Essays on the Determinants and Impact of Private Contributions to Public Goods

This dissertation was typeset by the author using LaTeX.

Cover design: BIRDesign, Culemborg, the Netherlands (www.birdesign.nl)

Print: Haveka.nl, Alblasserdam, the Netherlands

ISBN: 978-90-9029526-8

Essays on the Determinants and Impact of Private Contributions to Public Goods

Essays over de determinanten en impact van private contributies aan publieke goederen

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols
and in accordance with the decision of the Doctorate Board.
The public defense shall be held on
Thursday 3rd of March 2016 at 13:30hrs

by

Frank Roeland Hubers
born in Papendrecht, the Netherlands

Erasmus University Rotterdam

Doctoral Committee

Promotor:	Prof.dr. H.D. Webbink
Other members:	Dr. J. Delfgaauw
	Prof.dr. R. Dur
	Prof.dr. R. Ruben

Preface

A few years ago, I worked as a cultural anthropologist for the international aid organisation Oxfam Novib, in those days one of the largest charity organisations in the Netherlands. As most charity organisations, Oxfam Novib largely depended on voluntary donations and government subsidies. One major problem we were facing was that the public had become increasingly sceptical of the impact of international aid, which affected both public approval and funding. This critical attitude of donors was in contrast with the decades before, where the good intentions of charity organisations were considered more important than their results. Working for the evaluation department, it was our task to provide evidence about the impact of development projects. Did the projects alleviate people out of poverty? But despite the fact that dozens of evaluations had been carried out over the years, most of them lacked the rigorous approach necessary to make claims about causal effects. Many charity organisations in the Netherlands currently deal with this problem: A public that has become increasingly suspicious about the efficacy of their interventions, and a lack of empirical evidence to show that their interventions have a positive impact on society.

As a cultural anthropologist I am trained to use qualitative research techniques - in-depth interviews, participatory observation - and analysing qualitative data. Although these methodologies are very useful for conducting research on cultures, believe systems and describing small societies, they are unfit to evaluate the impact of a certain intervention. One of the most prominent problems with qualitative studies is the lack of a counterfactual. In other words: What would have happened to the beneficiary, had he or she not participated in this project? This question cannot be answered by merely interviewing the people that participate. Most people simply have no idea what the effect of a certain policy has had on them. Nevertheless, many of the evaluation studies on the impact of development projects funded by Dutch organisations, are based merely on qualitative research methods.

In 2010 I started working for a newly founded knowledge center that focussed on this problem. The Erasmus Centre for Strategic Philanthropy (ECSP) was founded at the Erasmus University of Rotterdam, with the objective to gain a better understanding about the effectiveness of voluntary donations. One of the major aims of this centre was to generate academic knowledge about the impact of various interventions done by philanthropic organizations. ECSP offered me the opportunity to do research on the impact of philanthropic projects in the context of a PhD-position. Most importantly, I got the opportunity to do the type of research that aims at establishing the causal effects of public interventions, and that is based on credible research designs. This also implied that I had to be trained in the foundations of statistics and applied econometrics. This dissertation contains the major studies I carried out during the time I worked for ECSP.

I couldn't have written this dissertation without the continuous support of numerous persons. First of all, I would like to thank Dinand Webbink. Although it was probably not easy for him to guide an anthropologist into the world of applied econometrics, he supported me throughout to process, and were always there to offer me useful critical feedback. Dinand, I have really enjoyed working with you. Especially the field work trips were unforgettable. Whether they took place in the *favelas* in Rio de Janeiro, the gold mines in Ghana, or at the white beaches in Thailand, we always managed to find some time and space to sit down and relax with a few - buckets of - cold beers. Special thanks also goes to Charles Erkelens and Kellie Liket, without whom I wouldn't have started this PhD trajectory in the first place. Kellie, thank you for introducing me to ECSP, and Charles, thank you for providing me the opportunity to start as a PhD researcher for ECSP. Also, I would like to thank the rest of my former ECSP colleagues, including Karen Maas, Lonneke van der Waa, Lonneke Roza, Eva van Baren, Pushpika Vishwantan, Marjelle Vermeulen, Pamela Wiepking, Michiel de Wilde, Manuela Ettekoven, Mirjam Locadia and all the others for the interesting discussions and enjoyable meetings. Likewise, I thank my fellow PhD students for the fruitful discussions, in particular Sanne Blaauw, Ivan Lyubimov, Pieter Schoonees, Jonathan Attey, Irena Mikolajun and Violeta Misheva. I also thank my colleagues at the Department of Economics, in particular Otto Swank, Ankimon Vernede and Milky Viola Gonzalez, for their help and support. Furthermore, I thank the members of the PhD committee: Ruerd Ruben, Josse Delfgaauw, Robert Dur, Lucas Meijs, René Bekkers and Philip Hans Franses for their time and effort.

The second and the third chapter in this dissertation are conducted in cooperation with non-profit organisations and required fieldwork periods with intensive data-collection. The study in Chapter 2 is

based on a fieldwork period of five months in cooperation with the Dutch non-profit organisation Movies that Matter. During this period I conducted experiments with pupils from six secondary schools scattered around the Netherlands, which required much coordination and organisation. Luckily I received help from Hamza Aammari and Hicham El-Mahtouchi, who assisted me during the experiments and provided me with feedback throughout the process. Thank you very much, not only for your assistance, but also for the good conversations during the - often very long - train rides to the different schools we had to visit. I also thank Movies that Matter, in particular Margreet Cornelius, Ali Remmelts and Marcus Eshuis, for their feedback, cooperation and support. My gratitude also goes to the the teachers of the six secondary schools that made it possible for me to conduct this study, in particular Jan de Goede, Gerrit Jan Stoffer, Roelof Doorn, Mohammed Faizi Nazir, Christel Wets, Jen Haase and Henriette van Antwerpen.

The study in Chapter 3 is conducted in close cooperation with Solidaridad Netherlands and Solidaridad West-Africa, the non-governmental organisations that implement the Fair Trade program for gold in Ghana. The evaluation study was funded by the Dutch National Postcode Lottery. I thank the Dutch National Postcode Lottery for funding this study, and Solidaridad for their cooperation. In particular I like to thank Wilbert Brentum, who has been of great help during the implementation of the study in Ghana, as well as Francisca Hubbeek, Ester Prins and Carien Duisterwinkel for their feedback and support. Also I would like to thank Charles Abbey and his team of enumerators from the University of Mining and Technology (UMAT) in Tarkwa, Ghana. A final organisation I would like to mention here is Children Asking. Although the evaluation of their PEDE project in Brazilian shantytowns is not included in this dissertation, I have learned a lot during this experience, which also served to inspire later studies. Special thanks go to Eric Visser, Cristiane Markus and Isabela Souza.

I want to thank all my friends and family who directly or indirectly contributed to this dissertation. In particular I thank my parents, Joke Hubers and Joep van Heeswijk, who supported me in many ways. Also I thank Pepijn Jansen, Reinout van Santen, Job Harms, Hanneke van Diepen and Jeroen Verburg for reading and commenting on earlier versions of this dissertation, which improved my thesis considerably. Last but certainly not least, I would like to express my gratitude and love to my dear wife. Mink, you have always been by my side and were of great moral support. Thank you for all your love and support. Sorry for all the nights that I spent writing this thesis, but now this is finished, I can finally give you the attention you deserve.

Frank Hubers, December 2015

Contents

[2] Joint work with Dinand Webbink
[2] Joint work with Dinand Webbink; Published in IZA Journal of Labor Economics, 4:10 (2015)

List of Figures

List of Tables

How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortune of others, and render their happiness necessary to him, though he derives nothing from it except the pleasure of seeing it.

Adam Smith, The Theory of Moral Sentiments

1

Introduction

Almost two decades before he wrote his most famous work - *The Wealth of Nations* - Adam Smith, founding father of modern economic thought, already noted that human beings are not driven by self-interest alone. The billions of euros that are donated to charitable organisations annually worldwide, indicate that this is indeed the case. In the Netherlands alone, approximately 4.3 billion euros were donated to charity organisations in 2013[1], approximately 0.7 percent of the GDP.

Besides the significant amount of money involved in the sector, the apparent contradiction between the rational and selfish *homo economicus* and its willingness to voluntary donate a part of its income attracts an increasing academic interest in the field of philanthropy. One of the main premises of classical economic theory is that humans are solely guided by self-interest. How can philanthropy exist if people are rational self-interested beings? Economists generally consider the act of philanthropy as voluntary

[1] Based on data from the annual report "Giving in the Netherlands 2015" from www.geveninnederland.nl

private contributions to public goods, and the philanthropic organisation as the private providers of these goods. The most obvious reason that people decide to donate time or money to charity is that they simply benefit from the public good provided. These benefits can be clear and direct, for example in donating to a sports club, but they can also be less clear and more uncertain, for example donations to cancer research, from which the donor may or may not benefit in the future. Not all philanthropic giving is thus necessary driven by altruistic motivations. At the same time, there are numerous charities that do not appear to be beneficial for the donor, for example donations to victims of a famine in a country on the other side of the world, or most other cases of international aid. Are these donations then solely driven by altruistic motivations? For economists, altruism implies that one's utility is affected by the utility of another (Fehr and Schmidt, 2006). Many people are found to be inequality averse - implying that their individual utility is negatively affected by high levels of social inequality - which might explain those donations to international aid and the willingness to redistribute income. If these donations are driven by altruistic motivations we may consider social equality a public good. However, Andreoni (2006) distinguishes a third motive for people to donate, which is that people derive utility from the act of giving itself. Note that if an individual would merely be interested in the quantity of public good, he or she would be indifferent about the means through which it is provided. There are numerous empirical studies that show that willingness to contribute to charitable institutions is strongly determined by the moral satisfaction the individual gains from contributing to it (Andreoni, 1988, 1989; Kahneman and Knetsch, 1992). Andreoni (1990) describes this motivation as the 'warm glow of giving' or 'impure altruism.' Warm glow giving is of particular importance for charitable organisations with a large number of donors, for which the marginal contribution of the individual contribution is small. Economic theory predicts undersupply of public goods through voluntary donations if there would be no warm-glow giving.

Regarding the fact that charitable donations are tax-deductable, and many charity organisations in the Netherlands receive subsidies, an empirical approach to the efficacy of philanthropic interventions is a matter of public interest. Private donors have become increasingly sceptical about the impact of their donations. During the years that I worked for an international aid organisation, we intended to increase transparency about the impact of development aid, but we encountered the problem that still little was known about the impact of our interventions. Although dozens of evaluation studies had been conducted over the years, most of them lacked the rigorous approach necessary to make any claims about causal effects.

A good understanding of the role of philanthropy, the government and the market in public good provision is essential. This thesis presents four essays about the effectiveness and determinants of private public good provision. It contains four empirical studies assessing the impact of specific private contributions to public goods; or investigating the impact of specific determinants of philanthropic giving. These four studies use experimental and quasi-experimental research designs in which the outcomes of individuals that receive a treatment or compared with the outcomes of individuals in control groups that do not receive a treatment. By using transparent control groups I aim to estimate the causal effects of the interventions on the outcomes. The four studies in this thesis contribute to the small pool of knowledge about the impact of philanthropic interventions based on credible research designs. In addition, the application of these research designs can be considered as quite novel for the Dutch philanthropic sector.

However, it is difficult to assess the effect of specific factors on philanthropic giving or the impact of specific private contributions on public outcomes because these factors or contributions are not randomly allocated. In general, a multitude of factors might influence philanthropic giving. Moreover, the impact of philanthropic effects might be influenced by many other observed or unobserved factors. As such, estimates of the causal effect of specific determinants or of private contributions to public goods might be biased by omitted variables. Despite the increased public interest in the effectiveness of philanthropic giving in the Netherlands, so far only a very limited number of evaluations have been carried out with credible evidence of causal effects. Most impact evaluations of Dutch interventions are still based on anecdotal evidence. In this thesis I apply quasi-experimental designs to investigate both the factors that influence voluntary contributions to public goods, as well as the effects of specific interventions on society. Chapter 2 and 3 investigate determinants of philanthropic giving. Chapters 4 and 5 investigate the impact of private contributions to public goods.

The first two studies focus on giving behavior. What affects the decisions of people to donate money - or time - to charity? Chapter 2 investigates the effect of request size and mood on the willingness to donate. In Chapter 3, I examine the effects of showing inequality to adolescents on their willingness to donate and redistribute. Both studies make use of experimental data. Chapter 3 shows the results of a field experiment that I carried out in 6 secondary schools, in which the treatment - watching a film about inequality - was randomly allocated per classroom. The data for Chapter 2 were collected via online surveys, and the value of the independent variable 'request size' was determined by observable answers of the respondent to other questions. Since the request size was determined by a random procedure and by observable game play, we are able to obtain exogenous variation in request size. We also investigated

the effect of mood via a natural experiment in which the donations of supporters of a winning football team were compared with the supporters of the losing teams.

Chapter 4 investigates the Fair Trade program for small-scale gold miners in Ghana. The Fair Trade initiative is a form of private provision of labour standards in the gold mining sector, carried out by an international nonprofit organisation. Ideally this would be evaluated with a field experiment, but in the practical implementation of the Fair Trade project, this was not possible. Instead I used a difference-in-differences approach with two waves of data. With a team of enumerators I conducted two waves of interviews with miners that worked at treated mines, as well as miners that worked at nearby non-treated mines. The assumption is that the time trend in the control miners reflect the time trend for the treated miners, were they not treated. In the last chapter, Chapter 5, I investigate the long term effects of military conscription. The military draft is a more traditional way of public good provision, in which the government forces its citizens to contribute in the form of labour - instead of tax. This study examines the effects on wages and educational attainment and illustrates that there are hidden costs of conscription. The results are based on a natural experiment that exploits an exogenous shock in which one particular birth cohort was exempted from military service. This strategy is based on the methodology used by Imbens and van der Klaauw (1995). Assuming that there is no direct relation between being born in this particular year and one's income and educational attainment, other than through the exemption of military service, the differences between this birth cohorts provide an estimate of the causal effect of conscription.

1.1 Outline and summary

Does it matter whether charitable donations ask for small or large donations? In Chapter 2, which is shared work with Dinand Webbink, we investigate the effect of request size on the willingness to donate. We exploit exogenous variation in request size from a natural experiment. Participants in two online surveys earned money from incentivized games and were asked to donate the full amount to charity. This request size was determined by a random procedure and by observable game play. We find that an increase in the request size reduces the probability to donate. In our context however, charitable organizations would maximize their revenues by choosing the highest request size in the range of our data (16 euros). We also investigate the importance of mood by exploiting variation in mood due to the outcome of the Dutch soccer competition. We find that mood matters for charitable giving but does not

affect the relationship between request size and charitable giving.

Chapter 3 investigate if showing inequality affects redistributional social preferences of teenagers. Can altruistic giving be affected by external factors, like films or the media? I use data from a field experiment in which 541 pupils from 6 secondary schools participated. The participants of the experiment, all pupils between 12 and 16 years old, were randomly assigned to watch a film about people living in poverty, and invited to play incentivized dictator and ultimatum games. After the games the participants were given the opportunity to donate a proportion of the earned money to charity. I find that the teenagers that watch the films make more generous offers than the teenagers that did not. Both the offers in the games as well as the donations to charity increase after watching one of the films. One film also positively affects the demanded offers in the ultimatum game, indicating an increased sense of justice. This study provides evidence that social preferences can be affected by showing cases of people living in poverty.

Chapter 4 evaluates the impact of the implementation of Fair Trade labour standards on the lives of small-scale - or artisanal - gold miners. Artisanal gold-mining poses severe health and safety risks to the miners and their surroundings. Fair Trade labelling is an effort of solving an information asymmetry in which consumers have insufficient information about the quality of a product. Assuming that consumers are willing to pay a higher price for a product that is produced under socially responsible circumstances, the non-profit organisation implements these standards in artisanal miner organisations, hoping to improve the living conditions of the miners, and selling the gold for a higher price in the market to cover the cost. To evaluate the effect of this intervention on the gold miners, I use a difference-in-difference approach, in which respondents that work for Fair Trade mines are compared over time with respondents that work for comparable control mines. The data were collected through more than 1900 interviews with artisanal miners from 10 artisanal mining sites in Ghana, conducted in two waves. I find that the implementation of the Fair Trade standards have a positive effect on the health of gold miners. The likelihood of miners getting ill or injured decreases by 23 percent. Moreover, the school attendance of children in the households increases by 10 percent as a result of the intervention. However, the intervention has a negative effect on income: the estimated proportion of miners that live above USD 2.50 per day decreases by 5 percent as a result of the intervention. Hence, there appears to be a trade-off in implementing the Fair Trade standards. A last important finding is that the turnover rate of gold miners is high - with a median employee tenure of one year. This limits the impact of the intervention.

In Chapter 5, which is shared work with Dinand Webbink and was recently published in IZA Journal

of Labor Economics[2], we explore the hidden costs of a specific case of compulsory public good provision. This study investigates the long-term effects of peacetime military conscription on educational attainment and earnings, by exploiting a policy change that exempted a complete birth cohort from military service. We find that compulsory military service decreases the proportion of Dutch university graduates with 1.5 percentage points from a baseline of 12.3 percent. In addition, being a conscript reduces the probability of obtaining a university degree with almost four percentage points. The effect of military service on earnings is also negative and long lasting. Approximately 18 years after military service, we still find a negative effect of 3 to 4 percent. The effect of conscription on educational attainment does not fully explain the wage reduction.

[2]IZA Journal of Labor Economics, 4:10 (2015)

2

How much to ask for charity? Experimental evidence of the effect of request size on small charitable donations *

2.1 Introduction

Each year private individuals donate substantial amounts of money to charitable organizations. For instance, in the US private donations amount to 2% of GDP and in the Netherlands approximately 5 billion Euros are yearly given by individuals. This probably explains the growing interest of the economic literature in the determinants of charitable giving. The literature has investigated various determinants

*Joint work with Dinand Webbink

of giving such as changing the price of giving, the method of solicitation, signaling devices, revealing the donors identities, and the communication of social information (List, 2011). A recent study looks at the effect of suggesting a specific amount during solicitation (Edwards and List, 2014) and builds on earlier work in social psychology (Weyant and Smith, 1987; Doob and McLaughlin, 1989). These studies suggest that the amount of money that is asked matters for charitable giving, but the evidence is yet quite limited and the findings do not seem to be consistent.

This study investigates how asking for higher donations affects charitable giving. We estimate the causal effect of request size on charitable giving by exploiting exogenous variation from a natural experiment. Participants in two online surveys could earn money from incentivized games. At the end of the survey they were asked to donate the full amount to charity. As it was not possible to donate a fraction of the earned amount this implies that the amount that was earned is equal to the request size. The amount that was won was determined by a random procedure for choosing the question that would be paid out and the game play of the respondent. The variation in the earned amount (request size) ranges between 0 and 16 euros and can be considered to be exogenous as we can control for differences in 'game play.' We use this variation for estimating the effect of request size on the probability to donate and for determining the request size that yields the highest revenues for charitable organizations. In the second part of the paper we investigate the importance of mood for charitable giving. We exploit the variation in the mood of the respondents induced by the outcome of the Dutch national soccer competition. Part of our data were collected among fans of various competing soccer clubs immediately after the decisive game of the competition. In the analysis we compare the charitable giving of the fans of the winning club (Ajax Amsterdam) with the charitable giving of the fans of the clubs that did not win the competition.

We contribute to the economic literature by investigating the effect of request size in a natural setting. The variation that we can exploit ranges from 0 to 16 euro's and this variation also offers the opportunity to investigate the optimal request size. Previous studies merely investigated the effect of either a low or a high suggested amount in the context of field experiments and were not informative about the issue of the optimal request size. Our data also provide the opportunity to study the impact of mood on philanthropic giving and the interaction between mood and request size. The variation in the mood of individuals in our study comes from a natural event, whereas previous studies used artificial settings for changes in mood.

We find that the request size matters in the willingness to donate: increasing the request size by one euro reduces the likelihood of receiving a donation by approximately 1.8 percentage point. However, this

does not imply that fundraising organisations are better off asking for the smallest possible donations, as suggested by Weyant and Smith (1987); Weyant (1984); Cialdini and Schroeder (1976). The optimal request size is greater than 16 euro and hence lies outside our range of data. We estimate the optimal request size - using a quadratic equation - to be approximately 31 euro. A second conclusion of this paper is that a good mood positively affects an individual's willingness to donate. This is in line with the conclusions of other studies on this topic (Kirchsteiger et al., 2006; Capra et al., 2010; Capra, 2004). We do not find a evidence of an interaction effect between mood and request size.

The remainder of this study is organized as follows. The next section will provide a literature review about the relationship between price, request size and charitable giving. Section 2.3 explains our empirical strategy. Section 2.4 provides a description of the data and shows summary statistics. Section 2.5 shows the main estimation results and in Section 2.6 we investigate the robustness of our results. Section 2.7 concludes.

2.2 Previous studies

Economists have so far paid little any attention to effects of request size in charitable giving. Instead they focussed on the price of donation. Economists regard non-profit organisations as private suppliers of a public good; and donation as voluntary contributions towards the output quantity (or quality) of this good. If a donor is mainly interested in the output of the public good, the request size should not matter. A large number of empirical studies investigate the effect of price on the willingness to give, by focussing on exogenous changes in the relationship between the output quantity and the price for which it is available. Although the exact output quantity is mostly unknown to the individual donor, the price can be manipulated, either by experimental setting or via government policies. Changes in tax policies - the extent to which donations are tax deductible, affecting the price for which a public good quantity is availabe - have served as exogenous shocks in a number of studies. (Boskin and Feldstein], 1977; Feldstein and Taylor, 1976; Feldstein and Clotfelter; Clotfelter, 1980; Auten et al., 2002). These policies may affect both the total income of an individual as the price of donation, these studies mostly both estimate income elasticity as well as the price elasticity of donations. Although the estimated price elasticity of donations vary between studies - ranging from 2.0 Boskin and Feldstein] (1977) to 0.40 (Auten et al., 2002) - all studies conclude that a discount on the price of donations via tax deductions increases the amount donated.

A recent wave of studies examined the effect of price on donations with an experimental approach. Karlan et al. (2011) investigate the effect of matching grants with a field experiment in which mail solicitations were sent to 20,000 prior donors to a charity. Matching grants, a fundraising strategy in which potential donors are told that their donations will be matched by an equal or higher donation, is a strategy comparable to the tax deductions in the sense that the potential donor gets a 'discount' on giving. They find only weak evidence that these matches affect the willingness to donate. Earlier studies found positive effects of matching grants. The study of Meier (2007), based on a similar field experiment finds the exact opposite effect of Meier (2007): The individuals that receive a match grant offer were more willing to donate than the control group, but this reversed after the experimental period. The individuals that did not receive the matching showed more willingness to donate in the period after the matching offer ended. Karlan and List (2007), carrying out a field experiment in which contributions were solicited from 50,000 supporters of a liberal organization, found that the announcement of a matching gift significantly increases the probability that an individual donates by 22% and the revenue per solicitation by 19%. The difference in estimates between the studies can be caused by the type of organisation or the period of the year the experiment took place. Rondeau and List (2008) investigated both the effect of challenge and matching grants on the probability of donating. A challenge grant is an unconditional grant - whether the individual decides to donate or not - and hence more comparable to the government grants to charity - in the perspective of the donor. Using both a field experiment as well as a lab experiment they find no evidence of the effect of matching grants, similar to Karlan et al. (2011).

All empirical studies above suggest that price matters in charitable giving. The willingness to donate increases when the donor can "buy" a higher quantity of public good for the same price. The fact that donors are sensitive about the price does necessary mean that they are also sensitive for the request size. In fact, if the potential donor is rational and would only care for the quantity of public good provided, the request size would not matter -e.g. have no effect - at all. On the other hand, if voluntary donations are solely driven by the motivation to provide a public good, government donations to charity - whether through tax reductions or subsidies - would crowd out private giving perfectly (Warr, 1982). Andreoni (1988, 1989) show empirically that this is not the case and that, in fact, the total donations to charity increase result of government grants. Andreoni (1990) describes this motivation as the 'warm glow of giving' and considers it "impure altruism due to its selfish nature. Kahneman and Knetsch (1992) show that the willingness to contribute to public goods is strongly determined by the moral satisfaction the individual gains from contributing to it. Since this moral satisfaction is not necessarily connected to

any potential public good output, but to the act of giving itself, it is likely that the the willingness to give is also affected by the amount that is solicited. We may thus expect that if the moral satisfaction becomes available for a lower price - a lower request size - the demand will increase. This would imply that requesting for a smaller amount would positively affect the number of donations.

The empirical literature on the relationship between request size and willingness to donate is limited, with the majority of the studies focussing on the effect of specific suggestions rather than the effect of an increase in request size itself. The earliest studies are conducted with door-to-door collectors for charity organisations and focus on the effect of suggesting a small donation - the solicitor using the phrase "Even a penny counts" or "Even a penny will help." Most studies find a positive effect on the willingness to donate. Moreover, while the willingness to donate increases, the average donation doesn't, indicating that charity organisations would be better off asking for small donations (Weyant, 1984; Cialdini and Schroeder, 1976; Reingen, 1982). Weyant and Smith (1987) further investigated this effect with an direct mailing experiments, in which one group was asked for a small donation (suggested donations: \$5, \$10, \$15), the other for large donations (suggested donations: \$50, \$100, \$150) and one control group in which no specific amounts were suggested. In line with the earlier studies, they find that smaller suggestions increase the proportion of donors that donate whereas larger suggestions decreases the number of donations. They conclude that charitable organisations will be better of requesting smaller amounts, since the average amount donated is similar in both groups. However, a replication of this study by Doob and McLaughlin (1989), finds opposing results. They find no difference in the willingness to donate between recipients that are requested either a small or a large donation, which makes them conclude that charity organisations will be better off if they increase their request sizes. The most recent study on the effect of suggestions shows the results of a field experiment in which alumni were asked for a donation for their university (Edwards and List, 2014). In one of the treatment conditions recipients were suggested to donate a certain amount (\$20), while in another no amount was suggested. Edwards and List (2014) find that suggesting an amount positively affects the proportion that decides to donate. Respondents are also more likely to donate the suggested amount, but less likely to donate more. They interpret this result as a consequence of the recipients receiving a fixed amount of utility from donating at least as much as the suggestion, but little utility from exceeding it.

This study will further investigate how the request size affects the willingness to donate. A second objective is to investigate how positive emotions can affect charitable giving. Recent studies investigating the relationship between mood and economic behavior indicate that charitable donations are affected

by mood. The studies provide evidence that happy emotions positively affects charitable giving whereas sad emotions have a negative effect on charitable giving (Kirchsteiger et al., 2006; Capra et al., 2010; Capra, 2004; Gummerum et al., 2010). The evidence is based on the outcomes of laboratory experiments in which the willingness to donate is approximated by the respondent's allocations in a one-shot dictator game. The experiments use mood inducement methods mostly film shorts with an emotional content to randomly induce participants with a happy or sad mood. In our setup we exploit an exogenous shock that serves as a mood inducer. We are unaware of any studies that apply a similar strategy to estimate the effect of emotions on charitable giving.

2.3 Empirical strategy

In this study we investigate the effect of request size on charitable giving. During the online survey the respondents could win money by answering questions about economic preferences. The respondents acquired information about the sum of money that was won and were then asked (online) to donate this amount to charity. The respondents only had two options, either donate or keep the full amount. Because of this specific setup of the survey, the amount that was won can be considered the request size for charity. We can estimate the effect of the request size on the willingness to donate with the following equation:

$$D_i = \alpha + \beta_1 R_i + BX_i + \varepsilon_i \tag{2.1}$$

in which D_i is a dummy variable equal to 1 if individual i donated the amount to charity or equal to 0 if she decided to keep it; and R_i the request size which equals the amount won by participating in the survey; X_i is a vector of control variables of individual i and ε is the error term that includes all unobserved factors. Estimation of Equation (2.1) will yield the causal effect of the request size if the request size is not correlated with unobserved factors. This conditional independence assumption seems plausible in the context of this study. The amount that was won was determined by the answers of the respondents to the questions about economic preferences and by a random procedure for choosing the question that would be paid out to the respondent. Although the answers of the respondents are not random, we can include these answers as controls in our estimation model. By controlling for differences in 'game play' the amount that was won (the request size) is likely to be random. We check this assumption by performing various balancing tests. In Equation (2.1) the estimated parameter β reflects the increase

of the probability of donating when the request size is increased with one euro.

To answer the question whether higher solicitations increase revenues for charity organisations, we will estimate the following equation:

$$Y_i = \delta + \gamma_1 R_i + \gamma_2 R_i^2 + \Gamma X_i + \eta_i \tag{2.2}$$

in which Y_i is the total revenue for the charity. We can calculate the optimal request size by taking the first derivative of Equation (2.2) and setting it equal to zero.

A further question that we want to investigate in this study is whether there exists a relationship between the mood of the respondent and charitable giving. More specifically, we want to investigate whether the estimated effect of request size depends on the mood of the respondent. If the mood of the respondent could be directly observed we could estimate the following equation to address this question:

$$D_i = \alpha + \beta_1 R_i + \beta_2 M_i + \beta_3 R_i M_i + BX_i + \varepsilon_i \tag{2.3}$$

in which M_i is a measure for the mood of individual i and β_3 our parameter of interest, indicating whether the effect of request size on charitable giving depends on the mood of the respondent. An important problem in estimating Equation (2.3) is that it is difficult to measure the mood of the respondents. Previous studies have addressed this issue by using variation in mood induced by showing films with a specific content. In this study we exploit the variation in the mood of the respondents induced by the outcome of the Dutch national soccer competition. In the second dataset that we will use in this study (see next Section) the data were collected among fans of various competing soccer clubs immediately after the decisive game of the competition. In the analysis we compare the charitable giving of the fans of the winning club (Ajax Amsterdam) with the charitable giving of the fans of the clubs that did not win the competition. We estimate the following reduced form equation:

$$D_i = \alpha + \beta_1 R_i + \beta_2 Won_i + \beta_3 R_i Won_i + BX_i + \varepsilon_i \tag{2.4}$$

in which Won_i is a dummy variable equal to 1 if the team that individual i supports won the competition and 0 otherwise. We expect that the variable Won_i proxies a higher average mood. In the questionnaire additional questions were asked about the importance of soccer and the results of the favorite team, and the importance of the outcomes of the competition, which we will also use to link an

individual mood to the effect of request size on charitable giving. The parameter β_3 shows the effect of request size that depends on the mood of the individual.

2.4 Data

This study uses data from two studies on economic preferences. The data for the two studies were collected in April and May 2012 through online questionnaires by a private firm specialized in the use of online panels. Respondents were all members of regular panels of the firm. The first study focused on the relationship between religion and economic behavior. Respondents were randomly assigned to a treatment group that obtained an assignment for unscrambling sentences with a religious content. The control group received an assignment for unscrambling similar sentences without a religious content. The aim of unscrambling sentences with a religious content was to prime respondents into a religious state and to investigate whether this has an effect on their economic behavior measured in several experimental games. The respondents for this study were selected from the online panels based on information about their religious affiliation. Moreover, the sample focused on individuals aged 40 or above. The second study focused on the relationship between mood (happiness) and economic behavior. Respondents were asked to play several experimental games in the first days after the finish of the Dutch soccer competition. It was expected that winning the championship would have a positive impact on the mood of the fans of the new champion (Ajax Amsterdam) only. The mood of the fans of the rivals of the new champion (e.g. Feyenoord Rotterdam, PSV Eindhoven) was expected to be negatively affected or remain constant. For this sample only male respondents were selected especially those males living in the region of the main soccer clubs.

Respondents for the two studies were invited to participate in a survey by email. This email contained a unique link to the online-questionnaire. The respondents were not informed about the content before the questionnaire started, nor were they informed about the fact that they could earn money by playing experimental games. For the first study 716 people were invited to participate. Seven days after receiving the invitation participants who had not yet started the questionnaire were sent a reminder. For the second study 1,545 male individuals were invited. Two rounds of data collection yielded a total of 1,053 valid observations. For the analysis in this paper we excluded 6 observations that did not receive money for their participation because they have no money to donate to charity. This yields a dataset of 1,047 observations (374 from the first study and 673 from the second).

The data consist of various background characteristics, such as age, gender, educational attainment, income, family size and composition, that were collected when the respondents started to participate in the online panels. Moreover, the questionnaires for the two studies included respectively questions about religious participation or questions about being a soccer fan. The questionnaires also contained various incentivized questions to reveal economic preferences. They contained 6 questions in regard to risk preferences which were stated as "Would you rather receive 1 Euro with certainty, or have a 50% chance on receiving x euro with a 50% risk of receiving nothing?" Several amounts were given between 1.60 and 3.60 in equal increments. Both questionnaires included a version of the public good game, in which respondents could choose to allocate an amount between 0 and 5 euro - which they received in the beginning of the game to a public good. The allocations to the public good would be doubled and returned equally to all players. The questionnaire used for the second study also included time preference questions in which respondents were asked: "Which amount would reflect the same value for you now as the amount of x euro would for you in t days?" In total 9 time preference questions were included with three values for x (4.11, 8.31 and 10.14) and three values for t: 7, 14 and 28. All games and preference questions were incentivized. The respondent was told, just before the section of preferences started, that one of the questions in this section would be paid out. After playing these games the amount that was won was shown to the respondents on their screen. In the final stage of the questionnaire respondents were asked whether they would like to donate the amount that was won to charity. Through this question the amount won becomes equal to the request size, which is one of the main variables in this study.

Table 2.1 shows the background characteristics of the respondents of the two studies. The first column shows the demographics of Study 1, the second column shows the demographics of Study 2 and the last column the demographics of both studies combined.

The samples differ in composition because of the different focus of the two studies. The first study focussed on religion and economic preferences; Respondents were selected based on their religious affinity and age. This explains the high proportion of religious respondents compared to the second study. The second study focused on soccer fans and only included male respondents. This sample also includes a wider range of ages than the sample of the first study. The reason is that the first study has an age restriction (at least 40 years). The average amount that was won by participating was 7.41 Euros. More than half of the respondents (55 %) decided to donate their winnings to charity.

Table 2.1: Summary statistics for Study 1 and Study 2

	Study 1	Study 2	Combined
Amount won	7.33	7.46	7.41
	(2.76)	(2.91)	(2.86)
Proportion donated	0.61	0.52	0.55
	(0.49)	(0.50)	(0.50)
Male	0.53	1	0.83
	(0.50)	(0)	(0.38)
Age	54.48	47.26	49.86
	(7.19)	(15.84)	(13.82)
Educational attainment	13.34	13.66	13.54
	(2.91)	(3.15)	(3.07)
Happiness (Scale 1-10)	7.51	7.34	7.43
	(1.19)	(1.28)	(1.25)
Household income: Low	0.15	0.13	0.14
	(0.36)	(0.33)	(0.34)
Household income: Median	0.34	0.27	0.30
	(0.47)	(0.45)	(0.46)
Household income: High	0.47	0.55	0.52
	(0.50)	(0.50)	(0.50)
Household income: Unknown	0.04	0.05	0.05
	(0.20)	(0.23)	(0.22)
Family size	2.69	2.69	2.69
	(1.38)	(1.36)	(1.37)
Religion: Catholic	0.33	0.25	0.28
	(0.47)	(0.43)	(0.45)
Religion: Protestant	0.51	0.19	0.30
	(0.50)	(0.39)	(0.46)
Religion: Other	0.07	0.07	0.07
	(0.26)	(0.25)	(0.26)
Religion: None	0.08	0.50	0.35
	(0.276)	(0.50)	(0.48)
Risk attitude [2]	2.21	2.23	2.22
	(0.88)	(0.74)	(0.80)
Public Good Game input	3.27	3.19	3.22
	(1.58)	(1.59)	(1.58)
Time preference	-	0.97	0.97
	-	(0.20)	(0.20)
Nr. of observations	374	673	1047
Date of first response	19 April 2012	3 May 2012	19 April 2012
Date of last response	20 May 2012	17 May 2012	20 May 2012

[1] Notes: Standard Deviations in parenthesis. Risk attitude is measured as the smallest amount of money with a 50% probability of receiving that the respondent prefers over a certain amount certain amount of 1 euro.

The dependent variable

The main dependent variable in this study is a dummy variable equal to 1 if the individual decides to donate the amount that was won to charity. The respondent has two choices: either to donate all the earnings he won or to donate nothing at all; it is not possible to donate a fraction of the earned amount. We also investigate the effect of the request size on the total amount that is donated. For this analysis we use the amount that was donated as the dependent variable. The total amount donated is equal to the amount won when the respondent decides to donate or equal to 0 otherwise.

The independent variable

Our main independent variable is the amount that the respondent 'won' by participating in the survey. This amount was determined by a random payout of one of the nine preference questions in the first study or by one of the sixteen preference questions in the second study. The amount won is endogenous as it depends on the answers of the respondents. However, the answers of the respondents are observed in the survey and by controlling for these answers the amount that was won can be considered as conditionally independent. Figure 2.1 shows the distribution of the amount won. We observe that the distribution peaks at 6 euro and that there is substantial variation between 1 and 15.6 euros. In the remainder of the article the variable 'Amount won will be referred to as 'Request size.

In our datasets we can distinguish three groups of control variables. The first group of control variables is related to the collection of the data and includes a dummy for the dataset and variables for the date and time that the respondent participated in the survey. The second group of control variables consists of the answers to the questions about economic preferences. This group of controls includes two variables for risk aversion: One indicating the smallest money amount with a 50% probability of receiving that the respondent prefers over a certain amount of one euro; and dummy variable indicating whether the respondent is consistent in his risk preference answers. [2] For the public good game we included a variable indicating the amount that an individual allocated towards the 'public good.' The time preference variable is a measure that indicates the average present value of the amount of 1 euro tomorrow. [3] Since the time preference question was only asked in the second study, we included a dummy

[2] Consistency with regard to risk implies that the respondent prefers all higher offers with the same risk - as the minimum amount - over the certain amount of 1 euro. Inconsistency hence implies that the respondent choose a 50% probability of receiving 2 euro over a certain amount of 1 euro, but also choose the 50% probability of receiving 3 euro over the certain amount of 1 euro.

[3] Or $timepref = (FV/PV)^{\frac{1}{t}}$ in which t is the number of days, FV the future value and PV the present value.

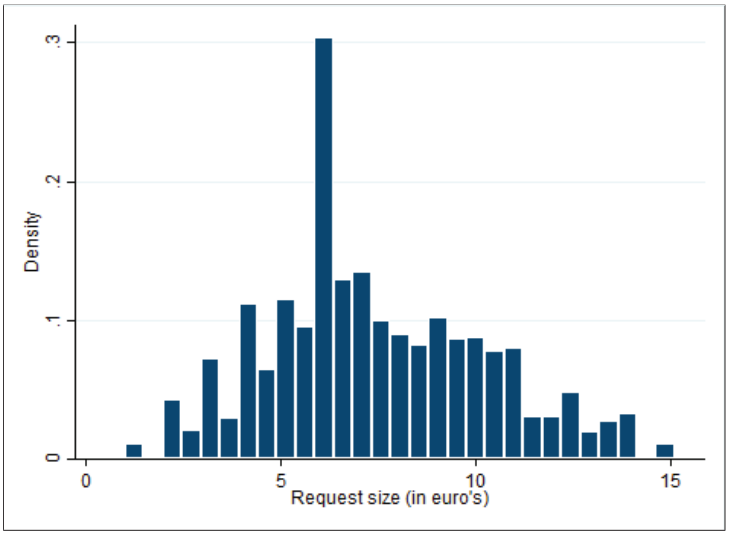


Figure 2.1: Distribution of variable 'Amount won'

variable equal to 1 if the respondent answered the time preference question - all respondents in the first study - or equal to 0 otherwise, and set the missing observations in the time preference variables to zero. The third group of control variables consists of the background characteristics: age, gender, educational attainment, family size, household income level and religion. The household income has been measured in three categories: low, median or high income.

Balancing test for 'Request size'

As a first step in the analyses we perform balancing tests to check whether the 'request size can indeed be considered as conditionally random. We regressed this variable on a set of background characteristics. Table 2.2 shows the results of these regressions. Columns (1) and (2) show the estimates for Dataset 1; Columns (3) and (4) for Dataset 2. Columns (5) and (6) show the results for the combined dataset. The odd columns show the estimates without controlling for the answers of the respondents on the questions about economic preferences (game play).

Table 2.2 shows that the request size is unrelated to the various observable characteristics of the respondents. This holds both for the models that control for 'game play' as for the models that don't control for 'game play.' This probably follows from the fact that the pay-out for the preference questions

Table 2.2: Balance test

	Dataset 1		Dataset 2		Total	
	(1)	(2)	(3)	(4)	(5)	(6)
Age	0.009	0.012	-0.010	-0.011	-0.006	-0.006
	(0.022)	(0.022)	(0.008)	(0.008)	(0.007)	(0.007)
Male	0.348	0.239			0.194	0.085
	(0.302)	(0.303)			(0.253)	(0.298)
Educational attainment	0.017	0.009	0.036	0.016	0.031	0.018
	(0.052)	(0.052)	(0.040)	(0.039)	(0.031)	(0.031)
Household income	0.041	0.008	0.190	0.146	0.128	0.085
	(0.209)	(0.207)	(0.172)	(0.170)	(0.133)	(0.131)
Household income unknown	0.997	1.132	0.351	0.077	0.448	0.303
	(0.880)	(0.871)	(0.638)	(0.630)	(0.511)	(0.507)
Family size	0.142	0.149	-0.096	-0.091	-0.012	-0.007
	(0.116)	(0.115)	(0.093)	(0.091)	(0.071)	(0.072)
Religious (dummy)	0.307	0.198	-0.112	-0.024	-0.049	-0.005
	(0.541)	(0.533)	(0.224)	(0.221)	(0.196)	(0.203)
Observations	374	374	673	673	1,047	1,047
Controls: Gameplay	No	Yes	No	Yes	No	Yes
F-value	0.62	0.62	0.71	0.55	0.67	0.34
F-test (Sig.)	0.740	0.743	0.644	0.773	0.700	0.937
R-squared	0.038	0.075	0.007	0.055	0.010	0.048

Notes: Each column contains regressions estimates of models that use the request size as dependent variable. Standard errors are shown in parentheses.

was randomly determined. These balancing tests support the conditional independence assumption with respect to the request size.

2.5 Main estimation results

This section will show the estimates on the effect of request size on charitable giving. First, we investigate the effect of changes in the request size on the probability of donating money to charity. Next, we investigate the optimal request size for charitable organizations. Third, we investigate whether a good mood increases the effect of request size on charitable giving.

2.5.1 The effect of request size on the probability of charitable giving

A first impression of the relationship between request size and charitable giving can be obtained from Figure 2.2 and 2.3. These figures plot the probability of donating against the request size. The lines show

the predicted values estimated by a LOWESS regression. Figure 2.2 shows the increase in the request size in intervals of one euro on the x-axis and the proportion of respondents that donated on the y-axis. Every dot in Figure 2.2 reflects the average proportion of the respondents that donated when winning an amount that lies in that interval. As shown in Figure 2.1 the proportion of individuals that won relatively small amounts of money or relatively large amounts of money is quite small. To adjust for this difference in the number of individuals Figure 2.3 shows the average proportion of the respondents that donated per vigntile (a quintile containing five percent of the observations). Every dot hence reflects the average amount won per vigntile against the proportion of respondents that donated. We observe a downward trend in both figures which suggests a negative effect of request size on the probability to donate. The decrease in the probability to donate seems to be limited on this range of request sizes. For instance, in Figure 2.2 we observe that the probability to donate is 63% for request size of 2.5 Euros, and 47% for a request size of 14 Euros. The predicted values in Figure 2.2 and 2.3 suggest that the effect of request size on the probability to give is not linear. We will further investigate this in the sensitivity analysis.

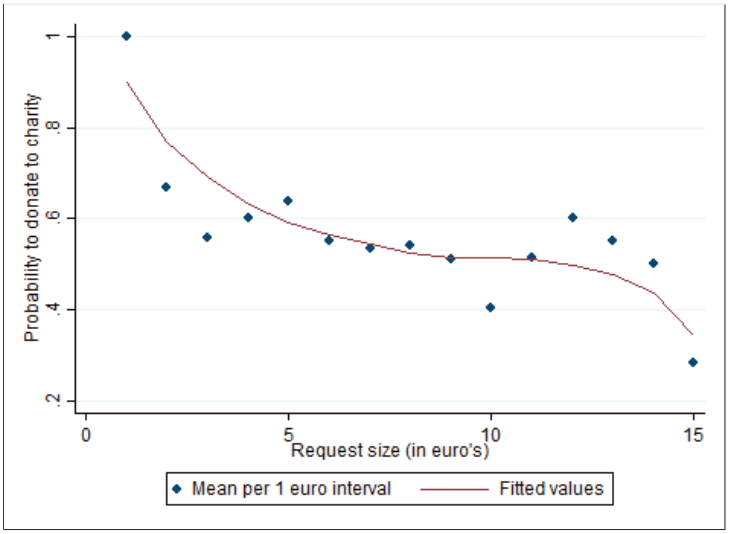


Figure 2.2: Relationship between the request size and the donations to charity (per interval)

Table 2.3 shows the OLS-estimates of the effect of a one euro increase in request size on the probability to donate based on the model as specified in Equation (1). Column (1) shows the results of Equation

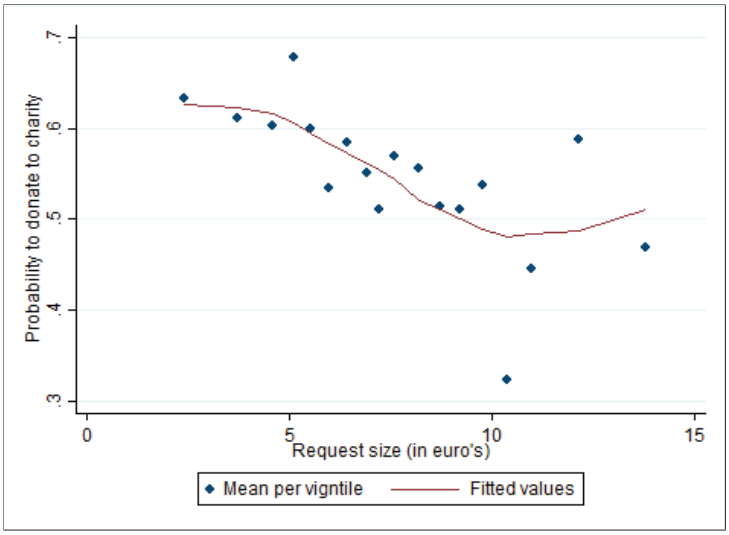


Figure 2.3: Relationship between the request size and the donations to charity (per vigntile)

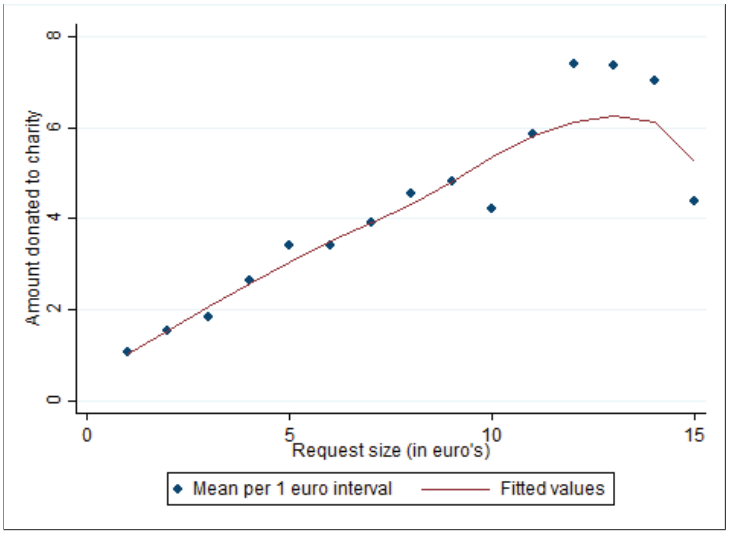


Figure 2.4: Relationship between the request size and the amount donated (per interval)

(2.1) without control variables; column (2) shows the results when controls are included for the survey type and time and date of participation. Column (3) also includes controls for risk, time and public good preferences. Hence, this model also deals with the potential endogeneity that arises from the fact that request size depends on these preferences. Column (4) shows the estimates when we include background characteristics of the respondent and survey details, but no controls for economic preferences. The specification used for column (5) includes the full set of controls.

Table 2.3: OLS estimates of request size on probability to donate

	Probability to donate				
	(1)	(2)	(3)	(4)	(5)
Request size	-0.016***	-0.016***	-0.018***	-0.016***	-0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Ln Request size	-0.116***	-0.115***	-0.122***	-0.115***	-0.121***
	(0.035)	(0.035)	(0.035)	(0.035)	(0.035)
Controls: Background respondent	No	No	No	Yes	Yes
Controls: Gameplay	No	No	Yes	No	Yes
Controls: Dataset, date and time	No	Yes	Yes	Yes	Yes
Observations	1,047	1,047	1,047	1,047	1,047

[1] Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

The estimates in Table 2.3 show that the request size matters for charitable giving. The probability to donate decreases with 1.6 to 1.8 percentage points when the request size increases with one euro. In addition, the log specification shows that an increase of the request size with 100% decreases the probability to donate with 12 percentage points. The estimated effects are robust to the specifications chosen in columns (1) to (5). The estimated effects are very similar when we use a non-linear model like a probit model. Hence, the main message from Table 2.3 is that the request size has a negative effect on the probability to donate but the effect seems not large. An increase of 10 Euros reduces the probability to donate with 16 to 18 percentage points (the probability to donate is 55 percent in the total sample).

2.5.2 Optimal request size

The estimation results in Table 2.3 show that an increase of the request size reduces the probability that individuals will donate to charity. This brings the question which request size would optimize the total amount of donations for the charities. For answering this question we regress the amount that was donated on the request size using a quadratic specification. The estimation results are shown in Table 2.4. Each column in Table 2.4 shows both the estimated effects of request size on the average amount

donated - with a quadratic specification - as well as the predicted optimal request size. Each column shows the estimates of a model with different control variables - similar as in Table 2.3.

Table 2.4: OLS estimates of request size on average amount donated

	Average amount donated				
	(1)	(2)	(3)	(4)	(5)
Request size	0.472**	0.459**	0.549***	0.493**	0.571***
	(0.197)	(0.197)	(0.198)	(0.197)	(0.199)
Request size squared	-0.002	-0.001	-0.008	-0.004	-0.009
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Optimal request size:	105	174	35	67	31
Controls: Background respondent	No	No	No	Yes	Yes
Controls: Gameplay	No	No	Yes	No	Yes
Controls: Dataset, date and time	No	Yes	Yes	Yes	Yes
R-squared	0.090	0.096	0.113	0.107	0.123
Observations	1,047	1,047	1,047	1,047	1,047

Notes: Each column shows the quadratic estimates of request size on average total amount donated as well as the predicted optimal request size (in italic). Standard errors in parentheses (*** p<0.01, ** p<0.05)

Table 2.4 shows that if the request size increases by 1 euro, the average total amount donated by an individual increases by an approximate 47.2 to 57.1 cents (depending on the control variables added to the model). This effect only slowly declines when the request size increases. For each 1 euro increase of the request size the estimated effect decreases by an approximate 0.2 to 0.9 cents. The specifications suggest that the optimal request size lies outside the range of request sizes in our data - which implies that the optimal request size for our dataset would be the highest request size available in our data (15.6 euro). The low statistical significance of 'Request size squared' indicates that a linear model more accurately predicts the amount donated with regard to request sizes up to 15.6 euro. This can also be observed in Figures 2.4 and 2.5 that plot the average donation amount by the request size. The low statistical significance explains why the predicted request sizes of the quadratic model are so far apart when different control variables are added. The optimal request size based on quadratic estimates range from 31 to 174 euro. The most accurate prediction we can make based on our data is 31 euro, as shown in Column (5).

2.5.3 Does the effect of request size depend on mood?

Previous studies, often in laboratory settings, provide evidence that mood influences one's generosity. In this section we investigate whether the effect of request size on charitable giving depends on the mood of the individual that is solicited for a donation. We proxy mood with a variable that indicates whether an individual's favorite soccer team has just won the Dutch championship. Table 2.5 shows the OLS estimates for this regression. Panel 1 shows the estimates of models that include the mood variable 'Won' and various sets of controls. Panel 2 show the estimates when an interaction between mood and request size is included. Each column shows the estimates with different groups of controls. This is similar to Table 2.4 and 2.5 above, except that Column (4) and (5) now also include a variable indicating whether or not the individual is a soccer fan. Note that the estimates are based solely on the second dataset, which explains the smaller number of observations.

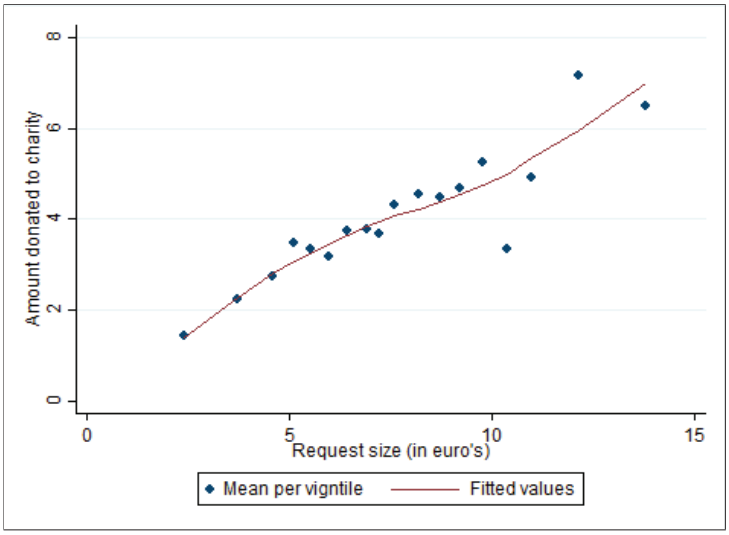


Figure 2.5: Relationship between the request size and the amount donated (per vigntile)

Table 2.5: OLS estimates of positive mood and request size on probability to donate

	Probability to donate				
	(1)	(2)	(3)	(4)	(5)
Panel 1					
Request size	-0.016**	-0.016**	-0.017**	-0.016**	-0.017**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Winning team	0.089**	0.089**	0.087**	0.082	0.088*
	(0.044)	(0.044)	(0.044)	(0.051)	(0.051)
Panel 2					
Request size	-0.017**	-0.017**	-0.018**	-0.017**	-0.018**
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Winning team	0.047	0.043	0.060	0.043	0.074
	(0.124)	(0.124)	(0.124)	(0.129)	(0.129)
Request size * Winning team	0.006	0.006	0.004	0.005	0.002
	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
Controls: Background respondent	No	No	Yes	Yes	No
Controls: Game play	No	Yes	No	Yes	No
Controls: Date and time	No	Yes	Yes	Yes	No
Observations	673	673	673	673	673

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

The estimates of the first panel in Table 6 show that being a fan of the team that won the championship has positive effect on the willingness to donate. The fans of the winning team are on average 9 percentage points more likely to donate than the other individuals in the sample. This effect remains significant at the 10% level when controlling for individual characteristics. This suggests that a good mood contributes to the willingness to donate to charity. The estimates of Panel 2 do not provide evidence for an interaction effect between price and mood. The estimated effects are close to zero and statistically insignificant. This indicates that although a positive mood will increase the probability that someone donates, it does not change the effect of request size. This implies that the optimal request size will not change due to a better mood.

2.6 Sensitivity analysis

In this section we investigate the sensitivity of the findings. First, we investigate whether the effects are similar in the two datasets. Second, we will investigate the linearity of the effect. Third, we will analyse the homogeneity of the effect with respect to gender, education, income and age.

2.6.1 By dataset

We will first investigate whether the effects are similar for the two datasets. Table 2.6 shows the results of this analysis. Columns (1) and (2) show the estimated effect on the probability to donate and columns (3) and (4) show the effect on the amount donated . The table shows both the results of the independent variable 'Request size' as 'Ln Request size' and for both the first as the second dataset. Columns (1) and (3) show the results from models without control variables, whereas the even columns show the estimates when including all three groups of control variables.

Table 2.6: OLS estimates of (ln) request size on probability to donate and amount donated, by dataset

			Probability to donate		Amount donated	
Indep. variable	Dataset	Obs.	(1)	(2)	(3)	(4)
Request size	1	374	-0.017*	-0.020**	0.473***	0.447***
			(0.009)	(0.009)	(0.072)	(0.074)
	2	673	-0.015**	-0.017**	0.422***	0.414***
			(0.007)	(0.007)	(0.053)	(0.054)
Ln Request size	1	374	-0.100*	-0.110*	2.948***	2.816***
			(0.059)	(0.059)	(0.463)	(0.464)
	2	673	-0.122***	-0.127***	2.666***	2.649***
			(0.0440)	(0.0444)	(0.355)	(0.358)
Controls: Background respondent			No	Yes	No	Yes
Controls: Game play			No	Yes	No	Yes
Controls: Data, date and time			No	Yes	No	Yes

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

Table 2.6 shows that estimated effects are very similar for the two datasets. Hence, the previous results are robust to using a sample of soccer fans versus a sample focused on religious behavior of individuals aged forty or above.

2.6.2 Linearity

The Lowess estimates in Figure 2.2 and 2.3 suggest that the effect of request size on the probability to donate is not linear. In the analysis below we will investigate the linearity of the effect estimating the local linear regression coefficients for four different intervals of the request size. The results of this analysis are shown in Table 2.7. Each row shows the estimated effect of a one euro increase in the request in the respective interval. Column (1) and (2) show the estimated effect on the probability to donate and Column (3) and (4) on the amount donated. The odd columns show the estimated results of a model without any added control variables, whereas the results in the even columns are based on the model with all control variables.

Table 2.7: Local linear regression estimates of request size on willingness to donate

			Probability to donate		Amount donated	
	Interval	Obs.	(1)	(2)	(3)	(4)
Request size:	$0 \leq x < 4$	92	-0.119*	-0.154**	0.465**	0.346
			(0.069)	(0.072)	(0.203)	(0.209)
	$4 \leq x < 8$	539	-0.026	-0.028	0.423***	0.413***
			(0.021)	(0.021)	(0.129)	(0.129)
	$8 \leq x < 12$	331	-0.036	-0.036	0.148	0.151
			(0.026)	(0.026)	(0.246)	(0.251)
	$12 \leq x < 16$	85	-0.078	-0.050	-0.491	-0.114
			(0.060)	(0.069)	(0.787)	(0.912)
Controls: Background respondent			No	Yes	No	Yes
Controls: Game play			No	Yes	No	Yes
Controls: Data, date and time			No	Yes	No	Yes

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

The effect of an increase in request size on donations appears to be greatest in the lower interval. Table 2.7 shows that, when the request size is between 0 and 4 euro, a one euro increase in the request size decreases the likelihood to donate with 15 percentage point. The effect in the other intervals lie between 3 and 5 percent. Column (4) shows that the the positive effect of increasing the request size on the total amount of donations decreases when when the request size grow larger. This indicates that the relationship between the two variables is quadratic (which validates the use of a quadratic model as we we did to optimise the request size in Section 2.5).

2.6.3 Heterogeneity

In this last subsection we investigate the heterogeneity of the effects. Table 2.8 shows the results of this analysis. The table is divided in four sections. Each section shows the estimates of Equation (2.1) on split samples that are based on age, gender, education and income, respectively. Column (1) and (2) show the estimates when the probability to donate is used as a dependent variable; Column (3) and (4) show the estimates with the average amount donated as dependent variable. As in the two tables above, the odd columns show the estimates of a model without any control variables, whereas the even columns show the estimates of the model when controlled for game play and individual background and data characteristics. For the first analysis we split our sample into two age groups that are approximately equal in size. The first quartile - age group "Young" - has an average age of 38 years old and a range between 16 and 51 years old; the second quartile contains the oldest half of our sample with an average age of 60 and a range between 52 and 89 years old, because one dataset consists exclusively of men. We also estimated the interaction effect of age (in whole years) and request size. In the third analysis we split up our sample in a group that is higher educated and a a group that is not. The group "High educated" consist of all participants that completed (applied) university education or higher; the group "Low educated" coniststs of all other observations. We also show the interaction effect of educational attainment in years and request size. The last analysis is performed on two groups based on income: those lower or equal to the median income, and those higher than median income. These groups are based on the question on household income, in which the participant was asked to indicate whether his/her household income is below the median, median or above the median income. For 51 of the 1,047 observations this information is unknown. Since we do not have more accurate data on income, the estimated interaction effect is based on the two categories.

Table 2.8: Hetereogeneity analysis: OLS estimates of request size on probability to donate and amount donated, per group

DV:			Probability to donate		Amount donated	
		Obs.	(1)	(2)	(3)	(4)
Analysis by age						
Age:	Young	498	-0.019**	-0.022***	0.402***	0.387***
			(0.008)	(0.008)	(0.063)	(0.063)
	Old	549	-0.014*	-0.017**	0.467***	0.439***
			(0.007)	(0.008)	(0.059)	(0.060)
Request size * Age (in years)		1047	-0.000	-0.000	0.001	0.001
			(0.000)	(0.000)	(0.003)	(0.003)
Analysis by gender						
Gender:	Male	870	-0.014**	-0.017***	0.458***	0.444***
			(0.006)	(0.006)	(0.047)	(0.047)
	Female	177	-0.029**	-0.026*	0.303***	0.324***
			(0.015)	(0.015)	(0.113)	(0.115)
Request size * Gender (female)		1047	-0.015	-0.016	-0.155	-0.164
			(0.016)	(0.016)	(0.127)	(0.126)
Analysis by educational attainment						
Education:	Low educated	571	-0.017**	-0.019**	0.451***	0.441***
			(0.008)	(0.008)	(0.060)	(0.061)
	High educated	476	-0.017**	-0.018**	0.416***	0.403***
			(0.007)	(0.008)	(0.062)	(0.063)
Request size * Education (in years)		1047	-0.000	0.000	-0.000	-0.001
			(0.002)	(0.002)	(0.015)	(0.014)
Analysis by income						
Household income:	Median and below	452	-0.025***	-0.027***	0.347***	0.347***
			(0.008)	(0.009)	(0.066)	(0.068)
	Above median	543	-0.015**	-0.016**	0.460***	0.460***
			(0.007)	(0.007)	(0.060)	(0.060)
Request size * Income (category)		995	0.010	0.008	0.112	0.096
			(0.011)	(0.011)	(0.089)	(0.089)
Controls: Background respondent			No	Yes	No	Yes
Controls: Game play			No	Yes	No	Yes
Controls: Data, date and time			No	Yes	No	Yes

Notes: Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

The results in Table 2.8 show that the request size affects the willingness to donate. Whether we split up up the groups in age, gender, education or income, the effect of an increasing request size on the likelihood to donate is negative and statistically significant for each different group. We find no differences in effect sizes with regard to age and educational attainment. With regard to income group, it appears that an increase in request size has a stronger effect on those with lower income and median incomes than those with higher incomes. Gender also appears to influence the effect of request size. Whereas an increase in the request size with one euro will decrease the proportion of men that will donate by 1.7 percent, it will decrease the proportion of women that donate by 2.6 percent. However neither the estimated interaction between gender and request size, nor the interaction between between income group and request size are statistically significant.

2.7 Conclusion

This study provides evidence that the request size matters in charitable giving. An increase in the request size negatively affects the willingness of an individual to donate to charity. Note that if a donating individual would only care for the quantity of public good provided, the request size should not matter. The individual would only be affected by the real price of a donation - the amount of public good available for the price. Previous studies provide evidence that the real price matters for charitable giving (Among others Feldstein and Taylor (1976); Feldstein and Clotfelter; Clotfelter (1980); Auten et al. (2002). So far the effect the evidence on how request size affects the willingness to donate was limited and the results mixed Edwards and List (2014); Weyant and Smith (1987); Weyant (1984); Doob and McLaughlin (1989); Cialdini and Schroeder (1976). The fact that the request sizes affects the willingness to give is another indication that people gain utility from the act of giving itself, rather than from the public good quantity alone.

We investigated the effect of request size on the willingness to donate, with data from two online experiments. In both experiments participants could win an amount between 0 and 16 euro which they could either keep or donate to charity. The amount was determined by a random procedure and the game play of the respondent. The variation in the earned amount (request size) can be considered to be exogenous as we can control for differences in 'game play.' We used the variation in the amount won to estimate the effect of request size on the willingness to donate. We estimate that an increase in the request size by one euro leads to a decrease in the willingness to donate by 1.8 percent. An 100% increase

in the solicited amount leads to an estimated decrease of 12 percent in proportion of donors. Although the request size clearly affects the likelihood to donate, we showed that fundraising organisations are not better off asking the smallest donations possible, as was suggested in the earlier studies of Weyant and Smith (1987); Weyant (1984); Cialdini and Schroeder (1976).

The optimal request size lies beyond the range of our data. Assuming that the willingness to donate is a quadratic function of request size, we estimate the optimal request size to be approximately 31 euro. There are however two reasons to be careful of generalizing this result. First, the low statistical significance of the quadratic function of request size up to 16 euro - the willingness to donate up to 16 euro is more accurately predicted by a linear model - make very accurate predictions difficult. Second, the participants decided about donating money they just recently won in the experiment. People might more generous with money they just received as part of an experiment as they are with money they have earned in the "real world." (Cherry et al., 2002; Reinstein and Riener, 2012). The optimal request size may hence be lower in a real world setting.

The effect of request size on the willingness to donate is quite consistent over different groups in the population in terms of age and education level. The populations with higher income appear to be less sensitive by an increase in the request size as are those with median and lower incomes. Moreover, women tend to be more sensitive to an increase of the request size than men. This study also provide new evidence that positive mood affects the willingness to donate. We exploit the variation in the mood of the respondents that was induced by the outcome of a Dutch national soccer competition. Part of our data were collected among fans of various competing soccer clubs immediately after the decisive game of the competition. We compare the charitable giving of the fans of the winning club with the charitable giving of the fans of the clubs that did not win the competition. Our analysis shows that the supporters of the winning club were 8 to 9 percentage points more likely to donate to charity, which shows that a positive mood positively affects charitable donations. We find no evidence of an interaction effect between mood and request size.

3

Does showing inequality affect preferences for redistribution? Evidence from a field experiment

3.1 Introduction

Charity organisations invest considerable amounts in fundraising campaigns that show the lives of people living in poverty. These campaigns are driven by the assumption that showing social inequality to potential donors will increase their willingness to donate. However, what motivates people to redistribute income is still poorly understood and so far there is no empirical evidence demonstrating that showing inequality affects the willingness to redistribute. Recent studies provide some evidence that redistributional preferences can be influenced by external factors. For example, studies show associations between national social values and the likelihood that one supports redistributive policies (Corneo and

Grüner, 2002; Alesina and Fuchs-Schündeln, 2007; Alesina and La Ferrara, 2005). Laboratory experiments demonstrate that mood and emotions affect an individual's social preferences (Kirchsteiger et al., 2006; Capra et al., 2010; Capra, 2004; Zak et al., 2007; Barraza and Zak, 2009; Andreoni and Rao, 2011).Social preferences, which determine the extent to which one's utility is affected by the utility of others, are important determinant in the willingness to redistribute.

This study provides evidence that showing inequality affects redistributional behavior, at least in the short run. I use data from a field experiment, carried out on six secondary schools, in which films are used as the main treatment variable. Participants either watched a film about a group of children in Africa or a documentary about young textile workers in China. They also played a series of incentivized economic games and were asked to donate a proportion of their earnings to charity. A random process determined whether the participants were to watch the film before or after playing the games. I use the allocations in the games and the voluntary contributions to charity as main outcome variables. The participants in the treated condition allocate higher amounts in both games, and demanded fairer allocations in the ultimatum game. The fact that the films positively affect generous behavior is also demonstrated by the donations to charity, which increases significantly. There are slight differences between the effects of the two films. The documentary about teenage laborers in China increases the demands in the ultimatum game, where the movie about Africa mainly increases the offers in the games.

This study contributes to the current literature on social preferences by examining the effect of showing inequality on redistributional preferences. To my knowledge, no previous studies have investigated this effect. A second contribution is that the participants in this experiment are between 12 and 16 years old and enrolled in vocational education. This makes them younger and lower educated than most participants in laboratory experiments. A third contribution to the literature is the fact that this study is carried out as a field experiment. Previous studies used lab experiments or observational data.

The remainder of this chapter is organized as follows. Section 3.2 discusses previous research on social preferences and redistribution. Section 3.3 shows the setup of the experiment and explains the econometric model. Section 3.4 shows the demographic characteristics of the participants and the balancing tests for randomization. In section 3.5 I will show the main results. In section 3.6 I investigate the heterogeneity of the effect. Section 3.7 concludes.

3.2 Previous studies

Most democratic societies have redistributional policies in place to protect the less fortunate. Also, many people are willing to make voluntary donations to charity organisations that support the poor. What explains this willingness to redistribute a proportion of income to an unknown other? To start with, it is useful to distinguish between motivations driven by self-interest and those that are more altruistic in nature. A rational selfish agent would support only those redistributional policies that make him better off. Redistributional policies thus would have more support among the poor than among the rich. High inequality in democratic societies will increase the proportion of people that are poor and hence increase the support for redistributional policies (Ravallion and Lokshin, 2000). Negative externalities of high income inequality can be another motivation for people to support redistributional policies. High income inequality might cause an increase in crime, affecting the safety of others in the society. A third motivation is that redistributive policies can have characteristics of a social insurance program, protecting people from the risk of falling into poverty. Piketty (1995); Cruces et al. (2013) show that someone's perceived position in the society and perceived social mobility strongly affect the support for redistribution. Those who have more negative perspectives on their own position will more often support redistributive policies.

One might argue that global income inequality has negative externalities (war, crime etc.), but these selfish motivations do not fully explain why so many people are willing to donate to charity organisations that target worldwide inequality. Voluntary donations are affected by, what Andreoni (1989) refers to as, 'impure altruism' or 'warm glow giving. Andreoni (1988) shows empirically that in cases where governmental public spending increases, the donations to charity organisations do not crowd out, which implies that voluntary giving can be - at least partially - explained by impure altruistic motivations. The act of giving may positively affect one's self-perception, his social status or simply provides a good feeling.

Self-interest alone cannot explain why people support redistribution. If an individuals preferences are partially influenced by the (perceived) utility of - known or unknown- others, we speak of social preferences. Those with higher inequality aversion derive more negative utility from social inequality. They will hence be more willing to redistribute part of their income to decrease social inequalities (Fehr and Schmidt, 2006). In the past decades a number of studies have been carried out to investigate which factors influence social preferences. Alesina and Fuchs-Schündeln (2007), Alesina and La Ferrara (2005)

and Bénabou and Tirole (2006) demonstrate that personal and collective beliefs about the causes for poverty influence the extent to which one supports redistribution. Those that believe that poverty or wealth is to a large extent caused by factors outside one's own control have a stronger tendency to support redistributive policies than those that believe that one's wealth is caused by one's own personal efforts and choices. Moreover, personal or public values about fairness influence these preferences (Corneo and Grüner, 2002; Fong, 2001).

Most of the empirical evidence on social preferences is obtained in laboratory experiments in which the outcomes of economic games are used as dependent variables. In the dictator game, first used in the study of Kahneman et al. (1986), one player has to decide how to allocate a fixed amount of money between himself and an unknown other player. Rational choice theory predicts that - if the player acts solely out of self-interest - he would decide to keep the whole amount to himself. Any giving in the dictator game can be considered as altruistic giving. Empirical studies show that the large majority of the people allocate some money to the other anonymous player (Kolm, 2006; Fehr and Schmidt, 2006). The ultimatum game, first applied by Güth et al. (1982) is similar to the dictator game. The first player, the proposer, is asked to divide a certain amount of money between himself and the other player. The second player, the receiver, then has to decide whether he accepts the offer. If the receiver accepts the offer the amount is allocated accordingly; if the offer is rejected neither player will receive anything. Under the standard assumptions that both proposer and responder are rational, the outcome of this game would be that the receiver would accept any positive amount offered. Hence the proposer would allocate the smallest amount possible. The empirical evidence however shows that most offers are between 40 and 50 percent of the amount (see Kolm (2006) for an extensive overview). Many proposers anticipate that low offers will be rejected with high probability . Fairness and equity concerns play a strong role in the decision making of both players in the ultimatum game (Hoffmann and Tee, 2006). Forsythe et al. (1994) shows that the offers in dictator games are on average lower than offers in the ultimatum game, but that a strong correlation also exists between the size of the allocations.

The use of economic games as a measure for social preferences is not uncontested; some scholars point out that the behaviour during the games may be an artifact of the experimental design. In both games, the proposer has only the option to give. List (2007) and Bardsley (2008) show that when an additional option to take from their opponent was given, many proposers chose to take money from their opponent instead of giving it. Dana et al. (2006) and Broberg et al. (2007) show that altruism may not be the core motive of giving in the dictator game. Rather, giving in the dictator game may be explained by

the participant's fear of hurting the expectations of their opponents. Andreoni et al. (2008) also point out that the lack of social distance between the researcher and the receiver - although the receiver is not known by the proposer he is known by the researcher - might partially explain the generous offers. The offers may not be driven by generosity alone, but may be fueled by social expectations. According to Frohlich et al. (2001) also the mere fact that the game is considered as a game by participants may influence decision making. Eckel and Grossman (1996) compare giving in the dictator game with giving to charity, and find that people are more willing to donate to a "deserving" charity than to an anonymous other player.

A number of experiments exists that examine the effect of mood on allocation in games, using film clips as mood inducers. These short film clips can subsequently be considered as an exogenous shock for mood (Bradley et al., 1996). Although the films should only affect the mood of the participants it is not completely clear how these studies avoid affecting other factors that may influence economic preferences. Ifcher and Zarghamee (2011) explores the effect of mood on time preference using short movie clips that put participants either in a happy or a neutral mood. They find that a positive mood significantly reduces time preference over money. Studies with a comparable setup find that good mood - evoked by film clips - affects giving behavior in gift-exchange games (Kirchsteiger et al., 2006). Capra et al. (2010) find that good mood leads to more generous bidding in a bidding game. Charness and Grosskopf (2001) show that a correlation exists between unhappiness and the willingness to lower another persons payoff. Capra (2004) further investigates this relationship and finds that mood affects giving in one-shot dictator and ultimatum games. Good mood has a positive effect on the amount that the player donates. Hence happiness or positive emotions may increase altruistic behavior. Zak et al. (2007) and Barraza and Zak (2009) demonstrate that empathy strongly affects social preferences. Participants that received an exogenous shock of the 'empathy' hormone Oxytocin gave significantly more during dictator and ultimatum games. Andreoni and Rao (2011) draw similar conclusions about this relationship.

Empirical studies show that age plays a role in social preferences. Benenson et al. (2007) studies allocations in a dictator game with children in the age from 4-9 years old and finds that age is an important factor for generous behavior. Older children propose significantly larger offers. The studies of Harbaugh et al. (2003) and Murnighan and Saxon (1998) investigate allocations in the ultimatum and dictator games with children in the age of 7 to 18 years old and derive the same conclusion. Younger children appear to be more strategic: they make smaller offers in the dictator game and make and accept smaller offers in the ultimatum game than older children. Their explanation is that altruistic behavior

is determined by norms and values and hence will be more developed at a later age. Hoffmann and Tee (2006) however find that adolescents make significantly larger offers in the ultimatum game and are less likely to reject offers than adults. Andreoni and Vesterlund (2001) demonstrate that gender also plays a role in social preferences. In general women tend to give higher offers than men in dictator games.

In sum, the willingness to redistribute is partially explained by social preferences. There is empirical evidence that shows that social preferences are affected by external factors, but, to my knowledge no studies exists that have explored the effect of showing inequality on these preferences.

3.3 Experimental design

3.3.1 Experimental setup

This study investigates whether social preferences can be influenced by showing films about social inequality. I use data from a field experiment that I carried out in cooperation with six secondary schools in the Netherlands. The participants, all pupils of these secondary schools, were to watch a film about social inequality and played a series of incentivized games of which the allocations would reflect their social preferences. The field experiment was carried out over nine days in the period between March and July 2013. The list of experiments in chronological order can be found in Appendix 3.A.

The schools voluntary participated in the field experiment under the condition that all pupils would eventually receive the same program. Under the ideal experimental conditions, a random process would determine whether the participants would watch a film or not. The impact of the film can then be estimated by taking the difference between the outcomes of those that watched the films and those who didn't. Since the schools were only willing to participate if all participants would eventually watch the film, I had to make minor changes to this setup. In the setup I used during this experiment all participants would eventually watch a film, but the timing of watching is randomly determined. The setup, which I will describe below, is valid under the assumption that changes in the outcome are not affected by the small changes over time.

Per school the participants were randomly allocated to a classroom where they watched the films, played the games, and filled out a questionnaire. The order in which these activities took place differed per classroom. About half of the participants played the games *after* watching the film, whereas the other half played the games *before* they watched the films. Since the participants were randomly allocated to

these classrooms, I can estimate the effect of the film on the allocations in the games by comparing the outcomes of the two groups.

The experiment was carried out in close cooperation with the staff of the secondary schools. Teachers informed the pupils about the experiment a few days in advance. This information was rather limited: The pupils were told that they would attend a class different from normal and that a film was involved. On the day of the experiment they were given more detailed information about the process. After the experiment, the participants were also informed about the objective of the research. All participants were between 12 and 16 years old at the time of the experiment and enrolled in the lowest track of Dutch secondary education VMBO (abbriviation of *voorbereidend middelbaar beroepsonderwijs*, literally: preparatory middle-level vocational education).[1]

The treatment consists of watching a film that displays a case of social inequality. I used two films during this experiment. The first film is the 2005 non-fiction film 'China Blue', a 55-minute documentary about the difficult labour conditions of factory workers in the Chinese textile industry. The documentary focuses on a group of female labourers, all teenagers, that have to cope with long working hours, low salaries and unsafe working conditions. The second film is the 2010 movie 'Africa United,' a 90-minute fictional film about a group of African children travelling from Rwanda to South-Africa to attend the World Cup event. The movie discusses topics as child prostitution, HIV/Aids and child soldiers. Despite these topics the overall tone of the movie is optimistic, with friendship as its central theme. The films were selected by the non-profit organisation Movies that Matter, a subsidiary of Amnesty International, specialised in human rights education with films. The children were supervised while watching the films to make sure they would not interact. Participants watched only one of the two films. Only one film was used per school.

The procedures during the session of the economic games were as follows. Before the start of the games, the instructors informed the participants about the rules. When the rules of the games are clear to all, an instructor randomly assigned participants a role (proposer or receiver) and formed couples. The identity of the opponent was unknown to the participants during the experiment, nor was it revealed after the experiment. During the first game, the dictator game, the participants that took the role of proposer were informed that they would receive 1 euro, and that their unknown opponent would receive nothing.

[1] In the Dutch secondary education system, pupils have to follow one of four tracks in secondary education after they complete primary education. The track chosen is usually based on the score on the standardized national test (Cito) that the pupils take in the last year of primary school. The VMBO track is the track for pupils with the lowest cognitive skills. It combines vocational training with theoretical education and lasts four years. The VMBO track is again divided into four different levels that differ in the ratio of practical vocational training and theoretical education. The participants are pupils in the first two years of this track, from all four levels.

The participant was then given the opportunity to allocate a certain amount of this euro (in partitions of ten cents) to the other player. The participants in the role of receiver could not act, but were asked what he or she expected to receive from the other player. The participants did not receive tokens or coins yet, but were told to write the allocations down on a sheet of paper. The answers remained anonymous and participants were unaware of the roles of other classmates. Participants were not informed about the outcomes of the dictator game before the start of the second game. The second game played was a version of the ultimatum game. Every participant played one turn in the role of proposer and one turn in the role of receiver. Participants had the same opponent as in the previous game. Participants in the role of proposer were informed that they would receive 1 euro and could allocate a certain amount to the second player. The participants in the role of receiver were asked what the minimum offer was to would be willing to accept. If the amount allocated by the proposer was less than the minimum amount demanded by the receiver, neither player would receive anything. If the amount was equal or higher the money was allocated accordingly.

After finishing the games, the participants filled in in a questionnaire. The questionnaire contained a request to give a proportion of the earnings from the games to charity. The participants could decide to either keep the whole amount to themselves, give half of the amount to charity or donate the full amount. At this point the participants were not yet informed about the outcomes of the games and hence did not know the size of the amount they earned. Those who choose to give some of their earnings to charity could choose between a number of causes. They had the choice between "a charity that helps children with a life threatening disease," "a charity that supports animals and the environment" and "a charity that supports poor people in developing countries." After all experiments were conducted the money collected was donated to three actual charity organisations.

3.3.2 Econometric model

During the experiment participants were randomly allocated to a condition in which they received the treatment - the film - either before or after they played the economic games. This means I have data of participants that just watched the film, and participants that did not (yet) watch the film. I am interested in the effect of the film on the allocations in the games and to the charity organisation. Comparing the outcomes of the two groups, would give an unbiased estimate of the effect of the film. However, neither the schools, nor the films, nor the day of the experiment were randomly selected. Schools participated on a voluntary basis. The film was selected prior the experiment in cooperation with the teachers of the

school that were involved. Only one film was used per school. The day (or days) on which the experiment would take place also depended on the availability of the teachers and pupils of the school. I will hence have to control for these factors by adding school-fixed effects and a variable indicating the day of the experiment. Hence I will apply the following model:

$$Y_{is} = \phi_s + \beta_1 d_i + \beta_2 T_s + BX_i + \varepsilon_i \tag{3.1}$$

In which Y_i is the dependent variable of interest; d_i is a dummy variable equal to 1 if individual i watched one of the films before the games or 0 otherwise; and X_i is a vector of child specific covariates. School-fixed effects are denoted by ϕ_s. T_s is a variable equal to 1 if the data was collected on the first day, or equal to 2 if it was collected during the second day at the particular school. Note that in most schools the experiment was carried out in one day only (see Annex 3.A for an overview of the experimental days per school). β_1 is the parameter of interest.

To examine the difference between the effects of the two films I apply the model below:

$$Y_{is} = \phi_s + \beta_1 Africa_i + \beta_2 China_i + \beta_3 T_s + BX_i + \varepsilon_i \tag{3.2}$$

In which $Africa_i$ is a dummy variable equal to 1 if the participant watched the film 'Africa United' and $China_i$ a dummy variable equal to 1 if the participant watched the documentary 'China Blue' and 0 otherwise. Note that the $Africa_i$ and $China_i$ are mutually exclusive.

3.4 Data

The experiment yielded a dataset containing 541 observations, collected at 6 locations over 9 days. Of the participants 229 (42%) were in the control condition; 156 (29%) watched the film about the conditions of textile workers in China; and 156 (29%) watched the film about children is circumstances of poverty in Africa.

As a first step, I investigate whether the pupils in the control and treatment condition are similar on observed characteristics. Table 3.1 shows the composition of groups of participants in both treatment and control condition and investigates the differences. The first column of Table 3.1 shows the demographic characteristics of the group of participants as a whole; the second and third column show the characteristics of participants in respectively the control and treatment condition. The last column

shows the estimated differences between the control and treated groups.

Table 3.1: Demographic characteristics of treatment and control groups

	(1) All $E[X]$	(2) Control $E[X\|d=0]$	(3) Treated $E[X\|d=1]$	(4) Estimated difference $E[X\|d=1]-E[X\|d=0]$
Age	13.94	13.96	13.93	-0.04
				(0.07)
Female	0.52	0.49	0.54	0.04
				(0.04)
Immigrant	0.19	0.19	0.19	0.00
				(0.03)
Home situation:				
Lives with both parents	0.74	0.72	0.75	0.03
				(0.04)
Lives with father only	0.02	0.02	0.02	0.00
				(0.01)
Lives with mother only	0.15	0.15	0.14	0.00
				(0.03)
Unknown	0.09	0.10	0.09	-0.02
				(0.03)
Cito score	532.6	533.4	531.8	-1.6
				(1.13)
Grade 1	0.45	0.45	0.46	0.01
				(0.04)
Grade 2	0.50	0.51	0.50	-0.01
				(0.04)
Grade 3	0.04	0.05	0.04	-0.01
				(0.02)
Level 1 (Lowest)	0.07	0.06	0.08	0.02
				(0.02)
Level 2	0.15	0.15	0.15	0.01
				(0.03)
Level 3	0.09	0.09	0.08	-0.01
				(0.02)
Level 4 (Highest)	0.69	0.70	0.69	-0.01
				(0.04)
Observations	541	229	312	

Notes: Variable Cito score is based on only 172 observations (Cito test is not compulsory). Two-tailed t-tests are used to determine whether the difference between the treatment and control condition are significant. Standard errors in parentheses. Statistical significance at the 1, 5, and 10 percent levels is indicated by ***, **, and *, respectively

The participants are on average just under 14 years old. They are enrolled in one of the first three classes of VMBO education, of which 69% is enrolled in the highest level.[2] The majority, 96 percent, attends either class 1 or 2. Gender is quite equally distributed in the sample, with 52 percent of the participants being female. Column (4) in Table 3.1 shows that the participants in the treatment and

[2]see footnote 1 for a description of tracks and levels in Dutch secondary school system

control condition do not vary on any of the observed characteristics, indicating that the randomization of the treatment is successful. I also regressed each control variable on the treatment while adding school-fixed effects, which did not show any significant differences on the observed characteristics.

The dependent variables

The main set of outcome variables are measures for social preferences based on the outcomes of the economic games. I will use four outcome variables: 1) The amount offered by the proposer in the dictator game; 2) The amount offered by the proposer in the ultimatum game; 3) The minimum amount demanded by the receiver in the ultimatum game; and 4)The proportion of total earnings donated to charity. All four outcome variables are discrete measures. The three variables based on the games have a range of 0 to 1 with intervals of .10, the outcome variable with regard to charity has a range 0 to 1 with intervals of .5. Before investigating the effect on the main outcome variables I conduct an induction test to determine whether the treatment (watching the film) is successful. I investigate this by using measures for inequality awareness and emotional attitudes (empathy, trust and happiness). The measure for inequality awareness is based on a series of statements about quality of life in China or Africa (depending on the film) and poverty in general. The participants were asked to give their opinion on these statements on a five-points Likert scale as part of the questionnaire. The statements and more background information on this measure can be found in Appendix 3.B. The measure for empathy is based on the empathy scales of (Davis, 1983), Bryant (1982) and Warden and Mackinnon (2003). Appendix 3.C provides more information about this scale. The measure for trust is based on questions from the General Social Survey (see Appendix 3.D). Happiness is based on the answer to the question "Generally speaking, how happy are you with your life at this moment?" on a scale from 0-10. This is a common approach to measure subjective wellbeing (Kahneman and Krueger, 2006).

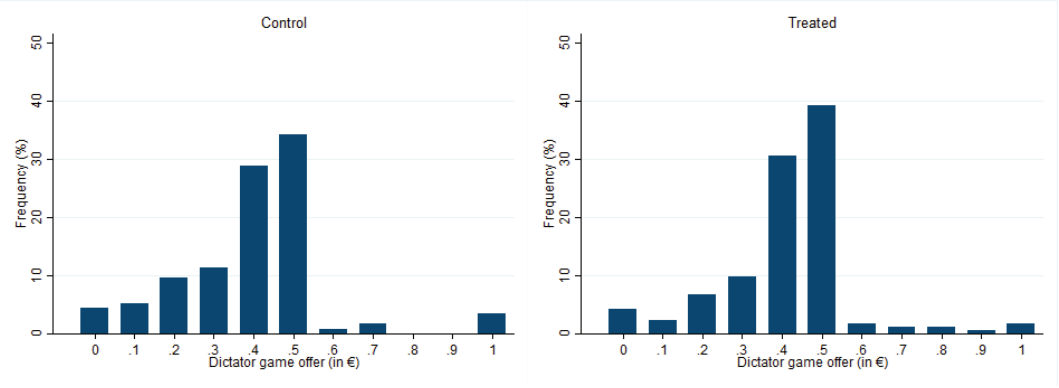
Table 3.2 below shows the allocations in the games and the summary statistics for poverty awareness and emotional attitudes of the participants. The first column shows the summary statistics of all participants. Columns (2) and (3) show the estimates of the participants in the control and treated condition respectively. The fourth column shows the difference between the two groups.

Table 3.2: Summary statistics of outcome variables

	(1) All $E[Y]$	(2) Control $E[Y\|d=0]$	(3) Treated $E[Y\|d=1]$	(4) Estimated difference $E[Y\|d=1]-E[Y\|d=0]$
Amount allocated in dictator game	0.41	0.40	0.42	0.02
				(0.02)
Amount allocated in ultimatum game	0.42	0.41	0.43	0.02*
				(0.01)
Minimum amount demanded in ultimatum game	0.38	0.36	0.39	0.03**
				(0.01)
Proportion of total earnings donated to charity	0.26	0.24	0.27	0.03
				(0.02)
Awareness score (0-1)	0.58	0.56	0.59	0.03**
				(0.01)
Empathy score (0-1)	0.66	0.65	0.67	0.02**
				(0.01)
Social trust score (0-1)	0.66	0.64	0.68	0.04*
				(0.02)
Self-reported happiness (0-10)	7.54	7.43	7.63	0.20
				(0.12)
Observations	541	229	312	

Notes: Dictator game contains less observations: n=277 (of which 163 in the treated and 114 in the control condition). Two-tailed t-tests are used to determine whether the difference between the treatment and control condition are significant. Standard errors in parentheses. Statistical significance at the 1, 5, and 10 percent levels is indicated by ***, **, and *, respectively

Table 3.2 shows that the allocations are higher in the treatment than in the control group, but only the allocations in the ultimatum game are statistically significant. The group that watched the film allocated on average 2 cents more in this game, but also demanded 3 cents more. Table 3.2 also shows that the measures for positive awareness, empathy and social trust were all positively affected by the film. I will further explore these measures in section 3.6.

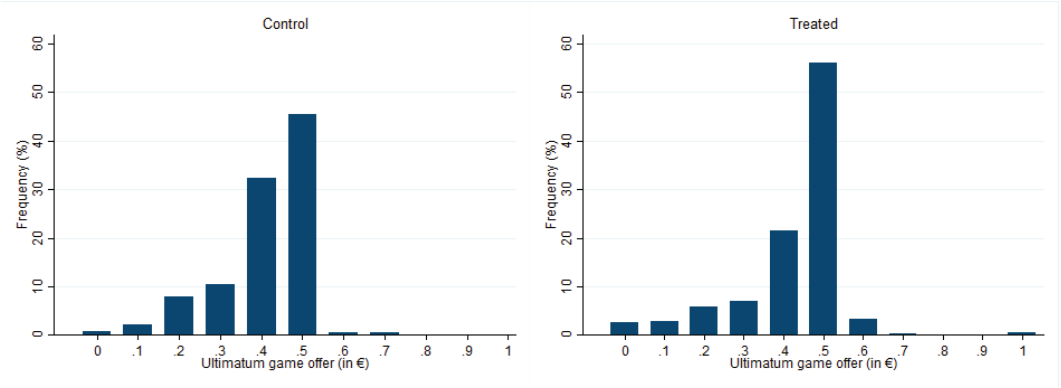


(a) Offers in dictator game (Control)

(b) Offers in dictator game (Treated)

Figure 3.1: Distribution of allocations in dictator game

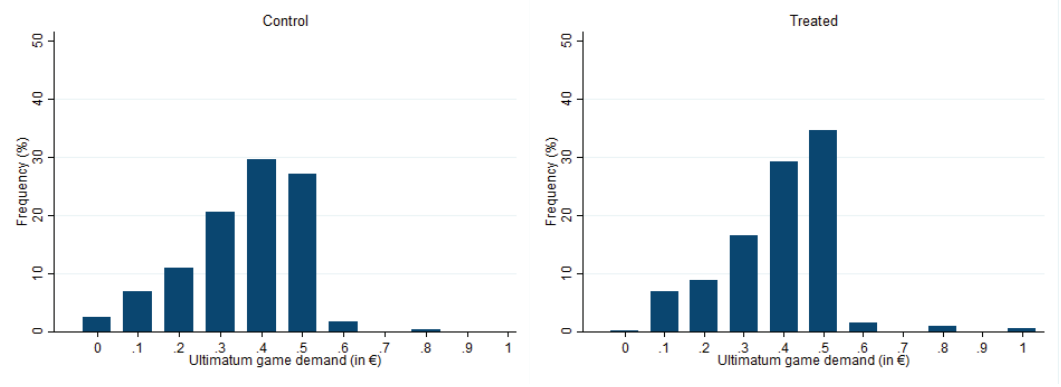
Figures 3.1 - 3.3 show the distributions of allocations in the games by treatment status (treatment and control). The offers of the proposer in the dictator game are shown in Figure 3.1. It shows that the vast majority of participants in the dictator game donates a proportion to the other player, with an average of 41 and a modal offer of 50 cents. Only 12 out of the 277 participants (4.3 percent) kept the full amount. Seven participants (2.5 percent) - donated the full amount to the other player. Approximately 6.5 percent of the participants donated more than half of the amount they received. The average offer is slightly higher in the treatment group. Since the dictator game was played once, only half of the total participants played the role of the proposer.



(a) Offers in ultimatum game (Control)

(b) Offers in ultimatum game (Treated)

Figure 3.2: Distribution of allocations in ultimatum game



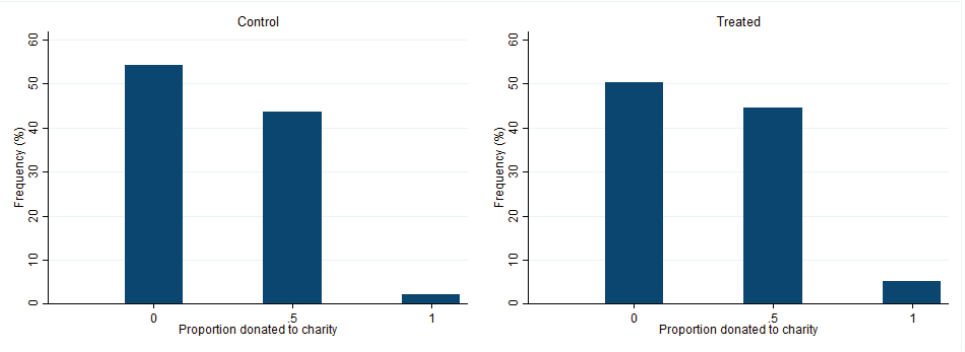
(a) Demanded in ultimatum game (Control)

(b) Demanded in ultimatum game (Treated)

Figure 3.3: Minimum demanded amount of responder in ultimatum game

Figure 3.2 shows the offers of the proposer during the ultimatum game. The average offer during this game is 42 cents, with a modal offer of 50 cents. More than half of the participants decide to split the amount equally with the other player: 51.6 percent offer an amount of 50 cents. The allocations in the treatment group are higher than the those in the control group. A larger proportion of the participants (56% versus 45%) opts for the fair allocation of 50 cents. Figure 3.3 shows the demanded offers of the receivers during the ultimatum game. The average demanded offer is 38 cents. There are some significant differences in demands between the treated and control condition. The average demand is 3 cents higher

in the treated condition. The modal offer changes from 40 cents in the control to 50 cents in the treatment condition.



(a) Donated to charity (Control) (b) Donated to charity (Treated)

Figure 3.4: Distribution of donations to charity

Figure 3.4 shows the proportion of the total earnings that participants donated to charity. Participants could either donate nothing, 50 percent or 100 percent of their earnings in the games to charity. The generous allocations during the dictator game are reflected in the donations to charity. Almost half of the participants (48 percent) donates a proportion of the earned amount to charity: 44 percent chooses to donate half, and 4 percent donates 100 percent of the earned amount. Overall 26 percent of the total earnings are donated to charity. Figure 3.4 clearly shows an increase of donations in the treated group. The proportion that decides to donate to charity increases from 50 to 54 percent. The proportion that donate the full amount to charity increases from 2 to 5 percent.

Independent and other variables

The independent variable of interest in Equation (3.1) is d_i, a dummy variable indicating whether the participant was in the treated group - and hence watched a film about poverty - or in the control group. Equation (3.2 differentiates between the treatment with the film 'Africa United' and the film 'China Blue,' indicated by variables $FilmA_i$ and $FilmB_i$ respectively. The model is specified with the following child-specific controls: gender, age, ethnicity (immigrant), grade, track and living situation. 'Level' indicates the track of Lower Vocational Training participant i is enrolled in. This is an indication of the cognitive skills of the student, with track 1 being the lowest of the four.

3.5 Main estimation results

This section shows the main results of this study and is divided in three subsections. The first subsection shows the results of the induction test. Here I investigate the extent to which the films affect poverty awareness and emotional attitudes. In the second subsection I will show the main estimates of the impact of the films on social preferences for redistribution - using outcomes of the games and the donations to charity as main dependent variables. The third subsection explores the difference in effect on social preferences between the two films.

3.5.1 Induction test

This section explores the effect of of the film of poverty awareness and emotions. Table 3.3 shows the responses on the six statements that are used as measures for inequality awareness. The first four statements are statements regarding the quality of life in either China or Africa (depending on the film). The last two statements are statements about poverty worldwide. Participants rank their l give a score between 1 and 5 on a Likert scale, of which 1 signifies 'strongly disagrees' and 5 'strongly agrees.' Columns (1) and (2) show the average scores for participants in respectively the treated and the control condition of the schools that watched the Africa film. Column (3) shows the average difference between participants in the two conditions. Column (4) - (6) shows the scores for the schools that watched the China film.

Table 3.3: Perception of inequality by treatment status

		Panel 1: Africa film			Panel 2: China film		
Nr	Statement	Control	Treated	Estimated difference	Control	Treated	Estimated difference
1	Africa/China is poor	4.22	3.85	-0.37***	2.58	2.99	0.41***
				(0.12)			(0.14)
2	Africa/China is modern	2.01	2.19	0.18	3.43	2.76	-0.67***
				(0.11)			(0.12)
3	Most children in Africa/China go to school	1.97	2.10	0.13	3.33	2.32	-1.01***
				(0.10)			(0.13)
4	The situation in Africa/China is getting better everyday	3.12	3.08	-0.05	3.91	3.49	-0.43***
				(0.05)			(0.12)
5	There is less and less poverty in the world	3.02	2.92	-0.10	2.84	2.77	-0.07
				(0.12)			(0.13)
6	Children in developing countries are likely to have a better future	2.41	2.37	-0.03	2.36	2.27	-0.07
				(0.15)			(0.14)

Notes: Estimates based on two-sided t-tests. Standard errors in parentheses. (*** p<0.01, ** p<0.05, * p<0.1)

The results in Table 3.3 indicate that the participant's perception is affected by the film about China. The participants that watched this film regard China as poorer and less modern than those that did not. They also think it less likely that most children in China go to school. There is no significant effect on the statements with regard to the overall situation in the world. Whereas the China film significantly enhances the perspective of the targeted geographical area as poor; the Africa film achieves the opposite. After watching the film the participants have a more positive view with regard to poverty in Africa than initially. The initial score on statement 1 is relatively high, indicating that the portrayal of Africa in the film is less negative than the perspective the participants initially had.

As a second step I investigate the extent to which the film affects emotional attitudes. Table 3.4 shows the estimations of regressing respectively empathy, trust and happiness on the treatment. The table also includes the estimations on the measure for poverty awareness (Columns (1) and (2)). Columns (3) and (4) show the estimates on empathy, Columns (5) and (6) on social trust and Columns (7) and (8) on self-reported happiness. The top panel shows the estimates of Equation (3.1), the bottom panel the estimates of Equation (3.2) - thus differentiating between films. All estimates in Table 3.4 include school-fixed effects. The even columns contain the estimates of the model with child-specific controls.

Table 3.4: Induction test: estimates of the effects on awareness and attitude

	Awareness		Empathy		Trust		Happiness	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel 1								
Film	0.040***	0.040***	0.022**	0.019**	0.043*	0.046*	0.207	0.207*
	(0.012)	(0.012)	(0.010)	(0.009)	(0.024)	(0.025)	(0.127)	(0.122)
Panel 2								
Film Africa	-0.018	-0.015	0.038***	0.035***	0.090***	0.091***	0.295	0.274
	(0.016)	(0.016)	(0.013)	(0.012)	(0.033)	(0.034)	(0.179)	(0.170)
Film China	0.108***	0.105***	0.002	-0.005	-0.021	-0.016	0.026	0.069
	(0.018)	(0.018)	(0.015)	(0.014)	(0.040)	(0.039)	(0.203)	(0.204)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
School fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	541	541	541	541	541	541	541	541

Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

The top panel shows that the films do not merely inform the participants about poverty, but also affect their moods. In line with the findings from Table 3.3, the results in the bottom panel of Table 3.4 show heterogenous effects of the films. The China film has a strong effect on awareness, but does not appear to affect any of the other emotional attitudes. The Africa film hardly appears to affect poverty awareness, but has a positive effect on empathy and trust. One of the explanations for this may be that the Africa film is a fictional story that intends to build up an empathic bond between the main characters and the viewers.

3.5.2 Effect on social preferences

This section discusses how the films affect the allocations in the games and the donations to charity. Table 3.5 shows the OLS-estimates of this analysis. Columns (1) and (2) show the estimates on the allocations in the dictator game. Columns (3) to (6) show the estimates when using the outcomes of the ultimatum game as dependent variable: Columns (3) and (4) show the estimates of the effect on the allocations of the proposer, and Columns (5) and (6) on the demanded amount of the receiver in the ultimatum game. Columns (7) and (8) show the estimates on the donations to charity. The odd columns show the estimates of the school fixed effect model without child-specific covariates, the even columns the estimates when those controls are added.

Table 3.5: OLS estimates of the effect film on social preferences (Equation (3.1))

	Dictator game		Ultimatum game				Charity	
	Player 1		Player 1		Player 2		Proportion donated	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Film	0.048**	0.048**	0.032***	0.032***	0.025**	0.024**	0.061**	0.053**
	(0.021)	(0.021)	(0.011)	(0.011)	(0.012)	(0.012)	(0.026)	(0.026)
Constant	0.349***	0.166	0.383***	0.247***	0.387***	0.418***	0.121**	0.226
	(0.033)	(0.135)	(0.018)	(0.083)	(0.018)	(0.077)	(0.058)	(0.172)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
School fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277	277	541	541	541	541	541	541
R-squared	0.181	0.199	0.094	0.104	0.072	0.090	0.045	0.098

Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

The three measures for altruistic behavior - the allocations in the two games and the donations to charity - are all affected by the films. The participants that take the role of proposer in the dictator game allocate 5 cents more to the other player after watching a film. Compared to the the baseline of 41 cents (see Column (2) of Table 3.2), this implies an estimated effect of 12.5 percent. The film positively affects the allocations in the ultimatum game. Proposers in the ultimatum game would give approximately 3 cents more than those who did not watch the film (7.5 percent). Columns (5) and (6) show that the film also affects the minimum amount demanded by the receiver. Those who watched the film demand a higher offer of approximately 2.5 cents (7 percent). The proportion donated to charity is positively affected by the films: the participants that watched the films give approximately 4 percentage points more than those who did not. Adding individual controls merely affects the estimates.

In sum, showing a film on inequality affects social preferences for redistribution.

3.5.3 Analysis by film

The analysis above focused on the effect of watching either one of the two films. Since the two films that were used in the experiment were quite different in style (documentary versus movie) and content (Chinese labourers versus African children) the films may have different effects. In this section I will investigate how the effects on social preferences differ between the two films. Table 3.6 shows the OLS estimates of Equation (3.2) in which a distinction is made between the two films. The structure is identical to Table 3.5 above.

Table 3.6: OLS estimates of the effect of film on social preferences, by film (Equation (3.2))

	Dictator game		Ultimatum game				Charity	
	Player 1		Player 1		Player 2		Proportion donated	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Film Africa	0.023	0.024	0.035**	0.034**	-0.000	-0.002	0.084**	0.077**
	(0.027)	(0.028)	(0.015)	(0.015)	(0.017)	(0.017)	(0.033)	(0.032)
Film China	0.079**	0.078**	0.029*	0.029*	0.060***	0.061***	0.028	0.018
	(0.032)	(0.031)	(0.015)	(0.015)	(0.018)	(0.018)	(0.042)	(0.042)
Constant	0.333***	0.145	0.385***	0.249***	0.369***	0.392***	0.137**	0.250
	(0.033)	(0.137)	(0.018)	(0.083)	(0.019)	(0.077)	(0.059)	(0.172)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
School fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277	277	541	541	541	541	541	541
R-squared	0.187	0.204	0.094	0.104	0.081	0.101	0.047	0.101

[1] Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

Both films positively affect the offers in the games and the donations to charity. Those who watched one of the films increase their offer in the ultimatum game with 3 to 4 cents. The extent to which the films affect the offers in the dictator game and the donations to charity vary between the films, but the effects are clearly positive. There is however a significant difference in how the films affect the response of the receiver in the ultimatum game. Pupils that watched the China film demand on average 6 cents more that those pupils that did not, an increase of almost 17 percent, whereas I find no effect of the Africa film. The offer demanded by the receiver is considered to be strongly influenced by notions of fairness. It hence appears that the China film thus clearly affects the participant's sense of fairness, which the Africa film does not.

3.6 Hetereogenity Analysis

In this section I will investigate whether there are heterogonous effects that may lead to re-interpretation of the results above. I will focus in particular on the potential influential factors age and gender. Earlier research shows that age and gender are important determinants for social preferences. When children grow older, they will exhibit more social behavior (Benenson et al., 2007; Harbaugh et al., 2003; Murnighan and Saxon, 1998) and women are generally more generous then men (Andreoni and Vesterlund, 2001). The main estimates in Table 3.5 take these factors into account by adding them as control variables in the regression - shown in the even columns. It is also not unlikely that films have different effects when children get older; or differently affect boys than girls. Table 3.7 and Table 3.8 shows the estimates of the effect of film on social preferences when interactions between respectively film and gender, and film and age are added to the model.

Table 3.7: OLS estimates of treatment interaction effects (gender)

	Dictator game		Ultimatum game				Charity	
	Player 1		Player 1		Player 2		Proportion donated	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Film	0.029	0.030	0.037**	0.037**	0.031*	0.032*	0.034	0.035
	(0.032)	(0.032)	(0.016)	(0.016)	(0.018)	(0.018)	(0.036)	(0.036)
Female	-0.011	-0.017	0.014	0.016	0.009	0.010	0.102***	0.103***
	(0.033)	(0.032)	(0.014)	(0.014)	(0.018)	(0.018)	(0.035)	(0.035)
Film x female	0.035	0.037	-0.009	-0.010	-0.012	-0.015	0.036	0.033
	(0.041)	(0.042)	(0.021)	(0.020)	(0.024)	(0.024)	(0.048)	(0.048)
Constant	0.358***	0.183	0.378***	0.241***	0.382***	0.410***	0.104*	0.243
	(0.034)	(0.134)	(0.019)	(0.083)	(0.020)	(0.077)	(0.060)	(0.172)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
School fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277	277	541	541	541	541	541	541
R-squared	0.184	0.202	0.096	0.104	0.072	0.091	0.090	0.099

[1] Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

I find no evidence for gender-specific effects. Table 3.7 shows that the behavior in the games does not differ much between the sexes; and that the film does not affect women very differently from men. There is a big difference between men and women in regards to the proportions donated to charity: Women donated 10 cents more than men did. The effect of the film is however constant between the sexes.

The results in Table 3.8 indicate that films are particularly effective in influencing the older children. This is most clearly visible in the allocations of the second player in the ultimatum game game. The overall effect of the film on the demand of the participants is -0.03, but the effect of the film increases on the older children with 0.03 per year. Note that the variable Age is adjusted to 12.0 - similar to the youngest participants in the sample. This implies that the films have a negative effect on the youngest participants in our sample and that the effect of the films become positive when the children turn older than 13 years of age. The films are thus mainly effective in affecting the sense of fairness of the older children. The estimates with regard to allocations in the dictator game and the donations to charity move in a similar direction, but are statistically insignificant.

Table 3.8: OLS estimates of treatment interaction effects (age)

	Dictator game		Ultimatum game				Charity	
	Player 1		Player 1		Player 2		Proportion donated	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Film	-0.000	0.006	0.029	0.027	-0.037	-0.035	0.009	-0.004
	(0.052)	(0.053)	(0.029)	(0.029)	(0.032)	(0.032)	(0.064)	(0.064)
Age	-0.009	-0.017	0.010	0.005	-0.012	-0.011	-0.048*	-0.052**
	(0.022)	(0.023)	(0.009)	(0.010)	(0.012)	(0.013)	(0.025)	(0.026)
Film x age	0.018	0.015	-0.003	-0.002	0.035**	0.033**	0.018	0.021
	(0.024)	(0.024)	(0.014)	(0.014)	(0.015)	(0.015)	(0.030)	(0.030)
Constant	0.473	0.393	0.242*	0.187	0.533***	0.583***	0.861**	0.880**
	(0.332)	(0.338)	(0.142)	(0.147)	(0.181)	(0.189)	(0.374)	(0.377)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
School fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	277	277	541	541	541	541	541	541
R-squared	0.170	0.188	0.078	0.087	0.076	0.093	0.042	0.092

Notes: Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1). The baseline of variable Age is adjusted to 12.0 years - similar to the youngest participants in the sample.

3.7 Conclusion

Do people become more willing to redistribute when they are shown the lives of people living in poverty? This study shows the outcomes of a field experiment in which films about people living in poverty were used as a treatment. It provides evidence that showing inequality affects redistributional behavior, at least in the short run. This shows that social preferences are not fixed and can be influenced externally. The field experiment was carried out on six secondary schools with participants in the age of 12 to 16 years old. The participants played a series of incentivized economic games and were asked for a donation to charity. The allocations in these games are used as measures for social preferences. A random process determined whether the participants were to watch the film before or after playing the games.Those who watched a film before playing the games, donated more to the other player in the dictator and ultimatum game, than those who watched the film afterwards. The donations to charity also became more generous after the game. This demonstrates that displaying inequality can invoke altruistic behavior and hence influences redistribution.

I used two different films for this experiment - the participants watched one of them - and their respective effects are slightly different. The first film, a documentary about the lives of teenage factory workers in China, not only affected generous behavior, but also strongly affected the notions of fairness of the participants - illustrated by their increased demands when playing the role of receiver in the ultimatum game. The other film, a movie about a group of teenagers in Africa, affected generous behavior in the games, but did not lead to higher demands in the ultimatum game. I also investigated the effect of the film on poverty awareness and on emotional attitudes (empathy, trust and happiness). The emotions of the participants that were shown the Africa film were affected (they became more empathic and had an increased trust level) but their perspective of poverty prevalence was not. The China film did not affect any of the emotions that were measured, but changed the perspective of the participants on the prevalence and severity of poverty. Whether this is the main reason that the participants that received the treatment demonstrated fairer behavior in the games is not clear. Another possible explanation is that the main theme in the documentary is injustice, which may logically influence behavior regarding to fairness.

This study demonstrates that showing social inequality can affect redistributional preferences in the short term. The participants played the games immediately after watching the film and it is possible that the effects fade away over time. The film affects social preferences either by informing the participants

about social inequalities they were unaware of before, or by affecting their mood, making their allocations in the games more emotionally driven. If the effect is merely caused by emotions that are evoked by the film, it is unlikely to last in the long run.

Appendix 3.A List of experiments

Table 3.9 shows the list of field experiments in chronological order.The first column contains the date the experiment took place; the second column the name of the town the school was located. Depending on the size of the school, the experiment took place in one or two days. For the schools in Hilversum, Dronten and Krommenie the experiment was spread out over two days. We worked together with six schools; the total data collection took 9 days. Column 3 shows the number of participants in the control group, whereas the columns 4 and 5 show the number of participants that watched film China Blue or the film Africa United respectively.

Table 3.9: List of Experiments

Date	School location	Treatment		
		Control	Film: China	Film: Africa
22-03-2013	Rotterdam	11	12	
27-03-2013	Hilversum	26	22	
09-04-2013	Goes	24		31
12-04-2013	Goes	29		41
13-05-2013	Dronten	25		48
16-05-2013	Dronten	50		36
27-05-2013	Gieten	20	40	
06-06-2013	Krommenie	18	40	
10-06-2013	Krommenie	26	43	
Total		**229**	**156**	**156**

In the schools with less than 100 potential participants the experiment could be carried out in one day. In the schools with more than 100 potential participants, the group of participants was split up in two, and the experiment was carried out over two days.

Appendix 3.B Measure for inequality awareness

The questionnaire included 18 statements about Africa and China. The participants indicate the extent to which they believed each statement was true on a five point Likert scale. We used 6 of these 18 statements - the ones that focus on aspects of poverty - as a measure for inequality awareness. These 6 statements are shown below. Although most of the statements focused specifically on Africa or China, 2 statements (Statement 5 and 6 in Table 3.10 below) are valid for both films.

Table 3.10: Poverty awareness

Nr	Statement	Response
1	Africa/ China is poor	+
2	Africa/China is modern	-
3	Most children in Africa/China go to school	-
4	The situation in Africa/China is getting better everyday	-
5	There is less and less poverty in the world	-
6	Children in developing countries are likely to have a better future	-

Notes: "Response" indicates whether agreement with the statement is either positively (+) or negatively (-) associated with poverty awareness.

Appendix 3.C Measure for empathy

Empathy is measured with scale based on 13 statements. It is based an adapted version of the empathy scale designed by Bryant (1982), which is based on the commonly accepted empathy quotient (Davis, 1983) but focussing specifically on children. I made two small changes to this scale. First, I reduced the number of questions. WhereBryant (1982) use a male and female version of each question I use gender neutral questions. Second, I use a 5-point Likert scale instead of binary (yes/no) scale (Similar to the scale used in Warden and Mackinnon (2003)). The variable empathy in this study is based on the total score of the 13 questions, in which 0 is the lowest and 1 the highest score possible. Table 3.11 shows the 13 statements (English translations) as well as the calculation of the total Empathy score.

Table 3.11: Empathy Scale

Nr	Statement	Response
1	People who kiss and hug in public are silly	-
2	I really like to watch people open presents, even when I don't get a present myself.	+
3	Seeing someone who is crying makes me feel like crying.	+
4	I get upset when I see someone being hurt.	+
5	Even when I don't know why someone is laughing, I laugh too.	+
6	It's hard for me to see why someone else gets upset.	+
7	It makes me sad to see a someone who can't find anyone to play with.	+
8	Grown-ups sometimes cry even when they have nothing to be sad about.	-
9	It's silly to treat dogs and cats as though they have feelings like people.	-
10	Kids who have no friends probably don't want any.	-
11	People who cry because they are happy are silly	-
12	I think it is funny that some people cry during a sad movie.	-
13	I am able to eat all my cookies even when I see someone looking at me wanting one.	-

[1] Notes: "Response" indicates whether agreement with the statement is either positvely (+) or negatively (-) associated with empathy.

The variable empathy is the total sum of score on statements 2 to 7; minus the total sum on questions 1 and 8 to 13. This score has then be then adjusted on a scale to 0 to 1 in which 0 is the lowest and 1 the highest score possible.

Appendix 3.D Measure for trust

The measure I use for trust is based on the combined score of three binary questions and uses a 0-1 scale (a higher score implying a higher level of trust). The questions are derived from the General Social Survey (GSS) and are standard questions for measuring social trust (Glaeser et al., 2000). The following three questions were used:

1. Would you say that most people will be honest to you? Or that they will take advantage of you if they get the chance?
   (a) Most people will treat me honestly
   (b) Most people will take advantage of me
2. Would you say that most people are helpful? Or that they will merely act out of self-interest?
   (a) Most people are helpful
   (b) Most people act our of self-interest
3. would you say that most people can be trusted? Or that you cant be too careful in dealing with people?
   (a) People can be trusted
   (b) You can't be too careful

For every answer (a) participants receive 1 point, for answer (b) 0 points. The variable Social Trust is the total score on the three questions above, adjusted to a 0-1 scale in which 0 is the lowest and 1 the highest score possible.

4

All that glitters is not gold: An evaluation of the impact of Fair Trade on artisanal gold miners in Ghana

4.1 Introduction

Fair Trade, the certification and labelling initiative supporting poor small-scale producers in developing countries to achieve better trading conditions, has recently expanded its horizon to the small-scale mining sector. Artisanal or small-scale gold mining is intrinsically linked to poverty and causes some serious environmental problems such as mercury pollution and land degradation. The process of gold extraction - which involves, among others, the use of mercury, explosives and working in poorly build

underground ghetto's - causes health risks for the miners and those living in the mining areas (Hilson, 2002; Hilson and Pardie, 2006; Hilson et al., 2007; Hilson, 2008; Banchirigah, 2006, 2008). The Fair Trade label for gold will serve as a guarantee for consumers that the product is produced under improved labour conditions. This study investigates the effects of implementing these labour standards on artisanal gold miners in Ghana. I use a difference-in-differences approach with two waves of data, in which gold miners from non-treated mining sites in the same area are used as a counterfactual.

Fair Trade is a labeling initiative directed at improving the lives of poor producers in developing countries. As is the case with other types of third party certification, Fair Trade labelling is an effort of solving an information asymmetry in which consumers have insufficient information about quality of a product. It is assumed that consumers are willing to pay a higher price for a product if it is produced under the right labour and environmental standards. Fair Trade focuses on agricultural products that are produced in developing countries, but of which consumption largely takes place in the geographical North. As part of the certification, small scale producers are offered better prices, stable market links and resources for social and environmental projects. Since the first official Fair Trade label was founded in 1988 - the 'Max Havelaar' label - the sales in Fair Trade have seen a steady annual increase, and the label has expanded to a wide variety of products. Gold is one of the most recent expansions.

Proponents of Fair Trade claim that the certification benefits small producers by providing higher incomes and greater income stability. Opponents have criticized Fair Trade in that it encourages producers to keep investing into commodities with a decreasing demand, instead of diversifying their activities. Collier (2007) called the the price premium in Fair Trade products a form of a charitable transfer that the producers receive "as long as they are producing crops that have locked them into this poverty." Despite the long history of Fair Trade the empirical evidence about its effects on income and wealth are limited (Dragusanu et al., 2014). Fort and Ruben (2008a), Sáenz Segura and Zuniga-Arias (2008) and Bacon (2005) empirically investigate the effect of Fair Trade on small coffee producers in respectively Peru, Costa Rica and Nicaragua, but find no significant effects on wealth or income of the producers. Zuniga-Arias and Segura (2008), Becchetti and Costantino (2008) and Ruben and van Schendel (2008), investigating the effect of Fair Trade certification on on banana producers in Kenya, Ghana and Costa Rica, do not find any effects with regard to income either. Fort and Ruben (2008b) on the other hand, find a positive effect of the certification on income and wealth of banana producers in Peru.

This study contributes to the literature by estimating the effects of improved Fair Trade labour standards on the livelihoods of small scale miners in Ghana. The current empirical literature on Fair

Trade focuses on agricultural products such as bananas and coffee. Miners are very different from the traditional agricultural target group, in particular when in it comes to mobility. Where miners can easily move from one area to the next, farmers are relatively immobile - and often need to sell their land, house and other modes of production before being able to move to another area. In our sample the median employee tenure in gold mines is 1 year. I find that this affects the impact of the Fair Trade intervention. The second contribution of this study is the the methodology. This study uses a difference-in-differences approach with two waves of data, of in total 1916 interviewed miners. As observed by Dragusanu et al. (2014) the current literature on the impact of Fair Trade exists mainly onf either qualitative studies or studies that apply matching techniques.

I find that the results of the Fair Trade intervention for gold are mixed. First, the intervention has a positive effect on the health of gold miners. The likelihood of miners getting ill or injured decreases by 5 percentage points against a baseline of 22 percent. This indicates that the intervention leads to a 23 percent decrease in the proportion of cases of illness or injuries. Second, I find positive effects on the miner's investments in human capital. The proportion of children in the households of miners that attends school increases by 0.08 against a baseline of 0.80; an increase of approximately 10 percent. Third, there appears to be a negative effect of the intervention on the income of gold miners. The estimated proportion of miners that live above USD 2.50 per day (two times the international poverty line) decreases by 5 percent as a result of the intervention. There are different potential explanations for this negative effect. It can be that the safer working conditions may negatively affect the revenue of gold mining. Certain risk-taking can increase the revenue of gold mining, despite its negative effects on health. Moreover, the additional required tasks that are part of the safety and environmental regulations - e.g. wearing protection gear, washing hands, cleaning surroundings - can reduce the revenue per hour. This is suggested by the increased number of working hours per miner in the treated mines. Also, the improved working conditions of the Fair Trade mines make them more attractive to work for, allowing them to pay miners a lower salary. A last finding is that the high turnover rates of miners affects the intervention. I find no indication of any positive effects of the intervention on miners with a tenure of one year or less, which indicates that it takes time for the effects on school attendance and health to materialise. However the negative effects on income are visible on the whole sample of miners - both low and high tenure.

The remainder of this chapter is organized as follows. Section 4.2 discusses the literature on Fair Trade. Section 4.3 provides background information about artisanal gold mining in Ghana, and a description

of the intervention. Section 4.4 explains the empirical strategy. Section 4.5 shows the demographic characteristics of the miners and a balance test on observed characteristics. Section 4.6 shows the main estimates of the intervention on wages, health, and investments in human capital. Section 4.7 contains sensitivity tests. Section 4.8 discusses the results and concludes.

4.2 Previous studies

The emergence of modern Fair Trade labels can be traced back to 1988, when a faith based NGO started an initiative with the objective to ensure that growers of crops in low-income countries were paid a "fair price for their products. This first Fair Trade label was called Max Havelaar, named after a fictional Dutch character who opposed the exploitation of coffee pickers in the Dutch colonies. The concept spread to other countries in the years after. The label is now applied to a wide variety of agricultural products, the most popular being bananas and coffee.

Fair Trade labelling is a type of third party quality certification, used to inform the consumer about characteristics of a product that he or she would otherwise be unable to identify. Labelling is an intervention typical useful in markets with information asymmetries. In markets where the producer has information - about the characteristics or quality of the product - that the consumer doesn't have, there will be undersupply of high quality products - assuming that the costs to produce a high quality product will be higher. A third party certificate provides objective information about the quality to the consumer, which can solve this information asymmetry. Fair Trade labelling differs from traditional labelling in the sense that it doesn't address an obvious information asymmetry about the quality and hence the value of the product, but rather about the circumstances in which the product was produced. In general, Fair Trade products sell at a higher price than comparable conventional products without delivering any apparent additional quality. Consumers should thus be willing to pay a higher price for a product only because of the better terms offered to the producer. The underlying assumption of Fair Trade labelling is that the consumer derives some utility from this altruistic gesture. The price difference between the conventional and the Fair Trade product that the consumer is willing to pay can thus also be regarded as a form of charity, in which the consumer gains utility from altruistic gestures (Andreoni, 1988, 1989).

The market for Fair Trade products has been growing steadily in the past twenty-five years, which is an indication that there is an additional value in the Fair Trade product that the consumer appreciates (Mohan, 2010). A number of studies have experimentally investigated the willingness to pay a higher

price for a product that is produced under better circumstances. A field experiment by Hainmueller et al. finds that a Fair Trade label increases the sale of coffee with 10 percent, but that the response to price increases differ between the higher and lower-priced coffee. The higher-priced coffee was hardly affected by an increase in the price, while the sales of lower-priced coffee declined by 30 percent after the price increased with 8 percent. With an experiment on eBay Hiscox et al. (2011) finds that consumers are willing to pay on average 23 percent more for a coffee that is labelled Fair Trade than for the regular product. A similiar experiment on eBay demonstrates that consumers are willing to pay more for a product that is auctioned on a charity auction, than an identical product on a non-charity auction. Consumers paid on average 6 percent more for the charity products (Elfenbein and McManus, 2010).

Despite the long history, there is only limited evidence available about the impact of Fair Trade certification. A large number of impact studies were carried out by the umbrella organisation Fair Trade International (FLO) and made publicly accessible online, but very few use a convincing methodology. Most of these are qualitative studies, based on anecdotal evidence, mostly acquired through interviews. Only a few studies have applied more rigorous standards to evaluate the impact of Fair Trade, most of them focussing on on coffee and banana producers. These two products are by far the most popular in the Fair Trade Brand. All of the available studies are carried out in countries in Africa and South-America.

Bacon (2005); Fort and Ruben (2008a); Sáenz Segura and Zuniga-Arias (2008) estimate the effect of Fair Trade certification on coffee producers in respectively Peru, Costa Rica and Nicaragua. Bacon (2005) finds no significant effects on income. Although he finds that the certified farmers receive higher prices for their products, their income and wealth are not significantly different from the non-certified coffee producers. The studies of Fort and Ruben (2008a) and Sáenz Segura and Zuniga-Arias (2008) come to similar conclusions. Neither find any effect of Fair Trade certification on income and wealth. A study by the same authors, investigating the effects of certification on banana producers in Peru, is more optimistic about the impact. This study finds that the households of certified producers are wealthier and have a higher income than the non-certified producers. The authors argue that this difference can be explained by the higher production of bananas in certified households, which would be a result of the improvements in production process that certification requires (Fort and Ruben, 2008b). Two other studies on Fair Trade bananas, in respectively Costa Rica (Zuniga-Arias and Segura, 2008) and Ghana (Ruben and van Schendel, 2008) did not find any effect on income or wealth. Becchetti and Costantino (2008), who evaluated the impact of Fair Trade certification on fruit and herb producers in Kenya, do not find any effect on income or wealth, but they do find that the subjective wellbeing of certified farmers

increases.

In sum, the empirical evidence on the impact of Fair Trade certification is inconclusive. Most of the studies do not find any effects on income or wealth.

4.3 Background

This section will provide a detailed description of the Fair Gold intervention. The first section gives background information about small scale gold mining in Ghana and its negative externalities. The second section provides a description of the intervention and includes an analysis of compliance.

4.3.1 Context

Ghana is listed in the top-10 of gold-producing countries worldwide; with over 200,000 people being employed in the industry (Hilson et al., 2007). Although the lion's share of the gold production is produced by one of the eleven large-scale mining companies operating in Ghana, the artisanal or small-scale mining sector employs most people.[1]In 2012 there were over 650 registered small scale mining companies in Ghana.[2] Two types of artisanal gold mining can be distinguished. The first type, dredging - the process by which gold is extracted from river sentiments - is illegal in Ghana. The second type, recuperation - the direct removal of metal-rich material from the ground - is permitted in Ghana under certain conditions.

Legal gold mining in Ghana requires a concession. Generally, each 25-acre concession employs five to twenty groups of artisanal miners, each consisting of five to ten workers. Equipment is typically confined to hammers, buckets, shovels, pickaxes and sluice boxes and is usually provided by the concession owner. Each concession features specialists, such as diggers, carriers and ore washers. The ore is obtained as either unconsolidated or consolidated material. If unconsolidated, it is taken to the sluice box. If consolidated; the ore is crushed and then milled into a powder form and taken to the sluice box. At this point the ore is put in slurry form and run over an inclined sluice box with the application of water. In the process, the heavy metals including the gold particles settle on the jute bags or blankets that have been spread along the length of the box. After this the concentrated ore goes through a secondary sluicing to make it more concentrated. The concentrate is then run through with mercury to obtain the amalgam - a mixture of gold and mercury (Appiah, 1998).

[1]Ghana Chamber of Mines report "Performance of the mining industry 2013" (wwww.ghanachamberofmines.org)
[2]Ghana Mineral Commission (www.ghana-mining.org)

This process of gold mining has negative externalities. Small-scale gold mining regions encounter serious environmental problems as mercury pollution and land degradation. Mercury poses a health risk not only for the miners, but also for the surrounding communities when land is contaminated. One of the major health risks is eating the fish in contaminated regions (Hilson and Pardie, 2006). Artisanal gold-mining causes significant damage to landscapes. It is responsible for the removal of vast quantities of surface vegetation and mass deforestation. Miners also typically abandon pits and trenches without properly reclaiming spoils. In turn this land degradation affects small farmers and communities, being forced to relocate after the land has become unworkable.

Besides the negative effects on the community and environment, the health risks on small scale miners themselves are even greater. Small-scale gold miners inhale noxious mercury fumes, generally handle mercury without gloves, and consume contaminated water with elevated concentrations of mercury. In addition, miners suck amalgam to remove excess mercury prior to roasting (Hilson and Pardie, 2006; Hilson et al., 2007). Once in the body, mercury poses a serious threat to humans. Symptoms include dizziness, fatigue, loss of appetite, headaches, convulsions, and eventually death. Besides the health risks related to mercury, accidents in mining poses another problem. Protective headgear and gloves may help preventing these risks.

A number of studies argue that artisanal gold mining is extricably linked to poverty (Hilson, 2002; Hilson and Pardie, 2006; Hilson et al., 2007; Hilson, 2008; Banchirigah, 2006, 2008). According to these studies, gold mining is one of the very few income generating options that the people in rural areas currently have - particularly in current mining areas such as Tarkwa. The large mining companies operating in Ghana have turned large areas into mining sites causing community dislocations and unemployment. Most of the households that were dislocated were originally peasants. The lack of alternative income-generating opportunities in the area forced many of them into artisanal gold mining. Hilson (2002) argues that the majority of the miners are 'trapped' in a cycle of poverty. A lack of equipment and a lack of opportunities to invest cause artisanal gold miners to be unable to escape poverty. Other studies on the other hand argue that the majority of artisanal gold miners are fortune seekers trying to get rich quickly (Childs, 2008).

The majority of miners is unregistered and works illegally. The exact numbers of illegal miners in Ghana are unknown, although Banchirigah (2008)) estimates it at almost one million (in 2006).

4.3.2 Intervention

The objective of the Fair Gold intervention is to improve the working conditions of artisanal gold miners and reduce the environmental, via Fair Trade labelling. The labelling organization FLO[3] is responsible for Fair Trade certification standards. FLO stipulates the minimum criteria as well as the contract and payment conditions that the trading process must fulfil, for a product to be labelled and sold as Fair Trade. The mining cooperatives and wholesalers that apply for the Fair Trade label will be charged certification fees by the the labelling organisations.

The Fair Gold certification guarantees that the gold has been "extracted, processed and traded in a fair and responsible manner." The criteria that need to be fulfilled to obtain the Fair Gold certificate focus on working conditions, the position of miners and the reduction of environmental damage. Two important criteria are safe working conditions and the responsible use of chemicals. Safe working conditions imply, for example, that miners will have to wear the appropriate personal protection gear when working with mercury. Other criteria are focus on the organization of miners, to increase their bargaining power with traders. Other important criteria deal with the rights of female labours and the prohibition of child labour.

After obtaining Fair Gold certification, gold miners will receive a ten percent premium on top of their wholesale price on all gold that is sold under the Fair Gold label. This premium is supposed to be invested in the local community. The certified mining companies have policies in place to ensure that this premium is invested democratically. Contrary to many of the agricultural products of the Fair Trade label, certified miners do not receive a guaranteed minimum price. Hence for the individual artisanal gold miner there is no apparent financial incentive to participate. In spite of this, the implementing non-profit organisation hopes that the new procedures and different way of working in the mines, and the increasing demand for Fair Gold products will lead to higher income of the miners.

In 2012 five artisanal gold mining companies in Ghana signed a contract in which they expressed their commitment to participate in the programme with the goal to obtain the Fair Gold certification within three years. Between 2012 and 2015 the organisation provided the mining companies with financial, technical and logistic support to enable them to comply to these standards. At the time of this evaluation none of the mining companies that participated in our study obtained the Fair Gold certificate yet. This study thus not investigate the effect of Fair Gold labelling per se, but rather the effect of working

[3]abbreviation of Fairtrade Labelling Organizations International, more generally known as Fairtrade International

according to those standards.

The organisation assist the five mining sites in obtaining the certification. As a first step, the organisation appointed a group of senior miners and a Fair Trade supervisor on each mining site. They are responsible for supervising the implementation of the Fair Trade standards. Management support and office equipment are provided in all mines from the start of the intervention. Second, on each mining site training sessions were provided on issues such as first aid, safety procedures, responsible mercury handling, labour rights and environmental management. Most of the training sessions follow a 'training of trainers' approach: Professionals are invited on the mining site to train the the senior miners and the supervisor. This group is then ought to train the rest of the workers in the mining site on regular intervals. Third, each mining site received free personal protection gear - helmets, safety shoes, mouthpieces, gloves etc. Fourth, female miners were organized in self-help groups. Female self-help groups - consisting of on average 30-50 female miners - were formed and supported by the organisation with additional training sessions on topics like children's care and alternative livelihoods. Fifth, a number of activities with respect to the children in the households of miners were carried out. Miners with school-going children received additional - financial - support and special training sessions were provided to emphasise the importance of children's education to miners. Construction of daycare centers was planned, but has not been finished yet at the time of this study.

The respondents were asked about their participation in training sessions and their access to certain facilities during the second wave of interviews. The compliance with the intervention can be roughly estimated by comparing the answers of the treated with the answers of the control mines. Table 4.1 shows the estimates of this analysis. Note that these questions were only asked during the second wave, hence no baseline data are available. The upper part of Table 4.1 shows the compliance with the different training sessions, the bottom part shows the extent to which respondents claim to have access to different types of facilities. Column (1) shows the means of the treated miners, Column (2) the means of the miners in the control group. The means can be interpreted as the average proportion of miners that either have participated in the specific training, or the average proportion that have free access to the facilities concerned. Column (3) shows the difference between the means of Column (1) and (2). I performed t-tests that were adjusted for clustering at mine-level to determine whether differences are statistically significant.

Table 4.1 shows that approximately 50 percent of the miners in the treated mines recall to have participated in training sessions on safety, first aid and environmental protection. This is approximately

Table 4.1: Compliance

	Treated (1) Y_1^D	Control (2) Y_1^C	Estim. Dif. (3) $Y_1^D - Y_1^D$
Trainings:			
Training on safety (procedures)	0.52	0.12	0.39**
			(0.16)
Medical training (First Aid)	0.49	0.09	0.41**
			(0.16)
Training about labour conditions	0.24	0.05	0.19*
			(0.10)
Training about environmental issues	0.47	0.07	0.40**
			(0.15)
Session about importance of children's education	0.37	0.08	0.29**
			(0.09)
Other:			
Access to gloves	0.33	0.15	0.18
			(0.17)
Access to helmets	0.38	0.05	0.33*
			(0.14)
Access to mouthpieces	0.33	0.07	0.26*
			(0.12)
Access to child care facilities	0.34	0.15	0.19
			(0.14)
Observations	495	458	953

Notes: Significance level tested with t-tests adjusted for clustering at the mine level. Standard errors in parentheses: ** p<0.05, * p<0.1

40 percent more than in the control mines. The proportion of miners that recall to have participated in a training session on labor conditions or about children education is lower - respectively 24 and 37 percent- but still significantly different from the control mines. One of the reasons that hardly 50 percent of the miners recall to have participated in one of the training session, is probably the high turnover rate of workers: 48 percent of the miners in the treated mines worked there for less than a year at the time of the interview.

The bottom part of Table 4.1 shows that between 33 and 38 percent of the treated miners have free access to the respective types of personal protection gear. Depending on the type of protection gear, this is between 18 and 33 percent more than in the control mines. Regarding the fact that the organisation freely distribute protection gear to each mining site on regular intervals, we would have expected this proportion to be higher. Focus group discussions with miners revealed that they experienced that the distribution of protection gear is insufficient. It appeared that many miners take the protection gear home, but that not all of them - in particular those that work only temporary in the mines - return it, causing a demand of personal protection gear that is greater than expected. Approximately one third of the miners in the treated mines claim to have access to child care facilities. This is 19 percent more than in the control mines. If I exclude the households without children, then the proportions of families with access to child facilities are 0.47 in the treated, and 0.28 in the control condition.

4.4 Empirical Strategy

The objective of this study is to evaluate the effect of the Fair Gold labour standards on the gold miners. In an ideal experimental setting the treatment would have been randomly allocated. Under these conditions we would have been able to estimate the impact of the intervention with the following model:

$$Y_i = \alpha + \beta Fair_i + \gamma X_i + \varepsilon_i \tag{4.1}$$

In which Y_i is the outcome variable of interest of individual i; $Fair_i$ a dummy variable equal to 1 if individual i worked according to the Fair Trade standards and equal to 0 if the individual otherwise. X_i is a vector of covariates reflecting background characteristics of individual i . Parameter β is the parameter of interest and reflects the effect of the treatment on the individual.

Equation (4.1) will give unbiased results only if there is no correlation between $Fair_i$ and ε_i. However, since the treatment could not be randomly assigned, it is unlikely that this assumption holds. The mines that invest in labour standards may attract more motivated or skilled miners, leading to an overestimation of the effect. It can also cause the opposite: the treated mines may scare away the more ambitious or risk taking miners, which will lead to an underestimation of the effect. Since these unobserved factors cannot be controlled for, the estimates of Equation (4.1) are likely to be biased.

This study makes use of a difference-in-differences design with two waves of data. The data were collected from a sample of miners from the five gold mines that participate in the Fair Gold program; and miners from five other artisanal gold mines - located in the same region - that serve as a counterfactual. Each wave contains data from approximately hundred miners per mining site. The first wave was collected shortly after the intervention started; the second wave 21 months later. Assuming that the time trend in the control group is similar to the time trend in the treated group, if they would not have received the treatment, the impact of the intervention can be estimated by comparing the changes in outcomes over time between the miners in the treated and the miners in the control condition:

$$\beta = (Y_{T=1}^{Fair} - Y_{d=0}^{Fair}) - (Y_{T=1}^{Control} - Y_{T=0}^{Control}) \tag{4.2}$$

Where $(Y_{d=1}^{Fair} - Y_{d=0}^{Fair})$ reflects the differences in outcomes over time in the treated group, and $(Y_{d=1}^{Control} - Y_{d=0}^{Control})$ gives the time trend in the control group. Parameter β gives the causal effect of the intervention.

For obtaining the difference-in-differences estimates, I use the following model:

$$Y_{it} = \phi_m + \lambda T_t + \beta(Fair_m \cdot T_t) + \gamma X_{it} + \xi Q_{it} + \varepsilon_{it} \tag{4.3}$$

In which T_t is a dummy variable equal to 0 if the data was collected during the first, or equal to 1 if it was collected during the second wave. ϕ_m is a mine-specific intercept. Q_i is a vector of variables reflecting the circumstances in which the interviews were conducted. Parameter β indicates the causal effect of the intervention. Standard errors are adjusted for clustering of error terms at the mine times wave level.

A comparison of the outcomes before and after the intervention and between treatment and control groups as in Equation (4.3) yields the average treatment effect of the intervention if the common trend

assumption holds (Angrist and Pischke, 2009). A standard approach for investigating the plausibility of the common trend assumption is to compare the pre-trends of the outcomes variables in the control and treatment mines. Unfortunately, these data are not available for the specific mines in Ghana. To improve the comparability of the control and treatment mines, and hence the plausibility of the common trend assumption, we selected control mines that are located close to the treated mines. In addition, In section 4.5 I examine the time trend in both groups on observable characteristics.

4.5 Data

The data were collected during two waves of interviews that were conducted with miners on the mining sites. The first wave took place in January 2013; the second in October 2014. For each wave a team of ten enumerators were selected [4] and trained to conduct the interviews. After receiving a day of training, the enumerators were sent to the mining sites in teams: Three teams consisted of two persons, one team of four. Each team conducted interviews in two mines (one treated, one control) or 4 mines (two treated, two control). The enumerators randomly selected approximately 100 miners at each mine to participate in the interviews. The team of enumerators during the second round deviated from the first group; Although some of them participated in both waves, most of the enumerators during the second round were new. The questionnaire was constructed in cooperation with the organisation that is responsible for the implementation of the Fair Gold program. I the second wave a number of questions were added to the questionnaire, most importantly a set of questions about whether the interviewee participated in specific activities. I will use this information to investigate the compliance to the treatment.

The intervention was carried out at five artisanal mining sites, which differ in size, ownership structure and type of gold mining activities. Two mining sites are owned by formal mining companies; one mining sites is owned by a sole concessioner who possesses the mining rights, one is family owned and one is owned by a cooperative group. In three of these mines the core activity is underground mining whereas the other mines focus on surface mining. In fact, what primarily makes the participating mining sites different from other artisanal mining sites is the willingness to participate in the Fair Gold program. Although the non-profit organisation offered to co-finance the investments necessary to fulfill the Fair Gold criteria, most mining companies were reluctant to participate.

The counterfactual consist of miners from artisanal gold mining sites that do not participate in the

[4] The enumerators were students of the University of Mines and Technology (UMAT) of Tarkwa Ghana

Fair Gold program. Together with the organisation I selected five artisanal mining sites that were eligible to participate in the Fair Gold program, but were not yet interested. The mining sites are geographically located close to the treated mines. Figure 4.1 shows a map of Ghana with the location of the ten sites. The red dots in Figure 4.1 show the location of the treated mines; the blue dots show the location of the control mines.

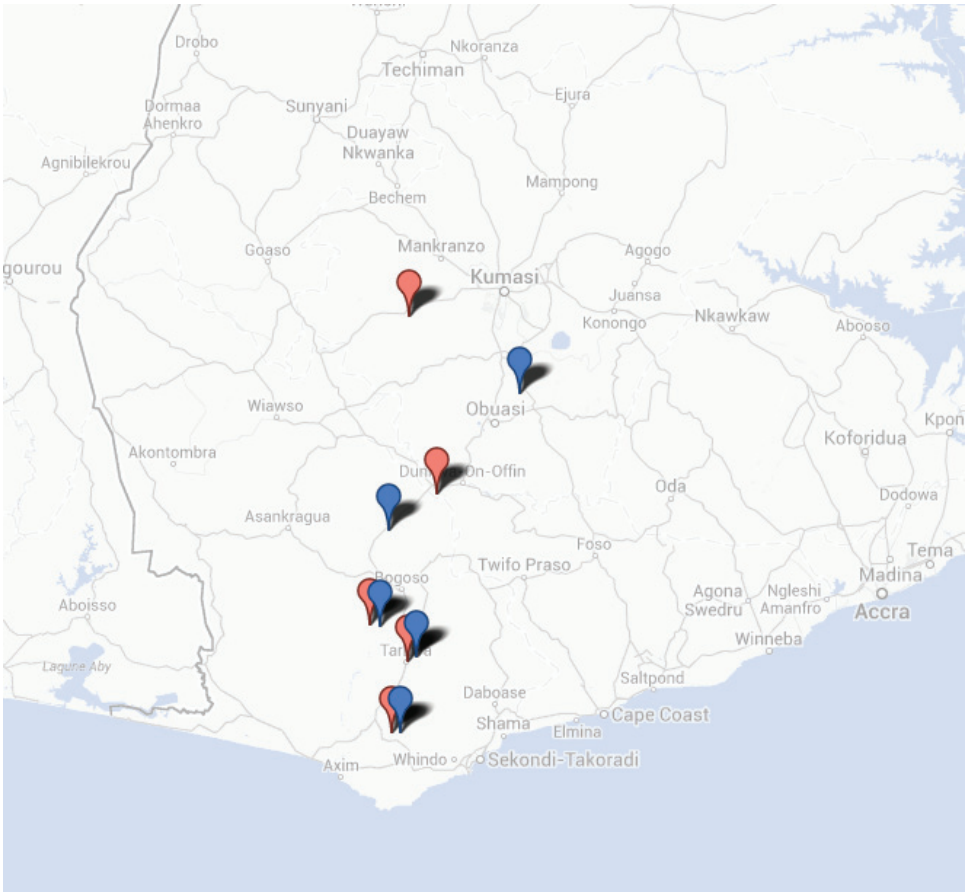


Figure 4.1: Location of treated (red) and control (blue) mines

Since the decision to participate lies with the management of the mining cooperations, the mining sites in the program are likely to be different from those in the control condition. The concession owners of the treated mines may be more motivated than the owners of the control mines to improve working

conditions; or the owners of the treated mine sites may be less risk averse, hence willing to take a risk to invest in the Fair Gold certification. I will investigate the extent to which this affects the selection of miners in the mining sites - on observable characteristics - further below.

Table 4.2 shows the background characteristics of the miners in both control and treated mines. Column (1) and (2) show the means of the first and second wave, respectively. Column (3) shows the difference between the two. Column (4) to (6) show the background characteristics for the control group. Column (7) shows the difference-in-differences estimates. This estimate can be interpreted as the change that specifically occurred to the miners in the treated mines.

Table 4.2: Demographic characteristics

	Treated			Control			Dif -Dif
	(1) Y_0^D	(2) Y_1^D	(3) $\Delta\bar{Y}^D$	(4) Y_0^C	(5) Y_1^C	(6) $\Delta\bar{Y}^C$	(7) $\Delta\bar{Y}^D - \Delta\bar{Y}^C$
Age	29.8	32.2	2.4	29.2	31.0	1.8	0.6
	(1.5)	(1.5)	(2.0)	(1.7)	(1.7)	(2.4)	(2.9)
Female	0.14	0.27	0.13	0.19	0.12	-0.08	0.21*
	(0.07)	(0.07)	(0.10)	(0.04)	(0.04)	(0.06)	(0.10)
Educational attainment	8.5	8.5	0.0	8.5	9.0	0.6	-0.5
	(0.4)	(0.4)	(0.6)	(0.5)	(0.5)	(0.7)	(0.8)
Educational attainment mother	3.8	3.8	0.1	3.3	3.9	0.6	-0.5
	(0.6)	(0.6)	(0.8)	(0.4)	(0.5)	(0.6)	(0.9)
Educational attainment father	6.2	5.0	-1.2	5.8	5.5	-0.3	-0.9
	(0.8)	(0.8)	(1.1)	(0.8)	(0.8)	(1.1)	(1.4)
Employee tenure (in years)	2.5	3.0	0.5	2.1	2.6	0.5	0.0
	(0.8)	(0.8)	(1.2)	(0.7)	(0.7)	(0.9)	(1.4)
Nr. of years ago moved to current town	10.3	11.5	1.2	10.3	10.8	0.5	0.7
	(0.9)	(0.9)	(1.2)	(1.0)	(1.1)	(1.5)	(1.7)
Risk aversion	1.8	1.6	-0.2	1.7	1.5	-0.2	0.1
	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
Household size	3.0	3.3	0.3	2.6	2.8	0.2	0.1
	(0.6)	(0.6)	(0.8)	(0.5)	(0.5)	(0.7)	(0.2)
Nr of children in the household (age<18yo)	0.9	1.1	0.2	0.7	0.8	0.1	0.1
	(0.3)	(0.3)	(0.4)	(0.2)	(0.2)	(0.3)	(0.4)
Observations	465	500	965	489	462	951	1916

Notes: Significance level tested with t-tests adjusted for clustering by mine times wave. Standard errors in parentheses: * p<0.1

Table 4.2 shows that, although there are small initial differences between the miners, the time trends in control and treated condition are similar on most background characteristics. A remarkable difference is the proportion of women working in the mines. The treated mines have attracted a larger proportion of women over the past two years than the control mines. This may be a direct or indirect effect of the intervention. The Fai Trade criteria emphasise the importance of women rights and the intervention contain activities that particularly benefit women - like the female self-help groups. The treated mines may thus have become more attractive for women to work for. On the other hand, women's jobs are typically confined to carrying rocks. This hadn't changed in the 21 months between the two waves of interviews.[5]. Another explanation might hence be that during the second wave the treated mines simply required more carriers, which is a temporal effect.

The average miner is in his/her late twenties or early thirties, part of a three-person household and has less than 1 child. An important observation is that the turnover rate of miners - in both treated and control condition - is relatively high. During the first wave, miners in the treated mines stated to have a tenure of 2.5 years; during the second wave this is 3.0 years. In the control mines employee tenure is 2.1 and 2.6 years respectively. Figure 4.2 shows the distribution of employee tenure per treatment condition, per wave, in boxplots. It shows that the variation in tenure is high; and that this variation is higher in the miners in the treated compared to the control condition. In the full sample the average tenure is 2.5 years with a median value of 1 year, implying that half of the miners work at the specific mine for less than a year. This variation in tenure can influence the effectiveness of the treatment. I will investigate this in the sensitivity analysis.

[5]The position of female miners were discussed in focus group discussions in the five mining sites in October 2014. In none of the five miners the perception of a women's job changed: In all mining sites women can only become 'carrier' or secretary

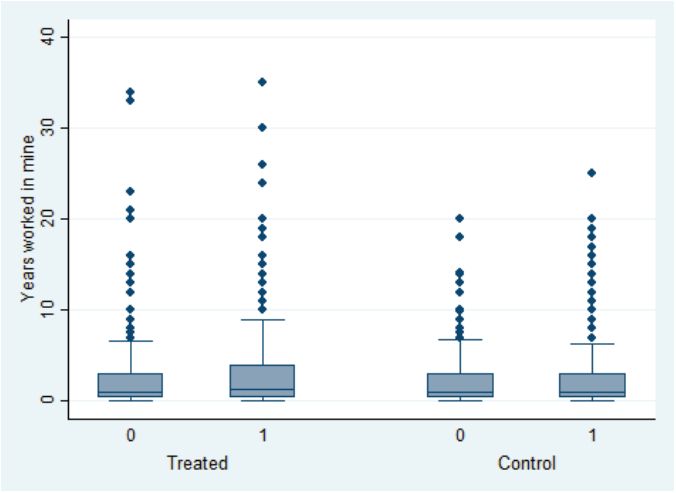


Figure 4.2: Employee tenure (in years) by condition and wave

4.5.1 Dependent variables

I investigate the impact of Fair Trade on three outcomes: income, health, and investment in human capital. The measures for these outcomes are all based on self-reported data from the questionnaire.

As a measure for income I use the natural logarithm of hourly wages. This measure is based on the miner's income in the 7 days prior to the date of the interview, and deduced from two questions in the questionnaire - the first about the amount of money that the respondent earned with mining in the past 7 days, and the second about the number of hours the respondent worked in the mine in the past 7 days. Income is difficult to measure, in particular in developing countries, and the fact that our measure is based on self-reported data raises concerns about measurement errors. Not all miners may be able or willing to accurately specify their income and purposely or involuntary over- or underestimate it. Although this measurement error should not affect the impact estimates - assuming outcomes in both conditions suffer the same measurement bias - it will affect the variation and hence the standard errors, increasing the likelihood of a false negative (Type II error). Another problem with this variable is that because of the short time frame of seven days it is more vulnerable to seasonal and random fluctuations. Although the seasonal fluctuations should be constant in both conditions in each wave, external fluctuations like weather conditions can significantly affect the estimates. For example heavy rain can make a mining site

inaccessible for a few days, significantly reducing the income of the respondent in that week.

Because of the potential measurement errors with regards to the log hourly wages, I also use another measure to determine income, known as the Progress out of Poverty Index (PPI). The PPI is a tool developed for microfinance institutions for the practical purpose of identifying poverty levels of their clients. The tool consists of ten questions, selected for their predictability of annual household income, for which the most recent household census in the country was used - in this case the Ghana Household Census of 2010. These 10 questions will provide a score between 0-100 that can be used to determine the likelihood that a certain household lives below the national poverty line. More information about the PPI index, as well as the questions it is based on, can be found in Annex 4.A. As a rule of thumb, the lower the score, the higher the likelihood that the household is poor.

I use three measures to determine changes in health. All measures are self-reported. The first is a dummy variable equal to one if the miner was sick or injured in the past 30 days and equal to zero otherwise; the second is the number of days the miner was ill or injured to the extent that he could not participate in daily activities. This second measure provides information about the severity of the illness. As a third outcome I use a binary variable indicating whether the miner had an accident in the mine in the past 30 days.

I have some basic data on all the members of the each respondent's household, including the children. As outcome variable for investment in human capital I use a ratio expressing the number children in the household that attend school as a proportion of the total children in the household. Household members are counted as a child as long as they are younger than 18 years old (18 being the age of maturity in Ghana). I use an additional outcome variable in which I focus only on the children up to 12 years old.

Table 4.3 shows summary statistics for the dependent variables of this study. Column (1) and (2) show the means of respectively the first and second wave; and Column (3) the difference between the two. Column (4), (5) and 6) show the background characteristics for the control group. Column (7) shows the difference-in-differences estimates.

Table 4.3: Summary statistics of outcome variables

	Treated			Control			Dif -Dif
	(1) Y_0^D	(2) Y_1^D	(3) $\Delta\bar{Y}^D$	(4) Y_0^C	(5) Y_1^C	(6) $\Delta\bar{Y}^C$	(7) $\Delta\bar{Y}^D - \Delta\bar{Y}^C$
Income							
Hourly wages (in GHS)	4.72	7.67	2.95	3.72	10.31	6.59	-3.64
	(2.28)	(2.21)	(3.18)	(2.73)	(2.80)	(3.91)	(4.72)
Income with mining in past 7 days (in GHS)	150.29	205.65	55.36	120.33	227.87	107.54	-52.18
	(39.91)	(39.59)	(56.22)	(30.22)	(30.96)	(43.26)	(64.06)
Nr of hours worked in past 7 days	44.5	48.9	4.4	47.5	40.9	-6.6	11.0
	(7.6)	(7.6)	(10.8)	(8.7)	(8.9)	(12.5)	(14.9)
PPI poverty index score (0-100)	60.6	57.5	-3.1	62.8	62.4	-0.4	-2.6
	(3.5)	(3.5)	(5.0)	(3.0)	(3.1)	(4.3)	(5.9)
Health							
Was ill/injured in past 30 days	0.32	0.20	-0.12	0.22	0.16	-0.06	-0.06
	(0.09)	(0.09)	(0.12)	(0.07)	(0.07)	(0.10)	(0.15)
Nr of days ill or injured in past 30 days	2.0	1.0	-1.0	1.4	0.5	-0.9	-0.2
	(0.7)	(0.7)	(0.10)	(0.4)	(0.4)	(0.6)	(1.0)
Human capital investments							
Prop. of children (age: 0-18) in hh in school	0.75	0.74	-0.01	0.79	0.60	-0.19*	0.18
	(0.06)	(0.06)	(0.08)	(0.07)	(0.07)	(0.09)	(0.10)
Prop. of children (age:0-13) in hh in school	0.72	0.73	0.01	0.76	0.52	-0.23**	0.13*
	(0.07)	(0.07)	(0.10)	(0.08)	(0.08)	(0.11)	(0.12)
Observations	465	500	965	489	462	951	1916

Notes: Standard errors - adjusted by mine times wave - in parentheses: * p<0.1

Table 4.3 shows how the outcome variables of treated and control miners changed over time. In the control mines we notice a trend in which the number of weekly working hours decreases, while weekly earnings increase. Although income also increases in the treated mines, the increase is much smaller than in the control mines. Moreover, the number of hours that miners work increases in the treated mines - whereas it decreases in miners in the control condition. Despite the fact that respondents are all people working on one of the mining sites, there is large variation in income. This is illustrated in Figure 4.3, which shows boxplots of the log hourly wages per treatment condition and wave. The PPI score appears to be more stable over time.

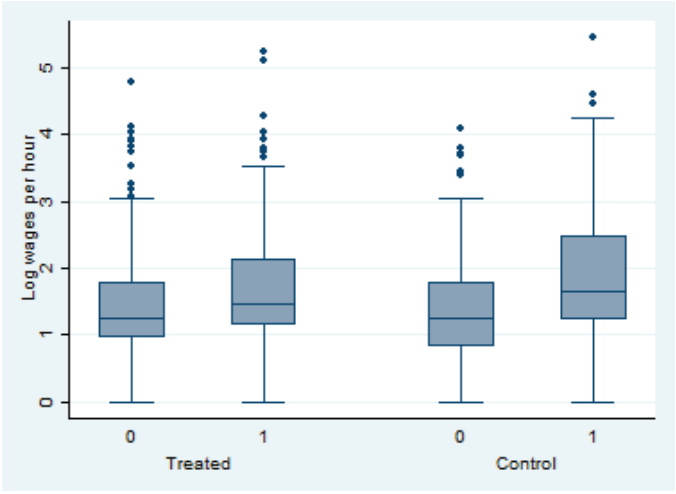


Figure 4.3: Log hourly wages by condition and wave

The proportion of miners that could not go to work because of illness or injury declined over time in both treated and control mines. This decline may be structural or caused by seasonal diseases. This decline appears to be greater in the treated than in the control mines. Table 4.3 also shows some remarkable effects of the treatment on the miner's investment in human capital. The proportion of children in the household going to school increases with 18 percentage points as a result of the treatment. This effect does not change much when focussing on the children in the household that are of school age. A closer look reveals that overall the proportion of children decreased over time in all mines in the control condition, but not in all mines in the treated condition. This decrease occurs in each specific control mine, but not in each treated mine. In Section 4.6 I will further investigate these results.

4.5.2 Other variables

I will further specify the model with control variables reflecting background characteristics of the respondent: gender, age, household size, educational attainment; and educational attainment of both mother and father. Educational attainment is expressed in years (reflecting the average number of years that are required to complete the particular level of education). I also include mine-fixed effects. I control for interview specific circumstances by including variables reflecting the date the interview took place and a dummy vector indicating the enumerator.

4.6 Main estimation results

This section shows the main results of the study. It is divided in three subsections, with each subsection showing the estimates of the three outcomes.

4.6.1 Effects on income

Table 4.4 shows the estimated effects of the intervention on income, using variables based on the respondents's earnings in the 7 days prior to the interview. Each row shows the estimates for a specific dependent variable: log hourly wages, log weekly earnings and number of hours worked, respectively. Column (1) shows the estimates when not including any control variables; Column (2) shows the estimates when we control for background characteristics of the respondent and Column (3) shows the estimates when we also control for data-specific controls (e.g. data of interview). Columns (4) to (6) show the estimates with mine-fixed effects, with Column (5) including controls for individual characteristics and Column (5) including controls for both individual as well as data characteristics. Standard errors are adjusted for clustering of the error terms at the mine $\times$ wave level.

Table 4.4: DD estimates of the effect of Fair Trade on income (self-reported)

	DD estimates					
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Log hourly earnings	-0.32	-0.35	-0.19	-0.25	-0.29	-0.23
	(0.45)	(0.43)	(0.22)	(0.40)	(0.39)	(0.21)
Log weekly earnings	-0.34	-0.27	-0.20	-0.42	-0.37	-0.31
	(0.55)	(0.51)	(0.31)	(0.41)	(0.40)	(0.30)
Number of hours worked in past 7 days	11.01	15.08	10.86	8.23	12.44	12.38**
	(14.94)	(13.65)	(6.42)	(13.52)	(13.03)	(5.19)
Mine fixed effects	No	No	No	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Data controls	No	No	Yes	No	No	Yes
Observations	1,916	1,916	1,916	1,916	1,916	1,916

Notes: Each cell shows the estimated effect of the Fair Gold intervention on dependent variables that are shown in the left column. Estimates with respect to Log hourly earnings are based on 1,886 observations (those that did not work in the past seven days were to be excluded). Standard errors - adjusted by mine times wave - in parentheses: ** p<0.05

The point estimates in Table 4.4 are statistically insignificant, but suggest a negative effect of the intervention on wages. The variation in wages is large, which is partially a result of measurement errors, resulting in high standard errors. The average number of working hours in the mine appears to increase as a result of the intervention.

Table 4.5 below shows the estimates of the effect of the intervention on the PPI poverty score. The table structure is similar to Table 4.4. The first row shows the point estimates with respect to the score. The second, third and fourth row show the estimates with respect to conversions of this score into poverty likelihoods. The second row shows the point estimates of the effect on the estimated proportion of miners living on at least \$3.75 per day, the second on the proportion living on at least \$2.50 per day and the last on the proportion living on at least \$1.25 per day - the universal poverty line.

Table 4.5: DD estimates of the effect of Fair Trade on income (PPI-score)

	DD estimates					
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
PPI score	-2.64	-2.63	-1.90**	-1.79	-2.95	-2.06***
	(5.97)	(2.16)	(0.68)	(4.29)	(1.87)	(0.27)
Estim. prop. of miners living on at least < USD 3.75 per day	-0.05	-0.05	-0.04***	-0.04	-0.06	-0.04***
	(0.12)	(0.04)	(0.01)	(0.09)	(0.04)	(0.00)
Estim. prop. of miners living on at least < USD 2.50 per day	-0.04	-0.04	-0.03*	-0.03	-0.05	-0.04***
	(0.11)	(0.04)	(0.02)	(0.08)	(0.03)	(0.01)
Estim. prop. of miners living on at least < USD 1.25 per day	-0.01	-0.01	-0.01	-0.00	-0.01	-0.01***
	(0.03)	(0.01)	(0.00)	(0.02)	(0.01)	(0.00)
Mine fixed effects	No	No	No	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Data controls	No	No	Yes	No	No	Yes
Observations	1,916	1,916	1,916	1,916	1,916	1,916

Notes: Each cell shows the estimated effect of the Fair Gold intervention on dependent variables that are shown in the left column. Estimates with respect to wealth are all based on the Progress out of Poverty Index (PPI). Standard errors - adjusted by mine times wave - in parentheses: *** p<0.01, * p<0.1

The results in Table 4.5 indicate that the intervention has a negative effect on the poverty index of approximately 2 points. The proportion of miners that live on at least 3.75 USD per day decreases by 4 percentage point. Against a baseline value of 0.56 this is a decrease of 7 percent. The proportion of respondents living on an income of more than 2.50 USD per day decreases by 4 percentage points against a baseline 0.79, indicating a 5 percent decrease. The proportion of those living above the international poverty line of 1.25 USD decreases by 1 percentage point against the baseline of 0.97 - a decrease of 1 percent.

In sum, using the PPI score, the estimates show a negative effect of the intervention on the income of miners. The estimates with regard to self-reported income point in the same direction, although the large variation in this variable complicate estimating the effects accurately. There are a few explanations for this negative effect on income. First, there might be a trade-off between working conditions and wages. The lower risks and more attractive working conditions may have made the treated mines more attractive to work for, decreasing the demanded wage compensation. Simultaneously it may have changed the composition of miners in the different types mines. Risk-avoiding workers will be attracted by the mines in the treated condition, whereas more risk-seeking miners will be attracted by the regular gold mines. A second explanation is that the implementation of safety regulations may have efficiency costs, decreasing overall benefits of gold-mining in the treated mines. Certain dangerous behaviors may increase the risks of accidents or illness, but can also increase the benefits of the ore extraction. The extra procedures with regard to safety and environment and the training sessions all cost time that would have otherwise been diverted to the core activity. The fact that the treatment increases the number of working hours of the miners may be an indication of this.

4.6.2 Effects on health

Table 4.6 shows the estimates with regard to health, using, respectively, the probability of being sick or injured, the number of days being sick or injured and the probability of having an accident in the mine in the 30 days prior to the interview as dependent variables. Table 4.6 has a similar structure as Table 4.4 and 4.5.

Table 4.6: DD estimates of the effect of Fair Trade on health

	DD estimates					
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Ill or injured in past 30 days	-0.05	-0.06	-0.04	-0.07	-0.06	-0.05**
	(0.15)	(0.10)	(0.05)	(0.08)	(0.06)	(0.02)
Nr. of days ill/injured in past 30 days (unable to participate in daily activities)	-0.21	-0.33	-0.17	-0.39	-0.47	-0.37
	(1.00)	(0.79)	(0.49)	(0.57)	(0.58)	(0.37)
Had accident in the mine in past 30 days	0.00	0.01	-0.01	0.00	0.01	-0.01
	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Mine fixed effects	No	No	No	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Data controls	No	No	Yes	No	No	Yes
Observations	1,916	1,916	1,916	1,916	1,916	1,916

Notes: Each cell shows the estimated effect of the Fair Gold intervention on dependent variables that are shown in the left column. Standard errors - adjusted by mine times wave - in parentheses: ** $p<0.05$

The intervention seems to improve the health of the workers. The intervention has a negative effect on the proportion of workers that were sick or injured in the past 30 days. This proportion decreases by 5 percentage point as a result of the intervention. Against a baseline of 0.22, this implies that the intervention decreases the likelihood to get ill or injured by 23 percent. The intervention also appears to have a negative effect on the number of days that the miners are ill or injured - a measure for the severity of illness or injuries- the effect is not statistically significant. The point estimates with regard to accidents in the mine are close to zero. In sum, the estimates in Table 4.6 suggest that the intervention has a positive effect on health of the respondents, but I find no evidence of a reduction of accidents in the mines.

4.6.3 Effects on human capital investment

This last subsection investigates if the intervention affects school attendance of children that are part of the miner's household. Table 4.7 shows the estimates of Equation (4.3), with the proportion of children in the household, younger than 18 years old and younger than 13 years old, respectively, that go to school as dependent variables. Table 4.7 has the same structure as Tables 4.4-4.6 above.

Table 4.7: DD estimates of the effect of Fair Gold on human capital investment

	DD estimates					
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Prop. of children (persons underage) in household attending school	0.18	0.19**	0.10**	0.15*	0.16**	0.08**
	(0.11)	(0.08)	(0.05)	(0.08)	(0.06)	(0.04)
Prop. of children younger than 13yo in household attending school	0.24*	0.25**	0.15***	0.20*	0.21***	0.11***
	(0.12)	(0.09)	(0.05)	(0.10)	(0.07)	(0.04)
Mine fixed effects	No	No	No	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	No	Yes	Yes
Data controls	No	No	Yes	No	No	Yes

Notes: Each cell shows the estimated effect of the Fair Gold intervention on dependent variables that are shown in the left column. Standard errors - adjusted by mine times wave - in parentheses: *** p<0.01, ** p<0.05, * p<0.1

The estimates in Table 4.6 show a positive effect of the intervention on school attendance. The proportion of children in the household that attend school increase by an estimated 8 percentage points - from a baseline of 0.80 - as a result of the intervention. This is based on the most accurate model of Column (6). The estimates remain relatively stable between the different models, with estimated effects between 8 and 19 percentage points. The estimates shown in Column (2), (3), (5) and (6) include controls for household size and the number of children in the household. The fact that the estimates remain stable when these controls are included, show that the estimated effect is not just a result of an increase in the number of children in the household - which is slightly higher in the treated mines (as shown in Table4.2). Focussing on the proportion of children in the age of 12 years that attend school, gives similar results. The estimated effect is 11 percentage points against a baseline of 0.76; an increase of 14 percent. The effect is thus slightly higher on the younger children. In Table 4.3 I showed that the time trend with regard to school attendance is negative for the control condition. The proportion of children that attends school drops from 79 to 60 percent. For the treated households this figure hardly changed over time. It appears that the Fair Trade intervention helped in keeping the children of miners in school.

4.7 Heterogeneity and sensitivity of the results

In this section I will further investigate the main estimation results. In the first subsection I will examine the effect of the intervention by mining site. Second, I will examine the extent to which employee tenure plays a role in the effect of the intervention. This gives us a indication about the short and long term effects of the intervention. It will also provides an indication of the effect of the high turnover rate of miners on the effect on the intervention. In the third subsection I will examine if the fact that the gender composition of the mining sites changed over time may affect the results.

4.7.1 Investigating the effects by mining site

In this subsection I investigate the effect of the intervention for each mining site separately. I conduct the difference-in-differences analysis on five different samples, with each sample consisting of the respondents of one of the five treated mines and the respondents of all control mines. Table 4.8 shows the results of this analysis. Each row shows the estimates for the four major outcome variables (Log hourly wages, PPI-score, likelihood of being sick or injured, and proportion of underage children in the household

attending school); each column shows the estimated effect of the intervention on one of the mining site. Each estimate is derived with a difference-in-differences model that includes mine-fixed effects and individual and data controls. Each column also indicates the start date of the treatment as well as the employee tenure of the mine concerned. The duration of the treatment decreases from Column (1) to Column (5).

Table 4.8: Estimated effects of the intervention on main outcome variables, by mining site

	DD estimates per mine:				
	Mine 1	Mine 2	Mine 3	Mine 4	Mine 5
Dependent variable	(1)	(2)	(3)	(4)	(5)
Log hourly wages	0.36***	-1.04***	0.20**	-0.78***	0.63***
	(0.09)	(0.07)	(0.07)	(0.07)	(0.08)
PPI-score	-1.47	-0.65	-3.04**	-0.12	-1.57**
	(1.09)	(0.78)	(1.14)	(0.60)	(0.56)
Being sick or injured (in past 30 days)	-0.10	-0.06	-0.03	0.04	-0.03
	(0.11)	(0.04)	(0.06)	(0.08)	(0.03)
Prop. of children (persons underage) in household attending school	0.46***	0.22*	0.20***	0.09	0.04
	(0.07)	(0.11)	(0.05)	(0.06)	(0.07)
Date start treatment (mm-yy)	06-12	06-12	06-12	07-12	11-12
Employee tenure (in years)	5.3	3.6	3.3	1.0	1.1
Mine fixed effects	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes
Data controls	Yes	Yes	Yes	Yes	Yes
Observations	1,121	1,125	1,161	1,159	1,154

Notes: Each cell shows the difference-in-defences estimates of the effect of intervention on dependent variables (shown in the left column) in each of the treated mining sites separately. Estimates are obtained by comparing the miners in a specific treated mine with the miners in all control mines. Each column shows the estimated effects of that specific mining site. Standard errors - adjusted by mine times wave - in parentheses: *** p<0.01, ** p<0.05

A first observation is that the wages appear to be very differently affected by the intervention between the different mining sites. The first row of Table 4.8 indicates that hourly wages - based on the 7 days prior to the interview - are positively affected in the first, third and fifth mining site, and negatively in the other two mining sites. Although it might be the result of the intervention, a more likely explanation it that this is the result of fluctuations in wages between the mines. Wages can be affected by something local such as weather conditions - e.g. heavy rain can make mining sites temporarily inaccessible - which can seriously affect the difference between two waves within one mine. An indication for this is the fact that the intervention has a consistent negative effect on the PPI-score for all mines.

With regard to the other outcomes the effects appear to be quieted consistent. The intervention appears to have a positive effect on school attendance in all five mining sites, although the effect size differs. The effects of the intervention in mining site 4 and 5, that started their participation in the project a few months later than the other three mines and are both not located in the Tarkwa region, appear to be smaller than on the other mines. This may also be a result of employee tenure, which is much lower than in the first three mines. This will be examined in the next subsection.

4.7.2 Tenure and duration of treatment

In this section I investigate the extent to which the high turnover rate affects the results, by examining the differences in outcomes between the respondents that started working in the mine relatively recently and those with a longer working history. I split the respondents up in two groups, based on the median value of tenure, with the group 'low tenure' consisting of those that have been employed in the mine for one year or less; and the group 'high tenure' consisting of those that have been employed for the specific mine for more than one year. I will also investigate the interaction effects between the intervention and employment tenure in years.

Table 4.9 shows the results of this analysis. Column (1) shows the estimated effects of the intervention on the group of miners working in the mine for one year or less; Column (2) shows the estimates on miners that work in the mine for more than a year; Column (3) shows the interaction effects between the binary variable "high tenure" - equal to 1 if the respondent worked in the mine for longer than a year or 0 otherwise - and the difference-in-differences estimates.

Table 4.9: Estimated effects of the intervention on main outcome variables, by tenure

	DD estimates		
	Employed $\leq$ 1yr	Employed $>$ 1 yr	Interaction
Dependent variable	(1)	(2)	(3)
Log hourly wages	-0.25	-0.17	0.14
	(0.27)	(0.17)	(0.17)
PPI-score	-2.13**	-2.70***	-2.08
	(0.95)	(0.93)	(1.64)
Being sick or injured (in past 30 days)	-0.03	-0.20***	-0.09
	(0.06)	(0.05)	(0.07)
Prop. of children (persons underage) in household attending school	-0.01	0.18***	0.22*
	(0.11)	(0.06)	(0.11)
Individual controls	Yes	Yes	Yes
Mine fixed effects	Yes	Yes	Yes
Data controls	Yes	Yes	Yes
Observations	974	942	1,916

Notes: Standard errors - adjusted by mine times wave - in parentheses: *** p<0.01, ** p<0.05, * p<0.1

The estimates in Table 4.9 indicate that employee tenure affects the results. Miners that have worked for more than a year in the treated mine, have a smaller likelihood to become sick or injured: The probability to become sick or injured decreases by 20 percentage points in the 'high tenure' group, whereas for the 'low tenure' miners this negative effect is 3 percentage point (and statistically insignificant). Human capital investment of respondents that have worked for a treated mine for more than a year, also increases: the proportion of children in the household that attend school increase by 18 percentage point. In the 'low tenure' group this result is close to zero. Both groups experience a decrease in income as a result of the intervention.

The analysis above is an indication that the treatment becomes more effective if the miners receive it for a longer period of time. The positive effects on human capital investment and health to materialise are visible only for the miners that work for a treated mine for at least a year. The negative effects on income are however also experienced for the miners that just start working in the mine. Note however, that this may be partially a selection effect. Miners that are e.g. risk-seeking might have left the mine within the first year, leaving only the risk-averse behind in the treated mines.

4.7.3 Influence of the change in gender composition

Section 4.5 showed that the gender composition of miners changed significantly over time in the treated mines, which may affect our results. In this subsection I hence investigate the heterogeneity of the effects of the intervention with respect to gender. Table 4.10 shows the results of this analysis. Column (1) shows the difference-in-differences estimates of the effect of the intervention for a sample of men; Column (2) shows these estimates for a sample of women; and Column (3) shows the estimates of interacting the difference-in-differences estimates with a gender (=*female*) variable. The latter can be interpreted as the additional effect of the intervention on women. Each row shows the effect on one of the four main outcome variables.

Table 4.10: Estimated effects of the intervention on main outcome variables, by gender

	DD estimates		
	Men only	Women only	Interaction
Dependent variable	(1)	(2)	(3)
Log hourly wages	-0.17	-0.50***	-0.06
	(0.22)	(0.15)	(0.20)
Log weekly earnings	-0.27	-1.02***	-0.20
	(0.32)	(0.25)	(0.36)
PPI-score	-1.67**	-2.61*	0.12
	(0.74)	(1.30)	(1.97)
Being sick or injured (in past 30 days)	-0.07	-0.04	-0.04
	(0.04)	(0.10)	(0.08)
Prop. of children (persons underage) in household attending school	0.05	0.35***	0.07
	(0.07)	(0.10)	(0.16)
Individual controls	Yes	Yes	Yes
Mine fixed effects	Yes	Yes	Yes
Data controls	Yes	Yes	Yes
Observations	1,564	352	1,916

Notes: Standard errors - adjusted by mine times wave - in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 4.10 shows that the direction of the effect of the intervention is similar for both genders: The estimated effects are negative with respect to income and the likelihood of illness/injuries and positive with respect to school attendance of children in the household. The intervention appears to have a more positive effect on women than on men with regard to the school attendance of children in the household. This increases by 35 percentage point among the female miners and only by 5 percentage point among the male miners. On the other hand, the negative effects on income also appear to be larger for the female miners. None of the differences between men and women are statistically significant.

In sum, although the effect sizes might vary between the genders, I find no evidence that the intervention affects women differently than men. The fact that the composition of the mines changed over time with regard to gender should hence not have a big effect on our estimates.

4.8 Conclusion

This paper investigated the effects of Fair Trade labour standards on income, health and human capital investments of small-scale gold miners in Ghana. I used a difference-in-differences approach, in which miners that work for treated mines - mines that are implementing the labour and environmental standards necessary for Fair Trade certification - are compared over time with miners who work for comparable non-treated mines. Data were collected over two waves from 10 mining sites - 5 treated and 5 control - in Ghana.

I find that the intervention positively affects the health of miners. The likelihood of a miner to fall ill or injured decreases by 5 percentage points. Most remarkable are the effects on school attendance. The Fair Trade standards have a positive effect on the household investments in human capital. I estimated that the proportion of children that attend school in the household of miners increases by 8 percentage point as a result of the intervention. On the other hand, I find a negative effect on income. The estimated proportion of miners living on at least USD 2.50 per day (based on the PPI-score) decreases by 5 percent as a result of the Fair Trade standards. This negative effect is visible for both genders. Hence, there appears to be a trade-off between the improved living standards as health and investment in human capital on the one hand, and income on the other. Another important finding of this study is that the turnover rate of miners is high - in both control and treatment mines - and that this affects the results. The average employee tenure is 2.6 years, with a median of 1 year, which means that approximately half of the miners work at a mining site for less than a year. Comparing the effect of the intervention

on the group of miners with less than one year of experience in the mine with the miners with a more than a year of experience, shows that the 'low tenure' miners hardly experience the positive effects of the intervention. They do not experience the positive effects on health and school attendance that senior miners do. This is an indication that the duration of the Fair Trade intervention matters. Miners who will experience the Fair Trade intervention for a longer period, will also experience more positive effects.

In sum, I find that the Fair Trade standards have a positive effect on the health of the gold miners and also has a positive effect on the children of the miners. On the other hand, the interventions appears to reduce the income of the gold miners. The latter effect might be explained by three factors. First, the improved labour conditions in the Fair Trade mines may have made the mines more attractive to work for. The reduced health risks, better facilities - in particular for women - and free provision of personal protection gear may have increased the willingness of miners to work for this mine, lowering wage demands. Second, the policies and standards to reduce risks in the miners may have directly affected wages. Following the safety standards will require extra time - e.g. wearing protection gear, attending training sessions - that may have otherwise been used for the core activity. Also the safety policies may limit specific activities that are profitable but high in risk. Hence, the intervention might have reduced the productivity in the treated mines. Third, the study investigates the effect of implementing Fair Gold labour standards on artisanal gold miners, and thus not yet of the label itself. The five mining sites that participated have not received the Fair Trade label yet - although they are currently under review. The additional benefits of receiving this label - in terms of increased sales and/or the additional premium - may reduce the negative effect on income.

Appendix 4.A Progress out of Poverty (PPI)

Progress out of Poverty Index™ for Ghana

Indicator	Value	Points	Score
1. How many members does the household have?	A. Seven or more	0	
	B. Six	6	
	C. Five	8	
	D. Four	11	
	E. Three	15	
	F. Two	23	
	G. One	31	
2. Are all children ages 5 to 12 in school?	A. No	0	
	B. Yes, or no children ages 5 to 12	4	
3. What is the highest grade completed by the female head/spouse?	A. No female head/spouse	0	
	B. None or pre-school	4	
	C. Primary or middle	7	
	D. Any JSS, SSS, S, L, U, or higher	10	
4. Is the main job of the male head/spouse in agriculture?	A. Male head/spouse has no job	0	
	B. Yes, main job is in agriculture	8	
	C. No, main job is not in agriculture	10	
	D. No male head/spouse	10	
5. What is the main construction material used for the roof?	A. Palm leaves/raffia/thatch, wood, mud bricks/earth, bamboo, or other	0	
	B. Corrugated iron sheets, cement/concrete, asbestos/slate, or roofing tiles	3	
6. What is the main source of lighting for the dwelling?	A. Not electricity (mains)	0	
	B. Electricity (mains)	5	
7. What is the main source of drinking water for the household?	A. Borehole, well (with pump or not, protected or not), or other	0	
	B. River/stream, rain water/spring, or dugout/pond/lake/dam	5	
	C. Indoor plumbing, inside standpipe, sachet/bottled water, standpipe/tap (public or private outside), pipe in neighbors, water truck/tanker, or water vendor	7	
8. Does any household member own a working stove (kerosene, electric, or gas)?	A. No	0	
	B. Yes	10	
9. Does any household member own a working iron (box or electric)?	A. No	0	
	B. Yes	6	
10. Does any household member own a working radio, radio cassette, record player, or 3-in-1 radio system?	A. None	0	
	B. Only radio	2	
	C. Radio cassette but no record player nor 3-in-1 (regardless of radio)	6	
	D. Record player but no 3-in-1 (regardless of radio or cassette)	9	
	E. 3-in-1 radio system (regardless of any others)	14	
Microfinance Risk Management, L.L.C.,		Total score	

Figure 4.4: PPI Questionnaire

Category Likelihoods according to Ghana PPI™ Score

PPI Score	$1.25/Day/2005 PPP Poverty Line		$2.50/Day/2005 PPP Poverty Line		$3.75/Day/2005 PPP Poverty Line	
	Total Below the $1.25/Day/2005 PPP Line	Total Above the $1.25/Day/2005 PPP Line	Total Below the $2.50/Day/2005 PPP Line	Total Above the $2.50/Day/2005 PPP Line	Total Below the $3.75/Day/2005 PPP Line	Total Above the $3.75/Day/2005 PPP Line
0-4	40.3%	59.7%	70.1%	29.9%	100.0%	0.0%
5-9	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%
10-14	85.4%	14.6%	95.9%	4.1%	100.0%	0.0%
15-19	75.9%	24.1%	95.1%	4.9%	100.0%	0.0%
20-24	65.5%	34.5%	90.6%	9.4%	97.1%	2.9%
25-29	48.1%	51.9%	88.0%	12.0%	97.9%	2.1%
30-34	34.0%	66.0%	81.7%	18.3%	93.6%	6.4%
35-39	16.8%	83.2%	73.5%	26.5%	91.6%	8.4%
40-44	13.4%	86.6%	66.1%	33.9%	86.2%	13.8%
45-49	8.9%	91.1%	47.0%	53.0%	76.3%	23.7%
50-54	6.0%	94.0%	42.0%	58.0%	68.4%	31.6%
55-59	1.6%	98.4%	28.0%	72.0%	60.5%	39.5%
60-64	1.2%	98.8%	9.5%	90.5%	43.2%	56.8%
65-69	1.2%	98.8%	9.0%	91.0%	32.9%	67.1%
70-74	0.0%	100.0%	9.8%	90.2%	21.9%	78.1%
75-79	0.0%	100.0%	4.2%	95.8%	9.4%	90.6%
80-84	0.0%	100.0%	0.7%	99.3%	6.4%	93.6%
85-89	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%
90-94	0.0%	100.0%	0.0%	100.0%	3.3%	96.7%
95-100	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%

Figure 4.5: PPI Conversion Table 1

Category Likelihoods according to Ghana PPI™ Score

PPI Score	National Poverty Line		National Food Poverty Line		150% of the National Poverty Line	
	Total Below the National Poverty Line	Total Above the National Poverty Line	Total Below the National Food Poverty Line	Total Above the National Food Poverty Line	Total Below the 150% of the National Poverty Line	Total Above the 150% of the National Poverty Line
0-4	40.3%	59.7%	40.3%	59.7%	40.3%	59.7%
5-9	100.0%	0.0%	85.9%	14.1%	100.0%	0.0%
10-14	88.1%	11.9%	77.9%	22.1%	91.8%	8.2%
15-19	78.5%	21.5%	64.1%	35.9%	95.1%	4.9%
20-24	68.7%	31.3%	46.3%	53.7%	84.1%	15.9%
25-29	52.9%	47.1%	37.4%	62.6%	80.7%	19.3%
30-34	40.0%	60.0%	21.8%	78.2%	65.9%	34.1%
35-39	21.4%	78.6%	9.9%	90.1%	54.4%	45.6%
40-44	17.8%	82.2%	8.1%	91.9%	48.4%	51.6%
45-49	11.0%	89.0%	4.9%	95.1%	31.1%	68.9%
50-54	9.0%	91.0%	2.7%	97.3%	29.1%	70.9%
55-59	2.0%	98.0%	1.1%	98.9%	14.7%	85.3%
60-64	1.8%	98.2%	0.3%	99.7%	6.2%	93.8%
65-69	1.2%	98.8%	0.0%	100.0%	4.2%	95.8%
70-74	0.0%	100.0%	0.0%	100.0%	3.1%	96.9%
75-79	0.0%	100.0%	0.0%	100.0%	4.2%	95.8%
80-84	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%
85-89	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%
90-94	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%
95-100	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%

Figure 4.6: PPI Conversion Table 2

5

The long-term effects of military conscription on education and wages *

5.1 Introduction

Many countries have a military draft that compels large populations of young men to spend a substantial period of time in military service. Conscripts typically have to serve in the army in a period of their life in which decisions on human capital investments are taken. A compulsory military draft might harm investments in human capital and reduce life time earnings. A number of recent studies have investigated the effect of military service on educational attainment or the effect on wages. Remarkably, these studies do not provide a consistent picture of the effects of military service. For instance, military service seems to

*Joint work with Dinand Webbink; Published in IZA Journal of Labor Economics, 4:10 (2015)

decrease educational attainment in the UK (Buonanno, 2006) and Italy (Cipollone and Rosolia, 2007)[2], but increase completion of tertiary education in Germany (Bauer et al., 2014), France (Maurin and Xenogiani, 2007) and in the US (Card and Lemieux, 2001) because of draft avoidance behaviours. In addition, military service seems to reduce wages in the US (Angrist, 1990; Angrist and Krueger, 1994) and in the Netherlands (Imbens and van der Klaauw, 1995) but in Germany (Bauer et al., 2012) there is no effect on wages and in the US the negative effects seem to fade away over time (Angrist et al., 2011). Moreover, the importance of education as a mediating channel for the long run effects on earnings remains unclear (Bauer et al., 2012).

This paper aims to provide new evidence by investigating the long term effects of military conscription on educational attainment and wages in the Netherlands. Investigating the effects of military conscription is difficult because those that have served might differ from those that have not served. Despite the fact that military conscription was compulsory for all men in the Netherlands, only 40 percent of each birth cohort actually served in the military. Those who were recruited were selected from a larger population and the various decisions made in this selection process are unobserved. This might induce selection bias if we compare the outcomes of those who served in the military with the outcomes of those that did not serve in the military.

To address this selection problem we exploit a policy change that created a major difference in conscription between birth cohorts. In 1979 the age that Dutch young men were called for military service was lowered from 20 to 19. The direct consequence of this policy change was that a whole birth cohort was exempted from military service. We identify the causal effect of military service by comparing the long term outcomes of this exempted birth cohort with the outcomes of those born in the adjacent years. This local comparison enables us to generate two types of estimates. First, reduced form estimates of the difference between the exempted cohort and the adjacent cohorts that received the regular treatment of conscription provide a direct estimate of the societal costs of a system of conscription. The societal costs might consist of a lower educated population with a lower earnings capacity. Second, instrumental variable estimates show the effect of conscription for males that actually served in the army. These estimates for the compliers provide insight in the private costs of conscription. This approach is related to earlier work by Imbens and van der Klaauw (1995). They introduced an instrumental variable approach that exploits all variation in conscription between fourteen birth cohorts

[2] A recent study finds no effect on university enrolment of the abolition of compulsory military service in Italy (Di Pietro, 2013).

for obtaining estimates of the short term wages effects for conscripts. We also apply their approach for testing the robustness of our main results on the long term effects of conscription on three education outcomes and wages using micro-level data from 1997 to 2002. Moreover, data on educational attainment and earnings of women enable us to perform a placebo test about the influence of other time related confounding factors.

This paper contributes to the recent literature on the effects of military service. First, by exploiting variation in conscription that is transparent and arguably exogenous in a local approach we are able to obtain estimates of both the societal and private costs of a system of compulsory military service. Second, we are able to trace the effects of conscription on investment in human capital. In particular, we are able to investigate the effects of military conscription on completion of university education. We find that the system of compulsory military service decreases the proportion of university graduates with 1.5 percentage points from a baseline of 12.3 percent. In addition, being a conscript reduces the probability of obtaining a university degree with almost four percentage points. Third, we investigate the long term effects of conscription on the average societal wages and the wages of conscripts. We find that the system of military service reduces average societal wages with 1.5 percent. In addition, conscripts lose approximately four percent of their wages by serving in the military. This suggests that the negative effects of military service are long-lasting. Finally, we investigate to which extent the wage costs of conscription can be explained by the decrease in investment in human capital. We find that the effect of conscription on educational attainment does not fully explain the wage reduction. This suggests that conscription reduces individual earnings capacity also through other channels than a reduction in human capital.

The remainder of this chapter is organized as follows. The next section will provide background information about military conscription in the Netherlands. Section 5.3 discusses previous research on the effects of military conscription. Section 5.4 explains our empirical strategy. The data are described in Section 5.5. Section 5.6 shows the main estimation results and Section 5.7 shows the results of several sensitivity tests. Section 5.8 concludes.

5.2 Background

The Netherlands has had a system of military conscription for almost two centuries. It was introduced by Napoleon in 1810, when the Netherlands were under French rule. It was only in the 1990s that Dutch

politicians started seriously questioning the need for a large drafted army. The fall of the Berlin wall and the end of the Cold War had diminished the fear for a military invasion. Eventually military conscription was abandoned in the Netherlands in 1997.[3]

Military conscription is part of the political ideal of the nation-state, in which all citizens alike have the duty to protect their fatherland. The advantages of the draft system for the state are evident; the supply of cheap labour provides the state with a large army. However, military conscription system is also criticised. First, the draft system might induce an inefficient match of jobs and people. Recruits are rather arbitrary allocated to a job, without proper consideration of productivity, comparative advantages and outside options (Lau et al., 2004). This inefficiency might also be linked to the fact that serving in the military was for many recruits a long period of boredom (Duindam, 1995). The second implication of the draft is that it can be regarded as a hidden tax, unequally distributed among the population. The state forces young men to work for a certain amount of time for wages that are probably much lower than they could have earned in the labour market. This difference can be regarded as a tax for the young men that have to serve (Lau et al., 2004; Poutvaara and Wagener, 2007; Poutvaara and Wagener).

In this study we investigate the long term effects of the Dutch military draft system from the period between 1970 and 1990. Since the Netherlands were not involved in any wars during this period our analysis is about the effects of peace time conscription. During this period the military consisted for almost 50 percent of conscripts. At the age of 17 Dutch males received a letter from the Dutch Ministry of Defence about the registration as a recruit and duties and procedures of conscription[4]. The medical and psychological examination took place in the year the potential recruit turned 18. On average 30 percent of the potential recruits were immediately exempted from military service. The reasons to be exempted varied. These could be serious medical conditions or 'psychological instability' but also less obvious 'medical conditions' such as being at least 2 meters tall. The 'psychological instability' exemption became notorious, since many attempted to be exempted via this route by pretending to be mentally ill. Those who passed the medical examinations were considered fit to serve, and were supposed to wait until they were called for service. During this period the recruit could still opt for other reasons to be exempted. Some reasons to be exempted were being enrolled in tertiary education or being the sole breadwinner in the household. Although most exemptions were temporary, in practice they often became

[3] Although the Netherlands abandoned military conscription in 1997 every Dutch man still has a military obligation. Since 1997 however, recruits are no longer called to serve. We may regard this as a 100 percent exemption rate.

[4] This description refers to the the procedures and ages in regard to the military conscription after the policy change in 1977

full exemptions since men were rarely called for service after their 26^{th} birthday. A special case was the conscientious objector, who refused to perform military service out of principle. Conscientious objectors were granted an alternative social or civilian conscription. Those considered fit to serve were called for military service in the year they turned 19. The recruits called for service generally served in the military for 14 months. About 15 percent of the recruits were invited to be trained as officers. This meant slightly better pay, but also serving for 16 instead of 14 months.

From 1969 the military was confronted with a structural surplus of recruits, caused both by a decreased demand -a result of budget cuts which allowed a maximum of 45,000 recruits annually - and an increasing supply of recruits because of the post-war birth wave (Hoffenaar and Schoenmaker, 1994). A number of policies were implemented to reduce this surplus. The first was the introduction of the so called 'special exemption' category, which consisted of a random selection of men from those who were fit to serve. Being 'specially exempted' implied that the recruit was not called for duty, except in case of emergency. The second policy was a change in the conscription age, which was lowered from 20 to 19 from 1979 onwards. Although the official reason for this change in policy was to improve the transition from secondary school to military service, an important side effect was that birth cohort 1959 was completely exempted from military service. In 1978 all those turning 20 that year were called for military service (birth cohort 1958) while in 1979 all those turning 19 were called for service (birth cohort 1960). We will exploit this shock in our empirical model.

5.3 Previous studies

5.3.1 Conscription and educational attainment

Several studies have assessed the effects of military conscription on educational attainment. In general, the evidence seems mixed as some studies report a negative effect of military conscription whereas other studies find that military conscription induces an increase of enrolment in tertiary education. Keller et al. (2009) compare European countries since the 1960s and find a strong negative association between enrolment in tertiary education and the share of recruits in a country. Buonanno (2006) finds that that exemption from military service in the UK increases educational attainment with approximately a quarter of a year. However, for Italy no effect of military conscription on enrolment in tertiary education is found, except for recruits from lower socio-economic backgrounds (Di Pietro, 2013). Cipollone and Rosolia

(2007) exploit regional variation in military conscription in Italy and find that military conscription also has a negative impact on male high school graduation. Remarkably, they also find a positive effect on graduation rates of girls from the same areas and cohorts. Positive effects of military conscription on completion of tertiary education have been reported for France (Maurin and Xenogiani, 2007), for the US (Card and Lemieux, 2001) and for Germany (Bauer et al., 2014). These effects seem to be driven by draft avoidance behaviour.

5.3.2 Conscription and earnings

A number of studies have focused on the effect of military conscription on future wages. These studies find mixed effects, ranging from zero to a negative effect. Imbens and van der Klaauw (1995) find that being conscripted in the Netherlands reduces earnings ten years after conscription with almost 5 percent. Findings for the United Kingdom suggest that military conscription reduces wages with 4 to 7 percent (Buonanno, 2006). However, several other studies do not find an effect of military conscription on wages. For instance, Grenet et al. (2010) do not find a reduction of wages for conscripts in the United Kingdom. In addition, peace time conscription seems not to harm future wages in Germany (Bauer et al., 2014, 2012). Moreover, for Australian recruits no negative wage effects have been found of military conscription in the Vietnam-era (Siminski, 2013). For the US there seem to be a negative short term effect that fades out over time. Angrist (1990) shows that ten years after service during the Vietnam war veterans earn 15 percent less than those who didn't serve. However, in the long run these adverse consequences of conscription seem to fade out (Angrist et al., 2011). A study of the medium and long term effects of conscription in the Second World War finds no significant differences between veterans and non-veterans (Angrist and Krueger, 1994).

In sum, previous empirical studies suggests that military service, both in peace as in war time, has either no effect or a negative effect on future earnings. The negative effects might result from the effect of conscription on human capital. Conscription might diminish individual returns on human capital, which in turn affects future wages of a recruit (Lau et al., 2004; Poutvaara and Wagener, 2007; Poutvaara and Wagener). Young men are typically called for military service during a period of their lives that would otherwise be devoted to learning or gaining work experience. In addition, human capital accumulated before the draft might depreciate during military service. If military service causes a break in the educational career it is likely that more time will be needed for completion of tertiary education and might therefore reduce the probability of enrolment and completion of tertiary education. On the

other hand, the conscription system in the Netherlands gave recruits the opportunity to postpone their military service duties until they finished their tertiary education. In practice this often led to a full exemption as a recruit older than 26 was considered too old to serve in the military (Hoffenaar and Schoenmaker, 1994). Draft-avoidance behaviour may therefore just as well have led to an increase in enrolment in tertiary education in the Netherlands.

5.4 Empirical strategy

This study aims to estimate the effect of military conscription on education and wages. The most straightforward way to estimate the effect of conscription would be to regress the outcome variables on a dummy variable for serving in the military using the following equation:

$$Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \varepsilon_i \tag{5.1}$$

In which Y_i is the dependent variable, for instance earnings or education, S_i is dummy variable equal to 1 if individual i served in the military and 0 otherwise and X_i is a vector of covariates. The regression above would give an unbiased estimate if serving in the military could be considered as a random event. This is unlikely to be the case because of the complicated selection procedures and the various opportunities to be exempted from military service. For instance, individuals with specific - unobserved - skills may have been more successful in exploiting these exemptions than others. We may therefore not assume that S_i is uncorrelated with ε_i.

To address this endogeneity problem we exploit a policy change that exempted a complete birth cohort from being conscripted. From 1979 onwards the age that men were called for military conscription was lowered from 20 to 19. As a consequence, the complete birth cohort of 1959 was exempted from military service. Hence, this policy change creates a transparent counterfactual for regular cohorts that had to serve in the army. This variation can be considered as random because it is driven by a governmental decision which is likely to be unrelated with individual (unobserved) factors. We can assess the effect of conscription by comparing the outcomes of the cohorts born around 1959. For this comparison we estimate reduced form models that include a dummy for being born in 1959 ($bc1959_i$) and controls for a smooth function age ($f(A_i)$) and other covariates (X_i) :

$$Y_i = \delta_0 + \delta_1 bc1959_i + f(A_i) + \delta_2 X_i + \nu_i \tag{5.2}$$

The parameter δ_1 captures the effect of being born in 1959. This effect can be attributed to the system of conscription if we assume that the schooling and earnings capabilities of males born in 1959 are not different from the capabilities of males born in the adjacent years before and after 1959. If $bc1959_i$ is a dummy variable equal to 1 if individual i is born in 1959 and 0 otherwise, the estimates of δ_1 can be interpreted as the societal effect of not having a system of compulsory military service. For instance, if the average wage of males born in 1959 are higher than the wages of those born in adjacent years this can be interpreted as the societal costs of a system of conscription. The adjacent cohorts would have earned the same wages as males born in 1959 if there had not been a system of compulsory military service. It should be noted that the policy change exploited in this paper creates a discontinuity in the probability of serving in the army that had duration of one year only. After this year the probability of serving in the army returned to the regular level. In the standard regression discontinuity models, as for instance applied by Bauer et al. (2012), the probability of serving in the army only changes once. Our application provides the opportunity to use before and after cohorts for comparisons. This makes the estimates less vulnerable for functional form assumptions, for instance about the specification of age. In addition, we can apply a local approach by using discontinuity samples of birth cohorts around 1959 for our estimation.[5] Estimation of Equation (5.2) is based on the exclusion restriction that the change in outcomes between birth cohort 1959 and the adjacent birth cohorts can be fully attributed to military conscription. In our sensitivity analysis we will test this assumption by investigating differences between cohorts of women that were not subject to the conscription laws.

Imbens and van der Klaauw (1995) applied an approach in the proportion of men that served in a specific birth cohort ($\bar{S}_i$) was used as an instrument:

$$\begin{aligned} S_i &= \alpha_0 + \alpha_1 \bar{S}_i + f(A_i) + \alpha_2 X_i + \varepsilon_i \\ Y_i &= \beta_0 + \beta_1 \hat{S}_i + f(A_i) + \beta_2 X_i + \eta_i \end{aligned} \tag{5.3}$$

In innovative feature of the IV-approach of Imbens and van der Klaauw (1995) is that they were

[5] Gelman and Imbens (2014) advocate the use of local approaches in regression discontinuity models and show that estimates based on higher-order polynomials of the forcing variable can be misleading.

able to apply this approach without observing whether an individual had served in the army. Their key insight is that the expectation of S_i conditional on the proportion that served from their birth cohort and covariates is equal to the proportion of the birth cohort that actually served because all regressors remain constant within each cohort, hence $E(S_i|\bar{S}_i X_i) = \bar{S}_i$. This implies that $\alpha_1 = 1$, $\alpha_0 = \alpha_2 = 0$, and $\hat{S}_i = \alpha_1 \bar{S}_i = \bar{S}_i$. Hence, the causal effect of conscription can be estimated with the following equation:

$$Y_i = \beta_0 + \beta_1 \bar{S}_i + f(A_i) + \beta_2 X_i + \eta_i \tag{5.4}$$

Imbens and van der Klaauw (1995) estimated this equation using aggregated wage data (monthly birth cohort averages) from 1989 and 1990. We will estimate a local approach of the model of Imbens and van der Klaauw (1995) focussing on birth cohort 1959 and the adjacent cohorts. This approach is similar to Equation (5.2), but uses $\bar{S}_i$ instead of the variable $bc1959_i$. For the estimation we use discontinuity samples that include only one, two or three birth cohorts around 1959. In our sensitivity analysis we replicate the models of Imbens and van der Klaauw (1995) to test our main assumptions by performing 'placebo tests' using data of women.

In sum, we estimate reduced form models of the effect of conscription on education and wages (Equation (5.2)). These estimates can be interpreted as the societal impact of conscription. In addition, we estimate the long term effects for conscripts by using an instrumental variable approach (Equation (5.4)).

5.5 Data

The data that are used in the analysis come from the so-called Wage Structure Survey (LSO)[6] that includes individual-level data on wages, education and background characteristics. Data on wages are obtained through the annual survey on employment and wages among firms and partly through administrative data on the insured workers. This means that all information on wages comes from administrative sources (firms or administrative data on insured people). This dataset also contains information on gender, age and job characteristics. Data on education are obtained from the annual labour force survey and are matched with the wage data. This matched data-set is called the Wage Structure Survey. We obtained data from the surveys of 1995, 1996, 1997 and 2002. These yearly surveys consist of approximately

[6] "LSO" is an abbreviation of the Dutch word "Loonstructuuronderzoek"

125,000 individuals.

5.5.1 The outcome variable

For our first outcome variable 'educational attainment we can use three indicators: years of education and completion of two types of tertiary education (higher professional education or university education). Education is measured with a 16-point scale which is translated into years of education. Dutch tertiary education consists of two levels: higher professional education and university education. The first type of tertiary education has a vocational orientation and is considered to be the lowest level. University education has an academic orientation. Our second outcome variable is (log) annual wages. Our data contain the gross annual wages of 1995, 1996, 1997 and 2002. The data on wages of 1995-1997 have been converted from guilders to euros.

5.5.2 The independent variable

In the analysis we exploit variation in the proportion of conscripts between birth cohorts for identifying the effect of conscription. The year of birth has been registered in our data. In addition, we use information on gender and the year of the survey. Information about the proportion of males per birth cohort that served in the army was derived from Imbens and van der Klaauw (1995) who obtained this information from the Dutch Ministry of Defense. Data for the years since 1969 proved not to be available at the Dutch Ministry of Defense. The proportion of men that served in the army for the birth cohorts 1956 to 1969 is shown in Figure 5.1. The proportion of men that received a special exemption status is shown in Figure 5.2.

Summary statistics of the main variables are shown in Table 5.1 and Figures 5.3 and 5.4. These statistics also give a first impression of the effect of conscription on educational attainment and earnings. Table 5.1 shows summary statistics after pooling the four yearly LSO datasets. The four columns show the means for males born in 1959 and the means for the two, four or six adjacent male birth cohorts. We observe that males born in 1959 more often complete tertiary education, especially university education, than males born in the adjacent years (columns (2), (3) and (4)). In addition, we observe that the annual earnings of males born in 1959 are higher than the annual earnings of those born in the adjacent birth cohorts. Figure 5.3 shows the proportion of the Dutch population that completed tertiary education by birth cohort. The vertical bars are based on a regression of completion of tertiary education on birth

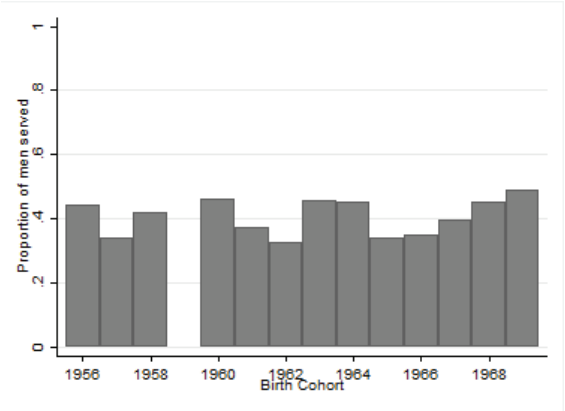


Figure 5.1: Proportion of Dutch men that served in the military by birth cohort

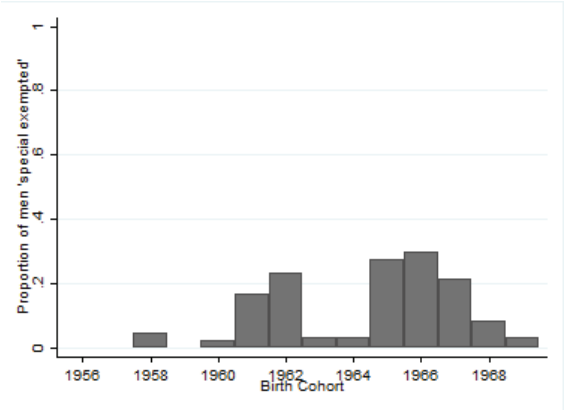


Figure 5.2: Proportion of Dutch men in the category 'Special Exemption' by birth cohort

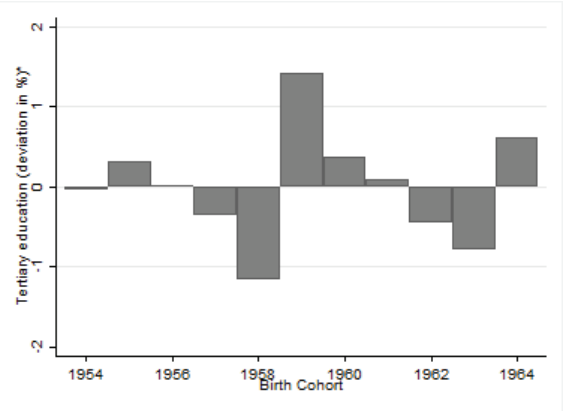


Figure 5.3: Completion of tertiary education of males by birth cohort adjusted for time trend

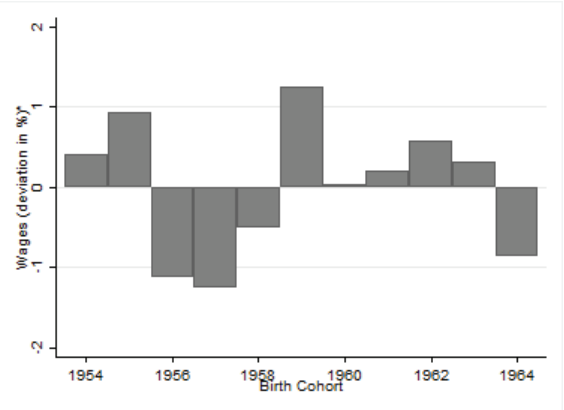


Figure 5.4: Log gross annual wages of males by birth cohort adjusted for time trend

cohort adjusted for a quadratic function of age. We observe that birth cohort 1959 deviates most from the expected value based on the age profile. Figure 5.4 shows the gross average wages of men per birth cohort. This graph is based on the combined LSO data of respectively 1995, 1996, 1997 and 2002. As in Figure 5.2, the bars in Figure 5.4 are based on a regression of wages on birth cohorts adjusted for a quadratic function of age. Figure 5.4 shows that the average annual income for those born in 1959 is above the earnings prediction based on a quadratic age curve. The summary statistics shown in Table 5.1 and in Figures 5.3 and 5.4 suggest a negative effect of conscription on educational attainment and earnings.

Table 5.1: Summary statistics

	(1)	(2)	(3)	(4)
Conscription	No	Yes	Yes	Yes
Birth cohort	cohort 1959	± 1 cohort	± 2 cohorts	± 3 cohorts
Age	38.1	38.0	38.0	38.1
Years of completed education	13.3	13.3	13.3	13.2
Tertiary education (%)	34.4	32.6	32.6	32.6
Higher professional education (%)	22.1	21.7	21.8	21.8
University degree (%)	12.3	10.8	10.8	10.8
Gross annual wages (in 1000 euro)	28.5	28.0	28.0	28.0
Served in the army (%)	0	44.1	39.9	39.4
Special exemption status (%)	0	3.5	5.9	7.7

Note: Columns (2), (3) and (4) contain respectively one, two or three cohorts before and after birth cohort 1959, but do not include cohort 1959.

5.6 Main estimation results

This section shows the estimates of the models that only exploit the variation due to the policy change that exempted all men born in 1959 from military service. The estimates are obtained by using local discontinuity samples of one, two or three years around the exempted cohort.

5.6.1 Conscription and educational attainment

Table 5.2 shows the estimated effects of military conscription on educational attainment using the local approach. We use three different schooling outcomes: years of completed schooling (top panel), obtaining a degree of higher professional education (middle panel) and obtaining a university degree (bottom panel). As mentioned before, tertiary education in the Netherlands consists of two levels: higher professional education and university education. We look specifically at the effects on tertiary education as the transition from secondary education coincides with the period that Dutch conscript are called to serve in the army. In addition, completion of tertiary education might be a route for avoiding military conscription. Each panel of Table 5.2 shows reduced form estimates and IV-estimates for discontinuity samples of one, two and three years around 1959. For each discontinuity sample we show the results of a linear and quadratic specification of age.

Table 5.2: Reduced form and IV estimates of the effect of military service on investments in human capital

	±1 birth cohort		±2 birth cohorts		±3 birth cohorts	
Years of completed education	(1)	(2)	(3)	(4)	(5)	(6)
Reduced Form	-0.019	-0.026	-0.041	-0.044	-0.059	-0.064
	(0.044)	(0.044)	(0.040)	(0.041)	(0.039)	(0.040)
IV	-0.043	-0.059	-0.064	-0.069	-0.084	-0.091
	(0.10)	(0.10)	(0.098)	(0.099)	(0.095)	(0.095)
Higher professional education (%)	(1)	(2)	(3)	(4)	(5)	(6)
Reduced Form	-0.003	-0.004	-0.003	-0.002	-0.003	-0.002
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
IV	-0.008	-0.009	-0.006	-0.005	-0.006	-0.005
	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)	(0.014)
University education (%)	(1)	(2)	(3)	(4)	(5)	(6)
Reduced Form	-0.015***	-0.016***	-0.015***	-0.015***	-0.015***	-0.016***
	(0.005)	(0.0051)	(0.0046)	(0.0046)	(0.0044)	(0.0044)
IV	-0.033***	-0.035***	-0.035***	-0.035***	-0.030***	-0.033***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Age polynomials	linear	quadratic	linear	quadratic	linear	quadratic
Observations	17,605	17,605	29,228	29,228	40,481	40,481

Notes: Each cell is based on a separate regression of education on military conscription as specified in Equations (5.2) and (5.4). Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

All point estimates in Table 5.2 suggest a negative effect of conscription on investment in human capital. However, the estimated effects on years of schooling and on the probability of obtaining a degree of higher professional education are statistically insignificant. The estimated effects on obtaining a university degree show a different picture; military conscription harms the probability of obtaining a university degree. Males born in 1959 are 1.5 percentage point more likely to obtain a university degree than those born in the adjacent three years. This implies that the system of military conscription causes a reduction of university education completion from 12.3 to 10.8 percent. If there had not been a system of military conscription the proportion of male university graduates in the adjacent three cohorts before and after 1959 would have been 1.5 percentage points higher. This reduction probably induces substantial societal costs as tertiary education is often considered to be important for social welfare and economic growth (Hanushek and Woessmann, 2008). The IV-estimates are obtained through a rescaling of the reduced form estimates with the proportion of conscripts per birth cohort. These estimates suggest that the impact for the compliers, those that would serve if they were not born in 1959, is approximately 3-4 percentage points. The estimates are robust for the discontinuity sample or the specification of age. The estimated effects do not change when we use non-linear models instead of linear probability models.

In sum, we find that military conscription seems to have a substantial societal cost by reducing the proportion of young males that obtain a university degree in the Netherlands. In addition, conscription also generates private costs by reducing the probability of completing university education for conscripts. Our results are different from earlier studies that found positive effects of military conscription on educational attainment because of draft avoidance behavior (Bauer et al., 2014; Maurin and Xenogiani, 2007; Card and Lemieux, 2001). However, other studies also found negative effects of conscription on educational attainment (Di Pietro, 2013; Buonanno, 2006). A factor that might be important in the Dutch context is that students that (temporarily) drop-out from their university education have to fulfill their military service before being allowed to re-enroll. This might have affected the proportion of university graduates as this creates an interruption in the educational career of these students.

5.6.2 Conscription and earnings

The estimated effects of military conscription on (log) annual wages are shown in Table 5.3 using the same models and estimation samples as in Table 5.2. The reduced form estimates show that males born in 1959 earn on average 1.5 percent more than those born in the adjacent one, two or three years. This implies that an abolishment of the system of compulsory military service would increase average male

earnings for these cohorts with 1.5 percent. At the individual level, the long term cost of conscription for compliers is approximately 3 to 4 percent. The average age in our sample is approximately 38 years, which implies that 18 years after conscription men that had to serve as recruits still earn 3-4 percent less than those who didn't. Imbens and van der Klaauw (1995) estimate a negative effect of 4-5 percent on wages 10 years after conscription.

Table 5.3: Reduced form and IV estimates of the effect of military service on log annual wages [1]

	± 1 birth cohort		± 2 birth cohorts		± 3 birth cohorts	
Log annual wages	(1)	(2)	(3)	(4)	(5)	(6)
Reduced form	-0.013*	-0.013*	-0.017**	-0.015**	-0.019***	-0.017**
	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
IV	-0.031*	-0.029*	-0.036**	-0.032*	-0.038**	-0.033**
	(0.017)	(0.018)	(0.017)	(0.017)	(0.017)	(0.017)
Log annual wages	(1)	(2)	(3)	(4)	(5)	(6)
Reduced form	-0.013*	-0.011	-0.015**	-0.013*	-0.016**	-0.013**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
IV	-0.028*	-0.026	-0.033**	-0.029*	-0.034**	-0.028*
	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)
Control: Years of education	0.051***	0.051***	0.049***	0.049***	0.049***	0.049***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log annual wages	(1)	(2)	(3)	(4)	(5)	(6)
Reduced form	-0.008	-0.007	-0.012*	-0.010	-0.014**	-0.011*
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
IV	-0.019	-0.017	-0.024	-0.021	-0.028*	-0.022
	(0.017)	(0.017)	(0.017)	(0.017)	(0.016)	(0.016)
Control: University education	0.348***	0.348***	0.339***	0.339***	0.331***	0.331***[1]
	(0.011)	(0.011)	(0.009)	(0.009)	(0.008)	(0.008)
Age polynomials	linear	quadratic	linear	quadratic	linear	quadratic
Observations	17,605	17,605	29,228	29,228	40,481	40,481

Notes: Each cell is based on a separate regression of wages on military conscription as specified in Equations (5.2) and (5.4). Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

5.6.3 Conscription, education and earnings

Next we investigate to which extent the negative effect of conscription on wages is driven by the effects on educational attainment. In Table 5.3 estimates are shown of the 'wage models that include years of schooling (middle panel) or university degree (lower panel) as additional controls. These controls aim to pick up the effects of conscription that work through education. Hence, the remaining effects of conscription on wages are not driven by differential investments in human capital. The estimates show that the estimated effect of conscription reduces after controlling for years of schooling or controlling for completion of university education. However, the effect of conscription on earnings remains substantial and statistically significant. This suggests that conscription reduces individual earnings capacities also through other channels than human capital obtained in the schooling system.

5.7 Sensitivity analysis

In this section we investigate the robustness of the previous findings. First, we will estimate the same models as Imbens and van der Klaauw (1995) using individual level data until 2002. These models not only exploit the variation from the birth cohort 1959 but also the variation in military conscription between other birth cohorts. Second, we re-estimate the main models for the sample of women. As women are exempted from military service we do no expect differences between birth cohorts. This analysis can be considered as a placebo test for our main estimates. Third, we will investigate how the drop in university graduates for birth cohort 1958 affects our estimates.

5.7.1 Replication of Imbens and van der Klaauw (1995)

Imbens and van der Klaauw (1995) exploited two types of variation. First, they used the variation in the proportion of conscripts from 14 birth cohorts. Second, they used the variation in the proportion that received a special exemption status. The special exemption status was first introduced for birth cohort 1958. Therefore, Imbens and van der Klaauw (1995) did not use the birth cohorts 1956 and 1957 in the analysis that exploited this type of variation. In our replication we use exactly the same variation and the same birth cohorts as Imbens and van der Klaauw (1995). Table 5.4 shows the estimation results of the effects of military conscription on educational attainment and earnings using the same specifications as Imbens and van der Klaauw (1995). The estimation sample consists of individuals from 14 birth cohorts.

The estimation results in the first row for each dependent variable (conscripted) exploit variation in the proportion of individuals within birth cohorts that actually served. The estimation results in the second row (status 'special exemption) exploit variation in the proportion of individuals within birth cohorts that did not serve due to a 'special exemption status. We estimated Equation (5.4) from Section (5.4) which included the proportion of young men that served or the proportion of young men that had a special exemption status within birth cohorts as the main independent variable. We estimated these models using a linear or a quadratic specification of age.

The estimation results in Table 5.4 confirm the previous findings. Serving in the army has a negative, but statistically insignificant, effect on years of schooling and graduation from higher professional education. Serving in the army reduces the probability of obtaining a university degree and also reduces long term earnings. These effects are statistically significant for both types of variation with one exemption. The size of the effects on university completion is somewhat smaller than the results from the local approach. The effects on wages are comparable when using a quadratic specification. In sum, exploiting all variation between birth cohorts gives similar results but the effects on university education are smaller. This implies that the variation used from cohorts further away from the policy change that affected the birth cohort of 1959 has less impact on the completion of university education. This variation is, however, less well understood than the variation used in the analysis of Section 5.6.

5.7.2 Placebo test on women

Conscription was compulsory for men, but not for women. If other factors are important for the schooling or wage differences between birth cohorts these factors might also affect schooling and earnings of women. To investigate the importance of other factors we re-estimate the main models from Tables 5.2 and 5.3 using a sample of women only. The results are shown in Table 5.5. The estimates of the effect on obtaining a university degree are close to zero and sometimes positive. This suggest that the difference in obtaining a university degree for males born in 1959 and males born in adjacent years is not driven by unobserved factors that increase university enrolment for those born in 1959. We also do not find statistically significant effects on female wages. Hence, we do not find evidence for the importance of unobserved factors that might have an effect on educational attainment or earnings.

Table 5.4: IV-estimates of the long term effect of conscription on education and earnings using the model of Imbens and van der Klaauw (1995)

Years of completed education	(1)	(2)
Conscripted	-0.080	-0.075
	(0.085)	(0.086)
Status 'special exemption'	0.046	0.046
	(0.051)	(0.051)
Higher professional education	**(1)**	**(2)**
Conscripted	-0.013	-0.012
	(0.013)	(0.013)
Status 'special exemption'	0.010	0.010
University education	**(1)**	**(2)**
Conscripted	-0.022**	-0.018*
	(0.0097)	(0.0097)
Status 'special exemption'	0.014**	0.015***
	(0.0058)	(0.0058)
Log annual wages	**(1)**	**(2)**
Conscripted	-0.078***	-0.036**
	(0.015)	(0.015)
Status 'special exemption'	0.014	0.019**
	(0.0090)	(0.0090)
Age polynomials	linear	quadratic

Notes: Each cell is based on a separate regression and uses variation in the proportion of conscripts or variation in the proportion having special exemption status between cohorts. The number of observations in the first regression is 73,892, for the second type of regressions this is 62,243 (the birth cohorts of 1956 and 1957 are not included). Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Table 5.5: Reduced form and IV estimates of conscription on sample of women (sensitivity)

	±1 birth cohort		±2 birth cohorts		±3 birth cohorts	
University education (%)	(1)	(2)	(3)	(4)	(5)	(6)
Reduced Form	0.007	0.009	0.004	0.005	0.003	0.004
	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)
IV	0.008	0.010	0.002	0.002	-0.001	-0.000
	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)	(0.011)
Log annual wages	(1)	(2)	(3)	(4)	(5)	(6)
Reduced Form	-0.013	-0.014	-0.007	-0.008	-0.003	-0.008
	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)	(0.011)
IV	-0.028	-0.031	-0.023	-0.025	-0.016	-0.025
	(0.029)	(0.030)	(0.029)	(0.029)	(0.027)	(0.028)
Age polynomials	linear	quadratic	linear	quadratic	linear	quadratic
Observations	12,761	12,761	21,448	21,448	29,960	29,960

Notes: Each cell is based on a separate regression of education and wages on military conscription as specified in Equations (5.2) and (5.4). Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

It should be noted that this placebo test is not completely clean as women may have been indirectly affected by military conscription. For instance, women might face less competition from men on the labor markets due to conscription. Moreover, for married women their might be substitution or income effects related to the direct effects of conscription on their husbands. These indirect effects will probably be more important for womens wages than for the education of women because of the timing of the effects.

5.7.3 A lower proportion of university graduates in birth cohort 1958

Our model relies on the assumption that men born in birth year 1959 are similar to those in the adjacent years. However, Figure 5.3 shows that the proportion of university graduates is remarkably lower for those born in 1958 than in adjacent birth cohorts. This may affect our estimates of the effect of conscription on university graduation. Therefore we conduct two sensitivity analyses. First, we re-estimate the effect of military service on the probability to complete university education while excluding birth cohort 1958 and including an earlier birth cohort. The estimates are shown in the upper panel of Table 5.6. Second, we pool the data of males and females and estimate specifications of Equations (5.2) and (5.4) that include a dummy for males and an interaction between birth cohort 1959 and males. Hence, this provides a difference-in-differences estimate of the effect of conscription on completion of education. The lower panel of Table 6 shows the difference-in-differences estimates for Equations (5.2) and (5.4). We observe that the estimates in both panels are very similar to the estimates in Table 5.2. This implies that the drop in the proportion of university graduates of birth cohort 1958 does not affect the main estimation results.

Table 5.6: Sensitivity analysis of the effect of conscription on university graduation

	±1 birth cohort		±2 birth cohorts		±3 birth cohorts	
Panel 1: Excluding birth cohort 1958	(1)	(2)	(3)	(4)	(5)	(6)
Reduced Form	-0.014***	-0.015***	-0.013***	-0.014***	-0.015***	-0.016***
	(0.0052)	(0.0052)	(0.0046)	(0.0047)	(0.0044)	(0.0045)
IV	-0.033***	-0.036***	-0.028**	-0.031***	-0.030***	-0.033***
	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)	(0.011)
Observations	17,527	17,527	29,059	29,059	40,367	40,367
Panel 2: Difference-in-differences estimates	(1)	(2)	(3)	(4)	(5)	(6)
Reduced form	-0.018**	-0.018**	-0.015**	-0.015**	-0.014**	-0.015**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
IV	-0.040**	-0.040**	-0.039**	-0.039**	-0.031**	-0.031**
	(0.017)	(0.017)	(0.016)	(0.016)	(0.015)	(0.015)
Age polynomials	linear	quadratic	linear	quadratic	linear	quadratic
Observations	30,366	30,366	50,676	50,676	70,441	70,441

Notes: In Panel 1 each cell is based on a separate regression of education on military conscription as specified in Equations (5.2) and (5.4) on a sample excluding birth cohort 1958 and including an earlier birth cohort. In Panel 2 each column shows the difference-in-differences estimates of the effect of conscription using the pooled sample of men and women. Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

5.8 Conclusions

This paper investigated the long-term effect of military conscription on educational attainment and wages by exploiting a policy change that exempted a complete birth cohort from military service. We compare the educational outcomes and earnings of the exempted cohort with the outcomes of men from adjacent cohorts. This local approach yields estimates of the societal costs of a system of military conscription and estimates of the private costs for individuals that had to serve in the military. Our approach is related to previous work by Imbens and van der Klaauw (1995) who investigated the effects of conscription on wages of conscripts until 1990.

We find that the system of compulsory military service decreases the proportion of university graduates with 1.5 percentage points from a baseline of 12.3 percent. In addition, being a conscript reduces the probability of obtaining a university degree with almost four percentage points. Our estimates also show that the system of military service reduces average societal wages with 1.5 percent. The private costs for conscripts are higher; they lose approximately 3 to 4 percent of their wages by serving in the military. The fact that the average man in our sample served in the army almost 18 years before, suggests that the negative effects of military service are long-lasting. Finally, we find that the effect of conscription on educational attainment does not fully explain the wage reduction. This suggests that conscription reduces individual earnings capacity also through other channels than a reduction in human capital.

This study provides a new piece of evidence about the hidden costs of conscription. Our estimates show that military conscription has long term negative consequences for completion of university education and for individual earnings. This implies that the costs of conscription are substantial, both at the societal level as at the individual level. Moreover, the private costs of conscription seem to be long-lasting.

Summary

This dissertation contains four empirical studies that investigate the impact of specific private contributions to public outcomes or the specific determinants of philanthropic giving. Although many studies exists in this field, only a very limited proportion of these actually contain an approach rigorous enough to determine causal relationships. In general, a multitude of factors might influence philanthropic giving; and the impact of philanthropic effects might be influenced by many other observed or unobserved factors. As such, estimates of the causal effect of specific determinants or of private contributions to public goods might be biased by omitted variables. The contribution of this dissertation is that all studies apply a quasi-experimental or experimental approach.

Chapter 2, which is shared work with Dinand Webbink, investigates the effect of request size on the willingness to donate. We exploit exogenous variation in request size from a natural experiment. Participants in two online surveys earned money from incentivized games and were asked to donate the full amount to charity. This request size was determined by a random procedure and by observable game play. We find that an increase in the request size reduces the probability to donate. In our context, charity organizations would maximize their revenues by choosing the highest request size in the range of our data (16 euros). We also investigate the importance of mood by exploiting variation in mood due to the outcome of the Dutch soccer competition. We find that mood matters for charitable giving but does not affect the relationship between request size and charitable giving.

Chapter 3 investigates if showing inequality affects redistributional social preferences of teenagers. Can altruistic giving be affected by external factors, like films or the media? I use data from a field experiment in which 541 pupils from 6 secondary schools participated. The participants of the experiment, all pupils between 12 and 16 years old, were randomly assigned to watch a film about people living in poverty, and invited to play incentivized dictator and ultimatum games. After the games the

participants were given the opportunity to donate a proportion of the earned money to charity. I find that the teenagers that watch the films make more generous offers than the teenagers that did not. Both the offers in the games as well as the donations to charity increase after watching one of the films. This study provides evidence that social preferences can be affected by showing inequality.

Chapter 4 examines the impact of the implementation of Fair Trade labour standards on the lives of small-scale - or artisanal - gold miners. Artisanal gold-mining poses severe health and safety risks to the miners and their family members. Fair Trade labelling is an effort of solving an information asymmetry in which consumers have insufficient information about the quality of a product. Assuming that consumers are willing to pay a higher price for a product that is produced under socially responsible circumstances, the non-profit organisation implements these standards in artisanal miner organisations, hoping to improve the living conditions of the miners, and selling the gold for a higher price in the market to cover the cost. To evaluate the effect of this intervention on the gold miners, I use a difference-in-difference approach, in which respondents that work for Fair Trade mines are compared over time with respondents that work for comparable control mines. The data were collected through more than 1900 interviews with artisanal miners from 10 artisanal mining sites in Ghana, conducted in two waves. I find that the effects of Fair Trade on the gold miners are mixed. On the one hand, the implementation of the Fair Trade standards have a positive effect on the health of gold miners: The likelihood of miners getting ill or injured decreases by 23 percent. Moreover, the school attendance of children in the households increases by 10 percent as a result of the intervention. However, the intervention has a negative effect on income: the estimated proportion of miners that live above USD 2.50 per day decreases by 5 percent as a result of the intervention. Hence, there appears to be a trade-off in implementing the Fair Trade standards. A last important finding is that the turnover rate of gold miners is high - with a median employee tenure of one year. This limits the impact of the intervention.

In Chapter 5, which is shared work with Dinand Webbink, we explore the hidden costs of a specific case of compulsory public good provision. It investigates the long-term effects of peace-time military conscription on educational attainment and earnings, by exploiting a policy change that exempted a complete birth cohort from military service. We find that compulsory military service decreases the proportion of Dutch university graduates with 1.5 percentage points from a baseline of 12.3 percent. In addition, being a conscript reduces the probability of obtaining a university degree with almost four percentage points. The effect of military service on earnings is also negative and long-lasting. Approximately 18 years after military service we still find a negative effect of 3 to 4 percent. The effect

of conscription on educational attainment does not fully explain the wage reduction.

Nederlandse samenvatting

Volgens de klassieke economische theorie laten mensen zich voornamelijk leiden door eigenbelang. De miljarden euro's die jaarlijks worden gedoneerd aan goede doelen lijken een ander beeld te schetsen. In Nederland werd in 2013 ongeveer 4,3 miljard vrijwillig gedoneerd aan goede doelen (ongeveer 0,7 procent van het BNP); In de Verenigde Staten besteedt men ongeveer 2 procent van het BNP jaarlijks aan filantropie. Er is een groeiende academische interesse in filantropie, die, naast het feit dat er een substantieel bedrag in de sector omgaat, te verklaren is door de ogenschijnlijke tegenstrijdigheid in de egocentrische *homo economicus* enerzijds en zijn bereidheid om vrijwillig een gedeelte van zijn inkomen op te geven anderzijds. Ook is er een groeiende publieke belangstelling in dit onderwerp. Donateurs zijn steeds sceptischer geworden over de impact van hun donaties aan goede doelen en eisen meer transparantie van charitatieve organisaties. Tevens zijn de donaties aan goede doelen vaak aftrekbaar van de belasting. Dit onderstreept het belang van een empirische benadering van de effectiviteit van filantropische interventies.

Hoe kan filantropie bestaan als mensen rationeel en zelfzuchtig zijn? Een eerste verklaring die de economische wetenschap ons geeft, is gerelateerd aan marktfalen. Economen beschouwen filantropische giften als private contributies aan publieke goederen. Filantropische organisaties zijn de private aanbieders van deze goederen. De meest voor de hand liggende reden dat mensen geld doneren aan goede doelen is omdat zij simpelweg baat hebben van het publieke goed. Deze baten kunnen zeer direct en zeker zijn, zoals wanneer men geeft aan religieuze instanties (kerken, tempels) of aan sportclubs, maar de baten kunnen ook onzekerder en indirect zijn, zoals bij de bijdragen aan bijvoorbeeld kankeronderzoek het geval is. Dit betekent dat filantropie niet per sé gedreven wordt door altruïstische motieven. Er zijn echter bepaalde giften - zoals bijvoorbeeld de donaties aan ontwikkelingshulp - waarbij de gever geen baat heeft van het publieke goed. Altruïsme kan hier een drijfveer zijn. Een donateur kan echter ook

nut ontlenen aan de handeling van het geven zelf, in de vorm van morele bevrediging. Dit wordt ook als onzuiver altruïsme beschouwd.

Het is van belang om beter te begrijpen hoe charitatieve organisaties bij kunnen dragen aan de provisie van publieke goederen. De impact van filantropische interventies is echter moeilijk vast te stellen, omdat ze wordt beïnvloed door geobserveerde en niet geobserveerde factoren. Hetzelfde probleem doet zich voor bij het vaststellen van determinanten van geefgedrag. Het simpelweg vergelijken van mensen die doneren met mensen die niet doneren, zal door selectie-effecten foutieve schattingen opleveren. Ondanks de toenemende publieke interesse in de effectiviteit van filantropie in Nederland, is er slechts een beperkt aantal impactevaluaties beschikbaar met een geloofwaardig design om causale verbanden vast te stellen. In Nederland zijn de meeste evaluaties nog slechts gebaseerd op anekdotisch "bewijs." Dit proefschrift draagt bij aan de kennis over publieke interventies en filantropie door gebruik te maken van experimentele en quasi-experimentele designs. Ik gebruik deze om zowel de factoren te onderzoeken die vrijwillige bijdragen aan publieke goederen kunnen beïnvloeden, als om de effecten van specifieke interventies vast te stellen op publieke uitkomsten. Hoofdstuk 2 en 3 onderzoeken determinanten voor filantropie. Hoofdstuk 4 en 5 onderzoeken de impact van specifieke contributies aan publieke goederen.

Hoofdstuk 2 (in samenwerking met Dinand Webbink) onderzoekt wat het effect is van de hoogte van het gevraagde donatiebedrag op de bereidheid van het publiek om te doneren. We maken hierbij gebruik van exogene variatie in de hoogte van donatieverzoeken die ontstaan is door een natuurlijk experiment. Deelnemers van twee online surveys konden geld verdienen met het spelen van economische spellen. Nadat hen op einde van de survey het gewonnen bedrag met medegedeeld, kregen ze tevens de mogelijkheid dit *volledige* bedrag te doneren aan een goed doel. De hoogte van dit bedrag werd bepaald door een gerandomiseerde procedure en de observeerbare beslissingen van de participanten in de spellen. We vinden dat een toename van het gevraagde donatiebedrag een negatief effect heeft op de waarschijnlijkheid dat iemand doneert. In onze context echter, zullen de liefdadigheidsorganisaties hun donaties kunnen optimaliseren door het hoogst mogelijke bedrag te vragen (16 euro). We onderzoeken ook het effect van emoties op de bereidheid om te doneren, door gebruik te maken van variatie in emoties door de uitkomsten van de Nederlandse voetbalcompetitie. We vinden dat positieve emoties de bereidheid om te doneren vergroten, maar we vinden geen effect op de relatie tussen de hoogte van het gevraagde bedrag en de waarschijnlijkheid dat iemand doneert.

Hoofdstuk 3 onderzoekt of het tonen van armoede en sociale ongelijkheid aan tieners invloed heeft op hun sociale voorkeuren. Kan altruïstisch geefgedrag beïnvloed worden door externe factoren, zoals films

of de media? Ik gebruik data van een veldexperiment waar 541 scholieren van zes middelbare scholen aan deelnamen. De participanten, allen tussen de 12 en 16 jaar oud, werden uitgenodigd om economische spellen te spelen, waarmee ze geld konden winnen. Aan ongeveer de helft van de participanten - via een gerandomiseerde procedure bepaald - werd voordat ze deelnamen aan de spellen een film vertoond over armoede en sociale ongelijkheid. Na het spelen van de spellen hadden de participanten de mogelijkheid een bepaald gedeelte van hun "gewonnen" geld konden doneren aan een goed doel. Mijn bevinding is dat de deelnemers, na het kijken van de film, guller zijn in hun offers, in zowel in de spellen als aan het goede doel. Dit hoofdstuk toont aan dat sociale preferenties beïnvloed kunnen worden door het tonen van ongelijkheid.

Hoofdstuk 4 evalueert de impact van de Fair Trade arbeidsstandaarden op het leven van kleinschalige mijnwerkers in de goudindustrie in Ghana. De kleinschalige gouddelving brengt een verscheidenheid aan problemen met zich mee, waaronder gezondheids- en veiligheidsrisico's voor de mijnwerkers en hun familieleden. Het Fair Trade label is een poging om een informatie-asymmetrie op te lossen, waarbij de consumenten onvoldoende informatie hebben over de kwaliteit van het product dat ze willen aanschaffen. De organisaties achter het Fair Trade label nemen aan dat een bepaalde groep consumenten bereid is een hogere prijs te betalen voor een product dat tot stand is gekomen onder maatschappelijk verantwoorde omstandigheden. De organisaties investeren daarom in een aantal goudmijnen in Ghana. In deze organisaties worden de arbeidsstandaarden ingevoerd, in de hoop leefomstandigheden van de (families van) mijnwerkers te verbeteren. Aangenomen wordt dat na certificatie deze goudmijnen op eigen benen kunnen staan door de toenemende vraag naar Fair Trade goud. Om het effect van deze interventie op de goudmijnwerkers te evalueren, gebruik ik een *difference-in-differences* benadering, waarbij ik mijnwerkers uit de Fair Trade mijnen door de tijd heen vergelijk met mijnwerkers uit vergelijkbare controlemijnen. De data zijn verkregen door meer dan 1900 interviews met goudmijnwerkers van tien goudmijnen in Ghana, uitgevoerd in twee rondes: de eerste in januari 2013, de tweede in oktober 2014. Mijn eerste bevinding is dat de interventie een positief effect heeft op de gezondheid van mijnwerkers. De kans dat de mijnwerker ziek wordt of gewond raakt daalt met 23 procent als een gevolg van de interventie. Een tweede positief effect is het effect op investeringen in menselijk kapitaal. Het aantal kinderen per huishouden dat naar school gaat stijgt met 10 procent door de interventie. De Fair Trade standaarden lijken echter een negatief effect te hebben op het inkomen van de mijnwerkers. Het percentage mijnwerkers dat leeft van inkomen van minimaal USD 2.50 per persoon per dag, daalt met 5 procent. Een laatste belangrijke bevinding is dat er een hoge omloopsnelheid van mijn mijnwerkers is in beide typen goudmijnen. Deze

hoge omloopsnelheid beperkt de impact van de interventie.

In Hoofdstuk 5 (in samenwerking met Dinand Webbink; gepubliceerd in IZA Journal of Labour Economics, 4:10 (2015)) onderzoeken we de verborgen kosten van een specifiek geval van verplichte publieke goed provisie. Deze studie onderzoekt de langetermijneffecten van de militaire dienstplicht in vredestijd, door gebruik te maken van exogene variatie die ontstond door een beleidswijziging waarbij een volledig geboortecohort (alle Nederlandse mannen geboren in 1959) vrijgesteld werd van militaire dienst. We vinden dat de verplichte militaire dienst heeft geleid tot een vermindering van het aantal afgestudeerden in Nederland. Het aantal afgestudeerden daalde met 1,5 procentpunt tegen een baseline van 12,3 procent. De militaire dienstplicht verkleinde de kans op het halen van een universitair diploma met 4 procentpunt. Het effect van militaire dienstplicht op inkomen is negatief, ook op de lange termijn. Ongeveer 18 jaar na de militaire dienstplicht is er nog steeds een negatief effect op inkomen zichtbaar van 3 tot 4 procent. Het effect van de dienstplicht op onderwijs kan dit effect op inkomen niet volledig verklaren.

Curriculum Vitae

Frank Roeland Hubers was born in 1982 in Papendrecht, the Netherlands. He studied cultural anthropology at Utrecht Unviersity from 2002 to 2006, obtaining his Bachelor degree in 2005 and his Master degree in 2006 (cum laude). In 2008 he obtained his diploma in the postgraduate dual program of international development studies at the Radboud University of Nijmegen. During the period 2008 to 2010 he worked for the Quality and Control department of Oxfam Novib, where he was responsible for monitoring and evaluation of development projects. In September 2010 he started as a researcher for the knowledge center ECSP and started his thesis at the Erasmus School of Economics in 2011. He is currently working as a lecturer in economics and anthropology at Webster University in Thailand.

Bibliography

Alesina, A. and Fuchs-Schündeln, N. (2007). Good-bye Lenin (or not?): The effect of communism on people's preferences. *The American Economic Review*, 97(4):1507–1528.

Alesina, A. and La Ferrara, E. (2005). Preferences for redistribution in the land of opportunities. *Journal of Public Economics*, 89(5):897–931.

Andreoni, J. (1988). Privately provided public goods in a large economy: the limits of altruism. *Journal of Public Economics*, 35(1):57–73.

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

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

Andreoni, J. (2006). Philanthropy. In Kolm, S.-C. and Ythier, J. M., editors, *Handbook of the Economics of Giving, Altruism and Reciprocity*, volume 2, page 1201 1269. Elsevier.

Andreoni, J., Harbaugh, W. T., and Vesterlund, L. (2008). Altruism in experiments. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics*, pages 134–38. Palgrave Macmillan.

Andreoni, J. and Rao, J. M. (2011). The power of asking: How communication affects selfishness, empathy, and altruism. *Journal of Public Economics*, 95(7):513–520.

Andreoni, J. and Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *The Quarterly journal of economics*, 116(1):293–312.

Angrist, J. D. (1990). Lifetime earnings and the vietnam era draft lottery: Evidence from social security administrative records. *The American Economic Review*, 80(3):313–336.

Angrist, J. D., Chen, S. H., and Song, J. (2011). Long-term consequences of Vietnam-era conscription: New estimates using social security data. *American Economic Review*, 101(3):334–38.

Angrist, J. D. and Krueger, A. B. (1994). Why do World War II veterans earn more than nonveterans? *The American Economic Review*, 12(1):74–97.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 41 William Street, Princeton, New Jersey 08540, 1 edition.

Appiah, H. (1998). Organization of small scale mining activities in ghana. *Journal of the Southern African Instute of mining and metallurgy*, 98:307–310.

Auten, G. E., Sieg, H., and Clotfelter, C. T. (2002). Charitable giving, income, and taxes: an analysis of panel data. *The American Economic Review*, 92(1):371–382.

Bacon, C. (2005). Confronting the coffee crisis: can fair trade, organic, and specialty coffees reduce small-scale farmer vulnerability in northern nicaragua? *World Development*, 33(3):497–511.

Banchirigah, S. M. (2006). How have reforms fuelled the expansion of artisanal mining? evidence from sub-saharan africa. *Resources Policy*, 31(3):165–171.

Banchirigah, S. M. (2008). Challenges with eradicating illegal mining in ghana: A perspective from the grassroots. *Resources Policy*, 33(1):29–38.

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

Barraza, J. A. and Zak, P. J. (2009). Empathy toward strangers triggers oxytocin release and subsequent generosity. *Annals of the New York Academy of Sciences*, 1167(1):182–189.

Bauer, T. K., Bender, S., Paloyo, A. R., and Schmidt, C. M. (2012). Evaluating the labor-market effects of compulsory military service. *European Economic Review*, 56(4):814–829.

Bauer, T. K., Bender, S., Paloyo, A. R., and Schmidt, C. M. (2014). Do guns displace books? the impact of compulsory military service on educational attainment. *Economics Letters*, 124(3):513–515.

Becchetti, L. and Costantino, M. (2008). The effects of fair trade on affiliated producers: An impact analysis on kenyan farmers. *World Development*, 36(5):823–842.

Bénabou, R. and Tirole, J. (2006). Belief in a just world and redistributive politics. *The Quarterly journal of economics*, 121(2):699–746.

Benenson, J. F., Pascoe, J., and Radmore, N. (2007). Children's altruistic behavior in the dictator game. *Evolution and Human Behavior*, 28(3):168–175.

Boskin, M. J. and Feldstein], M. (1977). Effects of the charitable deduction on contributions by low income and middle income households: Evidence from the national survey of philanthropy. *The Review of Economics and Statistics*, 59(3):351–354.

Bradley, M. M., Cuthbert, B. N., and Lang, P. J. (1996). Picture media and emotion: Effects of a sustained affective context. *Psychophysiology*, 33(6):662–670.

Broberg, T., Ellingsen, T., and Johannesson, M. (2007). Is generosity involuntary? *Economics Letters*, 94(1):32–37.

Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child development*, 53(2):413–425.

Buonanno, P. (2006). Long-term effects of conscription: Lessons from the UK. Working paper 2006/4, University of Bergamo, Department of Economics.

Capra, C. M. (2004). Mood-driven behavior in strategic interactions. *The American Economic Review*, 94(2):367–372.

Capra, C. M., Lanier, K. F., and Meer, S. (2010). The effects of induced mood on bidding in random nth-price auctions. *Journal of Economic Behavior & Organization*, 75(2):223–234.

Card, D. and Lemieux, T. (2001). Going to college to avoid the draft: The unintended legacy of the Vietnam War. *The American Economic Review*, pages 97–102.

Charness, G. and Grosskopf, B. (2001). Relative payoffs and happiness: An experimental study. *Journal of Economic Behavior & Organization*, 45(3):301–328.

Cherry, T. L., Frykblom, P., and Shogren, J. F. (2002). Hardnose the dictator. *The American Economic Review*, 92(4):1218–1221.

Childs, J. (2008). Reforming small-scale mining in sub-saharan africa: Political and ideological challenges to a fair trade gold initiative. *Resources Policy*, 33(4):203–209.

Cialdini, R. B. and Schroeder, D. A. (1976). Increasing compliance by legitimizing paltry contributions: When even a penny helps. *Journal of Personality and Social Psychology*, 34(4):599.

Cipollone, P. and Rosolia, A. (2007). Social interactions in high school: Lessons from an earthquake. *The American Economic Review*, pages 948–965.

Clotfelter, C. T. (1980). Tax incentives and charitable giving: Evidence from a panel of taxpayers. *Journal of Public Economics*, 13(3):319–340.

Collier, P. (2007). *The bottom billion: Why the poorest countries are failing and what can be done about it*. Oxford University Press.

Corneo, G. and Grüner, H. P. (2002). Individual preferences for political redistribution. *Journal of public Economics*, 83(1):83–107.

Cruces, G., Perez-Truglia, R., and Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics*, 98:100–112.

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2):193–201.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1):113.

Di Pietro, G. (2013). Military conscription and university enrolment: Evidence from Italy. *Journal of Population Economics*, 26(2):619–644.

Doob, A. N. and McLaughlin, D. S. (1989). Ask and you shall be given: Request size and donations to a good cause. *Journal of Applied Social Psychology*, 19(12):1049–1056.

Dragusanu, R., Giovannucci, D., and Nunn, N. (2014). The economics of fair trade. *The Journal of Economic Perspectives*, 28(3):217–236.

Duindam, S. (1995). *Militaire dienstplicht: Een economische analyse van de personeelscomponent in de krijgsmacht*. PhD thesis, University of Limburg.

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

Edwards, J. T. and List, J. A. (2014). Toward an understanding of why suggestions work in charitable fundraising: Theory and evidence from a natural field experiment. *Journal of Public Economics*, 114:1–13.

Elfenbein, D. W. and McManus, B. (2010). A greater price for a greater good? evidence that consumers pay more for charity-linked products. *American Economic Journal: Economic Policy*, 2(2):28–60.

Fehr, E. and Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism: Experimental evidence and new theories. In Kolm, S.-C. and Ythier, J. M., editors, *Handbook of the economics of giving, altruism and reciprocity*, volume 1, pages 615–691. Palgrave Macmillan.

Feldstein, M. and Clotfelter, C. Tax incentives and charitable contributions in the united states: A microeconometric analysis.

Feldstein, M. and Taylor, A. (1976). The income tax and charitable contributions. *Econometrica*, 44(6):1201–1222.

Fong, C. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public economics*, 82(2):225–246.

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

Fort, R. and Ruben, R. (2008a). The impact of fair trade certification on coffee producers in peru. In Ruben, R., editor, *The impact of Fair Trade*, chapter 3, pages 75–98. The Netherlands: Wageningen Academic Publishers.

Fort, R. and Ruben, R. (2008b). The impact of fair trade on banana producers in northern peru. In Ruben, R., editor, *The impact of Fair Trade*, chapter 2, pages 49–73. The Netherlands: Wageningen Academic Publishers.

Frohlich, N., Oppenheimer, J., and Bernard Moore, J. (2001). Some doubts about measuring self-interest using dictator experiments: The costs of anonymity. *Journal of Economic Behavior & Organization*, 46(3):271–290.

Gelman, A. and Imbens, G. (2014). Why high-order polynomials should not be used in regression discontinuity designs.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, 115(3):811–846.

Grenet, J., Hart, R. A., and Roberts, E. (2010). Above and beyond the call. Long-term real earnings effects of British male military conscription in the post-war years. *Labour economics*, 18(2).

Gummerum, M., Hanoch, Y., Keller, M., Parsons, K., and Hummel, A. (2010). Preschoolers allocations in the dictator game: The role of moral emotions. *Journal of Economic Psychology*, 31(1):25–34.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization*, 3(4):367–388.

Hainmueller, J., Hiscox, M. J., and Sequeira, S. Consumer demand for the fair trade label: Evidence from a multi-store field experiment. *Review of Economics and Statistics*, 97.

Hanushek, E. A. and Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3):607–668.

Harbaugh, W. T., Krause, K., Liday, S. G., et al. (2003). Bargaining by children.

Hilson, G. (2002). The environmental impact of small-scale gold mining in ghana: identifying problems and possible solutions. *The Geographical Journal*, 168(1):57–72.

Hilson, G. (2008). fair trade gold: Antecedents, prospects and challenges. *Geoforum*, 39(1):386–400.

Hilson, G., Hilson, C. J., and Pardie, S. (2007). Improving awareness of mercury pollution in small-scale gold mining communities: challenges and ways forward in rural ghana. *Environmental Research*, 103(2):275–287.

Hilson, G. and Pardie, S. (2006). Mercury: An agent of poverty in ghana's small-scale gold-mining sector? *Resources Policy*, 31(2):106–116.

Hiscox, M. J., Broukhim, M., and Litwin, C. (2011). Consumer demand for fair trade: New evidence from a field experiment using ebay auctions of fresh roasted coffee. *Available at SSRN 1811783*.

Hoffenaar, J. and Schoenmaker, B. (1994). *Met de blik naar het oosten: De Koninklijke Landmacht 1945-1990*. Historical Section of the Royal Netherlands Army, The Hague.

Hoffmann, R. and Tee, J.-Y. (2006). Adolescent–adult interactions and culture in the ultimatum game. *Journal of Economic Psychology*, 27(1):98–116.

Ifcher, J. and Zarghamee, H. (2011). Happiness and time preference: The effect of positive affect in a random-assignment exper. *The American Economic Review*, 101(7):3109–3129.

Imbens, G. and van der Klaauw, W. (1995). Evaluating the cost of conscription in the Netherlands. *Journal of Business & Economics Statistics*, 13(2):207–215.

Kahneman, D. and Knetsch, J. L. (1992). Valuing public goods: the purchase of moral satisfaction. *Journal of environmental economics and management*, 22(1):57–70.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *The Journal of Business*, 59(4):pp. S285–S300.

Kahneman, D. and Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *The journal of economic perspectives*, 20(1):3–24.

Karlan, D. and List, J. A. (2007). Does price matter in charitable giving? evidence from a large-scale natural field experiment. *The American Economic Review*, 97(5):1774–1793.

Karlan, D., List, J. A., and Shafir, E. (2011). Small matches and charitable giving: Evidence from a natural field experiment. *Journal of Public Economics*, 95(5):344–350.

Keller, K. R., Poutvaara, P., and Wagener, A. (2009). Does military draft discourage enrollment in higher education? Evidence from OECD countries.

Kirchsteiger, G., Rigotti, L., and Rustichini, A. (2006). Your morals might be your moods. *Journal of Economic Behavior & Organization*, 59(2):155–172.

Kolm, S.-C. (2006). Introduction to the economics of giving, altruism and reciprocity. In Kolm, S.-C. and Ythier, J. M., editors, *Handbook of the economics of giving, altruism and reciprocity*, volume 1, pages 1–122. Palgrave Macmillan.

Lau, M., Poutvaara, P., and Wagener, A. (2004). The dynamic cost of the draft. *German Economic Review*, 5(4):381–406.

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

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

Maurin, E. and Xenogiani, T. (2007). Demand for education and labor market outcomes lessons from the abolition of compulsory conscription in france. *Journal of Human Resources*, 42(4):795–819.

Meier, S. (2007). Do subsidies increase charitable giving in the long run? Matching donations in a field experiment. *Journal of the European Economic Association*, 5(6):1203–1222.

Mohan, S. (2010). Fair trade without the froth. *Institute of Economic Affairs Monographs, Hobart Paper*, (170).

Murnighan, J. K. and Saxon, M. S. (1998). Ultimatum bargaining by children and adults. *Journal of Economic Psychology*, 19(4):415–445.

Piketty, T. (1995). Social mobility and redistributive politics. *The Quarterly journal of economics*, 105(10):551–584.

Poutvaara, P. and Wagener, A. The political economy of conscription.

Poutvaara, P. and Wagener, A. (2007). To draft or not to draft? Inefficiency, generational incidence, and political economy of military conscription. *European Journal of Political Economy*, 23(4):975–987.

Ravallion, M. and Lokshin, M. (2000). Who wants to redistribute?: The tunnel effect in 1990s russia. *Journal of public Economics*, 76(1):87–104.

Reingen, P. H. (1982). Test of a list procedure for inducing compliance with a request to donate money. *Journal of Applied Psychology*, 67(1):110.

Reinstein, D. and Riener, G. (2012). Decomposing desert and tangibility effects in a charitable giving experiment. *Experimental Economics*, 15(1):229–240.

Rondeau, D. and List, J. A. (2008). Matching and challenge gifts to charity: evidence from laboratory and natural field experiments. *Experimental economics*, 11(3):253–267.

Ruben, R. and van Schendel, L. (2008). The impact of fair trade in banana plantations in ghana: Income, ownership and livelihoods of banana workers. In Ruben, R., editor, *The impact of Fair Trade*, chapter 6, pages 137–167. The Netherlands: Wageningen Academic Publishers.

Sáenz Segura, F. and Zuniga-Arias, G. (2008). Assessment of the effect of fair trade on smallholder producers in costa rica: A comparative study in the coffee sector. In Ruben, R., editor, *The impact of Fair Trade*, chapter 5, pages 117–135. The Netherlands: Wageningen Academic Publishers.

Siminski, P. (2013). Employment effects of army service and veterans' compensation: evidence from the Australian Vietnam-era conscription lotteries. *Review of Economics and Statistics*, 95(1):87–97.

Warden, D. and Mackinnon, S. (2003). Prosocial children, bullies and victims: An investigation of their sociometric status, empathy and social problem-solving strategies. *British Journal of Developmental Psychology*, 21(3):367–385.

Warr, P. G. (1982). Pareto optimal redistribution and private charity. *Journal of Public Economics*, 19(1):131–138.

Weyant, J. M. (1984). Applying social psychology to induce charitable donations1. *Journal of Applied Social Psychology*, 14(5):441–447.

Weyant, J. M. and Smith, S. L. (1987). Getting more by asking for less: The effects of request size on donations of charity1. *Journal of Applied Social Psychology*, 17(4):392–400.

Zak, P. J., Stanton, A. A., and Ahmadi, S. (2007). Oxytocin increases generosity in humans. *PLoS One*, 2(11):e1128.

Zuniga-Arias, G. and Segura, F. S. (2008). The impact of fair trade in banana production of costa rica. In Ruben, R., editor, *The impact of Fair Trade*, chapter 4, pages 99–116. The Netherlands: Wageningen Academic Publishers.