influences outcomes, it is likely that this is the case because people want to improve their social image. We will try to make this story more concrete by asking our respondents questions that reveal how much their social image had a role in their decision-making.

3 On real and hypothetical rewards

In general, varying conclusions have been drawn when examining possible differences in results when it comes to using real or hypothetical rewards to incentivize participants. For example, Hinvest and Anderson (2010) find differences for impulsive choices, but not for risky choices.

Locey et al (2011) seem to study a field a little closer to the dictator game. Results for real and hypothetical situations are compared for a repeated prisoner's dilemma game and a tit-for-tat game. On most measures, reward type (and discounting rates) did not make a significant difference in cooperation in these games.

Forsythe et al (1994) examine (among other questions) real and hypothetical situations for the dictator game. Differences are found, although results are conflicting for the experiments conducted in April and September. However, it is concluded that this is a result of the power of the test. For the closely related ultimatum game, even after additional testing, results remain inconclusive.

Ben-Ner et al (2008) seem the only ones purely focusing on the real/hypothetical reward discussion for the dictator game. Two situations are compared in their research: one where participants play a dictator game with real money, and one where participants indicate what they think they would give in a hypothetical dictator game. It turns out that outcomes are not significantly different. Additional investigation reveals that from information from self-report tests, differences can be observed if we distinguish two personality traits: agreeableness and extraversion. As we will do nothing more than testing for significant differences in giving amounts, this would not be a concern for us when we would not be using real money.

All in all, we conclude that there is not enough evidence to decide that real money is a must for researching dictator games. The general results differ per situation.

When looking at similar games, mostly there seem to be no differences. Forsythe et al (1994) find evidence, although some results are conflicting. The only paper purely focusing on the question we ask ourselves here (Ben-Ner et al (2008)), suggests using hypothetical rewards is suitable for answering our main research question. We will proceed to do so.

4 Treatments and hypotheses

As explained earlier, we will construct the following dictator game treatments:

- Treatment 1 (**MA**): Men playing the game anonymously.
- Treatment 2 (**FA**): Women playing the game anonymously.
- Treatment 3 (**MM**): Men playing the game under pressure of a male experimenter.
- Treatment 4 (**FM**): Women playing the game under pressure of a male experimenter.
- Treatment 5 (**MF**): Men playing the game under pressure of a female experimenter.
- Treatment 6 (**FF**): Women playing the game under pressure of a female experimenter.

In the past, women have been observed to give away more money in the dictator game, except for a very specific minority group in southwestern China. We expect to find that women will leave more money for the recipients. Therefore we hypothesize that we will find a gender difference between treatment **MA** and **FA**:

- Hypothesis 1 : Women in treatment **FA** will give significantly more money than men in treatment **MA**.

We will also compare all pressure-treatments with their anonymous counterparts. Although this is not needed to answer the main research question, this provides insights on the importance of social pressure. For all social pressure treatments we expect participants to give significantly more than in their anonymous counterpart.

- Hypothesis 2 : Men in treatment **MM** will give significantly more money than men in treatment **MA**.

- Hypothesis 3 : Women in treatment **FM** will give significantly more money than women in treatment **FA**.
- Hypothesis 4 : Men in treatment **MF** will give significantly more money than men in treatment **MA**.
- Hypothesis 5 : Women in treatment **FF** will give significantly more money than women in treatment **FA**.

Because in the past gender differences in behavior under pressure have seemed to be situation-specific, we have no expectation in favor of one of the genders when it comes to behavior under pressure of either a male or a female experimenter. Therefore we hypothesize that we will not find, apart from a possible ‘standard’ gender-difference found between treatment **MA** and **FA**, a difference in behavior under pressure between treatment **MM** and **FM**, and **MF** and **FF**.

- Hypothesis 6 : The difference between **MM** and **FM** will not be significantly different from the possible difference between **MA** and **FA**.
- Hypothesis 7 : The difference between **MF** and **FF** will not be significantly different from the possible difference between **MA** and **FA**.

Because we have, as already mentioned, not found any previous literature on a possible difference between male or female experimenters, we do not hypothesize a difference between the following treatments:

- Hypothesis 8 : We will not find a significant difference between treatments **MM** and **MF**.
- Hypothesis 9 : We will not find a significant difference between treatments **FM** and **FF**.

5 Experimental design

All participants were approached at the Erasmus University of Rotterdam. They were approached at different locations to try to prevent that they would all be active in the same field of study. Treatment **MA** and **FA** were the same in comparison to treatment **MM**, **FM**, **MF**, and **FF**, except for the fact that these participants did not feel social pressure from the experimenter. To achieve this, before filling in the survey, participants in treatment **MA** and **FA** were told that their answers could not be traced back to them. To achieve this, they were asked to put the answer sheet in the provided non-transparent envelope after they were done. After these short instructions, the experimenter made sure that the participant was alone while filling in the answer sheet, only to come back when the participant was finished to collect the closed envelope. The exact sheet given to the participants in treatment **MA** and **FA** can be found in Appendix A. The only difference between treatment **MA** and **FA** is that treatment **MA** consists of male participants and treatment **FA** of female participants.

Treatment **MM** and **FM** are equal to treatment **MA** and **FA** respectively, except for the fact that the participants were not told they were anonymous. In contrast, they were told that after they were done, the experimenter, a man, would ask them some questions about their decision. While the participant was filling in the first part, the experimenter made sure he was out of sight. However, before he left, the experimenter explicitly told that he was going to ask the participants additional questions about their decision afterwards. He did this to make sure that the participants knew that their answers would be intently observed, to let them feel social pressure. The additional questions were all asked with the intention of obtaining more information about the different motives for giving, and were asked on an additional answer sheet. The answer sheet contained the following:

(1) If someone decided to give €0 and keep €10:

Please indicate how much you agree with the following statements:

1. I would be more likely to give if more people would be watching.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

2. I would be more likely to give if I would know more about the other person.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

Or (2) If someone decided to give more than €0 and keep less than €10:

Please indicate how much you agree with the following statements:

3. I gave money because it makes me feel better about myself.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

4. I gave money because I felt I was supposed to do so.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

5. I gave money because I enjoy giving.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

6. I gave money because I care about the other person.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

7. I gave money because I think this is fair.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

8. I gave money because I think it would improve my image.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

With statements 1 and 2 we respectively tried to measure the importance of non-altruistic and altruistic reasons. We asked the non-givers to state their opinion only on these two statements, because it was hard to formulate statements that would reveal what *could* have been important for them (as apparently for them no motive had been important enough to give a positive amount). With statements 3 to 8 we respectively tried to measure the importance of self-image, experimenter pressure, warm-glow, altruism, fairness and social image.

It can be seen that the data obtained for motive-analysis is Likert-type data. A typical Likert-item is a single questionnaire item measuring an attitude by providing multiple categories to describe this attitude, like this:

- I gave money because this is fair

Disagree – Somewhat disagree – Neither agree or disagree – Somewhat agree – Agree

Note that this is a little different from what we present our participants with. We will try to explain the reason. An ever-lasting discussion on Likert-type data has been going on when it comes analyzing this data. The problem is that distances between numbers might not be equal. For example, one cannot be sure that the step from ‘Agree’ to ‘Somewhat agree’ is as big as the step from ‘Somewhat agree’ to ‘Neither agree or disagree’. Also, what one person might call ‘agree’ another person might call ‘somewhat agree’. This means that although two people might have the same attitude, they might state different numbers. As a consequence, some people argue that the data cannot be treated as interval data. Interval data allows for better comparisons between two outcomes, and is thus preferred. Although we could try to use a Rasch model to fix the problem (a Rasch model tries to transform ordinal data into interval data), we opted to present items the way we showed before:

-
- I gave money because I think this is fair.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

The reasoning is as follows. First of all we work with a relatively high number of 7 possible answers which reduces the problem. However, this is not the main point. The main point is that we did not work with categories. We did not mention a category for every number, trying to use the correct words in making the steps equally spaced. Two people agreeing exactly as much with a certain statement did not have the problem of

deciding whether this is called ‘agreeing’ or ‘somewhat agreeing’. We forced the people to treat every step as an equal one by just letting them pick a number in between the two extremes. This way we think we can treat the data as interval data.

When the experimenter would come back to observe the answers of the first answer sheet, the participant was given the additional answer sheet. The experimenter pointed out (depending on the answer on the previous sheet) which additional questions he wanted the participant to answer. To make sure the participant did not feel pressure while filling in this sheet (he or she could be reluctant to admit having felt pressure when this pressure is still present), he or she was asked to, after answering (so that answers could not be traced back to the participant), put this sheet in a provided envelope, together with the first answer sheet. The experimenter was again at a safe distance while the participant was answering, only to come back to collect the closed envelope and thank the participant. Answer sheets and additional questions given to participants in treatment **MM** and **FM** can be found in Appendix A.

Treatment **MF** and **FF** are exactly equal to respectively treatment **MM** and **FM**, except for the fact that the experimenter was a woman. This woman is a 22-year old master-student at the Erasmus University herself, and was instructed to act as similarly as possible as myself (the 23-year old male experimenter).

By construction of the experiment, it was not possible to ask the participants in the anonymous treatments the same questions on motive importances as the participants in the social pressure treatments. This would either be (1) decreasing their degree of anonymity, or (2) different from the way this was asked in the social pressure treatments. This means we will not be able to draw additional conclusions on how social pressure *changes* the importance of motives. However, we will be able to show relative importances of motives of people making their decision under social pressure.

6 Determining sample size

Deciding on sample size requires the following (Lenth (2001)):

- Deciding on the significance level (alpha).
- Deciding on the power (beta).
- Eliciting an effect size of interest.
- Obtaining information on expected variance.

Deciding on effect size appears the most difficult task in Lenth (2001). While Cohen (1977) opts for using standard effect sizes when you are experiencing difficulties (small, medium or large with corresponding definitions), Lenth (2001) argues that you should always decide on effect size based on your research and scientific questions. We agree with this last statement, and also add that we think it is not necessarily logical to define desired (meaningful) significant differences based on the variance of the observations. We decide to set the effect size at €1. This is in line with the fact that our respondents will be able to give money only in increments of €1. This also appears to be a meaningful difference for us, considering that the choices range from €0 to €10. This choice of effect size means as much as this: when the actual difference between means in two treatments is one dollar, we have a 95% chance of detecting it. If the actual difference is smaller, the probability of detecting it becomes lower. This does not mean we are unlikely to detect smaller differences. We are still likely to detect smaller differences, only not as likely as for our selected effect size.

We will be testing multiple hypotheses in this research, all on a 5% significance level. The reason we will not use a significance level of 10% is as follows: if we would use an alpha of 10% for the 5 hypotheses needed to answer our main research question (hypotheses 1, 6, 7, 8 and 9), the chance of wrongly finding at least one ‘significant’

effect in the case of no actual effects would become 41.0%. For a significance level of 5% this percentage drops to 22.6%. The corresponding percentage for the lowest significance level we consider (1%) is 4.9%. However, considering the domain of our research, we also do not want to become too conservative and adopt this low significance level of 1%. Making a trade-off between all factors involved, we decide on a significance level of 5%.

We decide on a power of 80%. This percentage is commonly used and we do not see a reason to deviate.

To estimate the expected variance, we have a look at observed variances in previous researches. We find \$10 dictator games in Forsythe et al (1994), Bolton et al (1998) and Hoffman et al (1994), all differing somewhat in terms of choice set and degree of anonymity. However, all standard deviations are similar, ranging from \$1.29 to \$1.79. We decide to compute the weighted average of the standard deviations as an estimate. This leads to a standard deviation of \$1.59 so we expect a variance of \$2.53 (actually €2.53 in our case).

We have examined all four factors influencing the sample size we need for this research. We will go on to compute this required sample using all numbers we decided on. In Chow et al (2002) we find a clear representation of the formula needed to compute the sample size that corresponds to a certain significance level, power, effect size and variance. For one-sided tests, this is formula 6.1 shown below. We will proceed to perform two-sided tests as well in this research, but some samples will be used for different tests and this is why we like to select equal sample sizes for every treatment.

Because most of our tests are one-sided, we opt for computing the sample size n per treatment based on one-sided tests:

$$\text{formula 6.1} \quad n = \frac{(Z_{0.05} + Z_{0.20})^2 \sigma^2}{\delta^2}$$

Where: $Z_{0.05}$ is the critical value obtained from the standard normal distribution corresponding to the chosen significance level, $Z_{0.20}$ is the critical value obtained from the standard normal distribution corresponding to the chosen power, σ^2 is the estimated variance and δ is the chosen effect size.

Computing the right sample size is now a matter of filling in the right numbers.

We find a value for n of 31.62. To obtain our desired power and significance we will proceed to collect information on a sample of 32 observations per treatment. Taken into account that we have 6 treatments this will add up to a total of 192 observations in this research.

7 Methods

The method section will be split up in two parts. One part will be about the method for answering the main research question. The second part will be about the motive-analysis on the additional answer sheet given to the 128 participants in treatments **MM**, **FM**, **MF**, and **FF**.

7.1 Hypothesis testing

Possible significant differences between average giving amounts across treatments can be tested with either a one-sided or a two-sided Mann-Whitney U-test. In both cases the test-statistic can be computed as follows, where this statistic will be compared with the critical value corresponding to either a two-sided or a one-sided test. The test-statistic is the smallest value of the following two:

$$\text{formula 7.1} \quad U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$\text{formula 7.2} \quad U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

In the above formulas, U_x is the test statistic, n_1 is the size of sample 1 (arbitrarily chosen), n_2 is the size of sample 2, and R_x is the sum of ranks in sample x .

7.2 Motive analysis

As explained in section 5, we were only able to obtain information on motive importances for the non-anonymous participants (treatments **MM**, **MF**, **FM** and **FF**).

For the participants in these treatments that gave some positive amount of money, we will compare the overall importance of *every* motive. We will compare the importance of a motive to all other average stated importances. Because we do not want to obtain a general result like ‘there are differences in importances’, but a result on the one on one relation of every possible combination of motives, we opt to perform 15

different Mann-Whitney U-tests (as described in section 7.1). However, to control for the amount of tests, we will consider a difference significant only when the p-value is lower than $p = \left(1 - 0.95^{\frac{1}{15}}\right) = 3.4 * 10^{-3}$ to make sure that the overall probability of wrongly concluding a significant difference remains 5%.

For the participants that gave some positive amount of money, we will also compare possible differences in the stated importances of motives per treatment. As we want to obtain information on every combination of motives again, we choose to perform 36 different comparisons, this time considering a difference significant only when the p-value of a t-test is lower than $p = \left(1 - 0.95^{\frac{1}{36}}\right) = 1.4 * 10^{-3}$

For the participants that decided to keep all the money, we were (as explained in section 5) only able to distinguish between the importance of altruistic motives versus non-altruistic motives. We decide to compare results with a single Mann-Whitney U-test on a 5% significance level.

8 Results

The mean amounts given in the six different treatments were as follows:

- Treatment 1 (**MA**): €2.03 (Males, anonymous)
- Treatment 2 (**FA**): €2.78 (Females, anonymous)
- Treatment 3 (**MM**): €1.84 (Males, Male experimenter)
- Treatment 4 (**FM**): €2.53 (Females, Male experimenter)
- Treatment 5 (**MF**): €2.03 (Males, Female experimenter)
- Treatment 6 (**FF**): €3.89 (Females, Female experimenter)

We can see that average values for the male participants range from €1.84 to €2.03 and so it seems like the male participants act similarly in all treatments. For female participants the averages range from €2.53 to €3.89 and so differences between treatments are bigger. The average of €3.89 in treatment **FF** is mainly responsible for this. We will check whether differences between treatments are significant. Also, average values for female participants seem to be higher than for male participants. We will start with investigating these differences by testing hypothesis 1.

8.1 Hypothesis testing

We will look at the results for our 9 hypotheses, starting with the anonymous treatments.

- Hypothesis 1 : The women in treatment **FA** will give significantly more money than the men in treatment **MA**.

In the anonymous treatments we observe women giving on average €2.78 while the average amount for men equals €2.03. However, mainly due to the big standard deviations in this game, the difference is not significant, as $p=0.058$. We do not find enough evidence to be sure enough of a pure gender-difference. However, this does not mean we can simply proceed to attribute possible significant differences across other male- and female-treatments completely to effects other than pure gender-differences.

The p-value is low enough to remain careful. The difference between **FF** and **MF** might be so of such magnitude that apart from a gender-dependent reaction to a female experimenter, we will have to consider a pure gender-difference partly explaining the effect.

- Hypothesis 2 : Men in treatment **MM** will give significantly more money than men in treatment **MA**.
- Hypothesis 3 : Women in treatment **FM** will give significantly more money than women in treatment **FA**.
- Hypothesis 4 : Men in treatment **MF** will give significantly more money than men in treatment **MA**.
- Hypothesis 5 : Women in treatment **FF** will give significantly more money than women in treatment **FA**.

The p-values for hypotheses 2-5 are as follows:

- Hypothesis 2 : $p = 0.610$
- Hypothesis 3 : $p = 0.637$
- Hypothesis 4 : $p = 0.413$
- Hypothesis 5 : $p = 0.037$

This means that analysis of all four treatments with experimenter pressure compared to the corresponding anonymous treatments leads to a significant value only for the **FF** treatment compared to the **FA** treatment (as $p=0.037$). Further analysis will give more insights on the reason for this.

- Hypothesis 6 : The difference between **MM** and **FM** will not be significantly different from the possible difference between **MA** and **FA**.
- Hypothesis 7 : The difference between **MF** and **FF** will not be significantly different from the possible difference between **MA** and **FA**.

Men facing a male experimenter gave an average amount of €1.84 while women gave €2.53. The corresponding p-value equals 0.1797 and this means we have no reason to think that men and women react differently to a male experimenter. However, looking at hypothesis 7, we observe men giving €2.03 and women giving €3.89. The corresponding p-value equals 0.005 and so this difference is very significant. This might still (partly) be due to a pure gender effect of which we have found some indication. Results may also arise because of the different reactions of the genders towards the experimenters. A closer look at other results is needed for a more detailed conclusion of what seems to be the main driver here.

- Hypothesis 8 : We will not find a significant difference between treatments **MM** and **MF**.
- Hypothesis 9 : We will not find a significant difference between treatments **FM** and **FF**.

The p-value for hypothesis 8 is 0.6751. This means that men do not appear to behave differently in the treatments with the male or female experimenter. The p-value for hypothesis 9 equals 0.040. This means that for women the opposite is true. They appear to be influenced by the female experimenter. This also means that the fact that we found a significant difference between treatment **MF** and **FF**, seems mostly due to the fact that women behave differently under the different experimenters, and not

men. However, the p-value of 0.005 is (actually more than) 8 times smaller than the one of 0.040 we find here, and this suggests that the difference between **FM** and **FF** does not solely explain the difference between **MF** and **FF**. We rule out the fact that men also are influenced by the gender of the experimenter, as the p-value here is 0.6751. What seems to remain as a possible other reason for the big difference between outcomes in treatment **MF** and **FF** is a pure gender difference. Due to our chosen power, the probability of wrongly accepting our null hypothesis of equal means in anonymous treatments was 20%. We think it is likely that this happened when we look at our other results. If we were to pool would pool sample **MA+MM** and **FA+FM** (assuming no influence of the male experimenter), we find a p-value for a gender-difference of 0.020. When we add this to the observation that we found a difference between **MF** and **FF** that does not seem to be explained solely by the difference between **FM** and **FF**, it seems likely that pure gender-differences have had a role in the difference between **MF** and **FF** as well.

8.2 Motive results

We have asked all participants that gave a positive amount of money in one of our treatments with social pressure to state how much they agree with six different statements. This means we obtained information on what participants stated to be important while they were deciding under social pressure. Initially 76 people were in the sample, but two participants were removed from the sample because they failed to fill in the second part of the survey. This leaves us with 74 observations. Each of the statements was linked to one of the main motives for giving. The higher the stated number, the higher the importance of this motive. This way we can compare the overall (self-stated) importances of the motives. In the table below (split up in two parts

because of the size) we show the p-values of the significance tests comparing all motives. The average importance is shown between brackets:

	Self-Image (4.92)	Exp. Pressure (4.42)	Warm-Glow (4.87)
Self-Image (4.92)	-	$6.8 * 10^{-2}$	0.82
Exp. Pressure (4.42)	$6.8 * 10^{-2}$	-	$8.8 * 10^{-2}$
Warm-Glow (4.87)	0.82	$8.8 * 10^{-2}$	-
Altruism (3.63)	$2.1 * 10^{-5}$	$2.5 * 10^{-2}$	$6.3 * 10^{-5}$
Fairness (5.74)	$1.4 * 10^{-4}$	$1.0 * 10^{-6}$	$2.2 * 10^{-4}$
Image (2.99)	$1.7 * 10^{-11}$	$1.2 * 10^{-5}$	$1.0 * 10^{-10}$

Table 8.1.1: P-values for comparing overall self-stated average motive (and experimenter pressure) importances of the ‘givers’ with the Mann-Whitney test. Average values between brackets.

	Altruism (3.63)	Fairness (5.74)	Image (2.99)
Self-Image (4.92)	$2.1 * 10^{-5}$	$1.4 * 10^{-4}$	$1.7 * 10^{-11}$
Exp. Pressure (4.42)	$2.5 * 10^{-2}$	$1.0 * 10^{-6}$	$1.2 * 10^{-5}$
Warm-Glow (4.87)	$6.3 * 10^{-5}$	$2.2 * 10^{-4}$	$1.0 * 10^{-10}$
Altruism (3.63)	-	$2.3 * 10^{-11}$	$2.7 * 10^{-2}$
Fairness (5.74)	$2.3 * 10^{-11}$	-	$1.6 * 10^{-16}$
Image (2.99)	$2.7 * 10^{-2}$	$1.6 * 10^{-16}$	-

Table 8.1.2: P-values for comparing overall self-stated average motive (and experimenter pressure) importances of the ‘givers’ with the Mann-Whitney test. Average values between brackets.

Because we performed 15 different significance tests we only consider a difference significant when the corresponding p-value is lower than $3.4 * 10^{-3}$, as explained earlier. We see that the average stated importance of fairness is significantly higher than that of any other motive. This is very much in line with the fact that we found that 48 out of the 74 people that gave some amount of money gave €5, which we consider the most fair division. After this clearly most important (self-stated) motive, we find self-image and warm-glow to be not significantly different from each other (and

experimenter pressure), but both significantly different from altruism and image. As altruism did not significantly differ from image, we end up with the following importance ranking, with the notion that all differences are significant except for the difference between experimenter pressure and altruism:

1. Fairness
2. Self-Image, warm-glow, experimenter pressure
3. Altruism, Image

We also broke down the importance ratings per treatment, to look for possible differences here. Results are shown in table 8.2:

	Sample Size	Self-Image	Exp. Pressure	Warm-Glow	Altruism	Fairness	Image
MM	14	4.86	3.93	4.86	4.50	5.79	2.79
MF	16	5.19	5.00	4.25	3.81	5.88	3.06
FM	18	4.94	3.83	5.00	3.05	5.55	3.11
FF	26	4.77	4.46	5.19	3.46	5.77	2.96
Weighted average	74	4.92	4.42	4.87	3.63	5.74	2.99

Table 8.2: Average motive and experimenter pressure importance of the 74 'givers' per treatment

We do not find a significant difference when we compare the column-values per treatment. There is no average value for a motive significantly different from the value for this motive in another treatment. The biggest absolute difference (**MM**-altruism versus **FM**-altruism) leads us to a p-value of 0.054. Considering we performed 36 tests here, this is far from the value of $1.4 * 10^{-3}$ that would be of meaning for us here.

When we look at results for people that kept all the money for themselves (52 participants), we find them stating that they would be more likely to start giving when

an altruistic motive would arise, compared to a non-altruistic motive. As explained earlier, for this group we could only formulate the following 2 statements to distinguish between the importance of altruistic and non-altruistic motives:

(1) I would be more likely to give if more people would be watching

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

(2) I would be more likely to give if I would know more about the other person.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

Where agreeing with (1) would indicate a non-altruistic reason, and (2) would indicate an altruistic reason.

Average values were as follows, where the difference is very significant, as $p = 4.4 * 10^{-8}$.

	Average
Altruistic reason (broad definition)	5.96
Non-altruistic reason	4.04

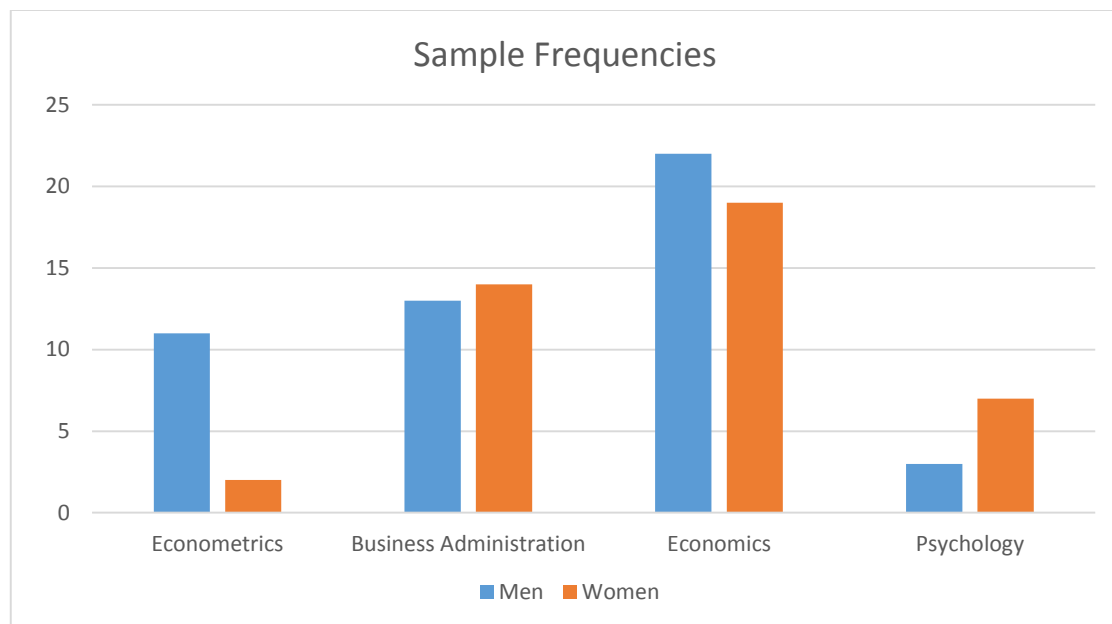
Table 8.3: Overall self-stated average motive importance of the 52 ‘keepers’

All in all we do not find a single difference in stated motive importance across treatments, but we are able to obtain a pretty clear view of the overall importance of every motive in our social pressure treatments.

8.3 The randomness of the sample

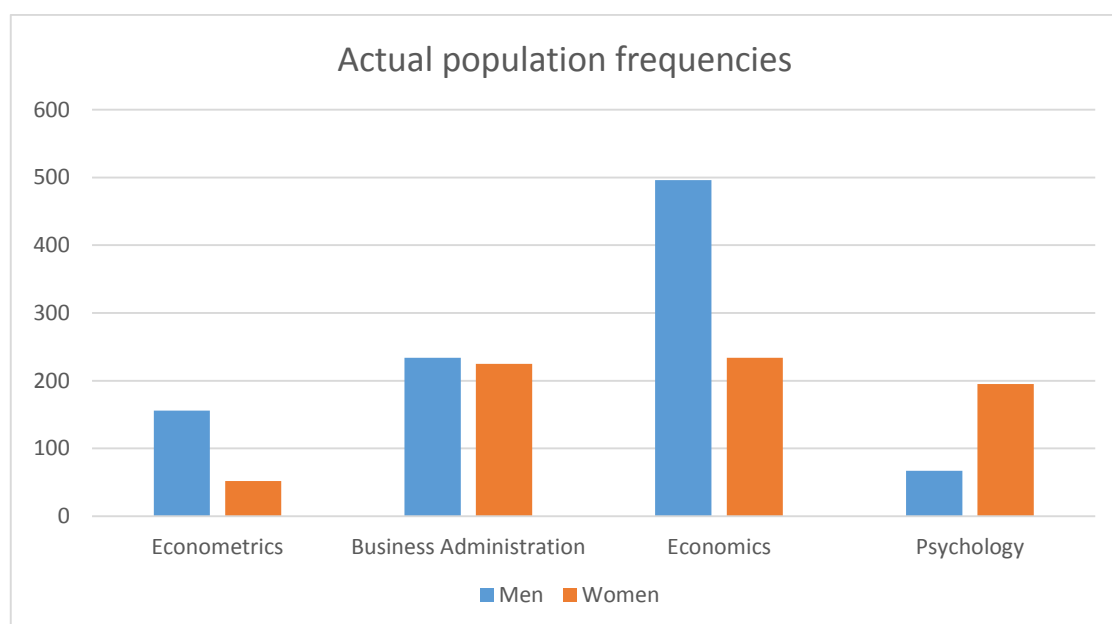
Although it would have been best to obtain an exact representation of the population, it can be the case that for some reason (time or place) by accident we selected a large proportion of a certain group of people, and this could be a driver of results. We asked

every participant for his or her field of study to check for this. In graph 8.1 below, we start by having a look at studies with a frequency of at least 10 in our sample:



Graph 8.1: Sample frequencies of studies with total frequency of at least 10

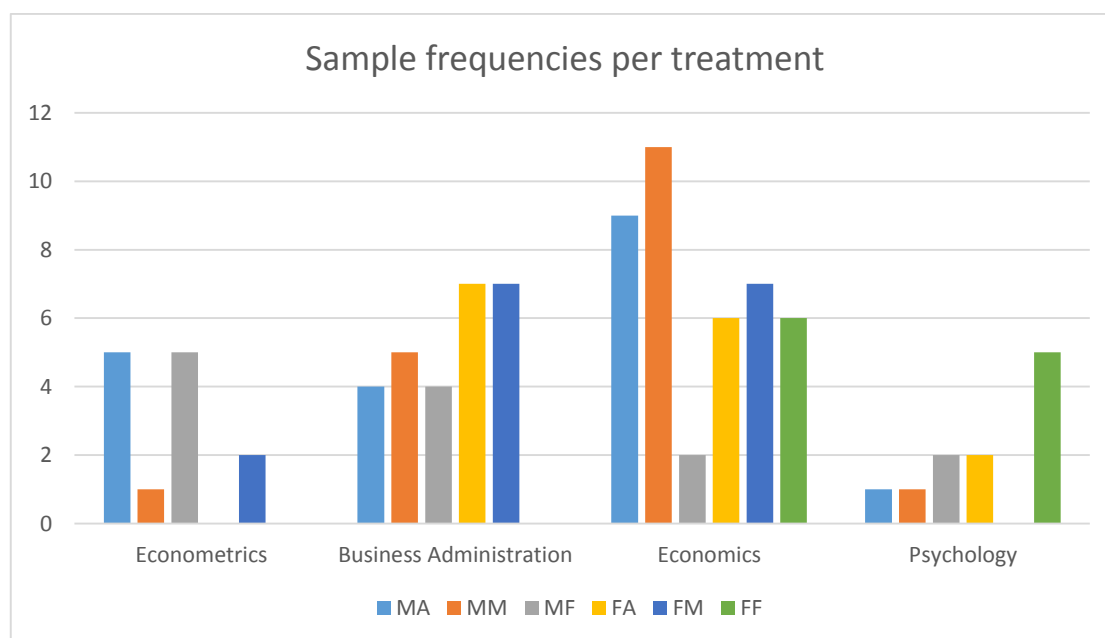
To decide on whether these frequencies reasonably represent the actual population, we have to compare them with the actual frequencies. For this we find registration numbers at <http://universitairebachelors.nl/studievergelijker/> from 2010 at the Erasmus University for each of these studies. Results are shown in graph 8.2:



Graph 8.2: Actual registration frequencies at the Erasmus University of Rotterdam as of 2010

Relative numbers of frequencies seem to fit very well, and none of this raises our concern. The only thing to point out seems to be the fact that within the economists we have asked a relatively high number of females (46.3%) compared to their presence in the actual population (32%). However, closer examination reveals that these women gave on average €2.63 (and the men €1.68). These numbers do not seem extreme enough to make overall results different from what we would expect from the general population. We do not consider the relatively high number of female economists a problem.

Although the total frequencies look fine, it can still be the case that somehow, in one of the treatments, a relatively large proportion of participants from a certain field of study is present. This group could potentially influence results, if they give or keep significantly more than the average person. To check for high frequencies per treatment we break down the total frequencies to the frequencies per treatment in graph 8.3 below:



Graph 8.3: Sample frequencies per treatment of studies with overall frequency of at least 10

We see a big proportion of economists in treatment **MA** and **MM**. We see that these groups gave an average of respectively €2.2 and €1.81. This is close to the respective

treatment-averages of €2.03 and €1.84. This means that it is not likely that these relatively big proportions of economists changed the results in a way that is not in line with the general population. Other frequencies are 7 at most and we do not consider this a potential problem.

All in all we did not find anything that can be a big concern for us. Overall frequencies seem to represent actual frequencies. More detailed analysis on treatment level shows that the relatively large proportions of economists in treatment **MA** and **MM** do not seem to be a problem. We have no reason to think that our sample does not accurately represent the actual population.

9 Conclusions

The most important conclusion is that the gender of the experimenter can have a significant influence on results. While behavior of men did not significantly differ across treatments, women gave significantly more when the experimenter was a woman compared to a man. Together with a pure gender effect of women in general giving more money (although we did not initially find such an effect, only after further investigation we concluded that such an effect was likely present), this led to the big significant difference between men and women with a female experimenter. We conclude that social experimenter demand effects differ for male and female experimenters, although we have not found literature ruling out the fact that participants can be influenced by experimenter-gender in a different way.

We also find fairness to clearly be the (self-stated) most important motive for giving in the pressure treatments. All other motives were significantly less important. Due to the way our own experiment was constructed, it was difficult for our research to add information on *how* social pressure influenced the importances of motives. However, combining our own findings with respectively the suggestions and findings of Hoffman et al (1994) and Andreoni and Bernheim (2009) we do proceed to conclude that social pressure also likely has increased the desire to behave fairly. When social pressure is present in the dictator game, people will act more fairly to improve their image. The fact that fairness is important is also in line with the fact that 48 out of 74 participants that gave money divided the total amount equally.

We also have an indication that altruism (in our case best described as caring about the other person) does not belong to the most important motives for giving. Participants stated fairness, self-image and warm-glow to be significantly more

important. This is in line with existing literature doubting the importance of altruism in seemingly altruistic acts.

When it comes to interpreting existing literature, we can say this was sometimes a little less straightforward than we might have expected. Significant part of this research turned out to be examining the complex definitions of and relationships between motives for giving. The main motives for giving in dictator games appeared to be very hard to examine in isolation. In addition, definitions did not always seem to be clear. The fact that answering the question ‘what is altruism?’ is a hard one, and the fact that different definitions are possible, suggests that some explanation on this might be useful at the start of a paper on altruism in dictator games.

Even when definitions are clear, it seems to be very hard to examine some of the motives in isolation, and this is not always made clear. Warm-glow effects can be separated from social image and social pressure motives, but not straightforwardly from the self-image motive. This might deserve more attention, as sometimes effects are assigned to warm-glow without mentioning this. Comparing conclusions between different researches then becomes dangerous territory, as warm-glow giving can be considered a form of (impure) altruism, and self-image concerns have of course very little to do with altruism.

In general, we think cautiousness (but clearness) is important. In addition, at the start of a research, some explanation on where to draw the lines when it comes to the motive of interest would be useful.

10 Discussion and limitations

The fact that the gender of the experimenter had a significant role in our research, suggests that one should be careful when comparing dictator game studies with differing experimenter-genders. In addition, this might raise attention to the fact that it is very well possible that also in other domains the gender of the experimenter could influence results, although this research provides no direct evidence for this.

When it comes to conclusions about the different motives for giving, we have to keep in mind that these are based on self-stated importance. This adds to the existing literature as we did not find such a way (asking directly) of measuring importance in previous researches. Although the fact that these importances are self-stated means they should be considered with care, results have often been highly significant.

However, as every study, this study has had its limitations, and we will list the most important ones. First of all, we were not able to obtain all 192 observations on one day. This means we cannot rule out the fact that for example weather conditions had an impact on results.

Also, in picking a male and a female experimenter, one cannot control for every other possible difference. It would have been best to pick 64 different men and women as experimenters, all to obtain one observation. This way all other differences would hopefully average out.

We also cannot be certain that (for women) giving increases when the experimenter is a woman because of the fact that they feel more pressure. Although we have not found evidence for a reason like this, one might think of the reason that people visualize the experimenter as the person they are giving money to. If they perceive the experimenter as being very nice and kind, then the reason might be a lot different than giving because of social pressure.

What remains is the fact that we found a significant difference in the dictator game between behavior of men and women with a female experimenter. This was due to the fact that women significantly increased giving in the treatment with the female experimenter compared to the male experimenter, in addition to a pure gender effect.

11 Future research

Future research might explore the exact reason behind the difference in female behavior under the different experimenters. Do participants feel more pressure when the experimenter is a female or can it be something else? Also, research might be extended to different domains to check whether the gender of the experimenter influences results elsewhere.

One could also try to visualize the complex decision-making framework in more detail by constructing treatments with the sole purpose of obtaining information on the exact path social pressure takes in influencing outcomes.

If results of this research were to be replicated, we would advise to obtain observations all on the same day, and vary between different male and female experimenters to average out other differences than gender.

12 References

- Andreoni, J. (1989). Giving with impure altruism: applications to charity and Ricardian equivalence. *The Journal of Political Economy*, 1447-1458.
- Andreoni, J. (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *The economic journal*, 464-477.
- Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics*, 293-312.
- 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., & Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5), 1607-1636.
- Arrow, K. J. (1972). Gifts and exchanges. *Philosophy & Public Affairs*, 343-362.
- Bardsley, N. (2008). Dictator game giving: altruism or artefact?. *Experimental Economics*, 11(2), 122-133.
- Batson, C. D. (1991). The altruism question: Toward a social-psychological answer. *Hillsdale, NJ: Erlbaum Associates*.
- Bellemare, C., Lepage, P., & Shearer, B. (2010). Peer pressure, incentives, and gender: An experimental analysis of motivation in the workplace. *Labour Economics*, 17(1), 276-283.
- Bénabou, R., & Tirole, J. (2005). Incentives and prosocial behavior. *National Bureau of Economic Research*, No. w11535.
- Ben-Ner, A., Kramer, A., & Levy, O. (2008). Economic and hypothetical dictator game experiments: Incentive effects at the individual level. *The Journal of Socio-Economics*, 37(5), 1775-1784.
- 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.
- Bicchieri, C. (2005). A taste for fairness. *Networks*, 5, 69-101.
- Brañas-Garza, P. (2007). Promoting helping behavior with framing in dictator games. *Journal of Economic Psychology*, 28(4), 477-486.

Brown, B. B. (1982). The extent and effects of peer pressure among high school students: A retrospective analysis. *Journal of Youth and Adolescence*, 11(2), 121-133.

Chow, S. C., Shao, J., & Wang, H. (2002). A note on sample size calculation for mean comparisons based on noncentral t-statistics. *Journal of biopharmaceutical statistics*, 12(4), 441-456.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. (rev. ed.). Lawrence Erlbaum Associates, Inc.

Crumpler, H., & Grossman, P. J. (2008). An experimental test of warm glow giving. *Journal of Public Economics*, 92(5), 1011-1021.

Eckel, C. C., & Grossman, P. J. (1996). Altruism in anonymous dictator games. *Games and economic behavior*, 16(2), 181-191.

Eckel, C. C., & Grossman, P. J. (1998). Are women less selfish than men?: Evidence from dictator experiments. *The economic journal*, 108(448), 726-735.

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

Gong, B., Yan, H. and Yang, C. (2010). Gender Differences in the Dictator Experiment: Evidence from the Matrilineal Mosuo and the Patriarchal Yi. *Journal of Economic Behavior & Organization*, 83, 59–65.

Grossman, Z., & van der Weele, J. (2013). Self-image and strategic ignorance in moral dilemmas. *Working paper, University of California at Santa Barbara*.

Hinvest, N. S., & Anderson, I. M. (2010). The effects of real versus hypothetical reward on delay and probability discounting. *The Quarterly journal of experimental psychology*, 63(6), 1072-1084.

Hoffman, E., McCabe, K., Shachat, K., & Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3), 346-380.

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

Kerr, B., Godfrey-Smith, P., & Feldman, M. W. (2004). What is altruism?. *Trends in ecology & evolution*, 19(3), 135-140.

Koch, A. K., & Normann, H. T. (2008). Giving in dictator games: Regard for others or regard by others?. *Southern Economic Journal*, 223-231.

Korenok, O., Millner, E. L., & Razzolini, L. (2009). Altruism and Warm Glow in Dictators' Giving. *Working paper, Department of Economics, Virginia Commonwealth University*.

Lenth, R. V. (2001). Some practical guidelines for effective sample size determination. *The American Statistician*, 55(3), 187-193.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, 1(3), 593-622.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3), 482-493.

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards. *Judgment and decision making*, 6(6), 552.

Luccasen, A., & Grossman, P. J. (2013). A robust test of warm glow giving and spiteful pleasure in a real donation experiment with and without earned endowments. *Monash Economics Working Papers 29-13, Monash University, Department of Economics*.

Margolis, H. (1984). Selfishness, altruism, and rationality. *University of Chicago Press*.

Murnighan, J. K., Oesch, J. M., & Pillutla, M. (2001). Player types and self-impression management in dictatorship games: Two experiments. *Games and Economic Behavior*, 37(2), 388-414.

Oechssler, J. (2010). Searching beyond the lamppost: Let's focus on economically relevant questions. *Journal of economic behavior & organization*, 73(1), 65-67.

Ploner, M., & Regner, T. (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization*, 93, 374-383.

Rosenblat, T. S. (2008). The beauty premium: Physical attractiveness and gender in dictator games. *Negotiation Journal*, 24(4), 465-481.

Saad, G., & Gill, T. (2001). The effects of a recipient's gender in a modified dictator game. *Applied Economics Letters*, 8(7), 463-466.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs*, 317-344.

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17*(3), 222.

Taylor, E. D. & Wong, C. A. (1996). Gender differences in the impact of peer influences and peer orientation on African American Adolescent's school value and academic achievement. *Paper presented at the Sixth Biennial Meeting of the Society for Research on Adolescence, Boston, Massachusetts.*

Tonin, M., & Vlassopoulos, M. (2011). An experimental investigation of intrinsic motivations for giving. *Theory and Decision, 76*(1), 47-67.

Tonin, M., & Vlassopoulos, M. (2013). Experimental evidence of self-image concerns as motivation for giving. *Journal of Economic Behavior & Organization, 90*, 19-27.

Studievergelijker universitaire bachelors. (2010). Vergelijking alle bachelors. *Information Retrieved June 2, 2014, from* <http://universitairebachelors.nl/studievergelijker/>

Wilson, D. S. (1992). On the relationship between evolutionary and psychological definitions of altruism and selfishness. *Biology and Philosophy, 7*(1), 61-68.

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

Zizzo, D. J. (2011). Do dictator games measure altruism?. Handbook on the economics of philanthropy, reciprocity and social enterprise, *Luigino Bruni and Stefano Zamagni, eds., Edward Elgar, Forthcoming.*

Zizzo, D. J., & Fleming, P. (2011). Can experimental measures of sensitivity to social pressure predict public good contribution?. *Economics Letters, 111*(3), 239-242.

Appendix A

The following 3 pages respectively show the answer sheet that was given to participants in the anonymous treatments (**MA** and **FA**), the answer sheet that was given to participants in the social pressure treatments (treatments **MM**, **MF**, **FM** and **FF**) and the additional answer sheet that was given to participants in the social pressure treatments (treatments **MM**, **MF**, **FM** and **FF**).

Economics Experiment

Male / Female

Age:

Field of study:

Imagine yourself in a situation where you have been given €10. In this situation you have been paired with another person. This person is a student but you don't know, and will not find out, who. You can keep money or give money to this person, all or any portion of €10. You will have no interaction with this person in the future.

You may give money only in increments of €1. You may give away an amount ranging from €0 to €10. The decision of how much to give is entirely yours.

Please carefully decide on the amount corresponding to what you would do in this situation. After answering, please proceed to put this sheet in the provided envelope, and close the envelope. All of your answers are entirely anonymous and the researchers have no way of linking them to you or to anybody else in this experiment.

Your decision is to give €___ to this person and to keep €___ to yourself, for a total of €10.

Economics Experiment

Male / Female

Age:

Field of study:

Imagine yourself in a situation where you have been given €10. In this situation you have been paired with another person. This person is a student but you don't know, and will not find out, who. You can keep money or give money to this person, all or any portion of €10. You will have no interaction with this person in the future.

You may give money only in increments of €1. You may give away an amount ranging from €0 to €10. The decision of how much to give is entirely yours.

Please carefully decide on the amount corresponding to what you would do in this situation. Remember, after you have answered, the experimenter will ask you some questions about your decision.

Your decision is to give €___ to this person and to keep €___ to yourself, for a total of €10.

Economics Experiment

After answering, please proceed to put this sheet with the previous answer sheet in the provided envelope without showing the experimenter, and close the envelope. You will receive no additional questions.

If you decided to give €0 and keep €10:

Please indicate how much you agree with the following statements:

- I would be more likely to give if more people would be watching.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

- I would be more likely to give if I would know more about the other person.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

If you decided to give more than €0 and keep less than €10:

Please indicate how much you agree with the following statements:

- I gave money because it makes me feel better about myself.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

- I gave money because I felt I was supposed to do so.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

- I gave money because I enjoy giving.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

- I gave money because I care about the other person.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

- I gave money because I think this is fair.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

- I gave money because I think it would improve my image.

Strongly Disagree -----1-----2-----3-----4-----5-----6-----7----- Strongly Agree

Thank you for your time!
