STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
659 Degree project in economics
Spring 2016

CROWDING-OUT IN VOLUNTARY WORK:
A randomized experiment on the effect of monetary incentives on the motivation of altruistic behaviour

Julia Lundholm (23125)

Abstract

The current refugee crisis Europe is facing has proven how important volunteers are, raising the question: what motivates voluntary work? This study investigates whether the theories on the crowding-out effect of monetary rewards found in the context of blood donation, mainly those presented by Titmuss and Bénabou and Tirole, can be generalized for altruistic behaviour and applied on voluntary work with refugees. A randomized experiment using an Internet survey was conducted on students to study the effect of monetary incentives on the supply of volunteering tutors for refugee children. The study also tested whether there are any differences in volunteering behaviour between the genders as previous studies on gender differences in prosocial behaviour have produced varying results. The experiment found no support for the crowding-out effect, not even when separating the sample by gender, although women were found to be more inclined to help in general. These results indicate that altruism is context dependent, resulting in a risk when generalizing voluntary work and donation behaviour as altruistic behaviour. Instead, they should be considered as two separate acts driven by different degrees of intrinsic, extrinsic and reputational motivation and incentivized accordingly. However, due to limitations of the experimental design, more studies are needed before any general conclusions for voluntary work can be drawn.

Keywords: altruism, crowding-out, voluntary work, randomized experiment, gender differences in prosocial behaviour
JEL-Classifications: D03, D64, C90, J31

Supervisor: Federica Romei
Date submitted: 31 July 2016
Date examined: 1 August 2016
Examiner: Örjan Sjöberg

Acknowledgements

The author would first and foremost like to express gratitude to the supervisor Federica Romei for her guidance and constructive critique. The author would also like to thank Anna Dreber for her dedication and for believing in the study. The valuable input from Chloé le Coq, Eskil Forsell, Tore Ellingsen and Erik Lindquist is also highly appreciated. Finally, a big thank you to everyone who participated in the experiment, to Pimp My Grades and Studiefrämjandet whose interest in the study made the experiment possible and to Sara Myrenfors for her assistance with the collection of data.

Contents

1 Introduction 4

2 Theory and previous research 6

2.1 Altruism and prosocial behaviour 6

2.2 The crowding-out effect 7

2.3 Crowding-out in voluntary work and gender differences 8

3 The experiment 10

3.1 Experimental design 10

3.2 Implementation 12

3.3 Activity 13

3.4 Compensation 14

3.5 Participants 15

4 Econometric model 16

4.1 The crowding-out hypothesis 18

4.2 The charity hypothesis 18

4.3 The gender differences hypothesis 20

5 Data 21

6 Results 23

6.1 Descriptive statistics 23

6.2 Analysis of the crowding-out hypothesis 25

6.3 Analysis of the charity hypothesis 25

6.4 Analysis of the gender differences hypothesis 26

7 Discussion 28

7.1 Discussion on the crowding-out effect in voluntary work 29

7.2 Limitations of the experiment 31

8 Conclusions 34

References 36

Appendix I 38

Appendix II 44

Appendix III 45

1 Introduction

Volunteering is defined by the United Nation's General Assembly as "activities undertaken of free will, for the general public good and where monetary reward is not the principal motivating factor" (Wallace et al., 2015). In 2015, Europe witnessed the greatest refugee crisis since World War II and as the number of newcomers continues to grow in Sweden, the need for volunteers has increased (Solna Stad, 2016). With the increasing demand for voluntary workers, the key is to identify the drive forces behind volunteering. However, in order to understand the behaviour of volunteers, it is critical to understand what many argue constitutes the driving force behind volunteerism: altruism (Unger, 1991).

Altruism – helping others at your own cost – has gained the interest of many scientists as it by definition contradicts the fundamental assumption of Homo Economicus driven by self-interested utility maximization (Henrich et al., 2001). Many scientists have therefore tried to explain the motivation of altruistic behaviour. However, in the case of voluntary work, a more interesting approach is how volunteers are affected by monetary rewards in order to explain the behaviour of unpaid labour supply since they contradict the relative price effect that higher prices increase the supply (Frey and Jegen, 2001). A central concept within research on the effect of monetary incentives on altruistic behaviour is the crowding-out effect. The crowding-out effect argues that monetary compensation crowd out the motivation of prosocial and reduce the net supply of the act. There are two main approaches: *the crowding-out theory* that argues that the introduction of monetary incentives undermines the sense of civic duty and reduce the altruistic motivation (Titmuss, 1970), and *the signalling model of crowding-out* that in addition to altruism also accounts for concerns of self-respect or social reputation and argues that the greedy signal of monetary compensation conflicts the altruistic signal of the prosocial activities (Bénabou and Tirole, 2006).

There are several studies that have empirically tested for the crowding-out effect in different contexts of prosocial behaviour (Frey and Oberholzer-Gee, 1996; Gneezy and Rustichini, 2000; Fiorillo, 2011) but only one has experimentally tested both the crowding-out theory and the signalling model of crowding-out (Mellström and Johannesson, 2008). This experiment was however conducted in the context of blood donation and the study only found significant support for the crowding-out effect for women (Mellström and Johannesson, 2008).

There is no experiment that has empirically tested the two theories on crowding-out and in the context of voluntary work for refugee help. The question of interest is therefore whether the crowding-out theories that found support in the blood donation experiment can be generalized for altruism and be applied for voluntary work and refugee help. Even if blood donation and voluntary work both are prosocial behaviour that are argued to be motivated by intrinsic, extrinsic and reputational motivation, the degree of the types of motivation depends on the context and individual preferences (Bénabou and Tirole, 2006); blood donation is mainly intrinsically motivated (Andreoni, 1989) while voluntary work is driven by both intrinsic and extrinsic motivation (Fiorillo, 2011). The reputational motivation does however not depend on the altruistic act itself; it is rather dependent on the signalling possibilities that the context facilitates (Bénabou and Tirole, 2006). As long as the experimental design and conduct for blood donation and voluntary work are similar, differences between the two studies can be minimized to the change of activity.

Based on the blood experiment by Mellström and Johannesson, a randomized experiment will be conducted on university students to test how monetary compensation and the opportunity to signal altruism by donating the compensation affect their motivation to volunteer for refugee help. This experiment will also test for gender differences in volunteering behaviour as previous studies have identified gender differences in prosocial behaviour (see Mellström and Johannesson, 2008; Croson and Gneezy, 2009; Ellingsen et al., 2013; Boschini et al., 2012). If there exist any gender differences, it would be relevant to take into consideration when developing measures using incentives to influence volunteer's behaviour. By investigating the impact of monetary compensation on the motivation of volunteers helping refugees, the aim of this study is to find a better understanding of unpaid labour supply and how volunteers can best be incentivized.

2 Theory and previous research

2.1 Altruism and prosocial behaviour

The term altruism is defined as a social behaviour that benefits the recipient at the cost of the actor (West et al., 2006). This prosocial behaviour is commonly expressed as people donate blood, help strangers, give to charitable organisation or volunteer. Many of these behaviours cannot purely be explained by other-regarding behaviour. As a solution to this issue, Bénabou and Tirole developed a theory that prosocial behaviour includes both altruism and greed with concerns for one's self-respect or social reputation. Prosocial behaviour then reflect a mix of intrinsic, extrinsic and reputational motivation, which all depend on the context of the actor and the behaviour.

The term intrinsic and extrinsic motivations reflect altruism and originate from the field of psychology and as expressed through behaviours that are rewarded by internal or external rewards. An activity is intrinsically motivated if performing the act is rewarding in itself. The act is then internally rewarded as the reward comes from within. One common example of an internal reward is the so-called warm-glow effect (Fiorillo, 2011) defined as a 'feel-good feeling' that people receive when they feel that they 'have done their bit' or helped others (Andreoni, 1989). External rewards are received as by-products from the performed activity (Fiorillo, 2011). Monetary compensation is a commonly used external reward (Gneezy and Rustichini, 2000) but investments in human capital or social networks are also some examples (Fiorillo, 2011). The reputational motivation is included to reflect the greed of people who only perform altruistic acts because they want to be perceived as a better person in front of others or to prove to themselves that they are good people (Bénabou and Tirole, 2006).

Recent studies on prosocial behaviour have also identified differences in behaviour between genders but with varying results. Women are for example expected to give more (Aguilar et al., 2009), which they did in a double-blind dictator game[1] (Croson and Gneezy, 2009). The results on differences in prosocial behaviour do however become quite ambiguous when the effect of the experimental design was studied; women were more sensitive to the social

[1] A double-blind dictator game is an all-anonymous experiment in which the participants first are divided into pairs of dictators and recipients and then the dictator receives a certain amount of money that they can share the recipient if they want.

conditions (Croson and Gneezy, 2009) and the social framing of the experiment (Ellingsen et al., 2013) while men were more sensitive to gender priming (Boschini, 2012). Finally, women were also proven to be more sensitive to the crowding-out effect[2] of monetary compensation for blood donations, although tests for gender differences were not part of the original study (Mellström and Johannesson, 2006). Given the varying results, more research on gender differences in prosocial behaviour is different contexts is needed.

2.2 The crowding-out effect

"Arguably, '*the crowding-out effect*' […] is one
of the most important anomalies in economics."
Frey and Jegen, 2001
(Italics per the original.)

When studying the motivation of voluntary workers, a more interesting approach is to investigate the effect of monetary incentives since the supply of unpaid voluntary labour contradicts the principle of relative price, that a higher price yields an increased supply (Frey and Jegen, 2001). The effect of monetary incentives is a controversial topic in the research on prosocial behaviour started by the introduction of the crowding-out effect. There are two main approaches to the crowding-out effect but the general idea is that monetary compensation for prosocial activities will undermine the motivation of altruistic behaviour and reduce the net supply of the act (Titmuss, 1970; Bénabou and Tirole, 2006).

The crowding-out effect was first introduced in the crowding-out theory, which is based on intrinsic and extrinsic motivation. According to the crowding-out theory, external monetary rewards crowd out the intrinsic motivation that drives altruistic behaviour. This was intuitively explained in the context of blood donation that the compensation for the blood donation would undermine the social values and the sense of civic duty, and consequently leads to a net reduction in the supply of blood donation (Titmuss, 1970). The second theory is based on the idea that altruistic behaviour is not purely driven by altruism and that greed with concerns of self-respect or social reputation also motivates people to perform altruistic acts. The signalling model of crowding-out therefore includes intrinsic, extrinsic and reputational motivation. The purpose of including the reputational motivation was to address the signal-extraction problem that for

[2] The crowding-out effect argues that monetary compensation undermines the altruistic motivation of prosocial behaviour and reduces the net supply of the act.

example occurs when the altruistic signal of prosocial activity is in conflict with the greedy signal of the reward (Bénabou and Tirole, 2006).

Several studies have been conducted to identify the crowding-out effect in different contexts. The first study to find empirical evidence of the crowding-out theory was an experiment that tested whether monetary compensation would increase the local support for the construction of a noxious facility (Frey and Oberholzer-Gee, 1996). There are two studies that have focused on the crowding-out effect in the context of voluntary work, but they differ in conduct and results. One was an experiment testing the crowding-out theory on the performance of volunteers who collected donations for charities. The participants received different levels of monetary compensation and the results showed that the introduction significantly reduced the performance but also that that the performance increased with the size of the payment once compensation was given (Gneezy and Rustichini, 2000). The second study collected data from annual surveys performed on active volunteers and tested whether monetary compensation affected the number of hours volunteered. This study did however not find any support for the crowding-out theory (Fiorillo, 2011).

Only one study has experimentally tested both the crowding-out theory and the signalling model of crowding-out in altruistic behaviour. They were tested by randomized experiment that was performed on Swedish university students who either received no compensation, a compensation of SEK 50 or a compensation of SEK 50 with the opportunity to donate the amount to charity for completing a health examination and registering as a blood donor.[3] The results showed support for both theories; the supply of donors fell as the compensation was introduced but it went up again as the charity option was offered. However, these effects were only significant for women after separating the sample by gender (Mellström and Johannesson, 2008). No other study has empirically tested both theories on the crowding-out effect in one experiment or in the context of voluntary work for refugee help.

2.3 Crowding-out in voluntary work and gender differences

Both blood donation and voluntary work are prosocial behaviour driven by intrinsic, extrinsic and reputational motivation. The degrees of each type of motivation do however depend on the context and the personal preferences (Bénabou and Tirole, 2006). Blood donation behaviour is mainly intrinsically motivated by the warm-glow effect (Andreoni, 1989). Voluntary work is

[3] Before a person can become a blood donor in Sweden, one must first complete a health examination that consists of a health declaration and a blood pressure test and a blood test performed by a nurse (Mellström and Johannesson, 2008).

however argued to be both intrinsically motivated by the warm-glow effect and extrinsically motivated by external rewards such as investments in human capital that will increase future earnings or investments in social networking to create contacts that can be useful for future careers (Fiorillo, 2011). Since the reputational motivation is more about the signalling of altruism, rather than the actually being altruistic, the difference in the reputational motivation between blood donation and volunteering will not depend on the activity in itself but rather on the signalling possibilities that the context facilitates (Bénabou and Tirole, 2006). The experiment in this study will therefore be based on the experiment performed by Mellström and Johannesson in order to minimize the risk that the differences in the results will depend on differences in the signalling possibilities and not the characteristics of the different activities. This decision does come with the risk that potential limitations of the blood donation experiment also will apply for the experiment on voluntary work. Every decision of the experimental design is however carefully motivated under the section *3 Methodology*.

Finally, this study will also test for gender differences in voluntary behaviour. Previous studies on gender differences in prosocial behaviour have produced varying results (Aguilar et al., 2009; Croson and Gneezy, 2009; Ellingsen et al., 2013; Boschini et al., 2012) and the experiment conducted by Mellström and Johannesson (2008) showed that women were more sensitive to the effect of monetary incentives in the context of blood donation. If male and female volunteers are motivated differently, identifying these are of importance so that the incentives can be adjusted accordingly.

3 The experiment

In this study, a randomized experiment was performed to test whether the crowding-out theory of Richard Titmuss and the signalling model of crowding-out of Ronald Bénabou and Jean Tirole could be generalized for altruism and applied on voluntary work. This experiment is an extension of the Mellström and Johansson's experiment on blood donation and will be conducted in another context – more specifically volunteering for the current refugee crisis. In addition to their *crowding-out hypothesis* and *charity hypothesis*, this paper introduces *the gender differences hypothesis* that tests whether there are any differences between men and women in regards to motivation of altruistic behaviour and voluntary work.

3.1 Experimental design

The experiment was performed on students of Stockholm School of Economics (SSE). The participants received a link to an Internet survey in which they were asked to sign up for a voluntary activity held at the school. When the link was opened, the online survey program randomly assigned the participants to one of three groups that offered different compensations. The first group was a control group, in which no compensation was offered. The second and the third group were both treatment groups in which a compensation of SEK 50 was offered to the students who signed up. However, the participants in the second treatment group were also offered the opportunity to donate their compensation to the charity UNICEF. Table 3.1 illustrates the structure of the experiment.

Table 3.1: The experimental design

	Control Group	**Treatment Group 1**	**Treatment Group 2**
Randomized	X	X	X
Compensation		X	X
Charity option			X

A randomized experiment was chosen since it is argued to be the most rigorous method to measure casual effects. This argument is based on the two main properties of the method; first it

eliminates bias and secondly, it enables measurement of uncertainty. The first property regards the randomization process which ensures that all pre-existing systematic differences between the groups are eliminated and only pure chance decides how the groups are divided. However, this also means that the process could result in differences between the groups but since they are caused by pure chance, these differences would be random errors and not biases. Consequently, the unbiasedness of the results is a property of the randomization process and not a feature of its application on the given sample per se (Bloom, 2008). For this reason, a robustness check will be performed to validate the randomization process of this experiment. Once validated, one can conclude that the differences in signups between the groups are caused by the treatments and not by any other factors.

Secondly, the process also randomizes all uncertainty of the estimated coefficients for the given sample (their internal validity), meaning that confidence interval or inference tests can be used to account for all this uncertainty. Still, this does not mean that one can account for all uncertainty when generalizing the estimated coefficients beyond the given sample (its external validity). In order to do this, one must both randomly sample subjects from a known population and then randomly assign them to the experimental groups. Unfortunately, this is rarely possible in social experiments (Bloom, 2008) and it was not possible in this study. Due to practical difficulties, access and time restriction, the sampling of participants from the student population of Stockholm School of Economics could not be randomized. Since randomized sampling was not possible, the survey was instead sent out to as many students as possible, representing different classes and genders. Consequently, the external validity was reduced and the results cannot be directly generalized for the chosen population of interest. This must be considered when interpreting the results. In order to minimize the risk of selection bias, the participants received no information regarding the topic of the survey or explicit information that it was a randomized experiment.

When it comes to social experiments however, the generalizability of the results does not only depend on the external validity but also the ecological validity. Ecological validity refers to whether the estimated effect is representative of what really happens in everyday life settings, how realistic the behaviour in the experiment and the results are. It is not only the question of how probable the participants are that is of concern but also how probable the situation is. For higher ecological validity, it should be probable that it would be a representative from the specified population of the experiment who would face the real life situation in a context not too far from the context of the experiment. That means that both participants and context should be held constant when turning from the experiment scenario to reality. The often atypical

population of students and atypical laboratory situations of social experiments, used to improve the internal validity, usually lead to low ecological validity (Brewer & Crano, 2014). An important aspect when designing the experiment was therefore that the context would be as realistic as possible, in order to produce results that would be probable to obtain in real life situations as well. This was central when formulating the survey, deciding upon the activity and partners, and choosing the size of the compensation, which are further discussed in the following sections. Furthermore, since the chosen population is students from Stockholm School of Economics, it cannot be guaranteed that a replication on another subject pool would produce the same results. The results of this study can therefore not be generalized for other populations outside the sampling frame such as the Swedish population, not even at a student level. This is further discussed in section *3.5 Participants*.

3.2 Implementation

The experiment was performed on 301 students between 7 and 20 April 2016 through an Internet survey that was accessed via a link. The participants were only informed that the survey was a part of a bachelor thesis in Economics in order to reduce the selection bias and to avoid eventual speculations that could affect the results.

The design of the survey was the same for the groups in all aspects except for one – the compensation. All participants entering the survey were first welcomed and then instructed to not communicate with other participants before, under or after the survey in order to avoid any circulation of information on the different compensations. When continuing to the next page, the participants were randomized into one of the three groups (either the control group, the first treatment group or the second treatment group). In the next step, the volunteering activity and the organizing partners were introduced, and the participants were asked if they would be interested in signing up for the activity. The question clearly stated that if they were interested and had decided to sign up, they would be contacted again once the practical details were specified, in order to verify they would be available. The purpose of this setup was to make the sign-up not too binding, in order to minimize the risk of participants declining the invitation for practical reasons. However, the drawback of this setup is that some participants might not have taken the commitment seriously and signed up for the activity without the intention to actually participate. This is further discussed in the section *7 Discussion*. In this part, the participants in Treatment Group 1 were also offered a compensation of SEK 50, while the participants in Treatment Group 2 were offered a compensation of SEK 50 with the possibility of donation. Those who were in the control group were offered no compensation. Finally, the last part

consisted of three control questions regarding gender, age and employment. The gender information was mainly collected as data for the gender hypothesis but the information, together with data on age and employment, was also collected for robustness checks performed to validate the randomization process. In the closing slide, the students received gratitude for their participation and they were once again reminded of the importance of discretion. The complete survey can be found in *Appendix I*.

The experiment was partially anonymous in the sense that only those who signed up for the voluntary activity were asked to give their contact information while those who chose to not sign up did not have to leave any information. Even if it has been argued that double-anonymous conditions that eliminate signalling possibilities are necessary to identify people's true altruistic motives (Eckel and Grossman, 1996), the public property is not a problem in this experiment. The public property is rather a requirement to enable the test of the charity hypothesis and the signalling model of crowding-out. Moreover, the contact information was necessary for the payments of the compensations. The people who did not want to sign up were however allowed to stay anonymous. The purpose of this was to reduce the risk that uninterested people would exit the survey before completion and resulting in a missing value, as these missing values would be difficult to interpret in the analysis.

3.3 Activity

The voluntary activity in the survey was a tutoring evening for refugee children held at Stockholm School of Economics. The event was to be two hours long sometimes during the first half of May, before the students entered the exam period. An important aspect of the activity was that it had to be credible in order to avoid hypothetical signups and yet, at the same time, it could not be too demanding that it would trespass the threshold of commitment. Tutoring was chosen as the activity because it is a natural activity for students, requiring no previous experience or preparation. Tutoring is also an activity that could be hosted at the school, making it more accessible and further lowering the demands on the students.

To further increase the credibility of the activity, the organizational partners *Pimp My Grades* and *Studiefrämjandet* were included. Pimp My Grades is a project group within the Student Association of Stockholm School of Economics that hosts weekly tutoring sessions for immigrant children at an elementary school in the Stockholm suburb Hässelby. Pimp My Grades were also included because of its connection to the students at Stockholm School of Economics and the school facilities. With a familiar name, the hope was to convince the students that the activity was real and that the possibility of it taking place at the school was believable. The

partner Studiefrämjandet is an external educational association advocating children's right for education. Their role was to act as the connection to the refugee children, to cooperate with Pimp My Grades in organizing the tutoring evening and to make the event appear more seriously.

When the experiment was conducted, the tutoring evening was in the planning process. Both organizations had been contacted and they both had the ambition to make the event happen. Unfortunately, the partners were in the end not able to organize the event due to practical reasons and time restrictions. However, this was concluded after the survey was closed and it did not affect the credibility of the experiment.

3.4 Compensation

Monetary incentives are argued to crowd out the motivation of altruistic behaviour (Titmuss, 1970; Bénabou and Tirole, 2006). Later studies have however shown that the effect of the monetary rewards depends on its size. Gneezy and Rustichini found that monetary compensation crowds out the intrinsic motivation when uncompensated and compensated behaviour are compared. However, given that monetary rewards were offered, the performance of the participants increased (Gneezy and Rustichini, 2000).

The choice of the size of SEK 50 was based on two arguments. The first was that Mellström and Johannesson (2008) used SEK 50 and in order to be able compare the two experiments – the original one based on blood donation and the extension based on voluntary work – the same amount was preferred. The second argument concerned the participants' perception of the size, relating to Gneezy and Rustichini's relative effects (2000b). The aim was to find a sum that would not be too small that that it would become dismissible but at the same time stay at a symbolic level so that would not be compared to and be competing against an hourly wage of a hired tutor. As a result, the amount of SEK 50 was decided upon.

Finally, the details regarding the charity option were also chosen based on several considerations. First of all, UNICEF was chosen because it is a well-known organization in Sweden and an active actor in the current refugee crisis. The purpose of this is to avoid biases caused by participants declining the charity option due to a distaste or distrust towards the organization. Furthermore, by making the charity donation an option for the participants, the choice to decline the compensation and donate the amount becomes an active decision for the participants. The donation is then a good deed of the participant and not the conductor of the experiment making the signalling motivation stronger.[4]

[4] Only one person who signed up from the second treatment group decided to not donate the amount.

3.5 Participants

The experiment was conducted on students from Stockholm School of Economics (SSE). The final sample consisted of 301 participants from different programmes at SSE. The choice of students was partly due to practical reasons (with regards to resource constraints and time limits), but also a deliberate decision since the student population provides a homogenous sample. Only using students, and only students from a school with quite similar demography, improved the internal validity. With the randomization process and the homogenous population, the effect of endogenous variables was minimized. In other words, the possibility of the treatments being the only differences between the groups was improved.

With a low representation of the Swedish student population, the results of the study cannot be generalized to the whole student population or the general Swedish population, as discussed previously. For example, as students generally have lower income compared to the general population, students could be more responsive to monetary compensation, which in that case would bias the results against the crowding-out effect. Even if this is a limitation of the study, it is not considered as an obstacle since students are a common subject pool in social sciences and experiments due to its practical benefits. Moreover, the experiment of Mellström and Johannesson was also conducted on students. As long as one is aware of this limitation when interpreting the results, the chosen sample is considered as reasonable.

4 Econometric approach

To investigate how monetary incentives affect the motivation of altruistic behaviour in the context of voluntary work and whether the effect differs on gender three hypotheses were tested:

1. The interest in signing up for the voluntary activity is higher when compensation is not offered (the control group) compared to when SEK 50 is offered as compensation (the first treatment group).

2. Given that a compensation of SEK 50 is offered, the interest in signing up for the voluntary activity is higher when the possibility of donating the amount to a charity is available (the second treatment group) compared to when the charity option is not available (the first treatment group).

3. The interest in signing up for the voluntary activity is higher for women compared to men both in general and when comparing the effect of the compensation of SEK 50 (CG vs TR1) and the effect of the charity option (TR1 vs TR2) between gender.

The two first hypotheses, the crowding-out hypothesis and the charity hypothesis, were tested by Pearson chi-square tests on nonparametric contingency tables. This method was chosen because the data consists of categorical responses and because only two groups were tested at a time. The contingency tables illustrate the different fraction of successful and unsuccessful responses between the groups that were compared. The Pearson chi-square tests then tested whether these differences are statistically significant (Berenson et al., 2012). The fractions tested in this study are denoted as π_{CG}, π_{TG1} and π_{TG2}, for the control group, the first treatment group and the second treatment group, respectively.

The third hypothesis – the gender differences hypothesis – was tested with the Linear Probability Model (LPM). The LPM was used because the data consists of categorical answers to a yes/no question, leading to a binary dependent variable and because more than one dummy variable was tested. The Ordinary Least Square (OLS) method of analysis was used to estimate the coefficients, which are interpreted as how the probability of success is affected when the

independent variable of interest is changed by one unit (Wooldridge, 2013). In the case of this study, success is defined as a yes-answer, meaning that a participant has decided to sign up for the voluntary activity. The LPM was chosen as method of analysis since its results are easy to interpret. However, the weaknesses of the method are that it does not respect the fact that probabilities by definition must stay within 0 and 1, and that the construction of the model causes heteroskedasticity, unless the probability does not depend on the independent variable (Wooldridge, 2013). One can still use the LPM as long as these limitations are controlled for. The first problem can be solved by simply checking the predicted values to make sure that they are within the defined interval of probabilities (Wooldridge, 2013). This was done for this study and no extreme value was found. The second problem regarding the heteroskedasticity can be solved by using robust standard errors (Wooldridge, 2013), which was done in this study. Another method of analysis that is commonly used by social scientists for regressions with binary dependent variables is the logistic regression model. The coefficients estimated by this method are however presented as odds rations and they are more difficult and less intuitive to interpret. The logistic regression model was therefore only used as a robustness check to validate the results. The results from the logistic regression analysis are presented in *Appendix II*.

Before the gender hypothesis was tested, the LPM was also used to validate the results of the Pearson chi-square tests performed on the crowding-out hypothesis and the charity hypothesis. To investigate whether there are any gender differences in volunteering behaviour, a dummy variable for gender and interaction terms between gender and the two treatments (compensation of SEK 50 for the crowding-out hypothesis and the charity option for the charity hypothesis) were stepwise included in the regressions, as demonstrated below:

(1) $help = \beta_0 + \beta_1 TG1 + \beta_2 TG2 + u$

(2) $help = \beta_0 + \delta_0 female + \beta_1 TG1 + \beta_2 TG2 + e$

(3) $help = \beta_0 + \delta_0 female + \beta_1 TG1 + \delta_1 female * TG1 + \beta_2 TG2 + \delta_2 female * TG2 + \varepsilon$

Regression (1) was used to validate the results of the Pearson chi-square tests. In regression (2), the dummy variable female was included to test whether there are any general differences between the genders in inclination to sign up for the voluntary activity. Finally, regression (3) also included interaction terms between gender and the treatments to test whether the effects of the treatments depend on gender.

Since there were no strong priors on the sign of the coefficients, the conservative standard of inferential statistical analysis was followed and two-sided tests were applied. As a property of

the randomization process, the estimated coefficients are uncorrelated with the error terms, meaning that control variables are not necessary.

4.1 The crowding-out hypothesis

According to the crowding-out theory, the intrinsic motivation is reduced when monetary rewards are offered (Titmuss, 1970). In this study, that would mean that the motivation to sign up for the voluntary activity would decrease when the compensation of SEK 50 is introduced. In other words, the supply of volunteers would fall when payment is offered. Consequently, it is predicted that the fraction of signups in the control group will be higher than the fraction of signups in the first treatment group.

Table 4.1: 2 × 2 Contingency table for the crowding-out theory hypothesis

Sign up for the voluntary activity?	**Randomized group**	
	CG	TG1
Yes	π_{CG}	π_{TG1}
No	1 - π_{CG}	1 - π_{TG1}

Based on the contingency table above, the following null hypothesis is tested:

$$H_0: \pi_{CG} = \pi_{TG1}$$
$$H_1: \pi_{CG} \neq \pi_{TG1}$$

4.2 The charity hypothesis

According to the signalling model of crowding-out developed by Bénabou and Tirole (2006), the motivation to perform altruistic acts is based on a mix of intrinsic, extrinsic and signalling motivation. They argue that people who help others not always do it solely to do good but also to signal goodness with the intention to improve their reputation or self-image. As a result, the motivation to help others might be crowded-out if monetary compensation is accepted, since the compensation will make it more difficult to signal altruism. The compensation might instead change the signal to a negative one of greed or selfishness, as helping others in need is something that could be expected to be done regardless of a reward (Bénabou and Tirole, 2006). In the case

of this study, people receiving compensation would choose not to participate, since no signal[5] is better than a negative signal of greed or selfishness. However, if the opportunity to donate the compensation is available, the participants can once again signal goodness by giving away the amount to UNICEF, showing that it is not about the money. The altruistic signalling is rather strengthened as the donation is an additional act of altruism. That would mean that the motivation to sign up for the voluntary activity increases as the charity option is introduced.

In order to test the charity hypothesis, the fraction of signups in the second treatment group is compared to the fraction of signups in the first treatment group. That way, the effect of the charity option treatment (TR2) can be isolated from the compensation treatment (TR1). This is possible because according to the crowding-out theory, both treatments should have the same negative effect on the motivation as they both use external rewards to incentivize intrinsic motivation. It does not matter that the monetary compensation is donated to charity in the second group since the participants in the first treatment group technically also can donate the amount in private after the experiment. What matters is that intrinsic motivation is incentivized by external rewards. This is important for the signalling model of crowding-out. The donation act must be a part of the experiment and in public, together with the altruistic act, so that the participant can send a positive helping signal and not a greedy exploiting signal, which the charity option facilitates.

According to the signalling model of crowding-out, the option to donate the compensation should increase the motivation to sign up for the voluntary activity as the charity option facilitates the signalling of altruism. The fraction of signups is therefore predicted to be higher in the second treatment group compared to the fraction of signups in the first treatment group.

Table 4.2: 2 × 2 Contingency table for the charity hypothesis

Sign up for the voluntary activity?	**Randomized group**	
	TG1	TG2
Yes	π_{TG1}	π_{TG2}
No	$1 - \pi_{TG1}$	$1 - \pi_{TG2}$

Based on the contingency table above, the following null hypothesis is tested:

$$H_0: \pi_{TG1} = \pi_{TG2}$$
$$H_1: \pi_{TG1} \neq \pi_{TG2}$$

[5] There will be no signal for those who choose to not sign up for the voluntary activity as they remain anonymous.

4.3 The gender differences hypothesis

The research on gender differences in prosocial behaviour has produced varying results (Aguilar et al., 2009; Croson and Gneezy, 2009; Boschini et al., 2012; Ellingsen et al., 2013) but most importantly, the blood donation experiment conducted by Mellström and Johannesson (2009) showed that the crowding-out effect depended on gender. By introducing the gender differences hypothesis, this study will also test for any differences in prosocial behaviour between men and women. The hypothesis is tested by including a dummy variable for gender and interaction term between gender and the treatments. Based on the results of Mellström and Johannesson, the gender differences hypothesis predicts that women in general will be more inclined to help in general (the dummy coefficient δ_0 will be positive) and that women will be more sensitive to the treatments (δ_1 will be negative and δ_2 will be positive).

$$help = \beta_0 + \delta_0 female + \beta_1 TG1 + \delta_1 female * TG1 + \beta_2 TG2 + \delta_2 female * TG2 + \varepsilon$$

Based on the regression above, an F-test is performed on the null hypothesis with multiple restrictions to determine whether all three estimated coefficients are jointly insignificant:

$$H_0: \delta_0 = \delta_1 = \delta_2 = 0$$
$$H_1: \delta_0 \neq 0 \text{ and/or } \delta_1 \neq 0 \text{ and/or } \delta_2 \neq 0$$

5 Data

The data was collected through a randomized experiment performed on 301 participants. Only finished surveys were included in the tested sample. There were originally 310 observations in which the main question of whether the respondent would like to sign up for the voluntary activity was answered. However, nine of these surveys had not been completed to the end, as the last control variables regarding gender, age and employment had not been answered. Consequently, these observations were excluded from the sample. Missing values could become a problem if they are missing in a non-random fashion because systematically missing values could change the underlying population or affect the effectiveness of the estimations as the sample size decreases (Wooldridge, 2013). Since the attrition of observations occurred after the treatment question was posed, a potential systematic loss of observations could be a concern in this study, as it indicates causality with the treatment. Systematic attrition bias is however not considered to be an issue in this experiment since the observations are missing in a random fashion and since the excluded proportion is considered as small enough to not have a significant effect on the effectiveness. The losses of observations in each group, expressed as fractions, were as following: 1.020% (1 observation) in the control group, 2.778% (3 observations) in the first treatment group and 4.808% (5 observations) in the second treatment group.

Since a randomized experiment was used, the randomization process of the experiment ensures that there will not be any systematic differences between the groups. However, this does not exclude the possibility that pure chance could cause random differences. In order to validate the randomization process, data from the control questions was used to test for differences in the distributions of the different characteristics. The characteristics that were analysed were gender, age and employment. As the purpose of these variables and tests only was to validate the randomization process, and not to control for any potential omitted variable bias, the selection was based on variables that would vary within the studied population and were believed to be observable for all participants. If the distribution of these variables were to be even over the different groups, the randomization process could be assumed to be rather successful. There are other variables that could be used to validate the randomization process but gender, age and employment were chosen since they would be natural to include in a simple signup form, to keep the situation as realistic as possible and avoid raising any suspicion of an experiment. It was

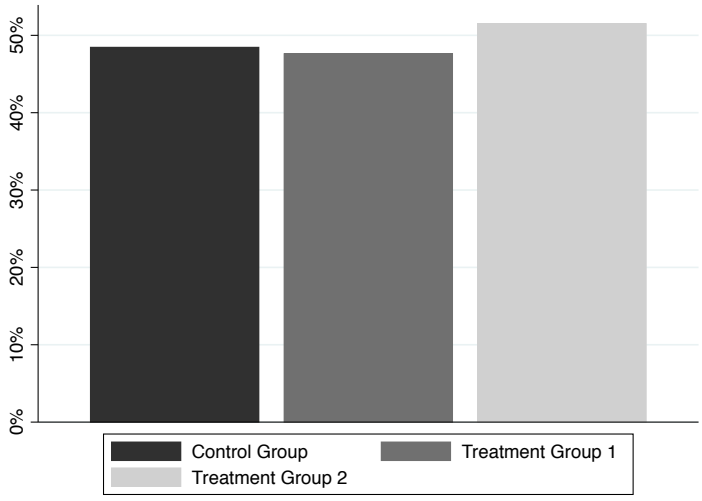
important that the survey was perceived as a real signup and not as a data collecting survey for a thesis, to make the experiment more representative of a real life signup and to improve the ecological validity. None of the distributions were significantly different between the groups. A full analysis can be found in the next section *6 Results* and the results from the t-tests can be found in *Appendix III*.

6 Results

6.1 Descriptive statistics

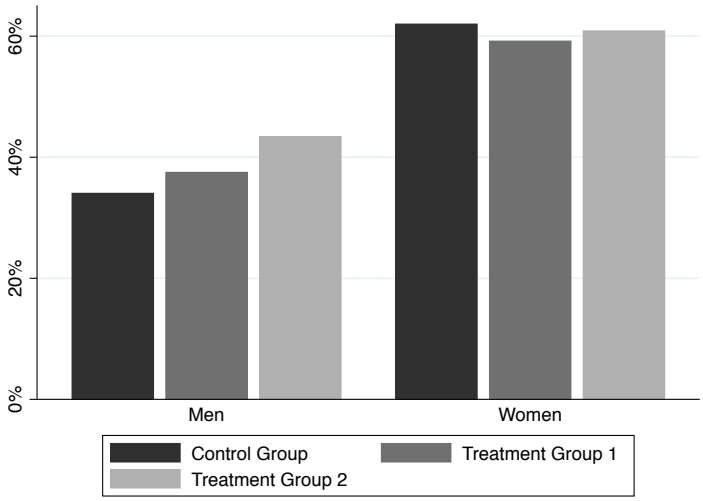
The experiment was conducted on 301 participants with 97 (32.23%), 105 (34.88%), and 99 (32.89%) in the control group (CG), the first treatment group (TG1) and the second treatment group (TG2) respectively. As Graph 6.1 shows, there are differences in the fractions of signups between the groups that are in line with the crowding-out hypothesis and the charity hypothesis. Inference tests on whether the fractions are significantly different are performed in the following sections 6.2 and 6.3.

Graph 6.1: Fractions of signups in each group



Graph 6.2 also indicates support for the gender differences hypothesis as the fraction of signups for women is higher than the fraction of men who sing up in general over the total sample. Inference tests are performed in section 6.4 to see whether the intercept and treatment effects for women are significantly different from zero.

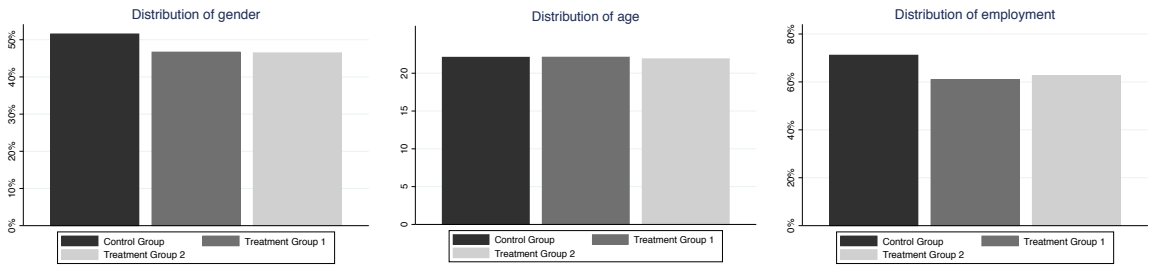
Graph 6.2: Fractions of signups by gender



In order to validate the randomization process, data was collected on the participants' gender, age and workload. This data was then analysed to confirm that the distribution of each variable was even over all groups. T-tests were performed for each variable and no significant differences were found. The regressions can be found in *Appendix III* and the descriptive statistics are presented below.

The proportions of females and males were even both in the total sample and over the three groups. Expressed as fractions of females over the total number of participants, the fraction in the total sample was 48.17% and the distribution over the groups was 51.55% in CG, 46.67% in TR1 and 46.46% in TR2. In the overall sample, the age of the participants ranged between 18 and 30 years old with a mean age of 22.04 years. The variation of the age in the sample was however small, with 87% of the respondents between the age of 19 and 24. The mean age in each group is 22.10 years (CG), 22.12 years (TG1) and 21.90 years (TG2). The distribution employment was also quite even between the groups. Two thirds of the participants had a part-time job (195 of 301 people, 64.78%) and the fractions in each group were 71.13% (CG), 60.95% (TR1) and 62.63% (TR2). The distributions of all three control variables are presented in Graph 6.3.

Graph 6.3: Distributions of control variables



6.2 Analysis of the crowding-out hypothesis

To test whether there is a significant difference between the fractions of signups in the control group, in which no compensation was offered, compared to the first treatment group, in which a compensation of SEK 50 was offered, a Pearson chi-square test was performed on the contingency table shown in Table 6.1.

Table 6.1: A 2 × 2 contingency table for the crowding-out hypothesis

	Frequencies			Fractions (π)	
	Total	Yes	No	Yes	No
CG	97	47	50	48.45%	51.55%
TG1	105	50	55	47.62%	52.38%
Pearson chi-square = 0.0141				**P-value = 0.906**	

The contingency table indicates differences in line with the crowding-out hypothesis, but these differences are proven to be too small as the chi-square test gave a p-value of 0.906 and the null hypothesis cannot be rejected at any conventional significance level. Any support for the crowding-out theory in the context of voluntary work could not be found in this experiment.

6.3 Analysis of the charity hypothesis

To test whether there is a significant difference between the fractions of signups in the first treatment group and the second treatment group, in which both were offered a compensation of SEK 50 but only the second treatment group was offered the opportunity to donate the amount, a Pearson chi-square test was performed on the contingency table shown in Table 6.2.

Table 6.2: A 2 × 2 contingency table for the charity hypothesis

	Frequencies			Fractions (π)	
	Total	Yes	No	Yes	No
TG1	105	50	55	47.62%	52.38%
TG2	99	51	48	51.52%	48.48%
Pearson chi-square = 0.3094				**P-value = 0.578**	

The contingency table indicates differences in line with the charity hypothesis, but the Pearson chi-square test gave a p-value of 0.578 and the null hypothesis cannot be rejected at any conventional significance level. Any support for the signalling model of crowding-out in the context of voluntary work could not be found in this experiment.

6.4 Analysis of the gender differences hypothesis

In order to test if there is a difference in altruistic behaviour between genders, the female dummy together with the interaction terms between gender and the two treatments, were tested with the Linear Probability Model. In order to test for gender differences, the three regressions presented in section 4 are performed; (1) a first one where only the treatment dummies were tested to validate the results from the Pearson chi-square tests, (2) a second one in which a dummy for gender was included to test whether there is a general difference in prosocial behaviour between men and women, and (3) a third one where the interaction terms between gender and the treatments finally included also investigate whether the treatments had different effects on the genders. The results are presented in Table 6.3 (regressions using logistic regression analysis are presented under *Appendix II*).

Table 6.3: Results from the regressions on treatment effects and genders differences

	(1)	(2)	(3)
VARIABLES	help	help	help
female		0.223***	0.280***
		(0.0566)	(0.0984)
		[0.000]	[0.005]
TG1	-0.00835	0.00255	0.0346
	(0.0707)	(0.0687)	(0.0956)
	[0.906]	[0.970]	[0.718]
TG2	0.0306	0.0420	0.0935
	(0.0718)	(0.0700)	(0.0980)
	[0.670]	[0.549]	[0.341]
female*TG1			-0.0627
			(0.138)
			[0.649]
Female*TG2			-0.105
			(0.140)
			[0.456]
Constant	0.485***	0.369***	0.340***
	(0.0510)	(0.0568)	(0.0698)
	[0.000]	[0.000]	[0.000]
Observations	301	301	301
R-squared	0.001	0.051	0.053

Robust standard errors in brackets
P-values in square brackets
*** p<0.01, ** p<0.05, * p<0.1

The first regression confirmed for the results presented in section 6.2 and 6.3 – support could be found for neither the crowding-out hypothesis nor the charity hypothesis. In the second regression, the tested gender dummy is significantly different from zero at the significance level of 1%. The third regression showed that the null hypothesis of no gender differences could be rejected at the significance level of 1% (the p-value of the F-test is 0.0062). However, looking at the tested coefficients separately, only the female dummy variable is found to be significantly different from zero at the significance level of 1% (the p-value is 0.005). With p-values of 0.649 and 0.456, neither the interaction term for the first treatment nor the second treatment is significantly different from zero. Considering the high p-values, far away from conventional significance levels, any speculation regarding trends becomes irrelevant. The conclusion of these results is that women were generally more prone to sign up for the voluntary activity, but they were not affected differently by monetary compensation. The logistic regression analysis in *Appendix II* produced the same results.

Finally, it can be noted that the R-squared for all three regressions are at a low level. However, given the purpose and design of the experiment, a low value of R-squared would be expected. The purpose of this analysis was to determine whether the treatments or gender had any impact at all on the participants' inclination to sign up for the tutoring event, not to identify the determinants of the inclination. Since a randomized experiment was used, the independent variables were dummies for the treatment groups and gender, and no control variables were included. It is therefore likely that there are other variables with explanatory value in the residual that inflate the residual sum of squares and consequently reduce the R-squared. Furthermore, the high p-values and the insignificant independent variables of the performed regressions make any discussion or speculation regarding the R-squared less relevant.

7 Discussion

The aim of this study was to investigate whether the crowding-out theory and the signalling model of crowding-out that previously have found support in an empirical experiment on blood donation behaviour (Mellström and Johannesson, 2008) could be extended to the context of voluntary work for refugees. Furthermore, as Mellström and Johannesson only found empirical support for their crowding-out hypothesis (Titmuss' crowding-out theory) and their charity hypothesis (Bénabou and Tirole's signalling model of crowding-out) for women, after separating the sample by gender, this study also posed the question whether there are any empirical differences in the motivation of altruistic behaviour between genders.

As the previous section *6 Results* showed, no support for neither the crowding-out hypothesis (that monetary compensation would reduce the fraction of signups) nor the charity hypothesis (that the opportunity to donate the amount, given that compensation was offered, would increase the fraction of signups) was found. The third hypothesis stated that there would be differences between the genders both in the general inclination for signing up (the female dummy variable) and the effect of the treatments of monetary compensation and the opportunity to donate (the interaction terms between gender and the treatments). When this hypothesis was tested, the null hypothesis could be rejected. However, the results were not as expected. The treatment effects were still insignificant, even when the sample was separated by gender as it was by Mellström and Johannesson. Only the estimated coefficient for the female dummy variable was significantly different from zero, more specifically positive with the size of 0.2896. This means being a woman increases the probability of signing up for the voluntary activity by 28.96%. In other words, women are in general more likely to help others by volunteering.

Based on the high p-values obtained from the hypothesis testing, this study does not find any empirical evidence supporting the effect of monetary incentives on the motivation of altruistic behaviour. This implies that the crowding-out effect previously found in the context of blood donation cannot be generalized for prosocial behaviour and applied in the context of voluntary work. The following sections will focus on interpreting the results, evaluating their reliability, examine their implications and identify areas of interest for future studies. The first section will discuss possible theoretical explanations to why conclusions regarding effect of monetary incentives on blood donation could not be applied to voluntary work. The second

section will mainly focus on the limitations of the design and conduct of the experiment to assess the reliability of the results and to which extent they can be applied.

7.1 Discussion on the crowding-out effect in voluntary work

The purpose of this paper was to investigate whether the theories of crowding-out can be extended from blood donation to voluntary work, thus generalized for altruistic behaviour. This study could not find any empirical evidence of crowding-out in the context of voluntary work and this section offers two possible explanations to why no crowding-out effect could be found as hypothesised. These explanations mainly discuss the interpretation and application of altruism, but they must however be read with caution as there are several limitations of the design and conduct of the experiment that must be considered. These limitations could also be an explanation to why no crowding-out effect was found and will therefore be presented in the next section. Potential explanations include:

1. **Blood donation and voluntary work are not perceived 'as altruistic'**

   Altruism is defined as helping others at your own cost, and both donation and volunteering are by definition considered as altruistic behaviour. However, both actions may not be considered as being 'as altruistic' in reality. With blood donation, you share something that you are given by nature for free[6] and the consequences of your actions could be saving lives. To accept a payment for such a thing may feel ethically wrong. Voluntary work however, is still performed work that might not feel as bad to get paid for, even if it is for the good cause of refugee children as in our experiment. If the voluntary work is too related to ordinary occupational work, a payment could also be rationalized. In this case for example, tutoring is a common part-time job for students and this volunteering activity could be compared to the occupation, making a payment seem fair. Finally, tutoring might also be considered a luxury good[7] instead of a vital need, and therefore not something that feels as wrong to get paid for.

   If tutoring for refugees is not considered as something that you do purely to help someone else in need, the connection to altruistic behaviour might not be as obvious and there might not be a crowding-out effect by the compensation. Consequently, the theories applicable on blood donations might not be applicable on voluntary work.

[6] Although personal costs of for example fear and time must be considered.
[7] Note the difference with basic education, which is considered a human right in Sweden.

2. **Altruism is context dependent**

Donation and voluntary work are both altruistic actions but they differ in the execution. Generally speaking, donation could be seen as a more passive action in which you share something physical,[8] while voluntary work is more of an active engagement in which you contribute with time and energy. Therefore, if you were to view donation and voluntary work as two separate actions and not jointly as altruistic, it would be reasonable that their motivations also consist of different proportions of intrinsic, extrinsic and reputational motivation.

From Titmuss' crowding-out perspective, it could be argued that blood donation consists of relatively more intrinsic motivation than voluntary tutoring for refugees. When you donate blood, you are unaware of the consequences of your contribution – you do not know who you are helping or what the blood will be used for. You simply do it for the good sake of it. It would then be reasonable to argue that blood donation is mainly motivated by intrinsic motivation; it is all about the act and not the results. When it comes to voluntary work, it may not always be just about the act itself. For example, as in this case with voluntary tutoring for refugee children, you might help them achieve a goal (improve in school), you gain some personal development as you actively engage with them (leadership, compassion, perspective) and you contribute to a societal problem (the refugee crisis). Then, it is more focus on the results – the end goal – compared to the case of blood donation, and it would make sense that voluntary work is relatively less driven by intrinsic motivation and relatively more by extrinsic motivation. If the proportion of intrinsic motivation is less to begin with, there might not enough to cause a notable crowding-out effect when monetary incentives are introduced.

Furthermore, from the perspective of the signalling model of crowding-out, it could be argued that donation is a more commonly used way to signal altruism to the people in your surroundings. Generally speaking, for those who perform an altruistic act solely to signal altruism, a donation through one transaction could be considered to be relatively easier and quicker than to participate in a voluntary activity.[9] When it comes to voluntary activities, you must be more engaged and it is more demanding in terms of time

[8] Although the personal investment differs depending on the type of the donation. In blood donation for example, the donor must invest at least half an hour of time and the loss of blood usually takes energy from you.

[9] Blood donation is however an exception since the donors must first complete a health examination that takes approximately fifteen minutes before becoming a registered regular donor. Once registered, the donation process is down to a single transaction that takes approximately half an hour. Even if the donor must consider personal costs, such as fear and energy losses, their engagement is quite passive at the clinic while the blood pump is working compared to the tutoring activity, in which the volunteer must interact with the student to solve and explain the academic problem for two hours.

and energy. As it is a less efficient way to signal altruism, there might not be as many people who participate in voluntary work for the signalling reason. If that is the case, the introduction of the donation possibility will not result in the same crowding-out effect as with blood donation, since the portion of people driven by reputational motivation will be relatively lower.

To conclude, blood donation and voluntary work might not be driven by the same relative proportion of intrinsic motivation and the signalling value might not be as high for voluntary work compared to donation, and it will therefore be no crowding-out effect to be found in the empirical data. Consequently, the theories of Titmuss and Bénabou and Tirole may for that reason not be applicable in the context of voluntary work. More empirical studies in other contexts are however necessary before any conclusions can be drawn.

7.2 Limitations of the experiment

When discussing the reasons for the lack of crowding-out in this study, it is important to recognize the limitations of the experimental design and conduct. This section will present four factors that must be taken into consideration when applying the results in other contexts. The identified limitations were however deliberate decisions based on practical restrictions and are also areas of improvement that should be considered when further research is conducted.

First of all, there is a risk that the conduct of the experiment was insufficient. In Mellström and Johannesson's blood donation study, the experiment was performed by professional scientists who visited the schools and came in direct contact with the participants. The presence of professional scientists may give a more formal approach, making the participants take the decision to help more seriously. Furthermore, the 'altruistic consequences' were immediate. Even if the participants technically only had to go through a medical examination, which meant no actual blood donation, the procedure was immediately after the survey and in connection with the experiment. With the direct consequences, the participants were forced to act on their decisions immediately and they could not sign up without the intention to actually go through with the act. Consequently, their design of the experiment may have given a more realistic impression.

In the experiment of this study, the participants were asked to sign up for a planned upcoming event through an Internet survey, which is not as personal and the consequences were further away in the future. Due to this difference, the participating students might not have taken the decision as seriously. There is a risk that they answered the survey only to finish 'just another

student survey' or that those who were offered compensation only signed up for the activity to 'get money for free', without a true intention to participate once it was time. Furthermore, the survey did not guarantee that the tutoring event would take place, it clearly stated that it was only in the planning process, although Pimp My Grades and Studiefrämjandet were involved with the aim to make it happen. This is an important difference, since signing up in this experiment did not directly demand participation like the blood donation experiment did. If the situation is not realistic enough, the decision might no longer be about an altruistic act and crowding-out theories might not be applicable. Even if the credibility was something that was of importance and of priority (for example, the experiment was designed with guidance from an associate professor with experience from behavioural economics, and two external actors were introduced), practical restrictions of resources and time made it difficult to address these differences. This is something that future research could aim to improve.

Secondly, the choice of students from Stockholm School Economics (SSE) limits the generalization of the study. As previously discussed, the choice was based on practical reasons, and is not considered as solely negative; as a consequence of the homogenous sample group, the internal validity of the results was approved. Even if the randomization process of the experiment ensures unbiased estimation, the homogeneity of the sample reduces the risk that the property of pure chance would create unbalanced groups, which randomization cannot affect. This internal validity does however come at the cost of the external validity. With such a homogenous sample, the results cannot be generalized to the whole student population or the Swedish people. The restriction to SSE also limits the sample size since the school is not that large in terms of number of students. Further research could therefore aim to find a larger and more representative sample group. For example, a next step could be to visit the largest universities from different parts in Sweden. That way, different geographical and socioeconomic groups will be represented and the larger sample size would improve the precision of the results.

A third factor that could limit the applicability of the results is the choice of refugee children. The refugee crisis has been a well-debated political issue in Sweden during the recent year. By having refugee children as the recipient of the voluntary work, the decision to sign up for the voluntary activity might no longer be about whether you want to help someone or nor, but rather a decision of whether you want to participate in the welcoming of refugees or not. The decision-making then shifts away from an altruistic act to a political statement. Previous research on the supply of volunteers has shown that the motivation to engage in volunteering activities is affected by historical events in society. These events are defined as volunteer activators (Penner, 2004) and studies on events such as the 9/11 attack (Penner, 2004) and the bombing of the

Murrah Federal Building in Oklahoma City (St John and Fusch, 2002) show that the event of disasters motivate people to volunteer. If the experiment had been conducted a year ago or a year from now, the results might be different.

Yet, the decision to have refugee children as recipients is motivated. The purpose was to increase the credibility of the voluntary activity. By having a current societal issue, the demand of help is believable and the probability that the participants would really believe that the activity was going to take place was improved. It is not that unexpected that this kind of an activity would be arranged at SSE and with the importance of the work, it is believable that it would take place. Having this trade-off in mind, future research could replicate the study at a different point of time and further expand this experiment into other contexts of voluntary work, in order to improve the strength of the results.

Finally, the fourth limitation identified is the choice of tutoring. Tutoring was chosen because it is a natural activity for university students and it was practical. The activity demanded no preparation or resources and it made the event accessible since the connection to education made it possible to host it at school property. It also enabled the collaboration with the tutoring group Pimp My Grades that was already established at the school. The risk of choosing this activity was however that it might be too closely associated with the occupation of tutoring. As mentioned, tutoring is a common part-time job for students. This could mean that the activity is more related to paid work than voluntary work, and taking it further away from altruism. A recent study has found that monetary incentives across work contexts improve performance, rather than causing a crowding-out effect (Kosfeld et al., 2016). This means that it might be the association to paid work instead of voluntary work that lead to the results of no crowding-out.

8 Conclusions

The recent refugee crisis in Europe has proven the importance of and need for volunteers. With an increasing demand of voluntary work, it is important to understand how volunteerism best can be incentivized. A randomized experiment was conducted to see if the crowding-out effect of monetary incentives previously found in blood donation behaviour could be generalized for a different type of altruistic behaviour, more specifically, if it can be applied to voluntary work in the context of tutoring for refugee children. The experiment tested the three following hypotheses: *the crowding-out hypothesis* that the introduction of compensation would reduce the motivation to volunteer, *the charity hypothesis* that the introduction of a donation possibility would increase the motivation to volunteer given that a compensation is offered, and *the gender differences hypothesis* that there would be differences in prosocial behaviour of men and women and if they should be incentivizes differently. The experiment found no support for the crowding-out effect, although the empirical analysis did show that women were more inclined to sign up for the tutoring event in general.

The main implication of this finding is that what motivates altruism may be context driven and hence vary depending on the amount of intrinsic, extrinsic and reputational motivation required for that specific altruistic act. This in turn may be affected by the level of active engagement and commitment required, the time and energy used, the subject the volunteer aims to help and the signalling effectiveness for the studied act of altruism. Hence, this study suggests that by simply generalizing all types of altruism and assuming they are driven by the same factors may result in less efficient measures aimed at increasing the supply of volunteers. Another implication is that volunteering behaviour differs between genders and that these differences may be relevant to be accounted for when developing these strategies. Finally, if altruism is context driven as suggested, the findings of this study would also have methodological findings, that not only concern the methods but also the central concepts used in social experiments studying altruistic behaviour. Further studies on the topic should therefore reconsider the theoretical application of altruism, and carefully take the whole context into account and try to find an experimental design and implementation which allow for context to stay constant.

As there are limitations of the experimental design, the reliability of the obtained results would benefit from more research. One suggestion is finding a larger and more representative

sample group by visiting the largest universities from different parts in Sweden. That way, different geographical and socioeconomic groups will be represented and the larger sample size would improve the precision of the results. Future research could also replicate the study at a different point of time and further expand this experiment into other contexts of voluntary work than tutoring. Conducting more studies and testing on other types of volunteer work, other volunteer objects, other time perspectives and other aid subjects can draw more reliable, general and pragmatic conclusions that ultimately can aid stakeholders in increasing the volunteer workforce.

References

Aguiar, F., Brañas-Garza, P., Cobo-Reyes, R., Jimenez, N. & Miller, L.M. 2009, "Are women expected to be more generous?", *Experimental Economics,* vol. 12, no. 1, pp. 93-98.

Andreoni, J. 1990, "Impure altruism and donations to public goods: a theory of warm-glow giving", *Economic Journal,* vol. 100, no. 401, pp. 464-477.

Andreoni, J. 1989, "Giving with impure altruism: applications to charity and Ricardian equivalence", *Journal of Political Economy,* vol. 97, no. 6, pp. 1447-1458.

Bénabou, R. & Tirole, J. 2006, "Incentives and prosocial behavior", *American Economic Review,* vol. 96, no. 5, pp. 1652-1678.

Berenson, M., Levine, D., Szabat, K.A. & Krehbiel, T.C. 2012, *Basic business statistics: concepts and applications,* 12th edition, Pearson Higher Education AU, Upper Saddle River, NJ.

Bloom, H.S. 2008, "The core analytics of randomized experiments for social research", *The Sage handbook of social research methods*, editors Alasuutari, P., Bickman, L. & Brannen, J., Sage, London, pp. 115-115-133.

Boschini, A., Muren, A. & Persson, M. 2012, "Constructing gender differences in the economics lab", *Journal of Economic Behavior & Organization,* vol. 84, no. 3, pp. 741-752.

Brewer, M.B. & Crano, W.D. 2014, "Research design and issues of validity", *Handbook of research methods in social and personality psychology*, 2nd edition, eds Reis H.T. & Judd, C.M., Camebridge University Press, New York, NY, pp. 11-26.

Croson, R. & Gneezy, U. 2009, "Gender differences in preferences", *Journal of Economic Literature,* vol. 47, no. 2, pp. 448-474.

Eckel, C.C. & Grossman, P.J. 1996, "Altruism in anonymous dictator games", *Games and Economic Behavior,* vol. 16, pp. 181.

Ellingsen, T., Johannesson, M., Mollerstrom, J. & Munkhammar, S. 2013, "Gender differences in social framing effects", *Economics Letters,* vol. 118, no. 3, pp. 470-472.

Fiorillo, D. 2011, "Do monetary rewards crowd out the intrinsic motivation of volunteers? Some empirical evidence for Italian volunteers", *Annals of Public and Cooperative Economics,* vol. 82, no. 2, pp. 139-165.

Frey, B.S. & Jegen, R. 2001, "Motivation crowding theory", *Journal of Economic Surveys,* vol. 15, no. 5, pp. 589-611.

Frey, B.S., Oberholzer-Gee, F. & Eichenberger, R. 1996, "The old lady visits your backyard: a tale of morals and markets", *Journal of Political Economy,* vol. 104, no. 6, pp. 1297-1313.

Gneezy, U. & Rustichini, A. 2000, "Pay enough or don't pay at all", *Quarterly Journal of Economics,* vol. 115, no. 3, pp. 791-810.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. & McElreath, R. 2001, "In search of Homo Economicus: behavioral experiments in 15 small-scale societies", *American Economic Review*, vol. 91, no. 2, pp. 73-78.

Internationell Människohjälp*, Visa ditt stöd för flyktingarna* [Homepage of Internationell Människohjälp], [Online]. Available: http://manniskohjalp.se/volontar/visa-ditt-stod-flyktingarna [2016, April 30].

Kosfeld, M., Neckermann, S. & Yang, X. 2016, "The effects of financial and recognition incentives across work contexts: the role of meaning", *Economic Inquiry*, DOI 10.1111/ecin.12350.

Mellström, C. & Johannesson, M. 2008, "Crowding out in blood donation: was Titmuss right?", *Journal of the European Economic Association,* vol. 6, no. 4, pp. 845-863.

Penner, L.A. 2004, "Volunteerism and social problems: making things better or worse?", *Journal of Social Issues,* vol. 60, no. 3, pp. 645-666.

Solna Stad 2016, 2016-07-04-last update*, Flyktingar - så kan du hjälpa* [Homepage of Solna Stad], [Online]. Available: https://www.solna.se/sv/om-solna/nyheter-om-solna/flyktingar-sa-kan-du-hjalpa/ [2016, July 15].

St John, C. & Fuchs, J. 2002, "The heartland responds to terror: volunteering after the bombing of the Murrah Federal Building", *Social Science Quarterly,* vol. 83, no. 2, pp. 397-415.

Titmuss, R.M. 1970, *The gift relationship: from human blood to social policy,* Allen & Unwin, London.

Unger, L.S. 1991, "Altruism as a motivation to volunteer", *Journal of Economic Psychology,* vol. 12, no. 1, pp. 71-100.

Wallace, T., Bhattacharjee, S., Chao, M., Devereux, P., El-Kholy, H. & Hacker, E. 2015, *State of the world's volunteerism report 2015*, The United Nations Volunteers (UNV) programme, Washington, DC.

West, S.A., Gardner, A. & Griffin, A.S. 2006, "Altruism", *Current Biology,* vol. 16, no. 13, pp. R482-R483.

Wooldridge, J.M. 2013, *Introductory econometrics: a modern approach,* Fifth international edition, South-Western, Cengage Learning, Mason, OH.

Appendix I

Survey: Control Group (CG)

Hello and welcome!

We are currently writing our bachelor thesis at Stockholm School of Economics (SSE) and as a first step we need to gather data from SSE students.

You will be asked **4 questions** that will take approximately **3 minutes** in total to answer.

IMPORTANT! In order for the data to remain valid, please do not talk to others about the content of this survey nor your answers.

Thank you for your participation!

***

Last autumn, the number of unaccompanied refugee children increased substantially. These children have the same rights as Swedish citizens to attain education. However, school is not always that easy, especially if you are also new in the country. Learning a new language, being surrounded by strangers, trying to fit in - sometimes you need a helping hand.

In order to contribute to the integration of these refugee children, a tutoring evening for unaccompanied refugee children is planned to be held at SSE. The refugee children will come to our school and receive help from our students. The tutoring evening is planned to take place in the beginning of May and will be two hours long.

This event will be a collaboration between **Studiefrämjandet** and **Pimp My Grades**, who are both interested in supporting the event. **Studiefrämjandet** is an educational association advocating children's right for education. **Pimp My Grades** is part of the Student Association of Stockholm School of Economics (SASSE) and are hosting weekly tutoring sessions.

In hope of making this evening possible, we are now looking into the recruitment of SSE students interested in helping these children.

Would you like to help tutor refugee children at this event? As soon as time and date are set, you will be contacted with further information and be asked to confirm if you will be available or not.

- Yes/No

- If yes:

  Please enter your contact information:

  Name:
  Mobile:
  Email:

***

Finally, please fill in some general information about yourself.

- Please enter your age:
  XXX

- Please enter your gender:
  Male/Female

- Do you have a part-time job?
  Yes/No

- If yes:
  Please indicate the workload per week:

  - Less than 10 hours/10 hours/More than 10 hours

***

Your participation is highly appreciated, thank you very much!

REMEMBER! In order for the data to remain valid, please do not talk to others about the content of this survey nor your answers.

Survey: Treatment Group 1 (TR1)

Hello and welcome!

We are currently writing our bachelor thesis at Stockholm School of Economics (SSE) and as a first step we need to gather data from SSE students.

You will be asked **4 questions** that will take approximately **3 minutes** in total to answer.

IMPORTANT! In order for the data to remain valid, please do not talk to others about the content of this survey nor your answers.

Thank you for your participation!

***

Last autumn, the number of unaccompanied refugee children increased substantially. These children have the same rights as Swedish citizens to attain education. However, school is not always that easy, especially if you are also new in the country. Learning a new language, being surrounded by strangers, trying to fit in - sometimes you need a helping hand.

In order to contribute to the integration of these refugee children, a tutoring evening for unaccompanied refugee children is planned to be held at SSE. The refugee children will come to our school and receive help from our students. The tutoring evening is planned to take place in the beginning of May and will be two hours long.

This event will be a collaboration between **Studiefrämjandet** and **Pimp My Grades**, who are both interested in supporting the event. **Studiefrämjandet** is an educational association advocating children's right for education. **Pimp My Grades** is part of the Student Association of Stockholm School of Economics (SASSE) and are hosting weekly tutoring sessions.

In hope of making this evening possible, we are now looking into the recruitment of SSE students interested in helping these children.

Would you like to help tutor refugee children at this event? By signing up, you will receive **SEK 50**. As soon as time and date are set, you will be contacted with further information and be asked to confirm if you will be available or not.

- Yes/No

- If yes:

  Please enter your contact information:

  Name:
  Mobile:
  Email:

  This information will also be used in order to give you your compensation when the signups are closed.

***

Finally, please fill in some general information about yourself.

- Please enter your age:

XXX

- Please enter your gender:
  Male/Female

- Do you have a part-time job?
  Yes/No

- If yes:
  Please indicate the workload per week:

  - Less than 10 hours/10 hours/More than 10 hours

***

Your participation is highly appreciated, thank you very much!

REMEMBER! In order for the data to remain valid, please do not talk to others about the content of this survey nor your answers.

Survey: Treatment Group 2 (TR2)

Hello and welcome!

We are currently writing our bachelor thesis at Stockholm School of Economics (SSE) and as a first step we need to gather data from SSE students.

You will be asked **4 questions** that will take approximately **3 minutes** in total to answer.

IMPORTANT! In order for the data to remain valid, please do not talk to others about the content of this survey nor your answers.

Thank you for your participation!

***

Last autumn, the number of unaccompanied refugee children increased substantially. These children have the same rights as Swedish citizens to attain education. However, school is not always that easy, especially if you are also new in the country. Learning a new language, being surrounded by strangers, trying to fit in - sometimes you need a helping hand.

In order to contribute to the integration of these refugee children, a tutoring evening for unaccompanied refugee children is planned to be held at SSE. The refugee children will come to our school and receive help from our students. The tutoring

evening is planned to take place in the beginning of May and will be two hours long.

This event will be a collaboration between **Studiefrämjandet** and **Pimp My Grades**, who are both interested in supporting the event. **Studiefrämjandet** is an educational association advocating children's right for education. **Pimp My Grades** is part of the Student Association of Stockholm School of Economics (SASSE) and are hosting weekly tutoring sessions.

In hope of making this evening possible, we are now looking into the recruitment of SSE students interested in helping these children.

Would you like to tutor at this event? By signing up, you will receive **SEK 50** that you can choose to keep or you can donate it to **UNICEF**. The money will be donated when the signup is closed. As soon as time and date are set, you will be contacted with further information and be asked to confirm if you will be available or not.

- Yes/No

- If yes:

  Please enter your contact information:

  Name:
  Mobile:
  Email:

  Further indicate whether you want to donate the amount or not.

- Donate:

  Yes/No

  If you chose to keep the compensation, this information will also be used in order to give you the amount when the signups are closed.

***

Finally, please fill in some general information about yourself.

- Please enter your age:
  XXX

- Please enter your gender:
  Male/Female

- Do you have a part-time job?
  Yes/No

- If yes:
  Please indicate the workload per week:

  - Less than 10 hours/10 hours/More than 10 hours

***

Your participation is highly appreciated, thank you very much!

REMEMBER! In order for the data to remain valid, please do not talk to others about the content of this survey nor your answers.

Appendix II

Logistic regression analysis

Table A.2: Logistic regression analysis of treatment effects and gender differences

	(1)	(2)	(3)
VARIABLES	help	help	help
help			
female		2.484***	3.161***
		(0.589)	(1.340)
		[0.000]	[0.007]
TG1	0.967	1.011	1.162
	(0.273)	(0.293)	(0.481)
	[0.906]	[0.970]	[0.716]
TG2	1.130	1.193	1.485
	(0.323)	(0.351)	(0.615)
	[0.668]	[0.547]	[0.339]
female*TG1			0.764
			(0.446)
			[0.645]
female*TG2			0.642
			(0.378)
			[0.452]
Constant	0.940	0.586**	0.516**
	(0.191)	(0.142)	(0.159)
	[0.761]	[0.028]	[0.032]
Observations	301	301	301

seEform in brackets
P-values in square brackets
*** p<0.01, ** p<0.05, * p<0.1

Note: The coefficients estimated by the logistic regression analysis are interpreted as odds ratios and will therefore differ in size compared to the coefficients estimated by the linear probability model. However, the inference tests still produce the same results.

Appendix III

Robustness checks of the control variables distributions

Table A.3: Inference tests of the control variables distribution over the groups

VARIABLES	(1) female	(2) age	(3) hasjob
TG1	-0.0488	0.0123	-0.102
	(0.0706)	(0.251)	(0.0673)
TG2	-0.0508	-0.204	-0.0851
	(0.0717)	(0.254)	(0.0683)
Constant	0.515***	22.10***	0.711***
	(0.0509)	(0.180)	(0.0485)
Observations	301	300	301
R-squared	0.002	0.003	0.009

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Note: The t-tests above produce no significant coefficients meaning that there is no significant difference in the fractions of females, mean age and the fractions of employment between the groups (mean values can be used for the fractions of females and employment as both are dummy variables with values of 0 and 1).