

MEASURES TO INFLUENCE
INDIVIDUAL BEHAVIOR:
FOUR ESSAYS IN ECONOMICS

Maximilian Linek

Dissertation submitted to the Hertie School
in partial fulfillment of the requirements for the degree of
Doctor rerum politicarum (Dr. rer. pol.)
in the Doctoral Programme in Governance.

Berlin, 2020

First advisor

Professor Dr. Christian Traxler
Hertie School
Friedrichstraße 180
10117 Berlin, Germany

Second advisor

Professor Eliana La Ferrara (Ph.D.)
Bocconi University
Via Roentgen 1
20136 Milan, Italy

Third advisor

Professor Leonardo Iacovone (Ph.D.)
Hertie School
Friedrichstraße 180
10117 Berlin, Germany

Summary

This dissertation in economics studies the effects of changing individuals' choice environments on their behavior. The choice environment is the setting within which an individual makes her decision. It is widely acknowledged that even inconspicuous changes to this environment can have strong effects on behavior. This notion has spurred large efforts in the public and private sector to design these environments and steer behavior. The dissertation studies four measures that change choice environments to achieve public and private objectives.

The first chapter investigates the effects of community meetings on behavior changes that were desired by the central government in Rwanda. It studies these effects with and without performance incentives for local leaders and sheds light on who controls meetings and how they operate. The second chapter studies expected effects of Basic Income on time use, deploying a large-scale survey experiment. The results show intended reallocations that may contribute to foster social cohesion in society. In addition, the chapter suggests its methodology as a new approach to explore policies that are difficult or costly to test. Finally, the third and fourth chapter study the effects of social information and defaults in online donations. The third chapter shows that the empirical regularity of conditional cooperation fails. The fourth and last chapter explores the complex relationship between different numbers and values of donation amount defaults.

Acknowledgements

First and foremost, I would like to thank my three supervisors Christian Traxler, Eliana La Ferrara, and Leonardo Iacovone. I am grateful to Christian for accepting me as a doctoral student and closely accompanying my studies from the beginning. His encouragement, guidance, and constructive criticism were invaluable, and this thesis has gained substantially from his comments and suggestions. I am indebted to Eliana for inviting me to visit her at Bocconi University and becoming my second supervisor. The exchange with her, her continuous support, and the hospitality at Bocconi were enormously inspiring, instructive and helpful. Last but not least, I would like to thank Leonardo for fruitful discussions and his excellent comments on my research.

I also would like to thank my friends and colleagues at the Hertie School. In particular, I would like to thank Paul Berenberg-Gossler, Vladimir Bogoeski, Franziska Loschert, Jonas Markgraf, Panayiotis (Panos) Nicolaides, and Nicolas Pardo. I am further grateful to the Ph.D. Team and administration at the Hertie School. They provided excellent support to pursue independent research at the highest standards.

In preparing this thesis, I have strongly benefited from collaboration with Christian Traxler. It was and continues to be a pleasure to work with him.

Finally, I would like to express my deep gratitude to my family. My partner Lisa always encouraged and supported through ups and downs of my doctoral studies, and my son Enno was a perfect balance to this work. I also would like to thank my parents, Monika and Jens, and my sister, Leoni, for their support at various stages in my life and while writing this thesis.

Financial support from the Hertie School, the German Academic Scholarship Foundation (Studienstiftung), and the Fritz Thyssen Foundation are gratefully acknowledged.

Maximilian Linek
Berlin, September 2020

Contents

Preface	1
1 Dictating Development? The Effects of Local Institutions under Upward Accountability	6
1.1 Introduction	6
1.2 Background	11
1.2.1 <i>Umuganda</i>	11
1.2.2 Performance Contracts	12
1.2.3 Development Targets and Popular Support	13
1.3 Data	14
1.3.1 Family Planning	14
1.3.2 Mosquito Bed Nets	17
1.3.3 Rainfall	18
1.4 Empirical Strategy	20
1.4.1 Rainfall and <i>Umuganda</i>	20
1.4.2 Exclusion Restriction	22
1.4.3 Specification	23
1.5 Results	24
1.5.1 Main Reduced-form Effects	24
1.5.2 Tracing the Effects	26
1.5.3 Robustness Checks	28

1.5.4	Alternative Channels	30
1.6	Mechanism	32
1.6.1	Evasive Behavior	32
1.6.2	Heterogeneity in Popular Support	34
1.7	Conclusion	34
	Appendix	36
2	Time Use with Basic Income: Evidence from a Large-Scale Survey Experiment	40
2.1	Introduction	40
2.2	Experimental Design	44
2.2.1	Assignment to Groups	44
2.2.2	Treatments	44
2.2.3	Outcomes	45
2.2.4	External Validity and Overoptimism	46
2.3	Data	48
2.3.1	Data Collection	48
2.3.2	Sample Selection	48
2.3.3	Integrity of the Experiment	49
2.3.4	Current Time Use	50
2.4	Empirical Method	51
2.4.1	Overoptimism Check	51
2.4.2	Specification	53
2.5	Results	54
2.5.1	Overall Effects	54
2.5.2	Main Effects	55
2.5.3	Validity Check	59
2.6	Conclusion	60
	Appendix	62

3	Are Wikipedia Users Conditionally Cooperative? Evidence from Fundraising Trials	73
3.1	Introduction	73
3.2	Trials and Data	75
3.3	Results	79
3.3.1	Variation in framing	79
3.3.2	Variation in baseline numbers	80
3.3.3	Direct variation in social information	80
3.4	Discussion	81
	Appendix	84
4	Defaults in Online Charity: Evidence from Wikipedia Germany	85
4.1	Introduction	85
4.2	Data	87
4.3	Donations without Defaults	90
4.4	Donations with Defaults	91
4.4.1	Trials on Desktop Website	91
4.4.2	Trials on Mobile Website	100
4.5	Conclusion	104
	Appendix	105
	Bibliography	106
	Annex: List of Papers	118

List of Tables

1.1	Summary Statistics of Family Planning Outcomes	17
1.2	Summary Statistics of Acquisitions of Mosquito Bed Nets	18
1.3	Summary Statistics of Rainfall on Saturdays	20
1.4	Main Effects	25
1.5	Tracing the Effects Under Performance Contracts	27
1.6	Bed Net Acquisition from Different Sources Under Performance Contracts .	31
1.7	Mechanism: Conception and Bed Net Acquisition by Altitude	33
2.1	Factors of Basic Income and their Levels	45
2.2	Activity Fields	46
2.3	Precise question and layout	46
2.4	Number of respondents in control and treatment groups	49
2.5	Differences in Characteristics	50
3.1	Overview of Trials	78
4.1	Overview Campaigns	89

List of Figures

1.1	Number of Contraceptive Adoptions and Conceptions over Time	16
1.2	Number of Mosquito Bed Nets Acquired in Different Months	18
1.3	Effect of Saturday Rainfall on Contraceptive Adoption over Time	28
1.4	Effects of Saturday Rainfall at Different Thresholds Under Performance Contracts	29
1.A1	Number of Contraceptive Adoptions and Conceptions in Different Months before the Interview	36
1.A2	Contraceptive Adoption Under Performance Contracts: Effects at Different Thresholds	37
1.A3	Bed Net Acquisition Under Performance Contracts: Effects at Different Thresholds	38
1.A4	Conception Under Performance Contracts: Effects at Different Thresholds .	39
2.1	Current Time Use	50
2.2	Intended Changes to Time Use in the Control Group	52
2.3	Intentions and Implementation in the Control Group	52
2.4	Overall Treatment Effects	55
2.5	Main Effects: Amount of Basic Income	56
2.6	Main Effects: Duration of Basic Income	57
2.7	Main Effects: Recipient of Basic Income	59
2.8	Treatment Effect Differences by Frequency of Having Achieved Plans . . .	60
3.1	Variation in framing	79

3.2	Variation in baseline numbers	80
3.3	Direct variation in social information	81
3.A1	Control banner of trial 1 (German original)	84
3.A2	Example donation banner (English)	84
4.1	Distributions of donations without defaults	90
4.2	Distributions of donations with defaults	92
4.3	Average donation amounts with defaults	93
4.4	Distributions of donations with different defaults	95
4.5	Average donation amounts with different defaults	96
4.6	Distributions of donations with different default values	98
4.7	Average donation amounts with different default values	99
4.8	Distributions of donations with different defaults (mobile)	101
4.9	Average donation amounts with different defaults (mobile)	102
4.10	Distributions of donations: Trial 12 (hk14_mob_3)	103
4.11	Average donation amounts: Trial 12 (hk14_mob_3)	103
4.A1	Example donation banner	105

Preface

Over the past decades, economics has been enriched by insights from sociology and psychology. The ideas, concepts, and empirical puzzles from these disciplines required economics to step outside its traditional bounds, set by a perfectly rational, individualistic world with complete information. Along this way, the design of the choice environment, which is the setting within which an individual makes her decision, has come to prominence. Today, it is widely acknowledged that even small changes to this environment can have strong effects on individual behavior. For this reason, public and private actors expend much effort to strategically shape these environments and steer behavior.

In this dissertation, I empirically study effects of changing individuals' choice environments on their behavior. My primary focus is on two factors in the choice environment that are usually difficult to control and, thus, particularly interesting to study. These are social interactions in groups, adopted from sociology, and heuristics under incomplete information, which originates from psychology. Each chapter investigates a different measure that deliberately changes the choice environment to steer behavior and achieve specific objectives. In the first two chapters, I study cases that explicitly alter incentives either within groups or for the individual to contribute to society. In the last two chapters, I analyze less intrusive modifications of people's choice environments. Specifically, I study two types of 'nudges' that aim to increase online donations.¹

With my dissertation, I contribute to our understanding of choice environments and their effect on individual behavior. The chapters study different measures to illuminate various aspects of decision-making with social interactions and under asymmetric information, and yield interesting policy recommendations. Beside studying measures that are designed to steer behavior, all chapters are also connected through the theme of social norms. While chapter 1 studies the creation and enforcement of new social norms in health, chapter 2 explores a specific policy that aims to reduce the bite of the existing norm to pursue gainful

1. A *nudge* is an approach to influence individual behavior without changing economic incentives or forbidding any options.

employment. Finally, chapters 3 and 4 both investigate effects that, to some extent, arise from uncertainty about norms of whether and how much to donate. I now provide an overview over the chapters and their most important findings, and draw a short conclusion thereafter.

Chapter 1 studies the effects of a local institution with and without upward accountability on individual behavior change in Rwanda. To achieve development, which often requires behavior change, governments and aid agencies have started to use local institutions as tools for implementation. The underlying idea is that local institutions make leaders downward accountable to the community and achieve behavior change by improving leaders provision of services to the population. However, leaders usually control local institutions and can also use them to promote specific types of behavior change in the population. By studying the effect of local institutions on behavior change before and after a reform that solely strengthened leaders accountability to the central government, I distinguish whether behavior change through local institutions arises under downward or upward accountability.

The local institution, which I study, is a community meeting that takes place across Rwanda on Saturdays and forms part of a traditional, mandatory community program called *Umuganda*. I investigate the effect of these meetings on two types of behavior change, using individual panel data. The two behavior changes are women's adoptions of modern contraceptive methods and households' acquisitions of mosquito bed nets. While being fundamentally different behaviors, both types of behavior change became incentivized targets for leaders after a reform that introduced performance contracts. To identify the causal effect of meetings on behavior change, I exploit rainfall on Saturdays. In the absence of meeting data, I estimate the reduced-form relationship. The exclusion restriction is that rainfall on Saturdays only affects behavior through attendances at meetings. Rainfall on other weekdays are natural placebo tests.

I find that Saturday rainfall significantly affects behavior change after the introduction of performance contracts. No effects are found before the reform or for rainfall on other weekdays. These results indicate that local leaders use community meetings to implement their targets. The same pattern for different behaviors allows to attribute the effects of meetings to the introduction performance incentives because *Umuganda* did not change and access as an alternative channel of rainfall can be ruled out. The results show, thus, that a central government can deploy local institutions for development by making the leaders, which control the institutions, upward accountable. However, I also find evidence suggesting that behavior change was involuntary. These findings challenge the commonly assumed functioning of local institutions and caution that breaches to personal freedoms can arise without necessary

checks and balances. In addition, they show the complementarity of performance incentives and local institutions, both of which try to solve the last mile problem.

Chapter 2 studies the effect of Basic Income (BI) on intentions to change time use deploying a large-scale survey experiment among 72,134 participants in Germany.² In the experiment, treated participants were instructed to imagine a specific BI scenario and think about their intended changes to current time use with this BI. In contrast, control participants had to think about intended changes to time use given their current situation. After treatment, all participants received the identical question to record their intended changes to time spent on seven activity fields. The seven activity fields are work, education, volunteering, sport, care, socializing, and hobbies. By comparing treated with control participants and participants with different BI scenarios against each other, we can learn about the likely effects and varying features of BI on behavior.

Three findings are particularly interesting. First, I find that, compared to the control group, treatment reduces intended time spent on work, sport, socializing, and hobbies, and increases intended time spent on education, volunteering, and care. This finding suggests that BI makes people reallocate time from the end points of a work-leisure spectrum to intermediate activities. In light of the specific activities, this reallocation can be considered socially desirable as it increases social cohesion. Second, I find marginally decreasing treatment effects of BI for amounts larger than 1,000 Euros per month. This finding has important implications for the design of future BI experiments. Finally third, all treatment effects are relatively small, suggesting that, overall, BI will not significantly change people's daily routines and the functioning of our society.

This survey experiment on BI is innovative for economics as it leverages subjective data for policy evaluation. It is guided by two notions. First, intentions predict behavior, and we can, thus, obtain realistic predictions of actual behavior with BI from stated intentions. The comparison against the control group arguably allows to extract this actual behavior from intentions. Second, we need a new approach to assess policies, like BI, that are difficult to test because they are, for example, too large or expensive. For these types of interventions, the study proposes a specific way how to aggregate the judgments of experts. Though the results should be taken with a grain of salt, they are important inputs to maximize explanatory power and meaningfulness of actual BI experiments.

2. Basic Income are unconditional cash transfers that are paid at regular intervals.

Chapters 3 and 4 are written jointly with Christian Traxler and study donations to Wikimedia Germany (WMDE).³ Chapter 3 studies whether donors to WMDE behave conditionally cooperative. Conditional cooperation is an empirical regularity that has been verified in many different settings, including donations. It states that prosocial behavior depends positively on the prosocial behavior of relevant others, empirically manifesting in cooperation. With respect to donations to WMDE, the concept of conditional cooperation suggests that information or framings that provide cues about the number of other donors should affect the individual donation decision.

We investigate conditional cooperation in online donations to WMDE by analyzing data from six trials. All trials test a control donation banner that indicates a low number or rate of donations (which is the number of donations relative to total users or banner impressions) against a treatment banner that indicates a larger number or higher rate of donations. These banners pop-up when browsing the Wikipedia website during a fundraising campaign and assignment to either treatment or control banner is stable. If donors to WMDE act conditionally cooperative, we would, thus, expect positive treatment effects on the decision whether to donate or not.

Our results reject the hypothesis of conditional cooperation in online donations to WMDE. In five trials, we find (partially significant) negative or null results, meaning that donation rates per banner impression decline or remain constant with treatment. Only one trial yielded a statistically insignificant positive effect on donation rates. We find no intensive margin effects from treatments, meaning that average donation amounts are statistically indifferent in all trials. These results strongly suggest the absence of conditional cooperation in donations to WMDE. We discuss three possible explanations for this absence as potentially following from a lack of pressure to conform, the failure of donations as a signal to others, and self-identity returns.

Chapter 4 studies the effects of donation amount defaults in the Wikipedia banners. Specifically, we analyze data from 12 trials that test different numbers of default options, different default values and combinations of both in donation banners. The analysis of these modifications is complicated due to the fact that variation in the number of defaults necessarily also entails variation in the offered default values. The overall objective of our analysis is to explore which modifications contribute to increase revenue relative to the number of banner impressions. To this end, we study effects on the decision to donate and on donated amounts.

3. Wikimedia Germany (WMDE) operates the Germany language website of the worldwide largest online encyclopedia Wikipedia.

Our results document strong intensive margin effects both on the distribution of donations and on the average donation amount. However, we only find scarce and unreliable evidence of effects on the decision to donate (extensive margin effects). In this respect, our evidence suggest that amount defaults can increase the sum of donations primarily through their effect on the average donation amount, but not through raising the number or rate of donations. Comparing the different banner variations in detail reveals a complex interaction between the number of defaults and default values. While some trials suggest that the number of defaults has a larger potential to increase revenue, other trials indicate that the choice of default values is more important. In future research, we plan to theoretically explore default effects and test the resulting hypotheses in further trials.

In conclusion, this dissertation explores four factors of choice environments that can be altered to steer individual behavior. Social interactions (chapter 1) and defaults (chapter 4) are among the more powerful determinants, whereas reduced work incentives (chapter 2) and social information (chapter 3) seem to have smaller effects. In this respect, more intrusive approaches that alter incentives, as in the first two chapters, are not necessarily more promising than smaller interventions that use nudges, as in the last two chapters. Finally, the dissertation also documents some pitfalls of modifying choice environments and how these can backfire.

Chapter 1

Dictating Development? The Effects of Local Institutions under Upward Accountability

1.1 Introduction

Local institutions are a key to development because they regulate individual behavior. After rampant corruption in many centralized states, governments and aid donors have embraced local institutions as instruments for managing decentralized development (Mansuri and Rao 2013). The underlying tenet is that citizen participation in local institutions regulates the behavior of leaders through social capital. Recent research does not verify this mechanism.¹ It commonly attributes development to downward accountability and the absence thereof to elite capture. However, if elites control local institutions, development may just as well arise from leaders' ability to mandate targeted behaviors from community members. To understand how decentralized approaches to development achieve outcomes, it is important to investigate the effects of local institutions jointly with accountability. I do so in the East African country Rwanda.

Rwanda is an interesting case for studying local institutions and accountability. After the 1994 genocide, a new government took power and rebuilt the country. During the late 2000s,

1. Research on local institutions in development has commonly attempted to strengthen bottom-up accountability through either providing additional information (Banerjee et al. 2010; Björkman Nyqvist et al. 2017; Björkman and Svensson 2009) or altering composition, involvement or capacity of people in the institution (Casey et al. 2012; Olken 2007; Pradhan et al. 2014). All of these studies essentially assume downward accountability of local institutions and track, at best, civic participation as a proxy.

this government achieved spectacular progress toward the Millenium Development Goals, many of which require behavior change. This success is widely attributed to citizen participation in local institutions (UNDP 2014). However, Rwanda is also an authoritarian state in which, by definition, leaders control local institutions. By exploring the contribution of a local institution to development progress in Rwanda, this chapter challenges the assumption of downward accountability in decentralized development. The findings arguably are relevant for understanding and designing development in many countries around the world as local institutions often have strong authoritarian elements even when the national state is considered a democracy.

In this chapter, I study a local institution that is used for development in Rwanda and link its effects on behavior change to accountability.² The local institution is a community meeting that takes place on Saturdays and forms part of a traditional, mandatory community program called *Umuganda*. I trace accountability of and in this institution through its effects on behavior change in three ways. First, I analyze behavior changes that were desired by the central government, but unpopular at the local level. Second, I compare the effects when local leaders are upward accountable to when they are not. And third, I study factors related to enforcement. My results show that community meetings only change behavior of ordinary people toward development desired by the government if local leaders are upward accountable. They also suggest that behavior change is largely involuntary, which confirms upward accountability also within communities.

In the analysis, I pursue the following three objectives. My first objective is to identify the causal effects of *Umuganda* meetings on two unrelated changes in individual behavior. The behavior changes, which were desired by the central government, but arguably unpopular, are modern contraceptive adoption and mosquito bed net acquisition. For identification, I exploit exogeneous weather fluctuations on meeting days. My second objective is to relate the effects to the accountability of local leaders. To do so, I compare the effects just before and after a reform that introduced performance contracts and, thus, increased local leaders' upward accountability to the central government. My third objective is to classify the mechanism of *Umuganda* meetings by whether it generates voluntary or involuntary behavior change. For this purpose, I analyze conception, i.e. women's timing of pregnancy, as a behavior change to evade modern contraceptive adoption and heterogeneity in popular support for mosquito bed nets.

2. 'Accountability' describes a relationship in which a principle delegates a task to an agent and aligns the agent's incentives through threat of enforcement. As a such, accountability is unobservable, but manifests in behavior.

I identify the causal effects of *Umuganda* meetings on behavior change through exogenous variation in rainfall over time. The underlying idea is that rainfall on a meeting day reduces attendance. As a consequence, the meeting is impaired or canceled and cannot assume its function as a forum for discussion and problem solving, which can generate behavior change. Although, I lack data on *Umuganda* meetings, it is well known that these take place on Saturdays. That means, I can isolate their effect from general rainfall effects in estimations of the reduced-form.³ My explanatory variables are the number of rainy Saturdays, Sundays, Mondays, etc. in a village and month that match my panel data on behavior change. The estimate on Saturday rainfall can be interpreted as the effect of a “failed” *Umuganda* meeting that has low or zero attendance. Rainfall on each of the other six weekdays serve as placebo tests and control for potential, general rainfall effects (e.g. related to income from agriculture).

The results show that in the first year after the introduction of performance contracts one failed *Umuganda* meeting in a month significantly reduces the probability of behavior change in the direction desired by the government in the same month. The relative effects are -18% for contraceptive adoption and -10% for bed net acquisition. These effects are large, considering that there could be up to 5 *Umuganda* meetings in a month. Rainy day counts for other weekdays are consistently insignificant. The results are very robust and hold for different rainfall thresholds that are used to define a rainy day. One concern may be that some other regular event is affected by Saturday rainfall. I will discuss this possibility and present evidence that rules out access as an alternative explanation for Saturday rainfall effects at least for modern contraceptive methods.

In the year before performance contracts, rainfall on Saturdays and any other weekday have no significant effect on behavior change. The difference in effects of Saturday rainfall with and without performance contracts is statistically significant for both contraceptive adoption and bed net acquisition. This same effect pattern suggests the introduction of performance contracts as the common underlying shift because both changes in behavior were incentivized, but are otherwise completely unrelated. Any alternative explanation must plausibly affect both outcomes, work through *Umuganda*, be nationwide, and have taken place around the same time. I am unaware of any other such alternative reform or shift. Put into context, the findings show that performance contracts, which strengthened upward account-

3. A related paper is Bonnier et al. (2020). These authors use the same identification strategy in cross-sectional data to estimate the effect of *Umuganda* on civilian participation in the 1994 genocide. Other studies in economics document and exploit the effect of rainfall on attendance at events (see e.g. Madestam et al. 2013; Fujiwara et al. 2016; Collins and Margo 2007).

ability, turned *Umuganda* meetings into an effective tool for implementing behavior changes that were desired by the central government.

Finally, I find evidence suggesting that *Umuganda* meetings generate involuntary behavior change after the introduction of performance contracts. A failed *Umuganda* meeting significantly reduces the probability of conception by 8%, which is an effect in the same direction as that for contraceptive adoption. To yield this same sign result, the function of conception as an evasive behavior must outweigh the mechanical reduction from meetings increasing contraceptive adoption. As will be explained in more detail, this finding strongly suggests that contraceptive adoptions require enforcement and, thus, are to a large extent involuntary. The analysis of heterogeneity in popular support for bed net acquisitions exploits the fact that the incidence of mosquitoes and altitude are inversely related. The results indicate that the effect of *Umuganda* meetings is larger in communities at high altitudes with few or no mosquitoes and low popular support for bed nets. This finding is consistent with the notion that low popular support requires more enforcement to achieve behavior change, further corroborating the claim that meetings generate involuntary behavior change.

The chapter contributes to the literature in several ways. First, it speaks to the literature on local institutions as instruments for development. Research in this field is predominantly experimental and attempts to empower citizens in project implementation through information and training (Olken 2007; Duflo et al. 2015; Björkman Nyqvist et al. 2017; Björkman and Svensson 2009). However, it is well known that elites often control local institutions (Reinikka and Svensson 2004; Anderson et al. 2015). Especially successful interventions may actually have supported leaders to monitor citizens rather than the other way around. Against this backdrop, Casey et al. (2012) unsuccessfully attempt to change local institutions, using participation as a proxy for downward accountability. While Banerjee et al. (2010) suggest the use of new institutions for development, Pradhan et al. (2014) indicate that success of existing institutions is linked to elite support and power. I contribute to this literature by challenging the commonly assumed downward accountability of local institutions. I document strong development through a local institution, community meetings, under upward accountability. Similar to Acemoglu et al. (2014), I also show that participation in local institutions fosters leaders' control over community behavior, invalidating it as an indicator for downward accountability. My findings are novel among microeconomic studies of local institutions, but consistent with cross-country evidence on development after decentralization (see e.g. Enikolopov and Zhuravskaya 2007).

The chapter also speaks to the literature on performance incentives in public administration (Finan et al. 2017). Performance incentives generally strengthen upward accountability.

Similar to the literature on local institutions, some research documents strong improvements in outcomes under performance incentives (Ashraf et al. 2014; Duflo et al. 2012), while other research finds only small, temporary effects (Celhay et al. 2018; Rasul and Rogger 2018; Olken et al. 2014). However, performance incentives can also have negative effects. For example, Khan et al. (2015) finds that performance incentives increase bribes to tax-collectors as it strengthens their bargaining power over taxpayers, and Dhaliwal and Hanna (2017) indicate that monitoring lowers job satisfaction and leads to evasion. I contribute to this literature in a similar way. My results show that leaders implement central government targets only when incentivized, but they also indicate that very strong incentives may fuel perverse measures, in this case severe restrictions of personal freedom and rights. By jointly studying accountability and local institutions, I connect the literature on performance incentives with that on local institutions. Both of these strands attempt to solve the same problem of accountability in delegated tasks. My findings indicate that local institutions and performance incentives are complements rather than substitutes.

Finally, the chapter relates to research on the “dark side” of local institutions.⁴ It is most closely related to Bonnier et al. (2020) who study the same institution, *Umuganda* meetings in Rwanda, in a different context. They find that during the time leading up to the 1994 genocide meetings were used by the old government for propaganda and to agitate people, resulting in larger civic participation in violence. Similarly, Satyanath et al. (2017) show that social clubs in Germany after World War I have spurred recruitment into the Nazi party. I contribute to this literature by providing evidence on involuntary behavior change through *Umuganda* meetings. In addition to negative consequences for outsiders, local institutions can be detrimental for the people in them. This finding speaks to Acemoglu et al. (2014) who suggest that local institutions are used by leaders for social control of civil society in Sierra Leone and relates to research on social sanctions (La Ferrara 2003; Karlan 2007; Miguel and Gugerty 2005).

The rest of the chapter is structured as follows. Section 1.2 provides background information on *Umuganda* as a local institution, on performance contracts, and on development and popular support of targets. Section 1.3 describes the data and its construction. Section 1.4 explains and discusses the empirical strategy. Section 1.5 presents the main results, and section 1.6 studies the mechanism of *Umuganda*. Finally, section 1.7 concludes.

4. I use “dark side” as an attribute for local institutions because participation in them is widely considered to represent social capital. This attribute was originally proposed by Putnam (2000) for social capital.

1.2 Background

1.2.1 *Umuganda*

Umuganda is a traditional, local institution in Rwanda that dates back to pre-colonial times.⁵ Having been used for political mobilization before the 1994 genocide (Bonnier et al. 2020), the new government of Rwanda suspended it thereafter. In 1998, however, this government reintroduced *Umuganda* nationwide to create socio-economic development (MINALOC 2011; RGB 2020). Since then, *Umuganda* was formalized in three stages. In November 2001, it was integrated into the government's Community Development Policy. In June 2005, its organization was harmonized by the National *Umuganda* Policy (MINALOC 2008). And finally, on November 17, 2007, *Umuganda* became a law (Organic Law N° 53/2007). The purpose of these policies was to embed the existing practice of *Umuganda* as a tool for policy-making into the public administrative structure. For my analysis, it is only important that no policy changed *Umuganda* in 2006.

Umuganda is and has most of the time in the past been a mandatory community program for all Rwandan adults on Saturdays. It is organized by a committee of village chiefs and consists of outdoor, physical labor (e.g. clearing bushes or cleaning roads) followed by a meeting (Uwimbabazi 2012). The local leaders typically announce *Umuganda* on the same day through word of mouth and loudspeakers mounted on cars (RGB 2014). During *Umuganda*, all shops must close and public transport stops. To enforce participation, local leaders have the discretion to fine absence by up to 5,000 Rwandan Francs, roughly 9 USD in 2007 and corresponding to half the monthly median wage (MINALOC 2007). Evidence suggests that many Rwandans participate involuntarily at *Umuganda* (Mukarubuga 2004; Uwimbabazi 2012; Purdeková 2011).

I argue that the meetings after physical labor during *Umuganda* affect behavior change. Local leaders use these meetings to mobilize, sensitize and support the population to collectively define and resolve their economic and social problems (MINALOC 2008). In practice, meetings amount to local leaders communicating top-down information about government programs and policies (Uwimbabazi 2012). They are also officially acknowledged by the government as a tool to implement development targets from performance contracts (RGB 2014). With respect to my outcomes, annual reports of Rwanda's Ministry of Health docu-

5. Similar local institutions are common in countries of the African Great Lakes Region, notably Burundi, Ethiopia and South Sudan, and have also have been proposed for other countries. In addition, mandatory community programs also existed in many Soviet countries. In Russia, an equivalent institution was called *Subbotnik*, derived from the word 'subbota' meaning 'Saturday'.

ment that both family planning and mosquito bed nets were regularly discussed and promoted at *Umuganda* meetings (MoH 2009, 2008).

Some dissent exists about the frequency of *Umuganda* after the genocide. Recent government documents and research suggest different numbers of Saturdays with *Umuganda* in a month. Organic Law N° 53/2007 states that it takes place only on the last Saturday (MINALOC 2007), and most official documents follow this representation. In contrast, research suggests that *Umuganda* is, in fact, held every week, as it was before the genocide (Purdeková 2011; Uwimbabazi 2012; NAR and Interpeace 2016). This frequency of *Umuganda* is also reported in the 2008 revised Community Development Policy, which is one of the few deviating government documents (MINALOC 2008). In my main specification, I exploit variation from rainfall on all Saturdays in a month. However, I also isolate and explore the effects of specific Saturdays, such as the last Saturday of every month.

1.2.2 Performance Contracts

In 2006, Rwanda's president, Paul Kagame, introduced Performance Contracts (*Imihigo*) in the public sector and beyond. On April 4, he signed the first contracts with all 30 district executives (the Mayors) to re-tie local government to central authority after decentralization. Immediately after, targets were passed down through cascading contracts to all levels of public administration and even further to the individual household (MINALOC 2010; Purdeková 2011). Contrasting the official portrayal as arrangements that reflect local priorities, performance contracts set targets top-down. In the first year after their introduction, three quarters of districts' targets were national policies and programs (GoR 2008), 71% were quantifiable and a majority was set at 100% (OSSREA 2007). Similar evidence exists at the household level. For example, in the government's 2010 Citizen Report Card survey, 78% of respondents state that they have not participated in formulating their own targets (Munyandamutsa 2011).

Two features of Rwandan performance contracts make them particularly effective for fast-track policy implementation. First, comparable units are regularly ranked against each other. And second, contracts set very strong social and material incentives for relative performance.⁶ The consequence is a rat race in which leaders try to outperform one another. Local leaders implement village targets by letting household heads vow contributions in front of

6. Within administration, rewards are commonly financial bonuses and promotions, whereas sanctions consist of removals from office and public shaming (Murray-Zmijewski and Gasana 2010). At district level, for example, approximately 75% of Mayors were removed from office between 2007 and 2009 due to poor performance (Scher and MacAulay 2010).

the community during *Umuganda* meetings (Bugingo and Interayamahanga 2010). Pledges are then recorded in a household's *Imihigo Booklet* and stamped upon completion. Stamped booklets serve as proof of "good standing" and are necessary to access certain government services like registering a marriage or birth (Sommers 2012; Uwimbabazi 2012). In addition, there are reports of fines, destruction of property and corporal punishment for refusing to contribute (Thomson 2008; Huggins 2009). Along this line, quantitative evidence suggests a large degree of compulsion in the implementation of performance targets at the local level (OSSREA 2007).

1.2.3 Development Targets and Popular Support

Rwanda's government targeted modern contraception and mosquito bed nets as key performance indicators for development. In April 2004, two years before the introduction of performance contracts, it set ambitious goals in both. The aim was to raise modern contraceptive prevalence among women of reproductive age from 4% to 20% until 2010 and the percentage of children sleeping under bed nets from a baseline of 18% to 70% (MoH 2004). Rwanda reached both of these targets. Between 2005 and 2010, modern contraceptive prevalence among women 15 to 49 years old increased from 5.6% to 25.2%, and the percentage of households with at least one mosquito bed net increased from 18% to 83%. During the same time fertility dropped from 6 to 4 children per woman (NISR et al. 2012; NISR and Macro 2006).⁷ Rwandan policy-makers attribute this fast-track development to performance contracts (Scher and MacAulay 2010).

Since the beginning, new users of modern contraception and households owning mosquito bed nets were explicit targets in performance contracts.⁸ However, evidence suggests that many Rwandans did not approve of these behaviors. With respect to modern contraception, strong pro-natalist social norms prevailed after the 1994 genocide, and women using contraception were stigmatized as prostitutes (Kraehnert et al. 2019; Berry 2015; USAID and MoH 2002; Farmer et al. 2015). In addition, access does not seem to have been holding back development. The 2005 Rwandan Demography and Health Survey, for example, documents that only 3% of women reported knowledge, access or cost as a reason for

7. Rwanda's fertility transition between 2005 and 2010 is one of the fastest in history and comparable to that in China. Similar to China, also Rwanda considered to limit the number of children per family by law, but legislation was never passed (News 2007). For more information on the Chinese transition and how it was achieved see Zhang (2017).

8. OSSREA (2007) compiles district level targets and Sommers (2012, Appendix) presents a village leader's contract for the first year of performance contracts (2006-07). Further information on district targets between 2009 and 2013 can be found in RGB (2014).

not using modern contraception, in contrast to 33% quoting pro-natalist and opposing attitudes (NISR and Macro 2006). In fact, the government's own assessment in March 2006, one month before performance contracts, was that 'Up to now there have been very few achievements in part due to a lack of advocacy at all levels of Government and civil society' (MoH 2006, p.16). I argue that performance contracts ensured this advocacy from leaders.

With respect to mosquito bed nets, the claim of low popular support is based on the fact that Rwanda is a high altitude country. Its lowest point is already at 995 meters above sea level and much of the population lives at altitudes where the risk of contracting malaria should be very close to 0% because mosquitoes cannot survive (Bodker et al. 2003). I will explore heterogeneity in altitude as a proxy for popular support in the analysis. One potential reason why Rwanda's government implemented mosquito bed nets, despite low demand in many parts of the country, may be the fact that it received large financial support for this task from international aid donors through the U.S. President's Malaria Initiative and the Roll Back Malaria Partnership.

1.3 Data

In this section, I describe the panel data on outcomes and rainfall. For the analysis, this data is matched through GPS coordinates and time. The panel data on outcomes of behavior change are constructed from retrospectively collected information, using dates and times that have been reported in cross-sectional Demography and Health Surveys (DHS) from Rwanda.

1.3.1 Family Planning

I use information from the 2010 Rwandan DHS to study women's adoptions of modern contraceptive methods and conceptions, i.e. when they become pregnant. The 2010 DHS interviewed 13,413 women who were between 15 and 49 years old and usual residents in 492 different communities. An integral part of each woman's questionnaire was a monthly calendar stretching from January 2005 to the date of the interview. In this calendar, interviewers recorded times of pregnancy and modern contraceptive use through retrospection. To ensure accuracy of the information, interviewers were required to ask a set of different questions in a recursive routine for each entry.

I construct my panel data on family planning outcomes from this retrospective calendar data. First, I build a panel data set indicating whether a woman is pregnant, using modern

contraception or neither. Second, I define the two outcomes of behavior change in family planning, contraceptive adoption and conception. These outcomes are indicators that take the value 100 for behavior change in a given month and 0 otherwise. This coding produces estimates in percentage points later in the analysis. Third, following DHS sampling rules, I set all information to missing for times when a woman is below age 15.

In an interview, months of pregnancy are the first entries recorded in the calendar. Hence, I first explain the coding of conceptions before that of contraceptive adoptions. To record a pregnancy spell in the calendar, interviewers mark the monthly date of birth or termination and write back the status of pregnancy until the sum of marked months equals the number of completed months reported by the woman. As a consequence, pregnancies that end, for example, with birth generally consist of spells of 9 months in the calendar. However, these entries and also those of pregnancies that end with termination are inaccurate in the date of conception by two month and for two reasons.

First, recording the month of birth or termination as a full month of pregnancy implicitly assumes that births or terminations occur on the last day of the month. This assumption is extremely unlikely and passes through to the start of pregnancy. As a consequence, almost all women should already be pregnant one month earlier, i.e. in the month before the start of pregnancy in the calendar. Second, recording only completed months misses month 0 of a pregnancy, which is the month of conception. Around the world, pregnancy is counted to last 40 to 41 weeks, starting on the first day of the last menstrual period. This duration translates to 10 months with the fertilization occurring within 1-3 weeks after starting to count. In this respect, conception as the time of sexual intercourse and decision-making actually takes place another month before the start of pregnancy as recorded in the calendar. Based on these two reasons, I code conception as an indicator that takes the value 100 two months before the start of every pregnancy in the calendar and 0 otherwise. For pregnancies that end in birth, this definition is largely identical to lagging an analogue date-of-birth-indicator by 10 months.

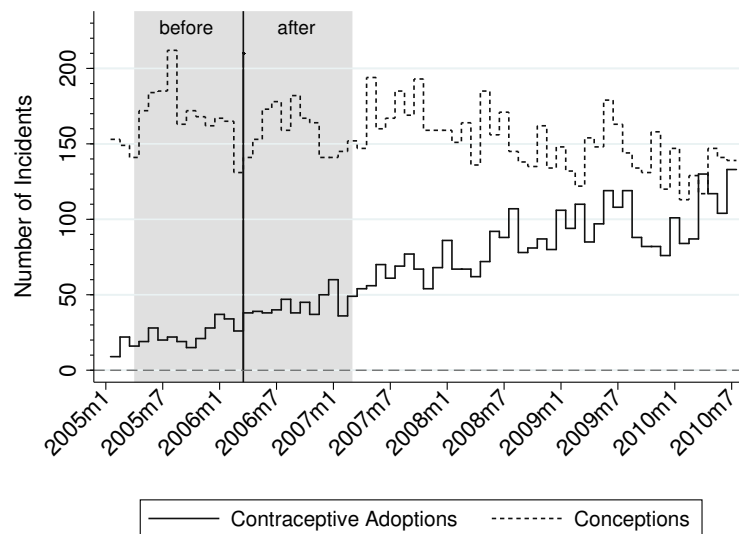
The coding of contraceptive adoption of a modern method is straightforward. Contraceptive adoption is an indicator that takes the value 100 on each start date of modern contraceptive use in the calendar and is 0 otherwise. For the definition of modern contraceptive methods, I follow the standard DHS classification.

For the main analysis, I split the data along the time dimension and restrict the two panel lengths. First, I use April 2006, which is the introduction date of performance contracts, to split the data into a Before and an After panel data set. The reason for this measure is that it simplifies the regression equation, used later to separately estimate the before and

after effects of *Umuganda* meetings. Second, I restrict the two panel lengths to 12 months before and 12 months after the introduction of performance contracts. The purpose of this measure is to narrow the time window around the introduction of performance contracts, which supports attribution of a change in effects to performance contracts.

Figure 1.1 presents the number of contraceptive adoptions and conceptions on each monthly date from February 2005 until July 2010. The solid black, vertical line marks the introduction of performance contracts at the beginning of April 2006. The grey shaded areas left and right of that line mark the lengths of the two panel data sets. No suspiciously high concentrations on certain dates can be observed for any of the two outcomes, suggesting that the calendar data is indeed accurate.⁹ Table 1.1 presents summary statistics of the outcomes for the before and after panel data. Only women with at least two observations are kept in each data set because the inclusion of women fixed effects in my regressions drops women with only one observation.

Figure 1.1: Number of Contraceptive Adoptions and Conceptions over Time



Notes: Based on 13,413 women between 15 and 49 years old and who are usual residents of interviewed households in the 2010 Rwandan DHS. The solid, vertical line marks the introduction of performance contracts in April 2006.

9. This observation is also confirmed when plotting the number of contraceptive adoptions and conceptions over months before the interview (see Figure 1.A1 in the Appendix).

Table 1.1: Summary Statistics of Family Planning Outcomes

<i>A. Before (Apr. 2005 - Mar. 2006)</i>	Mean	Min.	Max.	Obs.
Adoption	0.228	0	100	125,193
Conception	1.610	0	100	125,193
<i>B. After (Apr. 2006 - Mar. 2007)</i>				
Adoption	0.395	0	100	130,966
Conception	1.445	0	100	130,966

Notes: Based on 10,629 women in Before Panel and 11,103 women in After Panel. Only women with at least two panel observations are included in each data set.

1.3.2 Mosquito Bed Nets

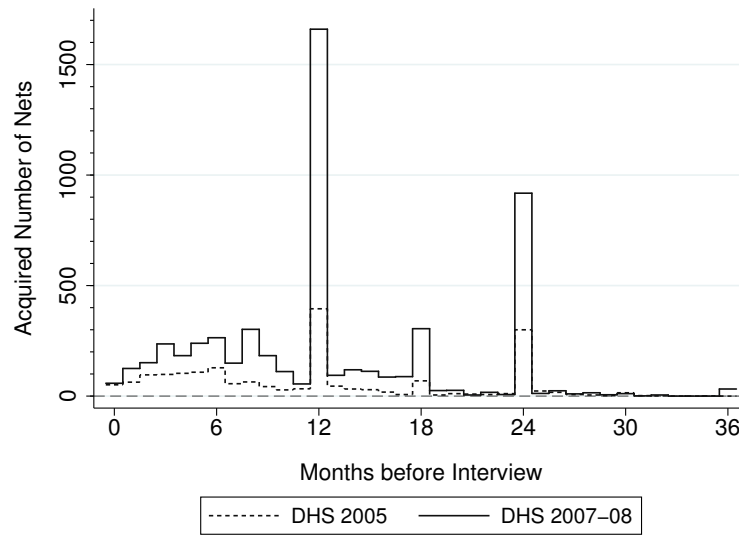
I use information from two Rwandan DHS to study households’ acquisitions of mosquito bed nets before and after the introduction of performance contracts. The 2005 DHS interviewed 10,146 households living in 456 communities with available GPS coordinates. The 2007-08 DHS interviewed 7,287 households in 246 geo-coded communities.¹⁰ Both surveys collect information about mosquito bed nets in households. For each bed net acquired in the past three years, the data records the number of months before the interview when a household obtained the net. In addition, information about the source of the bed net, i.e. from where it was obtained, is available for nets acquired within the previous six months before the interview.

Figure 1.2 presents the raw data of the total number of mosquito bed nets that were acquired in each month before the interview. The figure shows high numbers of acquisitions on months 12, 18 and 24 in both surveys, which indicates that reporting precision deteriorates for bed nets acquired 12 months and more before the interview. Without a routine of questions to ensure data accuracy, similar to that used for 2010 DHS calendar entries, the concentrations likely are due to rounding and imprecise recall. For this reason, I restrict my analysis to bed nets acquired in months 0-11 before the interview.

Based on the raw data, I construct for each DHS a separate, retrospective, household level panel data set spanning 0-11 months before the interview. The 2005 DHS provides data before the introduction of performance contracts and the 2007-08 DHS provides data for the time thereafter. My main outcome is an indicator that takes the value 100 if a household acquired one or more mosquito bed nets in a given month before the interview and 0 otherwise. To later explore heterogeneity in the source of bed nets, I create two additional

10. The analysis samples exclude 126 households from 6 communities without GPS coordinates in DHS 2005 and 90 households from 3 communities without GPS coordinates in DHS 2007-08.

Figure 1.2: Number of Mosquito Bed Nets Acquired in Different Months



Notes: Based on 10,146 and 7,287 households with GPS coordinates in DHS 2005 and DHS 2007-08 data.

outcome indicators of panel length 0-6 months for the 2007-08 data. The first indicator takes the value 100 if at least one bed net in a month before the interview was acquired at a Health Center and is 0 otherwise. The second indicator takes the value 100 if at least one bed net in a month was acquired from Other Sources (e.g. a shop or market) and is 0 otherwise. Table 1.2 presents summary statistics of the panel data on acquisitions of mosquito bed nets.

Table 1.2: Summary Statistics of Acquisitions of Mosquito Bed Nets

<i>A. Before (based on 2005 DHS)</i>	Mean	Min.	Max.	Obs.
Bed Net Acquisition	0.636	0	100	121,752
Acquisition from Health Center	0.275	0	100	71,022
Acquisition from Other Source	0.536	0	100	71,022
<i>B. After (based on 2007-08 DHS)</i>				
Bed Net Acquisition	1.969	0	100	87,444
Acquisition from Health Center	1.049	0	100	51,009
Acquisition from Other Source	1.057	0	100	51,009

Notes: Based on 10,146 and 7,287 households with GPS coordinates in DHS 2005 and DHS 2007-08 data. Panel lengths of variables: 12 months (0-11) for 'Bed Net Acquisition'. 7 months (0-6) for 'Acquisition from Health Center' and 'Acquisition from Other Source'.

1.3.3 Rainfall

I construct my rainfall measures from CMORPH rainfall estimates of the US National Oceanic and Atmospheric Administration's Climate Prediction Center. This data starts in

1998 and has two advantages. First, it has very high spatial and temporal resolution that captures the rainfall variations of Rwanda's many different micro-climates. A tile (data point) in the gridded map has a side length of approximately 8 km (0.073 degrees) and measures rainfall for a 30 minutes interval (Joyce et al. 2004). This resolution facilitates the confinement of rainfall to local communities and single days, with Saturday being the day of *Umuganda* meetings. Second, validation studies suggest that CMORPH rainfall estimates are particularly precise over complex terrain like Rwanda due to the morphing of satellite images and the exploitation of both infrared and microwave electromagnetic radiation (see e.g. Abera et al. 2016). While there will always be some measurement error in satellite rainfall data, this error should work against my findings as long as it is uncorrelated with the outcomes.

I construct my rainfall measures in two steps. First, I aggregate the data to daily estimates and extract rainfall in each community based on its GPS coordinate. Second, I create rainfall measures for each weekday (Mondays, Tuesdays, Wednesdays, etc.) that count the number of "rainy" days on that weekday in a month. A rainy day is defined as a day with rainfall above a certain threshold, and a month is either a calendar month or a month before the interview, depending on the time structure of the outcome data to be matched with. I use round number thresholds from 1 mm up to 10 mm rainfall. My preferred threshold choice is 3 mm rainfall, which I discuss in section 1.4.1. Measures with other thresholds are used to evaluate the robustness of my results to that choice.

For the analysis, rainfall data are matched to outcome data using community and month identifiers. Table 1.3 presents summary statistics of the number of rainy Saturdays in a calendar month for the 492 communities of the 2010 Rwandan DHS between April 2005 and March 2007. The statistics for rainfall on other weekdays and time periods are very similar.

Table 1.3: Summary Statistics of Rainfall on Saturdays

Rainy Saturdays (Definition)	Mean	Std. dev.	Min.	Max.	Obs.
# Sat.(Rainfall>1mm)	1.249	1.125	0	5	11,808
# Sat.(Rainfall>2mm)	0.951	1.017	0	5	11,808
# Sat.(Rainfall>3mm)	0.764	0.938	0	5	11,808
# Sat.(Rainfall>4mm)	0.620	0.838	0	4	11,808
# Sat.(Rainfall>5mm)	0.512	0.756	0	4	11,808
# Sat.(Rainfall>6mm)	0.433	0.695	0	4	11,808
# Sat.(Rainfall>7mm)	0.374	0.641	0	4	11,808
# Sat.(Rainfall>8mm)	0.325	0.597	0	4	11,808
# Sat.(Rainfall>9mm)	0.273	0.547	0	4	11,808
# Sat.(Rainfall>10mm)	0.239	0.509	0	4	11,808

Notes: Based on 492 communities in the 2010 Rwandan DHS between April 2005 and March 2007. # Sat.(Rainfall > Xmm) is the number of Saturdays with rainfall above X mm in a calendar month.

1.4 Empirical Strategy

To identify the effect of *Umuganda* meetings on behavior change, I use variation of rainfall over time. Without data on meetings, I estimate the reduced-form relationship. Rainfall on Saturdays proxies for low participation and cancellation (zero participation). This identification strategy rests on two assumptions. First, Saturday rainfall affects participation at *Umuganda* within a community over time (first stage). Second, the reduced-form effect of Saturday rainfall on behavior change operates only through this channel (exclusion restriction).

1.4.1 Rainfall and *Umuganda*

Saturday rainfall should strongly affect *Umuganda* because meetings and physical labor happen outside under the open sky and are usually only communicated on the same day through word of mouth or loudspeakers mounted on cars driving around (RGB 2014). Hence, both the program and the spread of information about it (e.g. the meeting point and time) are likely inhibited by rainfall. Ideally, I would like to test this first-stage assumption, but data on meetings does not exist. Bonnier et al. (2020) face this same obstacle and make two efforts to substantiate an existing first stage.

First, Bonnier et al. (2020) collect anecdotal evidence in the form of government and media reports on low participation at and cancellations of different *Umuganda* meetings and other events due to rainfall. Second, they estimate an effect of Saturday rainfall on participation at community meetings in neighboring Burundi, using Afrobarometer data. Burundi

shares with Rwanda the same colonial history and is comparable in many socio-economic characteristics. Most importantly, community service with a meeting, very similar to *Umuganda*, is also held every Saturday in Burundi. Its name there is *Ibikorwa rusangi*. Bonnier et al. (2020) find a statistically significant, negative relationship between self-reported frequency of attendance at these meetings and the number of rainy Saturdays in the year leading up to the interview. Both, the anecdotal evidence from Rwanda and the quantitative evidence from Burundi, directly suggest the existence of a first stage, i.e. an effect of Saturday rainfall on participation at *Umuganda* in Rwanda.

Several other studies document and exploit a negative relationship between rainfall and participation at different events. An important example is Madestam et al. (2013). They are the first to use a single, binary rainfall indicator to instrument for participation in Tea Party rallies in the US.¹¹ Similar to Bonnier et al. (2020), I follow this approach because I expect threshold effects, meaning that *Umuganda* is over proportionally affected or fails due to small dips in participation. Uwimbabazi (2012, p.216), for example, suggests this type of effect when she writes that at *Umuganda* ‘successful implementation of any policy can be affected by the absence of the full participation of those especially who should benefit from these policies’. Moreover, the expectation of threshold effects is supported by theory and evidence of collective decision-making and action (see e.g. Olken 2010; Dal Bó et al. 2010; Faillo et al. 2013).

As I cannot empirically determine the relationship between rainfall and participation at *Umuganda*, I choose the threshold that defines a rainy day based on established standards and reasoning. According to the American Meteorological Society, rainfall above 2.5 mm is classified as “moderate” and above 7.5 mm as “heavy” rain (AMS 2012). Madestam et al. (2013) use the first mark and exploit both, moderate and heavy rain, by defining a rainy day to count more than 2.5 mm (0.1 inches) rainfall in their main specification. Bonnier et al. (2020) deviate from this practice and use only heavy rain above a threshold of 10 mm. They motivate this choice with their anecdotal evidence on low participation and cancellations due to rainfall. For these cases, they find daily rainfall to have ranged between 1 mm and 18 mm with a median of 8 mm. However, all events with rainfall of 6 mm and more, except one, are reported as canceled. Hence, *Umuganda* very likely experiences reduced participation already at lower thresholds.

11. Most other studies use continuous measures of rainfall. For example, Collins and Margo (2007) use rainfall in April 1968 to instrument for participation in riots in the US. A large set of studies use rainfall to instrument for voter turnout on election day (see e.g. Fujiwara et al. 2016; Gomez et al. 2007; Hansford and Gomez 2010; Lind 2019; Fraga and Hersh 2010).

Two simple arguments with respect to the size of the complier group also suggest a threshold that is lower than that used by Bonnier et al. (2020). First, the size of the complier group, i.e. the number of people that do not attend *Umuganda* due to rainfall, may actually be already large at low levels of rainfall because most people do not like to attend *Umuganda*. Hence, rainfall reduces the cost of remaining absent rather than increasing the cost of attending. Specifically, I expect rainfall to protect against sanctions for absence because it is a ‘good reason’ in the sense that it is verifiable and because also other people will be absent. Both conditions make enforcement difficult. Second, as stated above, small dips in participation can let *Umuganda* fail in terms of being effective for policy implementation. Consequently, a comparably small complier group may in fact be already sufficiently large to cause strong or even complete impairment of *Umuganda* (e.g. because leaders or the community cannot make binding decisions). In my analysis, this claim should result in similar effect sizes when using different thresholds.

In my preferred specification, I use a threshold of 3 mm to define a rainy day because it is the closest integer number to the standard of 2.5 mm (0.1 inches). However, I show robustness of my results to thresholds between 2 mm and 10 mm rainfall.

1.4.2 Exclusion Restriction

My empirical strategy relies on two exclusion restrictions, one with respect to Saturday rainfall affecting outcomes only through *Umuganda* meetings and the other with respect to the introduction of performance contracts being the only change that altered the objectives of these meetings at the time.

The first restriction assumes that without *Umuganda* meetings rainfall on Saturdays does not affect behavior change. Two characteristics of my analysis limit the scope for a different channel other than *Umuganda*. First, any such channel would need to be time varying due to the inclusion of two-way fixed effects in all my regressions. Second, it would need to be specific to Saturday rainfall because rainfall regressors for every single other day of the week control for general rainfall effects and serve as natural placebo tests. Under these two constraints, an effect of Saturday rainfall must be generated by a reoccurring event on that day. In my analysis, I address this issue and can rule out any channel related to accessing modern contraception (e.g. market days, but also distribution of methods during *Umuganda*), which leaves very little scope for a channel unrelated to the social interactions during *Umuganda* meetings.

The second restriction builds upon the first and assumes that without the introduction of performance contracts *Umuganda* meetings do not affect my outcomes of behavior change. With my panel data, I can attribute the change in effects of meetings to the time when performance contracts were introduced. Hence, some other nationwide policy or change would need to have altered the practice or objectives of *Umuganda* meetings with respect to my outcomes and coincided in timing with the introduction of performance contracts. I am not aware of any such change. It is certain, however, that both my outcomes of behavior change were targeted under performance contracts and that *Umuganda* was used to implement targets.

1.4.3 Specification

To estimate the effect of *Umuganda* meetings on behavior change through OLS, I run variations of the following reduced-form regression:

$$y_{it} = \sum_{d=1}^7 \beta_d \text{rain}_{ctd} + \alpha_i + \tau_t + \varepsilon_{it}. \quad (1.1)$$

y_{it} is a binary indicator of behavior change of the observational unit i during month t . The unit of observation, i , depends on the outcome and is either an individual woman or household. Similarly, t may be either a monthly date or a month before the interview, depending on the panel structure of the outcome. rain_{ctd} is the number of days with rainfall above a specific threshold on weekday d in observational unit i 's community c during month t . Hence, the regression includes seven rainfall variables that count the number of rainy Mondays, Tuesday, Wednesday etc. in each community and month. In my preferred specification, a rainy day is defined by rainfall above 3 mm. α_i and τ_t are observational unit and monthly time fixed effects. In all estimations, I cluster standard errors at the community level because the community (or village) is the entity of *Umuganda* meetings and local leaders' performance targets. This allows the error term, ε_{it} , to be correlated both within communities and over time.

The coefficients, β_d , capture percentage point changes in the probability of behavior change in any given month following from an additional rainy day on the different weekdays in the same month. The interpretation as percentage point changes follows from the dependent variable, the indicator of behavior change, taking values of either 0 (no change) or 100 (change). Most interesting is the coefficient on Saturday rainfall, which can be interpreted as the effect of a failed *Umuganda* meeting. Rainfall on the other weekdays control for general

rainfall effects and are placebo tests. As I will show later, their inclusion in the regressions is unimportant for the results. With unit fixed effects, the coefficients β_d are identified from temporal variation in rainfall and behavior change.

To identify the effect of performance contracts on creating behavior change through *Umuganda* meetings, I estimate equation 1.1 with panel date before and after the introduction of performance contracts. Subsequently, I test whether the corresponding coefficient estimates from both regressions are statistically different. This test essentially evaluates the significance of the Differences-in-Differences.¹² I present estimates from separate regressions with before and after data to interpret each of the two point estimates on Saturday rainfall as the effect of a failed *Umuganda* meeting. Their difference, the Differences-in-Differences, are rather uninteresting. Only their statistical significance is relevant to show that performance contracts led to a change in the practice or objectives of meetings. For this reason, I directly present p-values of the Differences-in-Differences.

1.5 Results

1.5.1 Main Reduced-form Effects

Table 1.4 presents the relationship between the two outcomes, contraceptive adoption and bed net acquisition, and the total number of days with rainfall above 3 mm for each weekday in a month. Both outcomes were targeted by the government for development. The point estimate on Saturday rainfall can be interpreted as the effects of a failed *Umuganda* meeting, which is a meeting that is canceled or has too low attendance for effective decision-making.

Regression 1 uses a 12-months panel of women over the first year after the introduction of performance contracts, i.e. from April 2006 until March 2007. The reduced-form estimate on Saturday rainfall is strongly statistically significant at the 99% confidence level. It suggests that a failed *Umuganda* meeting reduces the probability that a woman adopts modern contraception in a given month by 0.071%. While this absolute effect seems small, the relative effect compared to the unconditional probability is –18%. Reassuringly, rainfall on any other weekday is insignificant.

Regression 2 estimates the same relationship for the year before the introduction of performance contracts, from April 2005 until March 2006. None of the coefficient estimates,

12. I conduct this test by including interactions of all regressors with an after-performance-contracts-dummy, $I(t \geq \text{Apr.2006})$, and estimating this expanded equation jointly with before and after data.

Table 1.4: Main Effects

Dependent variable: Panel data:	Contraceptive Adoption			Bed Net Acquisition		
	after	before	p-value	after	before	p-value
	(1)	(2)	(1) – (2)	(3)	(4)	(3) – (4)
# Sat.(Rainfall>3mm)	-0.071*** (0.023)	0.019 (0.025)	[0.008]	-0.203*** (0.070)	-0.019 (0.032)	[0.017]
# Sun.(Rainfall>3mm)	0.015 (0.027)	0.029 (0.022)	[0.688]	0.090 (0.074)	0.036 (0.035)	[0.512]
# Mon.(Rainfall>3mm)	-0.020 (0.032)	0.004 (0.028)	[0.575]	-0.115* (0.066)	0.009 (0.031)	[0.086]
# Tue.(Rainfall>3mm)	-0.034 (0.033)	-0.009 (0.021)	[0.522]	0.027 (0.070)	0.028 (0.038)	[0.989]
# Wed.(Rainfall>3mm)	-0.036 (0.032)	0.012 (0.021)	[0.213]	-0.019 (0.074)	-0.033 (0.037)	[0.869]
# Thu.(Rainfall>3mm)	-0.014 (0.027)	-0.002 (0.027)	[0.750]	0.065 (0.075)	-0.043 (0.033)	[0.184]
# Fri.(Rainfall>3mm)	0.037 (0.027)	-0.020 (0.023)	[0.106]	0.016 (0.066)	-0.021 (0.032)	[0.611]
Unit FE	Yes	Yes		Yes	Yes	
Time FE	Yes	Yes		Yes	Yes	
Observations	130,966	125,193		87,444	121,752	
R-squared	0.080	0.082		0.086	0.091	
Dep. var. mean	0.395	0.228		1.969	0.636	

Notes: *Imihigo* performance contracts were signed at the beginning of April 2006. 12-months panel data for the time after (before) that date is used in regressions 1 and 3 (2 and 4). The dependent variables, Contraceptive Adoption and Bed Net Acquisition, are monthly, binary indicators. # Sat.(Rainfall>3mm) is the number of Saturdays with rainfall above 3 mm in a calendar month (and similarly for all other weekdays). The unit of observation in regressions 1 and 2 is a woman, and a time step is a monthly date. Regression 1 uses data for April 2006 to March 2007. Regression 2 uses data for April 2005 to March 2006. The unit of observation in regressions 3 and 4 is a household, and a time step is a month-before-the-interview. Regression 3 uses data for 0-11 months before the DHS 2007-08 interview. Regression 4 uses data for 0-11 months before the DHS 2005 interview. Standard errors are clustered at community level. P-value: *** p< 0.01, ** p< 0.05, * p< 0.1.

including that on Saturday rainfall, is statistically significant at any conventional level. P-values for the differences in corresponding coefficient estimates between regressions 1 and 2 are presented one column to the right in Table 1.4. The difference in estimates on Saturday rainfall is statistically significant at the 99% confidence level. No other difference is statistically significant at the 90% level or below. This finding suggests that *Umuganda* meetings became an effective tool for the implementation of national family planning policy after the introduction of performance contracts.

Regressions 3 and 4 estimate an analogue relationships for the second government target, acquisitions of mosquito bed nets. The data structure in these two regressions is slightly different. Acquisitions are observed at the household level and the time dimension of the panel data are months-before-the-interview. Regression 3 uses 12-months panel data on bed net acquisitions 0 to 11 months before DHS 2007-08 interviews, which are data after the

introduction of performance contracts. Regression 4 uses equivalent data from DHS 2005, before the introduction of performance contracts. In regression 3, the reduced-form estimate on Saturday rainfall is highly significant at the 99% confidence level (similar to regression 1). None of the other weekdays is statistically significant at 95% confidence level or high. The relative effect of the point estimate on Saturday rainfall is -10% .

In regression 4, rainfall on all weekdays, including Saturdays, is statistically insignificant at any conventional level (as in regression 2). The p-values for the differences in corresponding coefficient estimates from regressions 3 and 4 demonstrate that only the effect of Saturday rainfall changed significantly at the 95% confidence level. Finding the same pattern of coefficient estimates for another targeted, but otherwise unrelated outcome suggests that in fact performance contracts are responsible for aligning *Umuganda* meetings with national policy. Over the observation period, the practice of *Umuganda* arguably did not change. However, performance contracts allowed the central government to suddenly set the agenda of meetings.

Magnitude The above coefficient estimates can be considered lower bound estimates of a failed *Umuganda* meeting. The estimates should be biased toward zero because the number of rainy Saturdays is an imprecise measure of the number of failed meetings in a month. Hence, relative effects of -18% and -10% suggest that meetings have a very strong effect on the timing of behavior change, especially when also considering that there may be up to 5 meetings in a month. The relative effect size is slightly larger compared to those found by Bonnier et al. (2020) who estimate the relationship between Saturday rainfall leading up to the Rwandan genocide and civilian participation rates in violence using cross-sectional data. For the period from October 1993 until March 1994, which is driving their results, they find that a rainy Saturday (defined by rainfall above 10 mm) reduced civilian participation by 10% compared to the unconditional mean.

1.5.2 Tracing the Effects

Table 1.5 traces the effects of Saturday rainfall on contraceptive adoption and bed net acquisition using the two 12-months panel data sets after the introduction of performance contracts. Regressions 1 and 3 show that only the coefficient estimates on the number of Saturdays with rainfall above 3 mm in the same month are highly significant at the 99% confidence level. All estimates on lagged Saturday rainfall can be considered placebo tests and are insignificant with one exception. The coefficient on the first lag in regression 1 is positive and

statistically significant at the 90% confidence level. This finding may indicate that Saturday rainfall dams up policy implementation and causes a catching up in the following month.

Table 1.5: Tracing the Effects Under Performance Contracts

Dependent variable:	Contraceptive Adoption		Bed Net Acquisition	
	(1)	(2)	(3)	(4)
# Sat.(Rainfall>3mm)	-0.068*** (0.023)		-0.192*** (0.073)	
First Sat.(Rainfall>3mm)		-0.064 (0.057)		-0.097 (0.161)
Second Sat.(Rainfall>3mm)		-0.117* (0.060)		-0.125 (0.144)
Third Sat.(Rainfall>3mm)		-0.085* (0.049)		-0.265* (0.150)
Last Sat.(Rainfall>3mm)		-0.068 (0.053)		-0.285*** (0.104)
L1.# Sat.(Rainfall>3mm)	0.046* (0.027)		0.016 (0.065)	
L2.# Sat.(Rainfall>3mm)	-0.020 (0.025)		0.089 (0.062)	
L3.# Sat.(Rainfall>3mm)	0.011 (0.028)		0.045 (0.064)	
Other Rainfall Regressors	Yes	Yes	Yes	Yes
Unit FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Observations	130,966	130,966	87,444	87,444
R-squared	0.080	0.080	0.086	0.086
Dep. var. mean	0.395	0.395	1.969	1.969

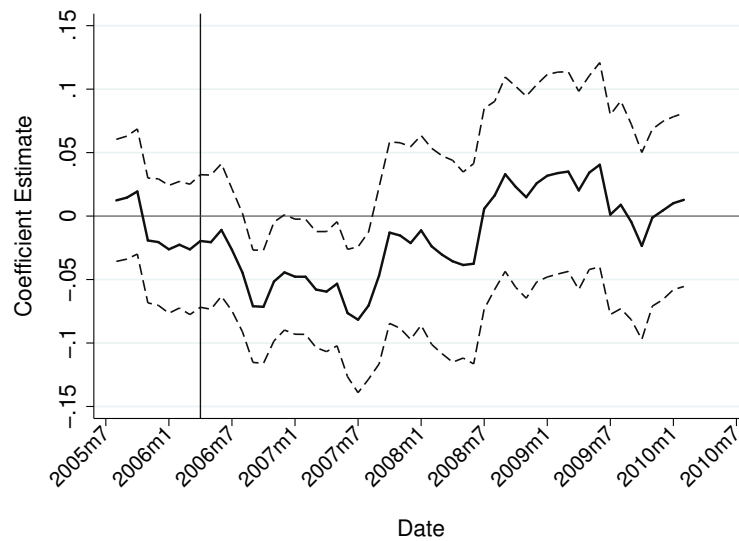
Notes: The dependent variables, Contraceptive Adoption and Bed Net Acquisition, are monthly, binary indicators. # Sat.(Rainfall>3mm) is the number of Saturdays in a month with rainfall above 3 mm. L1.# Sat.(Rainfall>3mm) is this variable lagged by one month (and similar for higher order lags). First Sat.(Rainfall>3mm) is a monthly, binary indicator which takes the value 100 if rainfall on the first Saturday of that monthly date is above 3 mm and 0 otherwise. Second, Third and Last Sat.(Rainfall>3mm) are corresponding indicators for rainfall on the other Saturdays in the calendar month. Other Rainfall Regressors are the numbers of days with rainfall above 3 mm for every other weekday. In regressions 1 and 2, the unit of observation is a woman, a time step is a monthly date and the data are for April 2006 to March 2007. In regressions 3 and 4, the unit of observation is a household, a time step is a month-before-the-interview, and the data are for 0-11 months before the DHS 2007-08 interview. Standard errors are clustered at community level. P-value: *** p< 0.01, ** p< 0.05, * p< 0.1.

Regressions 2 and 4 evaluate the relationship between the two outcomes and four binary indicators that respectively take the value 1 if rainfall on the first, second, third or last Saturday of a monthly date is above 3 mm and 0 otherwise. The effects of these specific Saturdays are similar to one another within the same regression. If at all, regression 2 suggests that central Saturdays of a monthly date are slightly more important for generating the overall effect of Saturday rainfall in a month, and regression 4 suggests that Saturdays toward the end of a monthly date may be more important. However, none of the differences of Saturday rainfall coefficients in the same regression is statistically significant and all estimates have the same

sign. This finding rather supports the evidence discussed in section 1.2 that *Umuganda* was held on multiple Saturdays every month.

Finally, the long panel data of contraceptive adoption allows me to study the effect of Saturday rainfall over time. Figure 1.3 presents estimates from rolling regressions over the 7th month of a 12-months rolling window. The solid black line connects the coefficient estimates and the dashed curves mark 95% confidence intervals. It can be observed that Saturday rainfall effects become negative and statistically significant upon or shortly after the introduction of performance contracts (vertical line on April 2006). The effect continues to persist for roughly 1.5 years after the introduction of performance contracts and then seems to fade away. This finding may mechanically arise if the fraction of women who are both not using modern contraception and being affected by *Umuganda* meetings decreases over time, which is very likely the case.

Figure 1.3: Effect of Saturday Rainfall on Contraceptive Adoption over Time



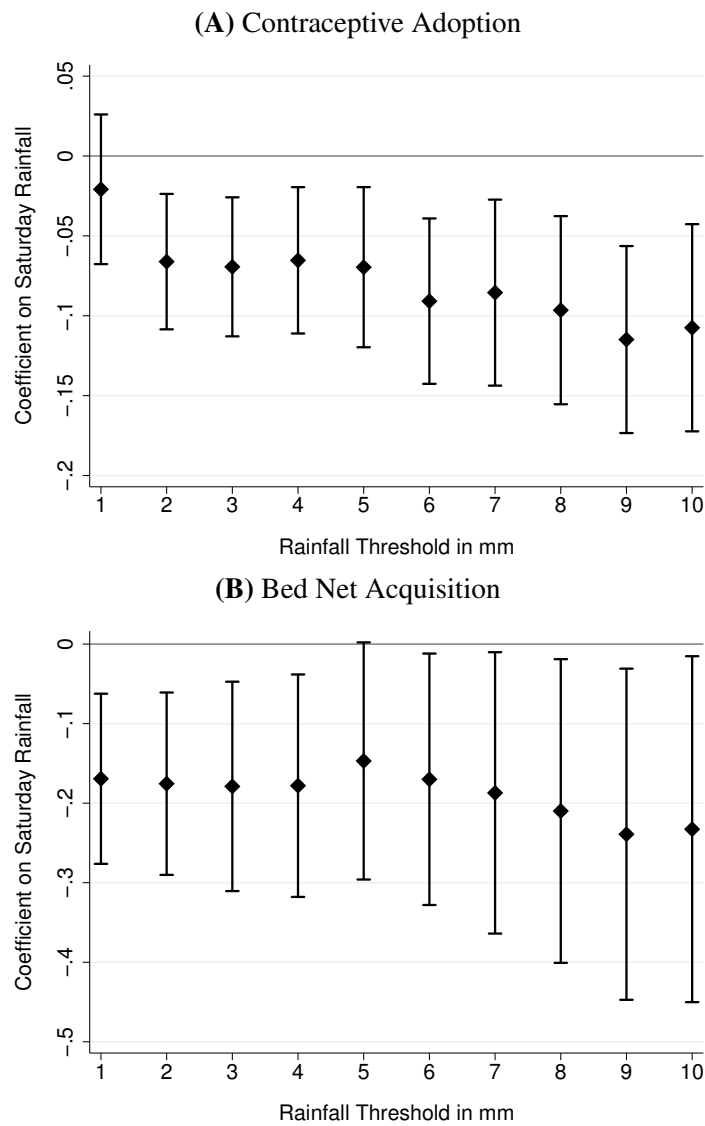
Notes: The figure presents rolling window coefficient estimates on # Sat.(Rainfall>3mm) (solid line) and 95% confidence intervals (dashed lines). The dependent variable is a monthly, binary indicator of contraceptive adoption. # Sat.(Rainfall>3mm) is the number of Saturdays with rainfall above 3 mm in a calendar month. All regressions include analogue rainfall regressors for the other weekdays. The rolling window size is 12 months. The regression estimates are displayed above the 7th month of the rolling window. The vertical line on April 2006 marks the introduction of performance contracts. Standard errors are clustered at the community level.

1.5.3 Robustness Checks

In this section, I present additional robustness checks. Figure 1.4 shows robustness of the results after the introduction of performance contracts with respect to two dimensions. First, the effects of Saturday rainfall are robust to using different thresholds to define a rainy day.

And second, they are largely unaffected by potential multicollinearity in the weekly rainfall variables. In Figure 1.4, each coefficient estimate (diamond) and 95% confidence interval (capped bar) is obtained from a separate regression of the outcome on the number of Saturdays with rainfall above a certain threshold, controlling only for unit of observation and time fixed effects, but not rainfall on other weekdays.

Figure 1.4: Effects of Saturday Rainfall at Different Thresholds Under Performance Contracts



Notes: The figures present the coefficients (diamonds) and 95% confidence intervals (capped bars) on the number of rainy Saturdays (# Sat.(Rainfall > X mm)) when varying the rainfall threshold in separate regressions. The dependent variables, Contraceptive Adoption and Bed Net Acquisition, are monthly, binary indicators. All regressions control for unit of observation and time fixed effects. Figure (A) uses monthly-date panel data of women for April 2006 to March 2007. Figure (B) uses months-before-interview panel data of households 0-11 months before DHS 2007-08. Standard errors are clustered at the community level.

Panel (A) shows that the effect of Saturdays rainfall on contraceptive adoption is statistically significant when using thresholds between 2 mm and 10 mm rainfall. Panel (B) shows that also the effect on bed net acquisition is statistically significant for 9 out of 10 thresholds at the 95% confidence level. The coefficient estimates with a 3 mm threshold are very similar to the estimates in Table 1.4, suggesting that multicollinearity between the rainfall regressors does not affect the results. In addition, the similar effect sizes across the different definitions of a rainy day are consistent with threshold effects in participation at *Umuganda*, as proposed in section 1.4.1.

Figures 1.A2 and 1.A3 in the appendix present results from equivalent regressions of the two outcomes on rainfall on each of the other weekdays. In this battery of 120 regressions only one coefficient estimate is marginally statistically significant at the 95% confidence level, which can be expected to occur by chance from this large number of multiple hypothesis testing.

1.5.4 Alternative Channels

Strong evidence suggests that access to contraception and bed nets does not generate my results. With respect to contraception, two complementary pieces of information rule out this explanation. First, hormonal contraceptives (injections, pills, IUDs and implants), as the most commonly used class of methods in Rwanda, were only available at health facilities, i.e. health centers and hospitals (USAID et al. 2011). And second, at the time of the analysis, health centers were closed and hospitals had high surcharges on weekends (Ueberschär 2018), preventing access on Saturdays.

My data supports this argument. In the 2010 DHS, 95% of hormonal method users (and 91% of any modern method users) report that their first source for the method was a health facility. Adoptions of these methods make up 90% of all adoptions and drive my results. Evidence that health facilities are closed on weekends comes from 52,539 vaccinations with their exact dates copied from children's health cards in the 2010 DHS. Only 3% of vaccinations took place on a Saturday or Sunday. Further support with respect to family planning is provided in the 2007 Rwandan DHS Service Provision Assessment. It documents (and these numbers are likely overreported) that most health facilities provided family planning services on five or less days a week (NISR et al. 2008), which most likely excludes weekends.

For mosquito bed nets, information on the source is available if the net is obtained up to 6 months before the interview. Table 1.6 presents results for the time after the introduction of performance contracts. The three outcomes are binary, monthly indicators that take the

value 100 if a bed net was acquired from a specific source and 0 otherwise. Regression 1 estimates the relationship between rainfall and bed net acquisition from any source (as in Table 1.4) on this short panel. All results hold and are very similar. Regressions 2 and 3 only use acquisitions from health facilities and other sources, respectively. The results clearly show that acquisitions from other sources, mostly shops and pharmacies, are generating the effect of Saturday rainfall. Hence, access to health facilities cannot be the channel.

Table 1.6: Bed Net Acquisition from Different Sources Under Performance Contracts

Dependent variable:	Bed Net Acquisition from ...		
	Any source	Health Facility	Other source
	(1)	(2)	(3)
# Sat.(Rainfall>3mm)	-0.211** (0.100)	-0.029 (0.070)	-0.177** (0.073)
# Sun.(Rainfall>3mm)	0.143 (0.093)	0.034 (0.064)	0.114 (0.073)
# Mon.(Rainfall>3mm)	-0.099 (0.083)	-0.015 (0.052)	-0.072 (0.062)
# Tue.(Rainfall>3mm)	0.031 (0.095)	0.010 (0.065)	0.024 (0.068)
# Wed.(Rainfall>3mm)	0.063 (0.100)	0.103 (0.066)	-0.055 (0.072)
# Thu.(Rainfall>3mm)	0.115 (0.114)	0.039 (0.075)	0.093 (0.085)
# Fri.(Rainfall>3mm)	0.014 (0.097)	0.034 (0.065)	-0.010 (0.072)
Household FE	Yes	Yes	Yes
Months-before-Interview FE	Yes	Yes	Yes
Observations	51,009	51,009	51,009
R-squared	0.147	0.148	0.150
Dep. var. mean	2.082	1.049	1.057

Notes: The dependent variables, bed net acquisitions from different sources, are monthly, binary indicators that take the value 100 if a bed net was acquired from that source and 0 otherwise. # Sat.(Rainfall>3mm) is the number of Saturdays with rainfall above 3 mm in a month before the interview (and similarly for all other weekdays). The data are for 0-6 months before the DHS 2007-08 interview. Standard errors are clustered at community level. P-value: *** p< 0.01, ** p< 0.05, * p< 0.1.

However, health facilities are the only source for the in Rwanda commonly used hormonal contraceptive methods. Consequently, it is reasonable to conclude that the same relationship pattern between the two targeted outcomes and Saturday rainfall must be generated by something else than access. I claim *Umuganda* meetings are generating the effects as they are known to regularly take place on Saturdays and the two analyzed behavior changes were explicitly discussed there.

1.6 Mechanism

Having documented a strong, robust effect of Saturday rainfall on two targeted behaviors, I now provide evidence indicating a mechanisms of *Umuganda* based on pressure. First, I explore conception as an evasive behavior that protects against pressure to adopt modern contraception. Second, I study spatial heterogeneity in the prevalence of mosquitoes as a proxy for popular support for bed nets.

1.6.1 Evasive Behavior

Enforcement creates evasion. With respect to contraception, conception can be considered an evasive behavior because pregnant women must not use it. If the mechanism of *Umuganda* is based on pressure then meetings should also increase conception, which is tantamount to a negative correlation between Saturday rainfall (failed meetings) and conception. However, the probability of conception depends mechanically and inversely on contraceptive use. With a negative effect of Saturday rainfall on contraceptive adoption, documented above, this connection predicts a positive correlation between Saturday rainfall and conception. Hence, we can infer the mechanism of *Umuganda* from the coefficient in a regression of conception on Saturday rainfall. A negative coefficient suggests a mechanism based on pressure, whereas a positive coefficient suggests a mechanism based on free choice.

Regressions 1 in Table 1.7 shows the relationship between conception and rainfall on different weekdays after the introduction of performance contracts. The coefficient estimate on Saturday rainfall is negative and statistically significant at the 95% confidence level. It suggests that one failed *Umuganda* meeting reduces the probability to become pregnant in the same month by 8%. This finding of a negative effect suggests that the mechanism of *Umuganda* is based on pressure. The statistical significance further indicates that a comparably large fraction of women chose conception as a behavior to evade contraceptive adoption. Otherwise, in the displayed average effect of Saturday rainfall on conception, the negative evasion effect would not outweigh the mechanical and positive effect from reduced contraceptive adoptions. Figure 1.A4 in the appendix presents the effects of each weekday at different rainfall thresholds from separate regressions. Panel (A) shows that the effect of Saturday rainfall is robust to using thresholds between 2 mm and 9 mm.

Regression 1 also displays an effect of Wednesday rainfall that is statistically significant at the 95% confidence level. However, Panel (E) in Figure 1.A4 shows that this effect is only statistically significant for rainfall thresholds 3 mm and below. As the effect does not persist

Table 1.7: Mechanism: Conception and Bed Net Acquisition by Altitude

Dependent variable: Panel data:	Conception		p-value (1) – (2)	Bed Net Acquisition	
	after	before		Median split by altitude high	low
	(1)	(2)		(3)	(4)
# Sat.(Rainfall>3mm)	-0.121** (0.048)	0.062 (0.066)	[0.024]	-0.256** (0.099)	-0.178* (0.100)
# Sun.(Rainfall>3mm)	0.050 (0.057)	-0.075 (0.059)	[0.143]	0.089 (0.107)	0.061 (0.113)
# Mon.(Rainfall>3mm)	0.063 (0.062)	0.016 (0.068)	[0.614]	-0.128 (0.079)	-0.094 (0.105)
# Tue.(Rainfall>3mm)	-0.088 (0.061)	-0.053 (0.066)	[0.705]	-0.019 (0.099)	0.064 (0.099)
# Wed.(Rainfall>3mm)	-0.145** (0.062)	-0.005 (0.058)	[0.096]	-0.017 (0.104)	-0.012 (0.106)
# Thu.(Rainfall>3mm)	0.024 (0.052)	-0.133* (0.070)	[0.079]	0.056 (0.095)	0.069 (0.119)
# Fri.(Rainfall>3mm)	0.075 (0.055)	-0.005 (0.062)	[0.323]	-0.033 (0.086)	0.064 (0.103)
Unit FE	Yes	Yes		Yes	Yes
Time FE	Yes	Yes		Yes	Yes
Observations	130,966	125,193		43,500	43,944
R-squared	0.072	0.070		0.083	0.087
Dep. var. mean	1.445	1.610		1.733	2.203

Notes: *Imihigo* performance contracts were signed at the beginning of April 2006. The dependent variables, Conception and Bed Net Acquisition, are monthly, binary indicators. # Sat.(Rainfall>3mm) is the number of Saturdays with rainfall above 3 mm in a calendar month (and similarly for all other weekdays). The unit of observation in regressions 1 and 2 is a woman, and a time step is a monthly date. Regression 1 uses data for April 2006 to March 2007. Regression 2 uses data for April 2005 to March 2006. The unit of observation in regressions 3 and 4 is a household, and a time step is a month-before-the-interview. Both regressions use data for 0-11 months before the DHS 2007-08 interview. Regression 3 only uses observations from villages located above median altitude, whereas regression 4 only uses observations from villages at and below median altitude. The median altitude is 1,670 meters above sea level. Standard errors are clustered at community level. P-value: *** p< 0.01, ** p< 0.05, * p< 0.1..

for larger thresholds, I believe this finding is spurious. The other Panels in Figure 1.A4 document that all effects of other weekdays at different rainfall thresholds are statistically insignificant.

Regression 2 in Table 1.7 shows the relationship between conception and rainfall on different weekdays before the introduction of performance contracts. None of the coefficient estimates is statistically significant at the 95% confidence level. The column to the right displays p-values of the differences in estimates between regression 1 and 2. Only the difference for Saturday rainfall is statistically significant at the 95% confidence level. This finding further corroborates my claim performance contracts turned *Umuganda* meetings into an ef-

fective tool to implement national development targets and that targets were achieved through pressure.

1.6.2 Heterogeneity in Popular Support

Regressions 3 and 4 in Table 1.7 explore heterogeneity in the effects of rainfall on bed net acquisition with respect to altitude. The underlying motivation is the well-known fact that the incidence of mosquitoes strongly decreases with altitude.¹³ This variation in the objective usefulness of mosquito bed nets should be strongly correlated with popular support because people are reluctant to invest time and money for something they do not need. With lower levels of support, more pressure (enforcement) is needed to create behavior change. Hence, if the mechanism of *Umuganda* is based on pressure, we can expect to find larger effects in high altitude areas.

In the 2007-08 DHS, the median community is located at an altitude of 1,670 meters. At this altitude, the risk of contracting malaria should be close to 0%, and we can expect extremely low support for targets in mosquito bed nets. Regression 3 and 4 in Table 1.7 estimate the relationship between acquisition of bed nets and rainfall on different weekdays after the introduction of performance contracts respectively using only communities located above and below median altitude. Saturday rainfall is negative and statistically significant at the 95% confidence level in regression 3 and at the 90% level in regression 4. While the difference in coefficients is not statistically significant, the results show a larger point estimate and relative effect for high altitude communities. This finding is consistent with pressure as the mechanism of *Umuganda*. In addition, it shows that Rwanda's government increased the prevalence of mosquito bed nets across the country without accounting for whether they were actually needed or not.

1.7 Conclusion

This chapter studies the role of community meetings and performance incentives in Rwanda to achieve development targets in health. Specifically, it investigates the effects of meetings on two changes in individual behavior that were unpopular among the population, but desired by the central government. The behavior changes are adoptions of modern contraceptive

13. In similar climate and terrain as in Rwanda, Bodker et al. (2003) study the incidence of mosquitoes in Tanzania using mosquito light traps. Their traps caught only 4 mosquitoes a year at altitude 1,700 meters above sea level, compared to 269 mosquitoes at altitude 1,000 meters and 3,282 mosquitoes at altitude 300 meters.

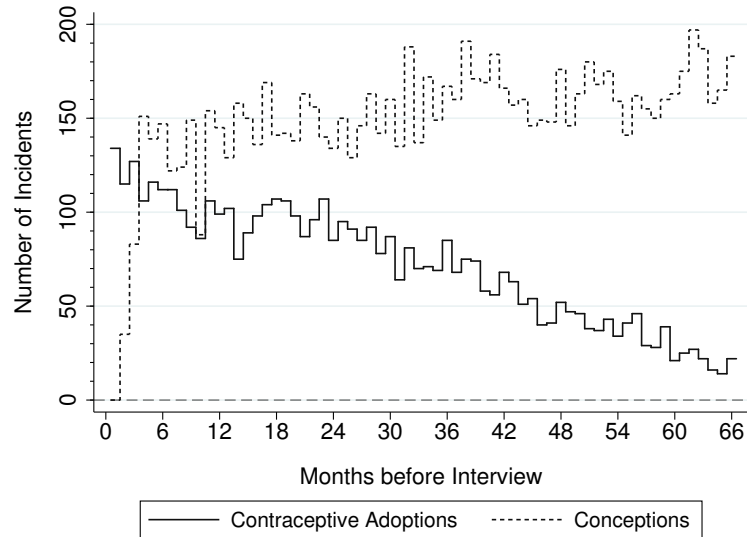
methods and acquisitions of mosquito bed nets. Identification comes from exogenous variation in meeting attendance over time induced by rainfall. This setup allows me to compare the effects before and after a reform that strongly increased upward accountability through performance contracts. I show that local leaders use community meetings on Saturdays to implement targets after the reform. Before the reform, these meetings have no effects. Similar effect patterns in the two unrelated but targeted behavior changes suggest that community meetings and performance incentives are complementary and form a governance system that can be used to implement a wide range of development goals. However, I find evidence that suggests that behavior change is involuntary.

These findings have two important implications. First, they challenge the commonly assumed downward accountability of local institutions when generating development. By showing fast-track development through a local institution under upward accountability, my findings warn that many successful community-based development programs may, in fact, have exploited upward accountability by helping leaders control behavior in their communities. Community-based projects need to be carefully designed in order not to thwart the actual objective of making development more democratic and inclusive. Second, my findings emphasize the importance of accounting for the institutional context in development projects. Performance incentives and community meetings are widely considered good policies on their own. However, their combination can have adverse consequences as suggested by the evidence of involuntary behavior change in Rwanda.

Finally, this chapter provides a new perspective on Rwanda's top-performance in many of the Millennium Development Goals and an explanation for its recent fertility transition. Its lessons may be valuable when assessing current and past developments in other countries, especially when these are authoritarian and aim at exploiting social capital for policy-making as e.g. China.

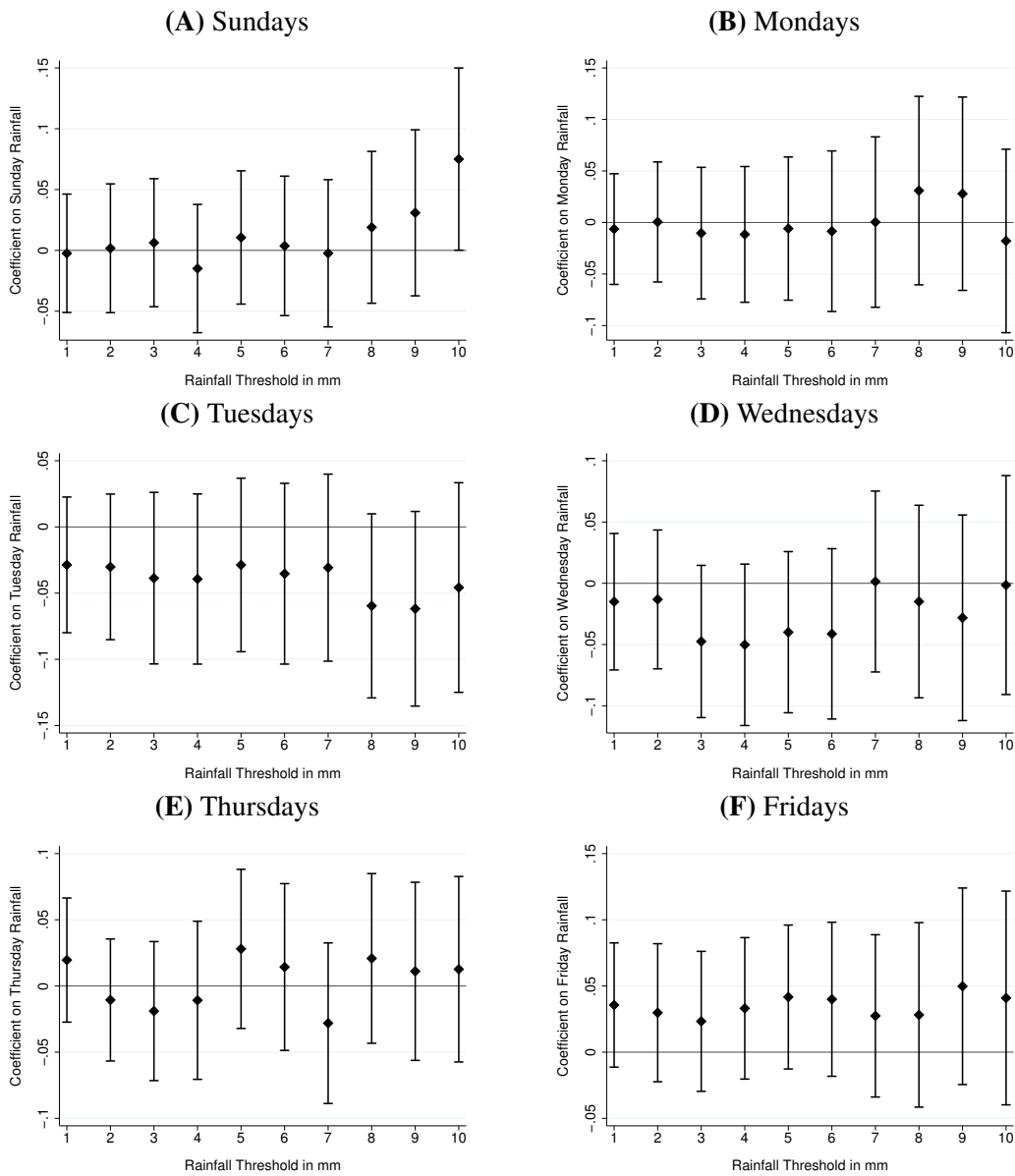
Appendix

Figure 1.A1: Number of Contraceptive Adoptions and Conceptions in Different Months before the Interview



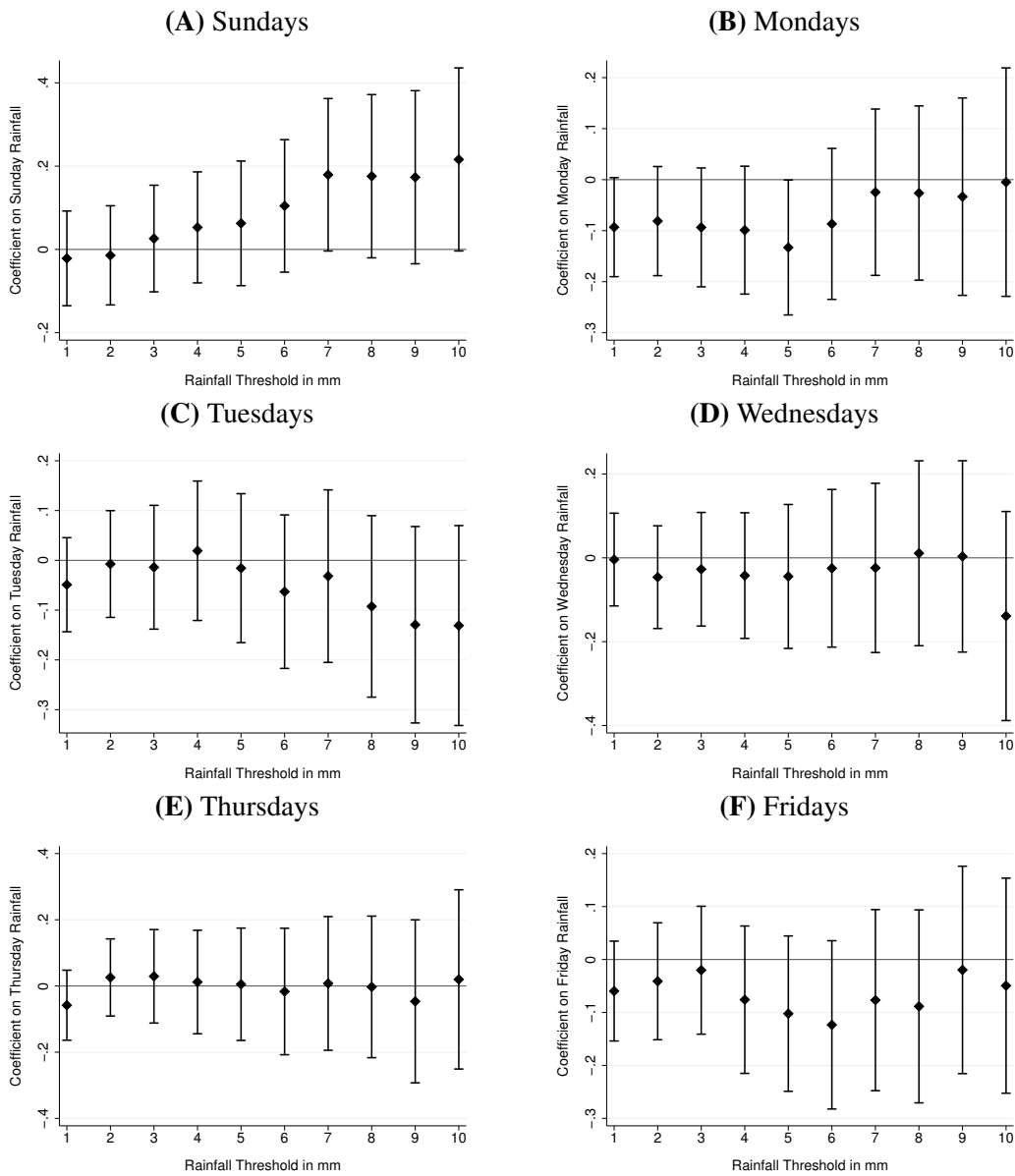
Notes: Based on 13,413 women between 15 and 49 years old and who are usual residents of interviewed households in the 2010 Rwandan DHS. The solid, vertical line marks the introduction of performance contracts in April 2006.

Figure 1.A2: Contraceptive Adoption Under Performance Contracts: Effects at Different Thresholds



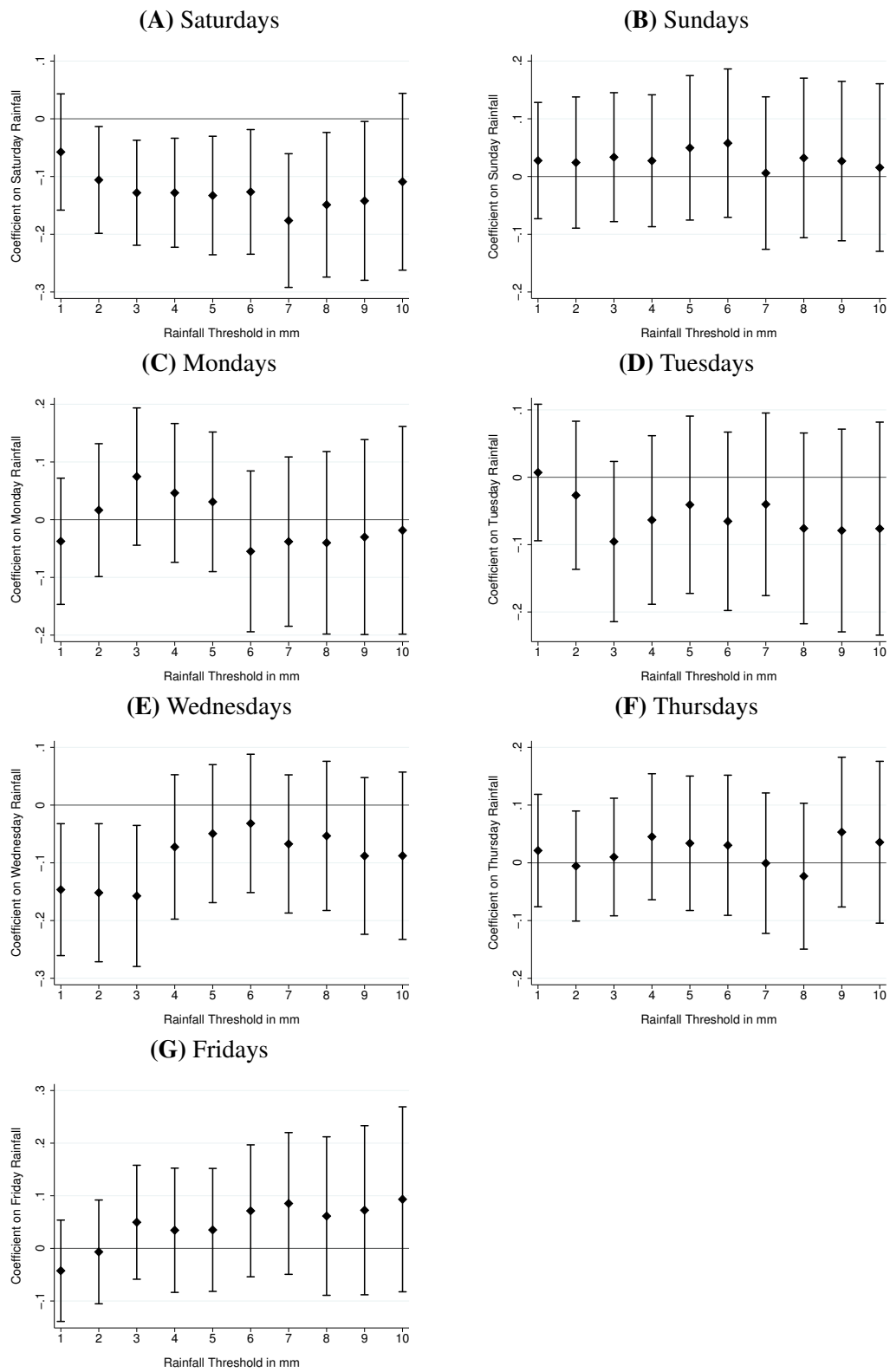
Notes: The figures present the coefficients (diamonds) and 95% confidence intervals (capped bars) on the number of rainy days on different weekdays when varying the rainfall threshold in separate regressions. The dependent variable, Contraceptive Adoption, is a monthly, binary indicator. All regressions control for women and monthly date fixed effects. The data are for April 2006 to March 2007. Standard errors are clustered at the community level.

Figure 1.A3: Bed Net Acquisition Under Performance Contracts: Effects at Different Thresholds



Notes: The figures present the coefficients (diamonds) and 95% confidence intervals (capped bars) on the number of rainy days on different weekdays when varying the rainfall threshold in separate regressions. The dependent variable, Bed Net Acquisition, is a monthly, binary indicator. All regressions control for household and months-before-interview fixed effects. The data are for 0-11 months before DHS 2007-08 interviews. Standard errors are clustered at the community level.

Figure 1.A4: Conception Under Performance Contracts: Effects at Different Thresholds



Notes: The figures present the coefficients (diamonds) and 95% confidence intervals (capped bars) on the number of rainy days on different weekdays when varying the rainfall threshold in separate regressions. The dependent variable, Conception, is a monthly, binary indicator. All regressions control for women and monthly date fixed effects. The data are for April 2006 to March 2007. Standard errors are clustered at the community level.

Chapter 2

Time Use with Basic Income: Evidence from a Large-Scale Survey Experiment

2.1 Introduction

What would people do when they receive cash with no strings attached? This question is at the heart of the ongoing debate on universal basic income. The concept of Basic Income (BI) is a guaranteed minimum income that is provided regardless of behavior, other income and means, and it is universal if everyone belonging to a geographic or political territory receives it. In its simplest form, BI is a regular, unconditional cash transfer. The idea of universal BI is not new, but ignites much hope and controversy over its prospects for the economy and society in general.¹ However, while the aggregate impact of universal BI is certainly important, the effects of BI on individual behavior need to be understood first. As high costs of experiments limit their number and size, a new approach is necessary to assess the effects of BI.

In this chapter, I study the effect of BI on intentions to change time use deploying a large-scale survey experiment. During the survey, treated participants were instructed to imagine a specific BI scenario and think about their intended changes to time use with this BI. In contrast, control participants were instructed to think about intended changes to time use given their current situation. This control group provides a suitable reference that allows to

1. Proponents commonly hope that universal BI secures domestic demand, spurs liberty, and improves social cohesion (Kasy 2018). In contrast, opponents criticize the idea for being too expensive, badly targeted and without incentives to work or contribute to society in some other way (Kearney and Mogstad 2019).

difference out overoptimism in intentions with BI in the treatment group, as shown later.² After treatment or control instructions, all participants received the same question to record their intended changes to time use with respect to seven different activity fields. The seven activity fields are work, education, volunteering, sport, care, socializing, and hobbies.

Treatment scenarios differed in the amount, the duration and the recipient group of BI. The amount of BI was either 500 Euros, 1,000 Euros or 1,500 Euros monthly transfer. The duration of BI, which is the time span over which transfers were promised, was either 1 year or 5 years. Finally, the recipient group of BI was either only the participant or everyone in the country. The combinations of characteristics give rise to 12 treatment groups. Beside these treatment groups and the control group, the experiment contains an additional “level group”. This group was asked to report current levels of time spent on each of the seven activity fields, which provides a baseline to the reported changes in the other groups. Random allocation of respondents to groups, though with different probabilities, allows to directly attribute any differences in intended changes to time use to differences in the BI scenarios.

The data consists of 72,134 respondents in Germany. To collect this data, I collaborated with the NGO *Mein Grundeinkommen*, meaning ‘my basic income’. This NGO promotes the introduction of BI in Germany and operates a crowd-funded lottery of monthly, unconditional cash transfers over the course of one year among its registered users. Participants in the experiment are registered users of *Mein Grundeinkommen* and, thus, familiar with the idea and concept of BI. More specifically, they can be considered experts that have previously imagined winning the BI lottery and the effect this would have on them. As a consequence, they should be particularly apt to imagine and evaluate the effects of BI scenarios in the survey.

The results can be summarized in five main findings. First, compared to the control group, treatment with any BI scenario reduces intended time spent on work, sport, socializing, and hobbies. In contrast, a BI scenario increases intended time spent on education, volunteering, and care. Finding negative effects in non-work activity fields is surprising as economic theory predicts that with rising levels of income people substitute working time for leisure. An explanation for this finding may be that overoptimism in the control group is in excess of that in the treatment group. Irrespective of the level of overoptimism in the control group, the pattern of effects suggests that recipients of BI intend to use it more strongly to in-

2. Overoptimism in intentions refers to participants overestimating the absolute size of changes that they will actually be able to realize. In the chapter, I use the word ‘overoptimism’ to verbally differentiate between the estimations made by participants while answering the survey and the econometric estimations run on that data.

crease their time spent on education, volunteering, and care compared to the other non-work activity fields. This finding indicates that BI may have socially desirable effects.

Second, the main treatment effects of BI scenarios on time use are generally small. The largest absolute effect is an intended 2.5 hours reduction of weekly working time when receiving a BI of 1,500 Euros. Compared against an average working time of 27 hours per week, this effect is a 9% reduction. The largest positive relative effects are in volunteering and education, which are estimated to increase by 32% and 12% with a BI of 1,500 Euros. These effects, however, are outliers and the average relative effect size of 1,500 Euros BI is 9%. Hence, the results suggests that an actual BI with cash transfers will have only small effects on time use and not significantly change people's daily routines.

Third, treatment effects show an interesting non-linearity in the amount of BI. The difference in treatment effects between a BI of 500 Euros and a BI of 1,000 Euros monthly transfers is larger than the difference in effects between a BI of 1,000 Euros and a BI of 1,500 Euros. This pattern is found across all activity fields and is consistent with marginally decreasing utility. Joint with small relative effects, this finding suggests that BI experiments should carefully weigh larger transfers against a larger recipient group. Importantly, however, all treatment effects of 500 Euros BI are either insignificantly different from zero or negative and, thus, opposite to reported intentions. If one discards negative effects, the results also suggest that there exists a threshold between 500 Euros and 1,000 Euros that BI needs to surpass to have any effect on time use.

Fourth, treatment with a BI scenario of 5 years has statistically significant, larger absolute effects than treatment with a scenario of 1 year. The difference in effects, however, is small. This suggests either that BI for more than one year has only moderate effects on time use or that participants do not account for adjustment costs in their intentions. Both explanations are in principle plausible. The first explanation suggests that it may be sufficient to make BI transfers for only one year to test the effects in an experiment. The second explanation suggests that the duration of BI may drive a wedge between initial intentions and subsequent implementation, which could have repercussions on well-being after an experiment.

Finally, treatment effects do not differ whether BI is said to be paid only to the participant or to everyone in the country. This finding may indicate that a temporary BI has no or only small equilibrium effects, as also suggested by Egger et al. (2019) and Haushofer and Shapiro (2016). Alternatively, it may be the case that participants simply do not account for equilibrium effects in their intentions in the survey experiment.

Overall, the above results should inform the design of future BI experiments with actual cash transfers. The main constraint in research on BI are the large financial resources necessary to issue regular, unconditional cash transfers. For this reason, the number and size of experiments will essentially be limited, making it all the more important to meticulously and prudently design each experiment. To improve the experimental design and the choice of outcomes, survey experiments as the one presented in this chapter are a valuable source of information that complements the predictions from economic theory. Similar arguments may suggest to take this approach also for the assessment other potential policies.

The chapter relates to two strands of the economic literature. First, it speaks to the literature on unconditional cash transfers. Since the Negative Tax Experiments in the USA and Canada during the 1970s' and 80s' (Pencavel 1986) only a handful of studies have researched unconditional cash transfers in developed countries (Jones and Marinescu 2018; Akee et al. 2010; Cesarini et al. 2017; Imbens et al. 2001; Kangas et al. 2020). These studies find that there are either no or very small effects on working time. The chapter contributes to this literature by also studying the effects on time spent on other activity fields. Moreover, the chapter analyzes the effect of different amounts, durations and recipient groups of BI in a consistent setting. In this respect, it is particularly suited to inform the design of future BI experiments.

Second, the chapter contributes to a small literature that uses ex-ante, subjective data to assess the likely effects of policies. Studies that take this approach elicit counterfactuals from stakeholders, experts or participants and use these to estimate treatment effects (Hirshleifer et al. 2016; McKenzie 2017; Groh et al. 2016). A common finding in that literature is that all groups overestimate actual treatment effects. However, DellaVigna and Pope (2018) show that estimates from different groups are similar, highly correlated with actual treatment effects and that the average expectation of the crowd outperforms almost every individual. Hence, these approaches are informative about the direction and the relative size of policy effects. The chapter contributes to this literature by extending the approach as a survey experiment with a pure control group that is used to difference out overoptimism.

The rest of the chapter is structured as follows. Section 2.2 explains the experiment. Section 2.3 describes the data and presents average current time use from the level group. Section 2.4 describes the empirical method. Section 2.5 presents the results. Finally, section 2.6 concludes.

2.2 Experimental Design

This section describes the experimental design and the considerations that guided it. The survey experiment was registered before data collection in the RCT Registry of the American Economic Association as trial AEARCTR-0003434 (Linek 2018). The full questionnaire in German language and an English translation are in the appendix.

2.2.1 Assignment to Groups

The online survey tool randomly assigned each survey session to one of 14 experimental groups, 12 treatment groups and 2 reference groups. Respondents in the treatment groups were asked to imagine a BI scenario that varied in its features across the groups. After this instruction, treated respondents were asked about intentions to change their current time use given the scenario. In contrast, respondents in the two reference groups were told to answer subsequent questions given their current situation.

The two reference groups differ in the type of outcome that was recorded. The first reference group, the *control group*, received the same questions on outcomes as the treatment groups, i.e. questions on intended changes to current time use. In comparison, the second reference group, the *level group*, was asked to report levels of their current time use as outcomes. The idea underlying this level group was to obtain information about average current time use of respondents without inflating the number of survey questions. The random assignment allows to extrapolate their responses to all other groups.³

The probability of assignment of each survey session to one group differed between treatment and reference groups. Assignment probability to a treatment group was 5%. Assignment probability to a reference group was 20%.

2.2.2 Treatments

Treatment was one of 12 BI scenarios that participants were asked to imagine after socio-demographic questions in the survey. A treatment scenario read as follows.

3. In principle, the experiment could have asked all respondents to calculate and report “new” levels after changes, holding the number of survey questions constant. Such questions, however, are cognitively much more demanding, and would likely have increased survey attrition and reduced reporting accuracy.

*Imagine your country runs a basic income experiment. **You have been randomly selected to receive a basic income of 1,000 EUR per month for 1 year.***

Please take a short moment and think about whether you intend to use your time differently with this basic income?

The treatment information (bold text) differed across treatment groups. Table 2.1 illustrates the values that each of the three factors of BI can take. The three factors are (i) the recipient group, (ii) the amount, and (iii) the duration of BI. Cross-randomization of these factors gives rise to the 12 treatment groups ($2 \times 3 \times 2$ factor values). To increase salience, all treatment information was colored red in the online survey.

Table 2.1: Factors of Basic Income and their Levels

Factor	Level
Recipient group	<ul style="list-style-type: none"> • <i>You were randomly selected to</i> • <i>All citizens</i>
Amount	<ul style="list-style-type: none"> • <i>500 Euros</i> • <i>1,000 Euros</i> • <i>1,500 Euros</i>
Duration	<ul style="list-style-type: none"> • <i>1 year</i> • <i>5 years</i>

Contrast with treatment groups, the control group received the following instruction: ‘*Please take a short moment and think about whether you intend to use your time differently in your current situation.*’ This sentence is similar to the last sentence of the treatment text, but differs in the qualifier ‘*in your current situation.*’

2.2.3 Outcomes

After receiving treatment or control instructions, participants were asked about intentions to change their current time use in seven activity fields. Table 2.2 presents the seven activity fields with the wording that was used in the survey.

The question on intended changes to current time use in the seven activity fields was posed to participants using the format presented in Table 2.3. For each activity field, respondents were asked to choose exactly one option of intended change to current time use

Table 2.2: Activity Fields

Activity field	Item in the survey
(1) Work	<i>Gainful employment</i>
(2) Education	<i>Education and continuous learning (incl. school, university)</i>
(3) Volunteering	<i>Volunteering</i>
(4) Sport	<i>Sport / fitness / gymnastics</i>
(5) Care	<i>Care for others (e.g. children, parents)</i>
(6) Socializing	<i>Leisure time with others (e.g. partner, friends)</i>
(7) Hobbies	<i>Time for yourself / hobbies (e.g. reading, gardening, watching TV)</i>

by marking the empty circle under the respective heading on the same line. The 9 headings were ‘20+ hours less’, ‘10-19 hours less’, ‘4-9 hours less’, ‘1-3 hours less’, ‘0 (no change)’, ‘1-3 hours more’, ‘4-9 hours more’, ‘10-19 hours more’, and ‘20+ hours more’. The circle under the heading ‘0 (no change)’ was preselected by default for all activity fields. As the list of activity fields is non-exhaustive, intended changes in time use did not need to sum to zero across the fields.

Table 2.3: Precise question and layout

<i>By how many hours would you increase or decrease the time that you spend on following activity fields in a normal week?</i>										
	20+ hours less	10-19 hours less	4-9 hours less	1-3 hours less	0 (no change)	1-3 hours more	4-9 hours more	10-19 hours more	20+ hours more	
• Activity field 1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
• Activity field 2	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	
...					...					
• Activity field 7	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	

For the analysis, I convert hour ranges to numbers of hours. Following the pre-analysis plan, each range is assigned its central number of hours, and the two open ranges, 20+ hours less / more, are assigned 30 hours.

2.2.4 External Validity and Overoptimism

External validity of the survey experiment is an important question. In the present context, external validity is the extent to which the survey experiment is capable to produce the same results as those that would be obtained from a BI experiment run on the same sample, but

with actual transfers. Obviously, without an experiment with actual BI, external validity cannot be established. However, this survey experiment is specifically designed to produce good approximate results from which we can learn about the effects of BI for the design of future, actual experiments.

Two main concerns with the survey experiment are (i) that participants may have difficulty to imagine the BI scenarios and (ii) that intentions differ from actions (see e.g. Bound et al. 2001; Bertrand and Mullainathan 2001; Manski 2004). With regard to the first concern, I intentionally ran the experiment on BI experts, as described in the next section. These experts are familiar with the idea and concept of BI and can be expected to have previously thought about its effects.⁴ As a consequence, they should be particularly apt to realistically imagine the BI scenarios. However, they may still be prone to the second type of error, which is that intentions differ from actions.

The survey and experimental design specifically address this second type of error. First, control and treatment groups were instructed to think about intended changes to time use without obvious restrictions. All instructions have positive connotations and all participants received the same question using the grammatical conjunctive to foster optimism and an enabling choice environment. In this way, the setup intentionally aimed at inducing overoptimism in the reporting of intended changes, meaning that participants report absolutely larger changes than they would actually be able to make with BI.⁵

With overoptimism present in all groups, the control group can be used to difference out the overoptimism in the treatment groups. For this empirical strategy to yield estimates that are free of overoptimism, the magnitude of overoptimism needs to be the same in the control and treatment group. Two features aid this objective. First, participants did not know that they formed part in a survey experiment. And second, each participant received only one instruction. That means, participants could not compare scenarios and adjust their answers to the variations in them. Consequently, overoptimism should be similar across all groups, especially across all treatment groups.

In section 2.4.1, I test for the presence of overoptimism in the control group. In section 2.5.3, I explore heterogeneity in treatment effects with respect to optimism to support the claim that not only the direction, but also the magnitude of overoptimism is the same in the control and treatment group.

4. My data shows that 74% of respondents read, listen or speak about BI at least once a month, and 27% do so at least once a week.

5. An implication of the outlined strategy is that reported intentions to change time use can be considered upper-bounds of changes to times use with actual BI.

2.3 Data

2.3.1 Data Collection

I collected data from registered users of the NGO *Mein Grundeinkommen*. *Mein Grundeinkommen*, which means ‘my basic income’, is a German NGO that promotes the introduction of unconditional BI. Its main activity is a regular, crowd-funded BI lottery among registered users. Winners of the lottery receive unconditional, monthly cash transfers of 1,000 Euros for one year.⁶ In October 2018, *Mein Grundeinkommen* sent out an email newsletter to all its 618,172 registered users at that time with a link to the online survey. The newsletter solely asked users to participate in a scientific online survey on BI. No further information was provided, except that the approximate duration of participation is 7 minutes and responses are anonymous. The survey closed three weeks later, counting 196,728 visits to the welcome page and 85,897 survey starts.

2.3.2 Sample Selection

For the analysis, I use completed surveys of individual respondents that live in Germany. First, I identify individual respondents through the unique combination of IP address, monthly birth date and sex. For these respondents, I only use their first survey start, which restricts the data to a single observation for each respondent and ensures unfamiliarity with the survey and experiment. This measure yields a sample of 83,602 respondents, which corresponds to a response rate of 13.5%. Second, I identify completed surveys by respondents having answered the second to last question, which is a simple yes/no question whether the respondent lives in Germany. 75,456 respondents answered this question, indicating 10% attrition. Finally, I restrict the sample to respondents living in Germany, which reduces the sample to 72,134 respondents.

Table 2.4 displays the number of respondents in the control and treatment groups. Reassuringly, the control group counts close to 20% and each of the 12 treatment groups counts close to 5% of observations. The level group consisting of 14,315 respondents is not included in the table.

6. Every real person can register with *Mein Grundeinkommen* on www.mein-grundeinkommen.de by providing a name, email and birth date. Until October 2018, 200 users had won the BI lottery since the start of *Mein Grundeinkommen* in 2014.

Table 2.4: Number of respondents in control and treatment groups

	Control group	you		all citizens		Total
		1 year	5 years	1 year	5 years	
Control group	14,478					14,478
500 euro		3,614	3,641	3,658	3,526	14,439
1000 euro		3,706	3,592	3,686	3,600	14,584
1500 euro		3,527	3,600	3,547	3,644	14,318
Total	14,478	21,680		21,661		57,819

Notes: The number of respondents in the level group counts 14,315 and is not included in the table.

2.3.3 Integrity of the Experiment

I use the following specification to test for differences in characteristics between control, level, and treatment group that were collected before treatment in the survey.

$$y_i = \beta_0 + \beta_1 L_i + \beta_2 T_i + \varepsilon_i$$

y_i is the characteristic of interest for individual i . L_i and T_i are binary indicators that, respectively, take the value 1 if the individual is in the level group or in the treatment group and are 0 otherwise. ε_i is an idiosyncratic error term. The omitted category are individuals in the control group. Thus, β_1 identifies the difference between individuals in the level and the control group, and β_2 identifies the difference between individuals in the treatment and the control group. Standard errors are clustered at IP addresses, which allows for correlation between respondents that used the same device to answer the survey. The number of IP addresses over the number of respondents is close to 0.937 in all regressions, demonstrating that most respondents used a personal device.

The results are shown in Table 2.5 for characteristics considered particularly interesting. Column (1) reports the mean and standard deviation of the outcome variable in the control group. Columns (2) and (3) report differences in outcomes between level and control group, and treatment and control group. Finally, columns (4), (5), and (6) report number of responses, and minimum and maximum values of the outcomes as additional summary statistics.

The results show that the characteristics are very balanced across the three groups. Statistically significant differences exist only in respondents' age between the control and treatment group and in the number of children in the household between the control group and the other two groups. These differences, while statistically significant, are very small. Considering the testing of 16 multiple hypotheses (8 outcomes \times 2 explanatory variables), two at significant effects at the 95% confidence level are inconspicuous. Thus, the results are proof of the integrity of the experiment.

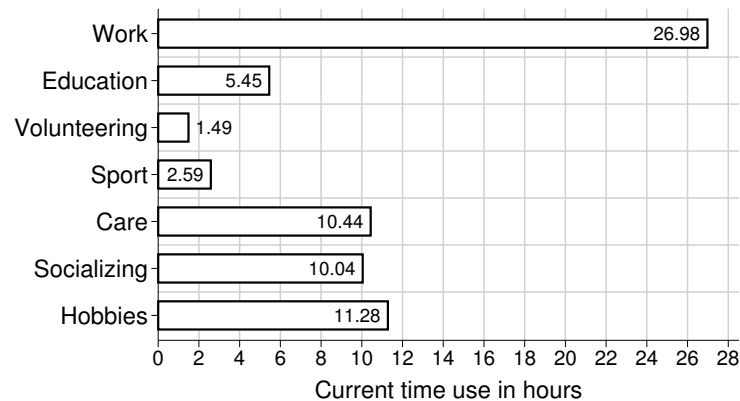
Table 2.5: Differences in Characteristics

	(1) Control mean (std. dev.)	(2) Level group	(3) Treatment group	(4) Obs.	(5) Min.	(6) Max.
Age	43.132 (12.926)	-0.068 (0.156)	-0.303** (0.126)	69,840	1	90
Female	0.620 (0.485)	-0.002 (0.006)	-0.002 (0.005)	71,610	0	1
Completed Abitur (highest German school degree)	0.517 (0.500)	-0.004 (0.006)	0.004 (0.005)	71,991	0	1
Employment is main source of income	0.718 (0.450)	0.008 (0.005)	0.001 (0.004)	71,975	0	1
Monthly disposable income (in Euro)	2278.504 (2659.923)	25.571 (32.967)	12.169 (26.803)	65,522	0	20,000
Number of children in HH	0.589 (0.952)	-0.024** (0.011)	-0.022** (0.009)	71,192	0	11
Life satisfaction (0 = low ; 10 = high)	6.107 (2.014)	0.032 (0.024)	0.025 (0.019)	71,966	0	10
Frequency of having achieved plans (1 = always ; 5 = never)	3.407 (0.841)	-0.015 (0.010)	-0.005 (0.008)	71,588	1	5

Notes: OLS estimates of differences in control, level, and treatment group. Column (1) reports the mean and standard deviation (in parenthesis) of the control group for a given outcome variable. Column (2) reports the difference in outcomes between level and control group, and column (3) reports the difference in outcomes between treatment and control group. Standard errors are clustered at the level of IP addresses and reported in parenthesis below the coefficient estimates. Columns (4), (5), and (6) report number of responses, and minimum and maximum values of the outcomes. Variables "Age" and "Monthly disposable income" are top coded at 90 years and 20,000 Euros. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

2.3.4 Current Time Use

Figure 2.1 presents current time use of respondents in the level group. Due to random allocation, these levels of current time use are a baseline against which we can compare intentions to change time use in the other groups.

Figure 2.1: Current Time Use

Notes: Average current time use of 14,315 respondents in the level group for each activity field.

Figure 2.1 shows that respondents spend on average 27 hours per week on work. They spend in total less than 10 hours per week on education, volunteering and sport, and more than 30 hours per week on care, socializing, and hobbies. In sum, the seven activity fields account for 67.4 hours in a normal week, meaning close to 10 hours of time every day. Considering that a very large fraction of the remaining time cannot be easily altered or delegated (e.g. time for sleep or eating meals), the seven activity fields arguably capture most activities across which we can expect substitutions in time use if people receive regular, unconditional cash transfers.

2.4 Empirical Method

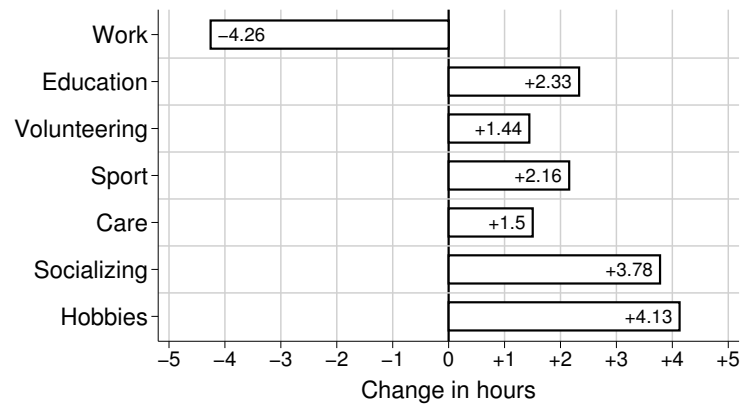
2.4.1 Overoptimism Check

Overoptimism is assumed to prevail in all groups, as explained in section 2.2.4. It raises intended changes to current time use above those that would be realized in an experiment with actual cash transfers. The data allows to check this assumption for the control group. Specifically, I assess whether their reported intentions to change time use are ‘realistic’.

Figure 2.2 presents average intended changes to time use in the control group. Participants in this group intend to increase time spent on each activity field except for work, on which they plan a reduction. The sum of changes over the activity fields amounts to an intended overall increase of 11.08 hours in a normal week. This increase corresponds to 16.5% of total time spent on the activity fields in a week, slightly more than the time spent on them on an average day. Such a large increase is unrealistic for participants in the control group, who cannot expect any external shock that allows them to gain this time from activities not included in the seven fields. Hence, this finding indicates that participants in the control group are overoptimistic in their intended changes to time use.

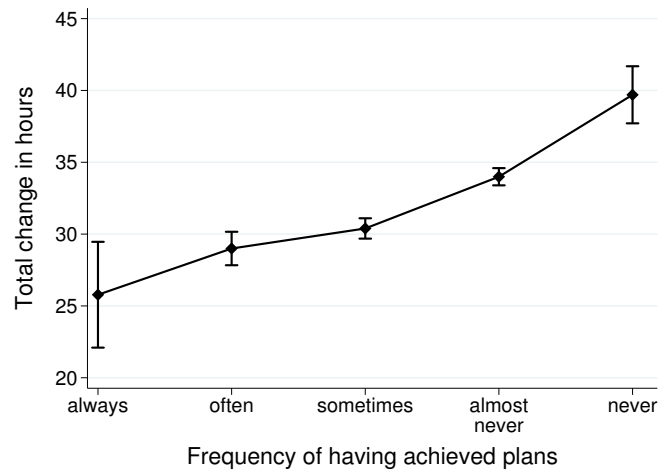
Figure 2.3 provides further support for the claim of overoptimism in the control group. It presents averages (diamonds) and 95% confidence intervals (capped bars) of the sums of absolute changes to time spent on activity fields by the reported frequency of having achieved plans in the four weeks before the survey. The sum of absolute changes to time is the *total change* and measures the size of intended changes to time use. Figure 2.3 shows a very strong positive relationship between this total change and the frequency of having achieved plans. Specifically, respondents that are less likely to have achieved their plans intend to make larger changes to their current time use. While understandable as a desire, implementation becomes

Figure 2.2: Intended Changes to Time Use in the Control Group



Notes: Sum of absolute changes to time spent with each activity field.

Figure 2.3: Intentions and Implementation in the Control Group



Notes: Average intended changes of time spent on each activity field in the control group.

increasingly unrealistic with more ambitious intentions and growing acknowledgment of past implementation failure. With 51% of respondents in the control group reporting to have achieved their plans never or almost never, the average respondent in that group can be considered overoptimistic in her intended changes to time use.

A second implication of the above is that the frequency of having achieved plans in the past can be considered a rank measure of overoptimism. Respondents that are less likely to have achieved their plans are more overoptimistic in their intentions. I will use this rank measure in section 2.5.3 to study whether the control group indeed differences out overoptimism in the treatment groups. Specifically, I study heterogeneity in treatment effects and test for constant effects across the different frequencies of having achieved plans. As intentions vary strongly with this frequency, finding constant treatment effects would suggest the

same variation in overoptimism in control and treatment groups. Hence, this finding supports the claim that the control group differences out overoptimism in the treatment group, which suggests that the results are informative of changes to time use in an actual BI experiment.

2.4.2 Specification

To analyze the average effect of treatment with a BI scenario, I estimate the following OLS regression equation for each outcome.

$$\mathbb{E}_i(\Delta y_i) = \beta_0 + \beta_1 T_i + \varepsilon_i \quad (2.1)$$

An outcome $\mathbb{E}_i(\Delta y_i)$ is individual i 's intended change to time spent on one activity field. T_i is a treatment indicator that takes the value 1 if individual i was treated with a BI scenario and 0 otherwise. ε_i is an idiosyncratic error term that I cluster at the IP address in all estimations. This clustering allows for correlations across different respondents who used the same electronic device to answer the online survey. The 57,819 respondents in the control and treatment groups that are included in these regressions form part of 54,805 cluster, indicating that most participants used a personal device. The coefficient β_0 provides the average intended change to time spent on the activity field in the control group. Coefficient β_1 is the average effect of being treated with any of the 12 BI scenarios.

I use modifications of equation 2.1 to estimate the main effects in the three factors (i) amount, (ii) duration, and (iii) recipient group of BI. Due to independent randomization in the three factors, the main effects of each factor are estimated with data from all control and treatment participants. The specification to estimate the three main effects in the amount of BI is

$$\mathbb{E}_i(\Delta y_i) = \beta_0 + \beta_1 T_{500,i} + \beta_2 T_{1000,i} + \beta_3 T_{1500,i} + \varepsilon_i. \quad (2.2)$$

$T_{500,i}$, $T_{1000,i}$, and $T_{1500,i}$ are treatment indicators that take the value 1 if individual i was treated with a BI scenario with monthly payments of either 500 Euros, 1,000 Euros or 1,500 Euros and are 0 otherwise. The coefficients β_1 , β_2 , and β_3 are the corresponding treatment effects. Analogue specifications are used to estimate the main effects in the other two factors, the duration and the recipient group of BI.

To estimate heterogeneity in treatment effects, I use the following regression equation.

$$\mathbb{E}_i(\Delta y_i) = \beta_0 + \beta_1 T_i \times X_i + \beta_2 T_i + \beta_3 X_i + \varepsilon_i \quad (2.3)$$

X_i is the characteristic across which I expect heterogeneous treatment effects, and the coefficient β_1 is an estimate of this heterogeneity. If X_i is a binary indicator with values 1 and 0, β_1 is the difference in the treatment effect among participants with $X_i = 1$ minus the treatment effect among participants with $X_i = 0$.

Finally, to account for testing the effect of BI on seven outcomes, I present Bonferroni-corrected confidence intervals for the 95% confidence level.⁷ These are presented in addition to standard 95% confidence intervals without adjustment.

2.5 Results

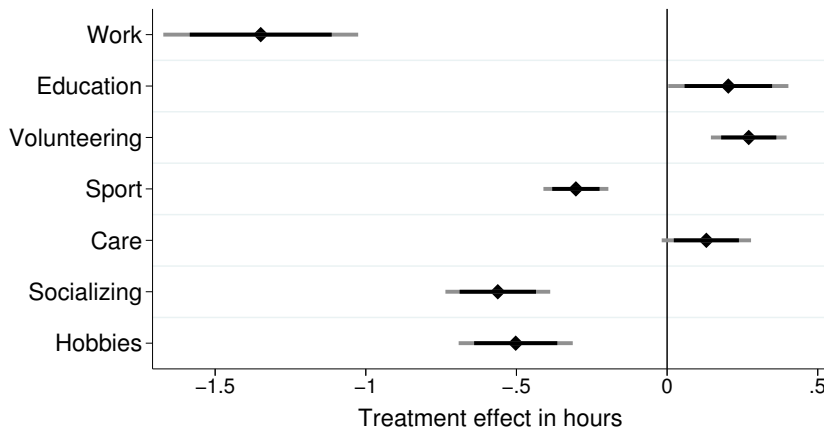
2.5.1 Overall Effects

Figure 2.4 presents the overall effects of treatment with any BI scenario on intended changes to time spent on each of the seven activity fields. These treatment effects are obtained from estimations of equation 2.1. The diamond shaped markers are point estimates, the black horizontal bars are 95% confidence intervals, and the grey horizontal bars are Bonferroni-corrected confidence intervals at that level, accounting for multiple hypothesis testing of the seven effects. Treatment with any BI scenario significantly affects intended changes to time spent on each activity field at the 95% confidence level. When correcting for multiple hypothesis testing, only the effect in care becomes insignificant with a p-value of 0.126.

The results are striking in at least two ways. First, all effects are small. The largest absolute effect of treatment with any BI scenario is a reduction of weekly working time by 1.35 hours, which corresponds to 5% of average working time (compare Figure 2.1). The effects on changes to time spent on the other six activity fields, i.e. non-working time, average 7.5% of time spent on an activity field. Importantly, each treatment effect is absolutely smaller than the intended change to time spent on the respective activity field in the control group (compare Figure 2.2). In this respect, treatment alters the sizes of intended changes, but not their directions. The small magnitudes suggest that BI has only small effects on time use.

Second, effects in non-work activity fields are positive and negative. This finding is surprising as economic theory predicts a substitution of working time for leisure with rising levels of income. While the effect on the intention to change working time is negative, as expected, only half of the effects in non-work activity fields are positive. Most interestingly,

7. The Bonferroni correction for seven multiple hypotheses of 95% confidence intervals corresponds to the 99.3% confidence interval based on the t-distribution.

Figure 2.4: Overall Treatment Effects

Notes: Average effects of treatment with any BI scenario (diamonds). Black horizontal bars are 95% confidence intervals. Grey horizontal bars are Bonferroni-corrected confidence intervals. All estimates are obtained from separate estimations of equation 2.1. Standard errors are clustered for IP addresses. Number of observations: 57,819. Number of clusters: 54,805.

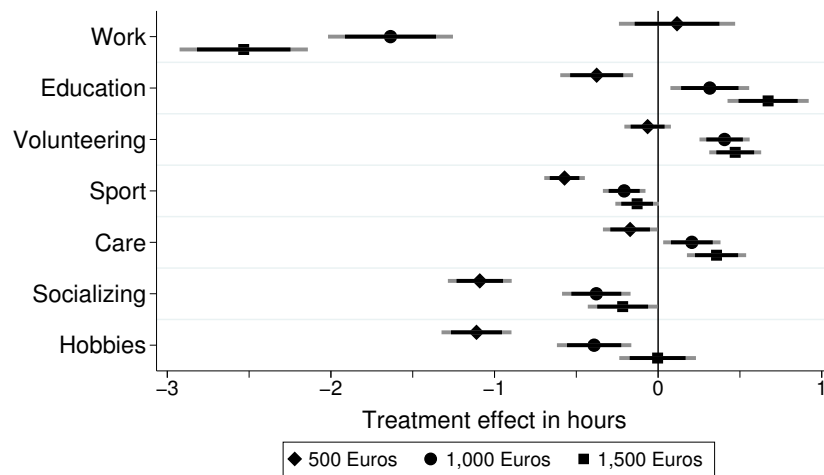
the effects on intended changes to time spent on activity fields that are clearly associated with leisure, i.e. sport, socializing, and hobbies, are all strongly negative, contradicting economic theory. In contrast, effects in the other non-work activity fields, education, volunteering, and care, which are more strongly associated with effortful engagement, are positive. This finding of both positive and negative treatment effects in non-work activity fields potentially arises from the control group's interpretation of the question on intended changes, which may have caused excessive overoptimism beyond that in the treatment group. In any case, the particular pattern across the activity fields suggests that BI may increase some feeling of responsibility in recipients to use this resource, paid for by the community, in a socially desirable way.

2.5.2 Main Effects

Amount of BI

Figure 2.5 presents the main effects of different amounts of BI. It displays point estimates on the treatment indicators for scenarios with monthly transfers of 500 Euros (diamonds), 1,000 Euros (circles) and 1,500 Euros (squares), obtained from estimations of equation 2.2. Black horizontal bars are 95% confidence intervals, and grey horizontal bars are Bonferroni-corrected confidence intervals, accounting for the testing of seven multiple hypotheses.

Figure 2.5 shows that treatment effects are generally small. Only five effects are larger than 10% of time currently spent on the respective activity fields. These are two negative

Figure 2.5: Main Effects: Amount of Basic Income

Notes: Main effects with respect to different amounts of BI in scenarios (diamonds, circles, squares). Black horizontal bars are 95% confidence intervals. Grey horizontal bars are Bonferroni-corrected confidence intervals. Estimates for each outcome are obtained from separate estimations of equation 2.2. Standard errors are clustered for IP addresses. Number of observations: 57,819. Number of clusters: 54,805.

effects in sport and socializing with 500 Euros BI and three positive effects in volunteering with 1,000 and 1,500 Euros BI and in education with 1,500 Euros BI. The relative size of the remaining treatment effects averages 4%. In absolute terms, all main effects of different amounts of BI are smaller than the intended changes to time spent on the respective activity fields in the control group (see Figure 2.2). That means participants treated with different amounts of BI all intend, on average, time reductions in work and time increases in every other activity field.

The intended changes of treated participants contain a non-linearity that is expected from marginal utility theory and evident in the treatment effects. Specifically, the difference in treatment effects of scenarios with 500 and 1,000 Euros BI is larger than the difference in treatment effects of scenarios with 1,000 and 1,500 Euros. This non-linearity is observable in every activity field. It suggests that the marginal effects of BI decrease after some threshold below 1,000 Euros of BI. At the same time, all treatment effects of 500 Euros BI are either insignificantly different from zero or negative, opposing reported intentions. If one discards negative effects, this finding would suggest a threshold effect of BI at some amount above 500 Euros.

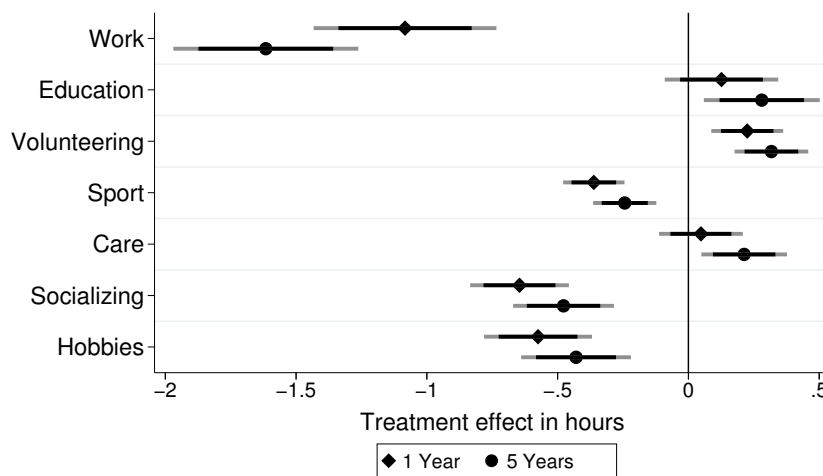
Overall, the findings are informative for future experimental research on BI in three ways. First, the generally small size of treatment effects advises a prudent power analysis and to potentially focus on a limited set of outcomes to avoid pitfalls from multiple hypotheses testing. Second, the onset of marginally decreasing effects after some threshold below 1,000

Euros suggests that BI experiments should carefully weigh larger amounts of BI against a larger recipient group when allocating limited financial resources. And finally third, the results suggest that BI should be larger than monthly transfers of 500 Euros. This last advice is also supported by the recent Finnish BI experiment, which finds effects on perceived well-being, but no effects on behavior with a BI of 560 Euros per month (Kangas et al. 2020).

Duration of BI

Figure 2.6 presents the main effects of different durations of BI. Diamond shaped markers are point estimates of treatment effects when monthly BI transfers are made over 1 year, and circle shaped markers are point estimates of monthly BI transfers over a 5-year period in a scenario. Black horizontal bars are 95% confidence intervals, and grey horizontal bars are Bonferroni-corrected confidence intervals of that level, accounting for seven multiple hypothesis tests. These estimates are obtained from estimating a variation of equation 2.2 that only contains two treatment indicators, one for each duration of BI in scenarios.

Figure 2.6: Main Effects: Duration of Basic Income



Notes: Main effects with respect to different durations of BI in scenarios (diamonds and circles). Black horizontal bars are 95% confidence intervals. Grey horizontal bars are Bonferroni-corrected confidence intervals. Estimates for each outcome are obtained from separate estimations of a variation of equation 2.2. Standard errors are clustered for IP addresses. Number of observations: 57,819. Number of clusters: 54,805.

Treatment effects in Figure 2.6 are larger in absolute terms for scenarios with 5 years compared to 1 year of BI. The difference in treatment effects on working time is 0.53 hours and strongly statistically significant, even after correcting for multiple hypothesis testing. In contrast, the differences in effects on time spent on non-work activity fields are much smaller and surprisingly similar, ranging between 0.09 to 0.17 hours. All of these differences in

non-work activity fields are statistically significant at the 95% confidence level, but become marginally significant and insignificant after correcting for multiple hypotheses testing.⁸

In conclusion, the findings suggest that the duration of BI above 1 year seems to be only a minor determinant of intended changes to time use. While this conclusion has important implications for BI experiments with actual transfers that also study intentions, caution may be warranted when outcomes are behaviors. The recorded intentions may not account for adjustment costs that hamper the implementation of intentions. This interpretation is supported by the similar size of differences in treatment effects across the non-work activity fields despite arguably large differences in adjustment costs. With large adjustment costs, even in only a few activity fields, the duration of actual BI experiments may have strong effects on behavior. In this respect, the duration of BI may drive a wedge between intentions and implementation with potential consequences for participants' well-being after an experiment.

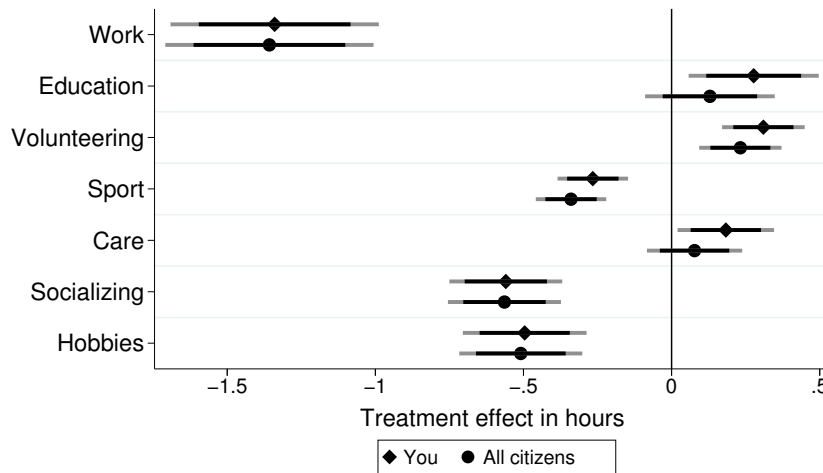
Recipient Group of BI

Figure 2.6 presents the main effects of different recipient groups of BI. Diamond shaped markers are point estimates of treatment effects when only the participant receives BI. Circle shaped markers are point estimates of treatment effects when all citizens in the country receive BI. Black horizontal bars are 95% confidence intervals, and grey horizontal bars are Bonferroni-corrected confidence intervals of that level. The estimates are obtained from estimation of a variation of equation 2.2 that only contains two treatment indicators, one for each recipient group in the BI scenarios.

In every activity field, the two treatment effects are statistically indifferent after correcting for multiple hypotheses testing. This finding indicates that participants do either not anticipate or cannot assess equilibrium effects from every citizen in the country receiving BI. In fact, recent research suggests that unconditional cash transfers have no or only minimal general equilibrium effects (Egger et al. 2019; Haushofer and Shapiro 2016). Hence, the size of the recipient group of temporary BI may not only be irrelevant for intentions, but for actual behavior as well.

8. The differences in treatment effects on time spent on education, volunteering, and hobbies become statistically insignificant when correcting for multiple hypotheses testing.

Figure 2.7: Main Effects: Recipient of Basic Income



Notes: Main effects with respect to different recipients of BI in scenarios (diamonds and circles). Black horizontal bars are 95% confidence intervals. Grey horizontal bars are Bonferroni-corrected confidence intervals. Estimates for each outcome are obtained from separate estimations of a variation of equation 2.2. Standard errors are clustered for IP addresses. Number of observations: 57,819. Number of clusters: 54,805.

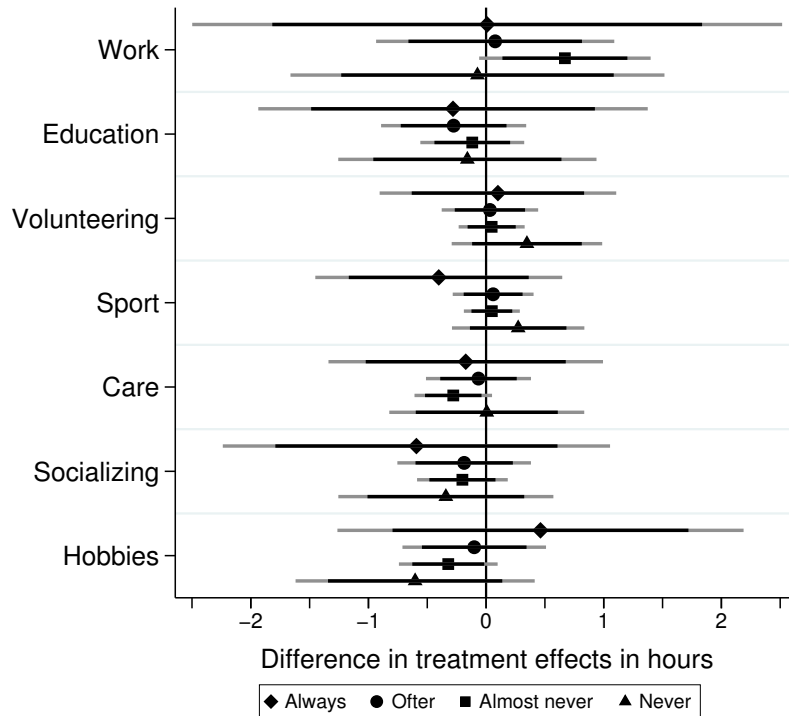
2.5.3 Validity Check

Figure 2.8 presents heterogeneity in treatment effects with respect to the frequency of having achieved plans in the past. Specifically, the figure displays differences in treatment effects (diamond, circle, square and triangles shaped markers) estimated by a variation of equation 2.3 that includes indicators and interactions for each frequency level. The baseline treatment effect is the effect of participants who achieved their plans “sometimes”. The treatment effects for the other frequency levels are evaluated against this baseline. The black horizontal bars are 95% confidence intervals, and the grey horizontal bars are Bonferroni-corrected confidence intervals that account, as before, for seven multiple hypothesis tests.

By and large, Figure 2.8 shows the absence of heterogeneity in treatment effects with respect to the frequency of having achieved plans. Point estimates are very close together at around zero and none of the differences is statistically significant after correcting for multiple hypothesis testing. This finding is instructive considering that there are large differences in intended changes to time use across the different frequencies of having achieved plans (see Figure 2.3).

The absence of differences in treatment effects means that intended changes to time use vary in the same way across the frequency levels in the control and treatment group. With the frequency of having achieved plans being a rank measure of overoptimism (as explained in section 2.4.1 above), this finding strongly supports the claim that the control group dif-

Figure 2.8: Treatment Effect Differences by Frequency of Having Achieved Plans



Notes: Differences in overall treatment effects by frequency of achieving plans. The reference treatment effect is the effect for participants that achieve their plans “sometimes”. Black horizontal bars are 95% confidence intervals. Grey horizontal bars are Bonferroni-corrected confidence intervals. Estimates for each outcome are obtained from separate estimations of a variation of equation 2.3. Standard errors are clustered for IP addresses. Number of observations: 57,819. Number of clusters: 54,805.

ferences out overoptimism of intended changes to time use in the treatment group. Consequently, estimated treatment effects from BI scenarios should be overoptimism-free and informative of treatment effects that can be expected in a BI experiment with actual transfers.

2.6 Conclusion

This chapter deployed a large-scale survey experiment to study intentions to change time use with BI. During the survey, treated participants were instructed to imagine a specific BI scenario and report their intended changes to time use with this BI. In contrast, control participants were asked to think about and report intended changes given their current situation. This control group provides a suitable reference that allows to difference out overoptimism in intentions with BI, which is the tendency to intend larger changes than those that can realistically be implemented. Comparing treated against the control participants thus yields

treatment effects that can realistically be expected from a BI experiment with actual cash transfers. Intended changes to time use were recorded with respect to the seven activity fields work, education, volunteering, sport, care, socializing, and hobbies. To study the effects of different dimensions of BI, scenarios varied in the amount, the duration and the recipient group of BI.

The main findings can be summarized in three points. First, treatment with a BI scenario reduced intended time spent on work, sport, socializing, and hobbies, and increased intended time spent on education, volunteering, and care, compared to the control group. The negative effects of BI in sport, socializing, and hobbies contradict economic theory and likely arise from excessive overoptimism in the control group. Irrespective of this, the results pattern indicates a stronger intended reallocation of time to education, volunteering and care compared to the other three non-work activity fields. In this respect, BI may have socially desirable effects.

Second, treatment effects are generally small. The largest absolute effect is an intended 2.5 hours reduction of weekly working time when receiving a BI of 1,500 Euros. This effect is a 9% reduction. Larger relative effects of a BI with the same amount are found only for education and volunteering. This evidence suggests that BI will not significantly change people's daily routines.

Third, treatment effects display a non-linearity in the amount of BI. For all activity fields, the difference in treatment effects of scenarios with 500 Euros and 1,000 Euros BI is larger than the difference in treatment effects of scenarios with 1,000 Euros and 1,500 Euros. This non-linearity is consistent with marginally decreasing utility of BI. At the same time, however, all effects from 500 Euros BI are negative or statistically indifferent from zero. If one discards negative effects, the results additionally suggest a threshold effect of BI between 500 Euros and 1,000 Euros monthly transfers.

Overall, the survey experiment suggests that future BI experiments should make monthly transfers that are larger than 500 Euros, but not necessarily larger than 1,000 Euros as the marginal utility of BI seems to decrease after that amount. In addition, the small effect sizes suggest that it may be advisable to focus on a limited set of outcomes to sidestep issues of multiple hypotheses testing. The combination of these recommendations calls for a prudent power analyses and carefully weighing larger transfers against a larger recipient group. Finally, future research should verify to what extent the source of monthly transfers, which in the survey experiment is the government, matters for the effects on different outcomes, and whether BI has indeed socially desirable effects.

Appendix

The next pages present the questionnaire in German language and a translation into English.

Fragebogen: Zeitnutzung mit Grundeinkommen **Questionnaire: Time use with basic income**

	Eröffnungstext	Introductory Text
A0	<p>Wir haben 14 bis 19 kurze Fragen an dich. Alle Antworten sind anonym und werden ausschließlich für die Forschung verwendet.</p> <p>Bitte benutze nicht den "Zurück-Button" deines Browsers, da sonst die Umfrage abbricht.</p>	<p>We would like to ask you 14 to 19 short questions. All answers are anonymous and will be used exclusively for research.</p> <p>Please do not use the „Back-Button“ of your browser as it will interrupt the survey.</p>
A	Teil 1:	Section 1
A1	<p>{ALLE}</p> <p>1. Wie zufrieden bist du gegenwärtig, alles in allem, mit deinem Leben?</p> <p>Frageart: Skala 0-10 0 = überhaupt nicht zufrieden ; 10 = sehr zufrieden</p>	<p>{ALL}</p> <p>1. How satisfied are you all in all with your life?</p> <p>Type of question: Scale 0-10 0 = not satisfied at all ; 10 = very satisfied</p>
A2	<p>{ALLE}</p> <p>2. Bitte denke einmal an die letzten 4 Wochen. Wie oft kam es in dieser Zeit vor, ...</p> <p>Liste:</p> <ul style="list-style-type: none"> • ... dass du dich gehetzt oder unter Zeitdruck gefühlt hast? • ... dass du dich ruhig und ausgeglichen gefühlt hast? • ... dass du nicht das geschafft hast, was du dir vorgenommen hast? • ... dass du dir Sorgen um deine wirtschaftliche / finanzielle Situation gemacht hast? <p>Frageart: Auswahl einer Option</p> <ul style="list-style-type: none"> • Immer 	<p>{ALL}</p> <p>2. Please think about the last 4 weeks. During this time, how often ...</p> <p>List:</p> <ul style="list-style-type: none"> • ... did you feel rushed or under time pressure? • ... did you feel calm and balanced? • ... did you not manage to achieve what you had planned? • ... were you worried about your economic / financial situation? <p>Type of question: Selection of one option</p> <ul style="list-style-type: none"> • Always • Often

	<ul style="list-style-type: none"> • Oft • Manchmal • Fast nie • Nie 	<ul style="list-style-type: none"> • Sometimes • Almost never • Never
A3	<p>{ALLE}</p> <p>3. Wie häufig kommt es vor, ...</p> <p><u>Liste:</u></p> <ul style="list-style-type: none"> • ... dass du deinen Kontostand anschaust / überprüfst? • ... dass du über Grundeinkommen liest, hörst oder sprichst? <p><u>Frageart: Auswahl einer Option</u></p> <ul style="list-style-type: none"> • Täglich • Mindestens 1 Mal pro Woche • Mindestens 1 Mal pro Monat • Seltener • Nie 	<p>{ALLE}</p> <p>3. How often do you ...</p> <p><u>List:</u></p> <ul style="list-style-type: none"> • ... check the balance of your bank account? • ... read, listen or speak about basic income? <p><u>Type of question: Selection of one option</u></p> <ul style="list-style-type: none"> • Daily • At least once a week • At least once a month • Less often • Never
A4	<p>{ALLE}</p> <p>4. Wie wichtig, würdest du sagen, ist dir beruflicher Erfolg?</p> <p><u>Frageart: Skala 0-10</u></p> <p>0 = überhaupt nicht wichtig ; 10 = sehr wichtig</p>	<p>{ALLE}</p> <p>4. How important would you say is career success to you?</p> <p><u>Type of question: Scale 0-10</u></p> <p>0 = not important at all ; 10 = very important</p>
A5	<p>{ALLE}</p> <p>5. Und wie wichtig, würdest du sagen, ist (oder war) beruflicher Erfolg für deiner Mutter und deinen Vater in ihrem eigenen Leben?</p>	<p>{ALLE}</p> <p>5. And how important would you say is (or was) career success for your mother and father in their lives?</p> <p><u>List:</u></p>

	<p><u>Liste:</u></p> <ul style="list-style-type: none"> • Wichtigkeit für deine Mutter • Wichtigkeit für deinen Vater <p><u>Frageart: Skala 0-10</u> 0 = überhaupt nicht wichtig ; 10 = sehr wichtig</p>	<ul style="list-style-type: none"> • Importance for your mother • Importance for your father <p><u>Type of question: Scale 0-10</u> 0 = not important at all ; 10 = very important</p>
B	Teil 2	Section 2
B0	<p>{ALLE}</p> <p>6. Bist du ein Crowdhörnchen? (Crowdhörnchen sind Menschen, die automatisch in regelmäßigen Zeitabständen für Mein Grundeinkommen spenden.)</p> <p><u>Frageart: Auswahl einer Option</u></p> <ul style="list-style-type: none"> • Ja • Nein • Weiß nicht / keine Angabe 	<p>{ALLE}</p> <p>6. Are you a Crowdhörnchen? (Crowdhörnchen are people who donate automatically in regular time intervals for Mein Grundeinkommen.)</p> <p><u>Type of question: Selection of one option</u></p> <ul style="list-style-type: none"> • Yes • No • Don' know / no answer
B1	<p>{ALLE}</p> <p>7. In welchem Monat und Jahr wurdest du geboren?</p> <p><u>Liste:</u></p> <ul style="list-style-type: none"> • Monat • Jahr <p><u>Frageart: Dropdown</u></p> <ul style="list-style-type: none"> • Monat: Januar, ..., Dezember • Jahr: 2018, ..., 1910 	<p>{ALLE}</p> <p>7. In which month and year were you born?</p> <p><u>List:</u></p> <ul style="list-style-type: none"> • Month • Year <p><u>Type of question: Dropdown</u></p> <ul style="list-style-type: none"> • Month: Januar, ..., December • Year: 2018, ..., 1910
B2	<p>{ALLE}</p> <p>8. Welches ist dein biologisches Geschlecht?</p>	<p>{ALLE}</p> <p>8. What is your biological sex?</p>

	<p><u>Frageart: Auswahl einer Option</u></p> <ul style="list-style-type: none"> • Weiblich • Männlich • Weder noch 	<p><u>Type of question: Selection of one option</u></p> <ul style="list-style-type: none"> • Female • Male • Neither
<p>B3</p>	<p>{ALLE}</p> <p>9. Hast du die deutsche Staatsangehörigkeit?</p> <p><u>Frageart: Auswahl einer Option</u></p> <ul style="list-style-type: none"> • Ja • Nein 	<p>{ALLE}</p> <p>9. Do you have German citizenship?</p> <p><u>Type of question: Selection of one option</u></p> <ul style="list-style-type: none"> • Yes • No
<p>B4</p>	<p>{ALLE}</p> <p>10. Welchen höchsten Schulabschluss hast du gemacht? (Bei ausländischen Abschlüssen, wähle bitte den ähnlichsten deutschen Abschluss aus.)</p> <p><u>Frageart: Auswahl einer Option</u></p> <ul style="list-style-type: none"> • Keinen Abschluss • Volks- / Hauptschule (DDR: 8. Klasse) • Realschule / Mittlere Reife (DDR: 10. Klasse) • Fachhochschulreife • Abitur 	<p>{ALLE}</p> <p>10. What is your highest school degree? (For foreign degrees please select the equivalent German degree.)</p> <p><u>Type of question: Selection of one option</u></p> <ul style="list-style-type: none"> • No degree • Volks- / Hauptschule (DDR: 8. Klasse) • Realschule / Mittlere Reife (DDR: 10. Klasse) • Fachhochschulreife • Abitur
<p>B5</p>	<p>{ALLE}</p> <p>11. Wie viele Personen gehören inklusive dir selbst zu deinem Haushalt? (Bitte wähle WG aus, wenn du in einer Wohngemeinschaft wohnst.)</p>	<p>{ALLE}</p> <p>11. How many people, including yourself, belong to your household? (Please select WG if you are living in a Wohngemeinschaft [shared flat].)</p>

	<p><u>Frageart: Dropdown</u></p> <ul style="list-style-type: none"> • WG • 1 Person (nur ich) • 2 Personen (inkl. mir) • ... • 10 Personen (inkl. mir) • 11+ Personen (inkl. mir) 	<p><u>Type of question: Dropdown</u></p> <ul style="list-style-type: none"> • WG • 1 Person (only me) • 2 Persons (incl. me) • ... • 10 Persons (incl. me) • 11+ Persons (incl. me)
<p>B6</p>	<p>{ALLE}</p> <p>12. Wie viele Kinder, die du erziehst oder erzogen hast, leben mit dir im selben Haushalt?</p> <p><u>Frageart: Dropdown</u></p> <ul style="list-style-type: none"> • 0 (keine Kinder) • 1 Kind • 2 Kinder • ... • 10 Kinder • 11+ Kinder 	<p>{ALLE}</p> <p>12. How many children that you raise or have raised are living with you in the same household?</p> <p><u>Type of question: Dropdown</u></p> <ul style="list-style-type: none"> • 0 (no children) • 1 Child • 2 Children • ... • 10 Children • 11+ Children
<p>B7</p>	<p>{ALLE}</p> <p>13. Woher beziehst du hauptsächlich dein Einkommen zum Leben?</p> <p><u>Frageart: Auswahl einer Option</u></p> <ul style="list-style-type: none"> • Aus Arbeit • Vom Staat (z.B. gesetzliche Rente, Arbeitslosengeld, Sozialhilfe) • Von der Familie (z.B. Eltern) oder durch die / den Partner*in • Ersparnes, Kapitaleinkünfte, Mieteinnahmen oder private / betriebliche Rente 	<p>{ALLE}</p> <p>13. What is your main source of income for living?</p> <p><u>Type of question: Selection of one option</u></p> <ul style="list-style-type: none"> • From Work • From the State (e.g. statutory pension, unemployment insurance, welfare) • From your family (e.g. parents) or through your partner • Saved money, income from capital, rental revenues or private / company pension • Other income source (e.g. scholarship)

	<ul style="list-style-type: none"> • Andere Einkommensquelle (z.B. Stipendium) 	
B71	<p>{WENN B71="Aus Arbeit"}</p> <p>14. Wie hoch war dein Einkommen vor Abzügen im letzten Monat? (Ohne Einmal- oder Sonderzahlungen. Abzüge sind z.B. Steuern, Krankenkassenbeiträge, Unterhaltszahlungen.)</p> <p>in Euro</p> <p><u>Frageart: Freie numerische Angabe</u></p> <p>Diese Frage ist wichtige. Wenn du diese Frage trotzdem überspringen möchtest, klicke auf weiter.</p>	<p>{IF B71!="From Work"}</p> <p>14. How much was your income before deductions in the last month? (Without one-time or extra payments. Deductions are e.g. taxes, health insurance fees, child support.)</p> <p>in Euro</p> <p><u>Type of question: Numeric response</u></p> <p>This question is important. If you would still like to skip this question click continue.</p>
B721	<p>{WENN B7="Aus Arbeit"}</p> <p>14. Wie hoch war dein Brutto-Einkommen (oder Gewinn vor Steuern) im letzten Monat? (Ohne Einmal- oder Sonderzahlungen oder Entgelte für Überstunden. Dein Brutto-Einkommen ist dein Arbeitseinkommen vor Abzügen wie z.B. Steuern, Krankenkassenbeiträge, Unterhaltszahlungen.)</p> <p>in Euro</p> <p><u>Frageart: Freie numerische Angabe</u></p> <p>Diese Frage ist wichtige. Wenn du diese Frage trotzdem überspringen möchtest, klicke auf weiter.</p>	<p>{IF B7=="From Work"}</p> <p>14. How much was your gross income (or profit before taxes) in the last month? (Without one-time or extra payments or overtime compensation. Your gross income is your income from work before deductions like e.g. taxes, health insurance fees, child support.)</p> <p>in Euro</p> <p><u>Type of question: Numeric response</u></p> <p>This question is important. If you would still like to skip this question click continue.</p>
B722	<p>{WENN B7="Aus Arbeit"}</p>	<p>{IF B7=="From Work"}</p>

	<p>15. Im Vergleich zu folgenden Gruppen, ist dein Einkommen aus Arbeit eher höher oder eher niedriger? (Bitte schätze, wenn du das Einkommen dieser Gruppen nicht kennst.)</p> <p><u>Liste:</u></p> <ul style="list-style-type: none"> • Im Vergleich zu Deinen Arbeitskollegen • Im Vergleich zu Deinen Freunden <p><u>Frageart: Auswahl einer Option</u></p> <ul style="list-style-type: none"> • Höher • Ein wenig höher • Gleich hoch • Ein weniger niedriger • Niedriger 	<p>15. In comparison to following groups, is your income from work rather higher or lower? (Please guess if you do not know the income of these groups.)</p> <p><u>List:</u></p> <ul style="list-style-type: none"> • In comparison to your work colleagues • In comparison to your friends <p><u>Type of question: Selection of one option</u></p> <ul style="list-style-type: none"> • higher • a little higher • equal • a little lower • lower
B723	<p>{WENN B7=="Aus Arbeit"}</p> <p>16. Wie viele Wochenstunden beträgt deine vereinbarte Arbeitszeit ohne Überstunden? (Wenn du keine vereinbarte Arbeitszeit hast, schätze die Anzahl der Stunden, welche du arbeitest.)</p> <p>Anzahl der Wochenstunden</p> <p><u>Frageart: Freie numerische Angabe</u></p>	<p>{IF B7=="From Work"}</p> <p>16. How many hours per week is your working time according to your contract? (If you have no defined number of working hours per week, state the number of hours which you usually work.)</p> <p>Number of hours per week</p> <p><u>Type of question: Numeric response</u></p>
C	Teil 3	
T1	<p>Zufällige Zuordnung zu einer von 14 Gruppen.</p> <p>Gruppen</p> <ul style="list-style-type: none"> • 12 Behandlungsgruppen • 1 Kontrollgruppe • 1 Level-Gruppe 	<p>Random assignment to one of 14 groups.</p> <p>Groups</p> <ul style="list-style-type: none"> • 12 Treatment groups • 1 Control group • 1 Level group

<p>C1</p>	<p>{WENN T1=="Level-Gruppe"}</p> <p>17. Bitte nimm dir kurz Zeit und überlege, ... Wie viele Stunden verbringst du in einer normalen Woche mit folgenden Bereichen? („0“ bedeutet, dass du in einer normalen Woche keine Zeit damit verbringst.)</p> <p><u>Liste:</u></p> <ol style="list-style-type: none"> 1. Berufstätigkeit 2. Aus- und Weiterbildung, Lernen (auch Schule, Studium, Promotion) 3. Ehrenamtliche Tätigkeiten 4. Sport / Fitness / Gymnastik 5. Betreuung anderer (z.B. Kinder, Eltern) 6. Freizeit mit anderen (z.B. Partner*in, Freund*innen) 7. Zeit für dich selbst / Hobbies (z.B. Lesen, Gärtnern, Fernsehen) <p>Anzahl der Wochenstunden</p> <p><u>Frageart: Dropdown</u></p> <ol style="list-style-type: none"> 1. 1 2. 2 3. 3 100. 100 	<p>{IF T1=="Level group"}</p> <p>17. Please take a short moment and think, ... How many hours in a normal week do you spend with following activity fields? („0“ means that in a normal week you do not spend any time with this activity field.)</p> <p><u>List:</u></p> <ol style="list-style-type: none"> 1. Gainful employment 2. Education and continuous learning (incl. school, university) 3. Volunteering 4. Sport / fitness / gymnastics 5. Care for others (e.g. children, parents) 6. Leisure time with others (e.g. partner, friends) 7. Time for yourself / hobbies (e.g. reading, gardening, watching TV) <p>Number of hours per week</p> <p><u>Type of question: Dropdown</u></p> <ol style="list-style-type: none"> 1. 1 2. 2 3. 3 100. 100
<p>C2</p>	<p>{WENN T1!="Level-Gruppe"}</p> <p>17. Wie viele Stunden würdest du in einer normalen Woche mehr oder weniger mit folgenden Bereichen verbringen?</p>	<p>{IF T1!="Level group"}</p> <p>17. By how many hours would you increase or decrease the time that you spend on following activity fields in a normal week?</p>

	<p><u>Liste:</u></p> <ol style="list-style-type: none"> Berufstätigkeit Aus- und Weiterbildung, Lernen (auch Schule, Studium, Promotion) Ehrenamtliche Tätigkeiten Sport / Fitness / Gymnastik Betreuung anderer (z.B. Kinder, Eltern) Freizeit mit anderen (z.B. Partner*in, Freund*innen) Zeit für dich selbst / Hobbies (z.B. Lesen, Gärtnern, Fernsehen) <p>Frageart: <u>Auswahl einer Option, Default 0</u></p> <ul style="list-style-type: none"> 20+ Stunden weniger 10-19 Stunden weniger 4-9 Stunden weniger 1-3 Stunden weniger 0 (keine Veränderung) 1-3 Stunden mehr 4-9 Stunden mehr 10-19 Stunden mehr 20+ Stunden mehr 	<p><u>Liste:</u></p> <ol style="list-style-type: none"> Gainful employment Education and continuous learning (incl. school, university) Volunteering Sport / fitness / gymnastics Care for others (e.g. children, parents) Leisure time with others (e.g. partner, friends) Time for yourself / hobbies (e.g. reading, gardening, watching TV) <p>Type of question: <u>Selection one option, Default 0</u></p> <ul style="list-style-type: none"> 20+ hours less 10-19 hours less 4-9 hours less 1-3 hours less 0 (no change) 1-3 hours more 4-9 hours more 10-19 hours more 20+ hours more
<p>D</p> <p>D1</p>	<p>Teil 4</p> <p>{ALLE}</p> <p>2 Fragen noch...</p> <p>18. Wohnst du in Deutschland? (Wenn du keinen Wohnort hast, beantworte die Frage bitte für deinen gegenwärtigen Aufenthaltsort.)</p> <p>Frageart: <u>Auswahl einer Option</u></p> <ul style="list-style-type: none"> Ja Nein 	<p>Section 4</p> <p>{ALL}</p> <p>2 questions left...</p> <p>18. Do you live in Germany? (If you do not have a residence, answer the question for your current whereabouts.)</p> <p>Type of question: <u>Selection of one option</u></p> <ul style="list-style-type: none"> Yes No

D21	<p>{WENN D1==„Ja“}</p> <p>Wie lautet die Postleitzahl deines Wohnortes? (Wenn du keinen Wohnort hast, beantworte die Frage bitte für deinen gegenwärtigen Aufenthaltsort.)</p> <p><u>Frageart: Freie numerische Angabe</u></p>	<p>{IF D1==„Yes“}</p> <p>What is the post code of your residence? (If you do not have a residence, answer the question for your current whereabouts.)</p> <p><u>Type of question: Numerical response</u></p>
D22	<p>{WENN D1==„Nein“}</p> <p>In welchem Land wohnst du? (Wenn du keinen Wohnort hast, beantworte die Frage bitte für deinen gegenwärtigen Aufenthaltsort.)</p> <p><u>Frageart: Auswahl einer Option</u></p> <ul style="list-style-type: none"> • Österreich • Schweiz • Sonstige-EU • Nicht-EU 	<p>{IF D1==„No“}</p> <p>In which country do you live? (If you do not have a residence, answer the question for your current whereabouts.)</p> <p><u>Type of question: Selection of one option</u></p> <ul style="list-style-type: none"> • Austria • Switzerland • EU other • Non-EU

Chapter 3

Are Wikipedia Users Conditionally Cooperative? Evidence from Fundraising Trials*

3.1 Introduction

Conditional cooperation is a well-documented empirical regularity (Fischbacher et al. 2001; Gächter 2007): for a non-trivial share of individuals, the inclination to act pro-socially depends positively on the cooperation of relevant others. Social information or framing that provide cues about others' behavior can therefore influence prosocial conduct at the extensive and intensive margin.

Consistent with this notion, numerous studies on charitable giving document conditionally cooperative donation behavior. Most of the evidence comes from trials that vary information about the amount relevant others were giving (e.g. Alpizar et al. 2008; Shang and Croson 2009).¹ Communicating a higher donation value (which may also serve as an anchor) typically triggers a positive effect on the donated amount.² Studies that vary information on the absolute number or the rate of donors are rare. The most well know is Frey and Meier (2004), who experiment in a student population, either stressing that 46% or 64% of their peers have contributed to charitable funds in the past. They report a positive, but statistically

*. This chapter is based on joint work with Christian Traxler, Hertie School, Berlin

1. Martin and Randal (2008) vary cues about both donation amounts and the number of donors.

2. These positive intensive margin effects tend to be negatively associated with the extensive margin: pushing up the average donation tends to reduce the number of donors. Note further that there are limits of which level of donations is perceived as relevant social information (Croson and Shang 2013).

insignificant treatment effect on donation rates. The effect only becomes significant after conditioning on students' past donations.

Despite this weak evidence on the power of providing “extensive margin” social information – and also neglecting the growing body of lab evidence that highlights the limits of social norm nudging (e.g. Bicchieri and Dimant 2019) – many practitioners seem to view belief management strategies as a universally successful policy tool to boost donation rates. This chapter challenges this view.

We present the results from a series of trials that were run as part of Wikipedia's online fundraising campaigns. While reading entries in the worldwide largest online encyclopedia, a banner might pop-up at the top of a user's browser. The texts of these banners, which all solicit donations, present direct or indirect social information about other users' propensity to donate. Wikimedia Germany (henceforth WMDE), which operates the German language website of Wikipedia, shared with us data on all their trials that systematically varied social information regarding donations at the extensive margin. In addition, we were allowed to implement a preregistered trial, which closely followed Frey and Meier (2004).

In total, we analyze six trials that produced nearly 23,000 donations summing up to 442,167 Euro. All trials used a control banner that points to a low number or rate of donations (the number of donations relative to total users or banner impressions). Treatment banners varied minor text parts pointing to higher numbers or higher donation rates. If Wikipedia users would act conditionally cooperative, one would thus expect positive treatment effects. The results reject this hypothesis. In five out of six trials, we find (partially significant) negative or null results: donation rates per banner impression decline or remain constant. Only one trial yielded a statistically insignificant positive effect on donation rates. At the intensive margin, treatments did not have any effect.

We discuss three possible explanations why the experimental variation failed to evoke conditionally cooperative donation behavior. First, conformity or social norms might generally be less relevant in online donations. Specific to our context, we note that all treatment banners convey a stronger prosocial norm to donate than control banners. At the same time, however, all banners typically communicate relatively low donation rates. The communicated numbers might therefore be perceived either as incongruous with the implicit normative expectation (Bicchieri and Xiao 2009) or as simply “too extreme” (Croson and Shang 2013).

Second, note that Wikipedia users are typically familiar with the quality of the online encyclopedia. Hence, the social information about others' donation behavior cannot provide

a signal about the quality of the provided public good (Vesterlund 2003; Potters et al. 2007). The absence of this channel again works towards a null result.

Third, one might argue that the self-identity return from donating might be in fact larger if relatively few others are donating (i.e., if users perceive the own donation as more important in maintaining Wikipedia). Hence, information that only few others act pro-socially might increase the self-image value associated with a donation (e.g. Bénabou and Tirole 2006, 2011). This channel could therefore explain why several trials document negative treatment effects.

While our setting does not allow us to discriminate among these different explanations, the evidence clearly challenges the wildly held prior that social information serves as a universally successful policy tool to foster prosocial conduct. In this vein, the chapter complements recent lab studies (Bicchieri and Dimant 2019), which also document limits in social information strategies.

The remainder of this chapter is structured as follows. Section 3.2 discusses the different trials and our data. Section 3.3 presents the results. Section 3.4 concludes.

3.2 Trials and Data

We study Wikipedia user's decision to donate after being exposed to a donation banner (see, e.g., Appendix Figures 3.A1 and 3.A2). During WMDE's trialing and fundraising period, a donation banner may slide down from the top of a user's browser window. A click on the banner opens another website that asks users to enter their payment details and to confirm the donation amount. WMDE provided us with data that track banner impressions and information on (completed) donations. Beyond this information, and different to other players in big tech, Wikipedia stores basically no data on their users.

Our analysis explores all randomized trials conducted by WMDE that systematically varied social information or (indirect) cues about other users' donation frequency. In addition, we designed, pre-registered and implemented our own trial.³ All trials tested a baseline banner ("control") against a variation ("treatment") that changed the banner text without affecting the size of the banner. Table 3.1 provides an overview of the six trials that we analyze. The trials, which were conducted between 2014 and 2018, vary in scale, ranging from 1.2 to 6.6 million total impressions per trial (approximately 50% of treatment and control conditions).

3. See AEA RCT Registry, November 13, 2018 <https://doi.org/10.1257/rct.3543-1.0>.

The first two trials test framing effects in the communication of a given piece of social information. The manipulations only affect one sentence of the banner text that compares numbers of donors and impressions. More specifically, trials 1 and 2 varied the framing of the (given) number of donors, referring to the number in a negative (“... but only N people donated ...”) or positive manner (“already N people donated ...”; see Table 3.1).

Trials 3 and 4 follow a similar logic. Both trials alter the same sentence that communicates a given number of donors relative to a larger or smaller baseline: the cumulative total (control) or the average daily number (treatment) of banner impressions (in millions). Thus, the same number of donors should appear larger in the treatment conditions as they imply donors to impressions ratios that are 20 (trial 3) or 40 times (trial 4) larger than those in the control conditions.

The last two trials directly vary social information. Trial 5 communicates the number of worldwide Wikipedia users and notes that “less than 0.1%” (control) or “less than 1%” (treatment) donate. Trial 6, which was designed by us. Following the spirit of Frey and Meier (2004), we communicated a lower and a higher number of donors: the control banner notes that 359.000 users donated last year; the treatment highlights “more than 400.000” annual donors “over the last years”. All trials provide direct or indirect cues about other users having either a relatively low (control) or relatively high propensity to donate (treatment). If Wikipedia users in Germany act conditionally cooperative, we should therefore observe a positive treatment effect: a higher number of donations in the treatment as compared to the control group banners.

Note further that all trials varied signals about the absolute number of donors or their share (relative to a baseline) rather than information about the donated amount (as in, e.g., Alpizar et al. 2008; Shang and Croson 2009). Consistent with this variation, our analysis therefore focuses on the decision to donate (i.e. the extensive margin). Specifically, we will compare donation rates, defined as the number of donations per 1,000 impressions. Effects on donation amounts (variation at the intensive margin) are examined, too.

Randomization was conducted via the webpage. During a trial, a certain fraction of users (typically 5%) are randomly sampled the first time a Wikipedia page is loaded. These users are then exposed, with equal chances, to either the control or the treatment banner of the trial.⁴ Users may see their banners repeatedly and might only donate after repeated exposure.⁵ This means that donation rates cannot be interpreted in terms of donations per

4. Using micro data on impressions, we confirmed that observables (e.g., time-stamps) are orthogonal to the treatment assignment.

5. The assigned treatment is stored locally (cookie). Unless this information is deleted, the same banner would be displayed (typically up to a maximum of 10 times) at the start of each page visit. Together with

user. As a consequences, our data do not allow to cluster standard errors at the user level; we only observe total impressions and donation outcomes. Hence, our inference will tend to over-reject the null of no effect.

the random sampling (which draws a share p of users, implying a p^2 chance of being re-sampled after deleting cookies and revisiting the website), this procedure strongly reduces the chance of exposure to different banners.

Table 3.1: Overview of Trials

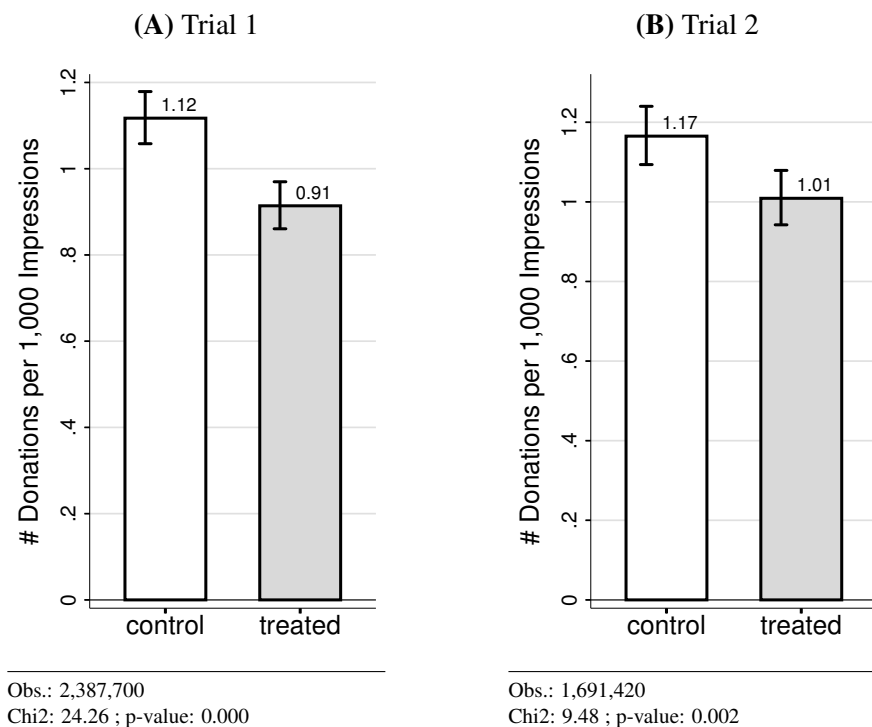
Trial	Group	Treatment text <i>Click on the texts below to display the full banner text online.</i>	Number of impressions	Number of donations	Total revenue (in EUR)	Average amount donated (in EUR)	Fundraising Year
Variation in framing (“only ...” vs. “already ...”)							
1	control	Our Donation Banner is viewed more than 20 Million times a day, but only 115.000 people have donate so far	1,189,600	1,329	27,012	20.32	2014
	treatment	Our Donation Banner is viewed more than 20 Million times a day. Already 115.000 people have donate so far	1,198,100	1,095	22,283	20.35	
2	control	Our Donation Banner is viewed more than 8 Million times a day, but only 28.300 people have donated so far	847,970	988	21,196	21.45	2015
	treatment	Our Donation Banner is viewed more than 8 Million times a day. Already 28.300 people have donated so far	843,450	851	17,223	20.24	
Variation in baseline numbers (cumulative vs. daily banner impressions)							
3	control	Our Donation Banner was viewed more than 184 Million times, but only 28.300 people have donated so far	3,268,600	3,383	69,072	20.42	2015
	treatment	Our Donation Banner is viewed more than 8 Million times a day, but only 28.300 people have donated so far	3,271,190	3,222	69,676	21.62	
4	control	Our Donation Banner is viewed more than 311 Million times a day, but only 28.300 people have donated so far	1,782,720	2,903	47,900	16.50	2015
	treatment	Our Donation Banner is viewed more than 7 Million times a day, but only 28.300 people have donated so far	1,782,740	2,916	46,474	15.94	
Direct variation in social information							
5	control	More than 450 Million people use Wikipedia, but less than 0.1% donates	616,400	1,881	38,259	20.34	2015
	treatment	More than 450 Million people use Wikipedia, but less than 1% donates	606,000	1,874	36,452	19.45	
6	control	In the last year 359.000 people donated for Wikipedia in total	1,010,430	1,210	23,357	19.30	2018
	treatment	Over the last years more than 400.000 people donated for Wikipedia on average	1,016,130	1,178	23,263	19.75	

3.3 Results

3.3.1 Variation in framing

Figure 3.1 presents the results from trial 1 and 2. Relative to the control, which communicated the number of donors in a negative frame (“...but only N people donated ...”), the positive frames from the treatment banners (“already N people donated ...”) had a negative effect on the donation rates in both trials. The rate dropped from 1.12 to 0.91 per 1,000 impressions (−18.2%) in trial 1 and from 1.17 to 1.01 in trial 2 (−13.4%). Keeping the caveat about inference in mind (see above), we note that both effects are statistically significant at the 1%-level. In both trials, the average amount donated is statistically indistinguishable between treatment and control. The higher total amount of donations in the control conditions reported in Table 3.1 is thus driven by extensive margin effects.

Figure 3.1: Variation in framing

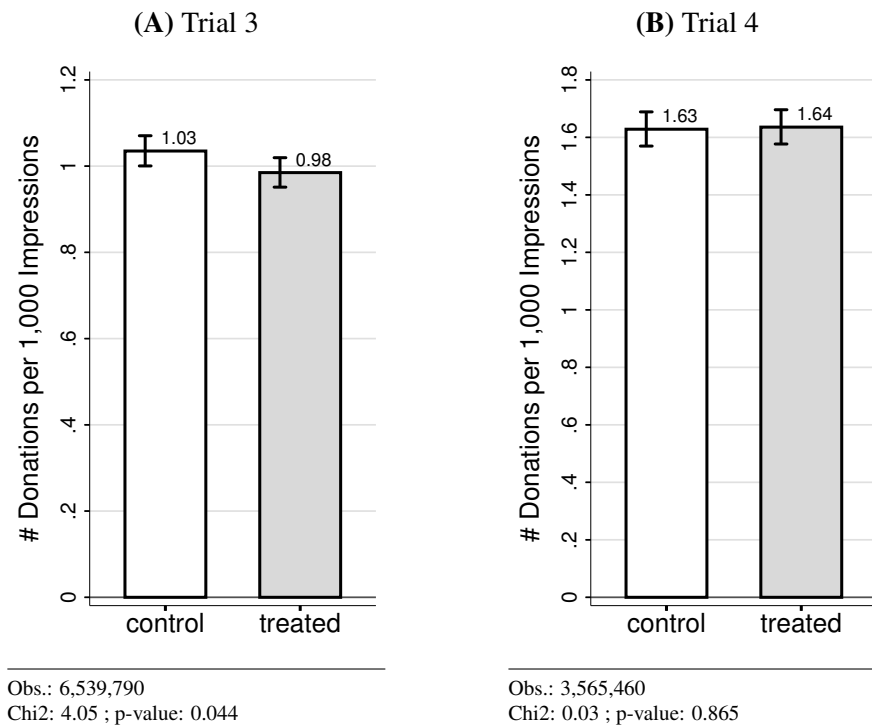


The findings are inconsistent with a conditionally cooperative donation behavior. Assuming that the positive framing provides a cue that other donors have a higher propensity to donate (relative to what is inferred from the negative frame in the control group banner), conditional cooperation would have implied a positive treatment effect. The data from trials 1 and 2 reject this case, indicating that, in fact, the positive frames decreased donation rates.

3.3.2 Variation in baseline numbers

Figure 3.2 presents the results from trials 3 and 4. Both trials contrasted a given number of donors with either large (control) or small numbers of impressions (treatment), implying a relatively smaller or higher propensity to donation. Neither trial 3 nor trial 4 document a positive treatment. For trial 3, the donation rate dropped from 1.03 to 0.98 per 1,000 impressions (-4.8% ; $p = 0.044$). In trial 4, we observe an insignificant increase from 1.63 to 1.64 ($+0.4\%$). The average amounts donated and revenues are very similar across treatment and control (see Table 3.1).

Figure 3.2: Variation in baseline numbers



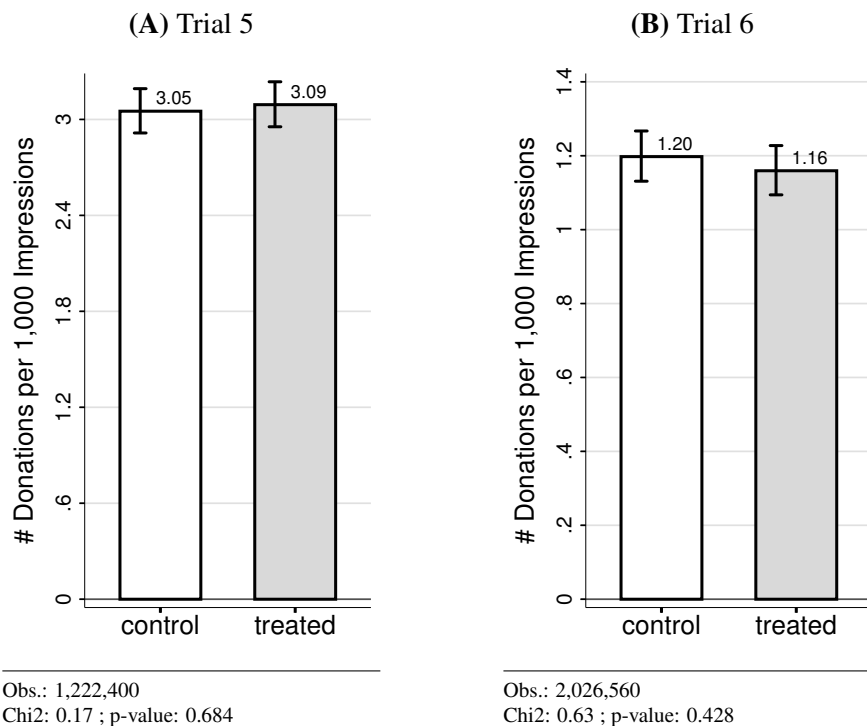
As long as the two treatments increase the perceived inclination to donate among other users, conditional cooperation would have again implied a positive treatment effect. However, this is not borne out by the data. In fact, the (weakly significant) findings from trial 3, point again into the opposite direction.

3.3.3 Direct variation in social information

Trials 5 and 6 directly varied social information. The former stressed that less than 0.1% (control) or less than 1% of users donated (treatment). The latter trial communicated a lower

(control) or higher absolute number (treatment) of donors. The results, which are presented in Figure 3.3, do not indicate any statistically significant increase in donation rates. In trial 5, we observe a small increase in the donation rate from 3.05 to 3.09 per 1,000 impressions (+1.3%).⁶ Finally, in the trial designed by us, we find an insignificant negative treatment effect: the donation rate drops from 1.20 to 1.16 (−3.2%). Average amounts donated are not statistically different between treatment and control in both trials, and revenues are very similar.

Figure 3.3: Direct variation in social information



The last two trials, which offer the most direct, straightforward variation in social information, provide again no evidence of conditional cooperation in donations among Wikipedia users in Germany.

3.4 Discussion

None of the six trials provides compelling evidence in support of conditional cooperation. On the contrary, several trials point into the opposite direction, indicating that signals

6. Donation rates in trial 5 are much higher than in the other cases. The difference is hard to interpret. It might be due to a different sample period or to the layout and size of the banners.

about lower donation rates increase Wikipedia users' inclination to donate. Why are these Wikipedia users not donating in a conditionally cooperative manner?

Let us first compare our findings with earlier results. Recall that most of the existing evidence on conditionally cooperative donations comes from trials that vary information about the donation amount (e.g. Alpizar et al. 2008; Shang and Croson 2009). Clearly, this is not comparable to our trials. One of the few pieces that experimentally varies social information about the rate of donors is Frey and Meier (2004).⁷ Among a student population, they either stress that 46% (control) or 64% (treatment) of students have contributed to charitable funds in the past. They find a statistically insignificant treatment effect: the contribution rate increases by 3.1%.⁸ While this is not too different from the null results observed in our trials 5 and 6, it conflicts with the negative effects observed in trials 1 and 2.

To explain the latter finding, it is worth considering the numerous channels that could drive donation choices. After all, conditional cooperation is an empirical pattern, not a theoretical concept. In fact, there is a vast range of mechanism that could induce conditional cooperation. Information about others' inclination to donate might, for instance, induce conformity pressure (Bernheim 1994) or strengthen the (perceived) social norm to support Wikipedia. Our null results might suggest that these channels are less relevant in the context of online donations. Alternatively, the presented social information might appear incongruous with the implicit normative expectation (see Bicchieri and Xiao 2009) or is "too extreme" (i.e., highlighting relatively low donation rates) in the sense of Croson and Shang (2013). The latter argument might also explain why we observe a weak, positive effect in trial 6, which communicates high absolute numbers.

Another possible driver of conditionally cooperative donations could be signals about the value or quality of the provided public good (Vesterlund 2003; Potters et al. 2007). However, since most Wikipedia users are familiar with the online encyclopedia, we doubt that this mechanism can operate in our context. The absence of this channel again works towards finding null results.

Information about others behavior could impact the self-image value associated with a donation (e.g. Bénabou and Tirole 2006, 2011). In fact, the self-identity return from donating might be larger if relatively few others are donating (if users perceive the own donation as more important in maintaining Wikipedia). Hence, this channel could in principle explain the negative treatment effects reported above.

7. Evidence from non-experimental (pre/post) variation is discussed in Heldt (2005).

8. As noted above, Frey and Meier (2004) do find a significant effect once they condition on student's past donation or after including student fixed effects.

Irrespective of the channels underlying the observed treatment (non-)responses, our findings provide field evidence that complements results from the lab (e.g. Bicchieri and Dimant 2019) and challenges the notion that the provision of social information is a universally successful policy tool: We do not find evidence on conditional cooperation in online donations to Wikipedia. Instead, our results indicates that social information approaches might even backfire.

Appendix

Figure 3.A1: Control banner of trial 1 (German original)

i Liebe Leserinnen und Leser: Verzeihen Sie die Störung. Einmal im Jahr bitten wir Sie um Ihre Unterstützung. Um Wikipedias Unabhängigkeit zu schützen, gibt es keine Werbung. Wir finanzieren uns durch Spenden von durchschnittlich 20 €. Jetzt sind Sie in Deutschland gefragt. Wenn alle, die das jetzt lesen, einen kleinen Beitrag leisten, ist unsere Spendenkampagne in einer Stunde vorüber. Über 20 Millionen Mal wird unser Spendenaufruf täglich angezeigt, aber nur 115.000 Menschen haben bisher gespendet. Wenn Sie Wikipedia nützlich finden, nehmen Sie sich bitte eine Minute Zeit, damit Wikipedia werbefrei und unabhängig bleibt. Vielen Dank!

Nur noch 0 Tage 8,3 Mio.

SPENDENKONTO Wikimedia Foerdergesellschaft BIC BFSWDE33BER IBAN DE33 1002 0500 0001 1947 00

einmalig regelmäßig

5 € 15 € 25 € 50 €

75 € 100 € 250 €

Buttons: Lastschrift, Überweisung, Kreditkarte, PayPal

Note: Control banner from trial 1.

Figure 3.A2: Example donation banner (English)

Dear readers:

I hope we're not disturbing you, but it is that time of year when we ask for your support. To protect our independence, we'll never run ads. Instead, we are financed by private donations, averaging €20 per donor. Now we are asking you in Germany to help out. If everyone currently reading this were to contribute a small amount, our fundraising campaign would be over in an hour. Our fundraising appeal is displayed over 8 million times a day, but currently only 28.300 people have donated. So if you find Wikipedia useful, please take a minute out of your day this Sunday to donate and give something back to Wikipedia. Thank you!

8,6 Mio. €

i DONATION ACCOUNT Wikimedia Foerdergesellschaft BIC BFSWDE33BER IBAN DE33 1002 0500 0001 1947 00

one-time recurrent

5 € 15 € 25 € 50 €

75 € 100 € 250 €

Buttons: Direct Debit, Bank Transfer, Credit Card, PayPal

Note: The figure is the control banner from campaign hk15_enorg_2. Campaign hk15_enorg_2 was run by WMDE on the English version of Wikipedia, but did not test conditional cooperation motives. Its banner is WMDE's translation of the text used in the control banner of trial 1. Notably, all campaigns analyzed in this chapter are from the German website of Wikipedia with donation banners in German language.

Chapter 4

Defaults in Online Charity: Evidence from Wikipedia Germany*

4.1 Introduction

Default options are the most powerful ‘Nudges’ in terms of affecting choices (Thaler and Sunstein 2008). A huge body of literature has emerged that studies defaults in many diverse settings (Jachimowicz et al. 2019). Johnson and Goldstein (2003), one of the earliest and more influential contributions in this field, document the role of defaults for the prevalence of organ donors in a society. When it comes to charity or monetary donations, however, the evidence is relatively scarce. This chapter seeks to contribute to this area by studying the role of defaults in charitable giving. More specifically, we study the role of defaults in shaping online donations to Wikipedia, the world’s largest encyclopedia.

We have obtained data on an extensive series of randomized (A/B tests) and observational trials conducted by Wikimedia Germany (henceforth WMDE). All these trials build upon banners that pop-up in a Wikipedia user’s browser. These banners ask for a donation and give users the opportunity to manually enter a donation amount. In addition, most banners also display numerous default values that can be selected by clicking on the respective button (e.g. 5, 15, 25 Euros). Building on the latter, WMDE’s trials tested the impact of (i) the number of default values, (ii) the specific values of the different defaults or (iii) a mixture along both dimensions.

*. This chapter is based on joint work with Christian Traxler, Hertie School, Berlin

Overall, our data cover almost 60,000 donations with a total volume of slightly more than 1 million Euro. A unique feature of the data is that WMDE ran the trials both on their desktop website, which opens for desktop computer users, and on their mobile website, which opens when users access Wikipedia with a mobile phone or tablet. This feature allows us to highlight differential effects of defaults on the donation decisions of desktop computer and mobile device users.

Our analysis documents several findings. First, donation banners that only offer an unconstrained, free choice of the donated amount result in ‘round number’ bunching (e.g. at 5, 10, 15, 20 Euro). This unconstrained setting also shows that mobile phone users donate on average lower amounts than desktop PC users. Second, adding donation defaults to the option to enter an unconstrained amount has a very strong impact on donated amounts. Vast shares of donations (in some trials more than 90%) correspond exactly to the default values. Third, beyond this pronounced intensive margin effect, we only find very modest and unsystematic evidence on extensive margin responses: the rate of donations relative to total banner impressions seems to be largely insensitive to the tested variations in defaults.

Fourth, several trials document a positive effect of default variations on the average amount donated and are accompanied by a null effect on the donation rate. Hence, defaults can boost the overall sum of collected donations. However, we also observe several trials that highlight various offsetting effects. For instance, including default values of 10/20/25 as compared to 10/15/25 ‘pulls up’ some donors to give 20 rather than 15 Euro but, at the same time, it also shifts some donation mass down from 15 to 10 Euro. The overall impact on average donations is thus often insignificant. Fifth, we find several different results for trials run on Wikipedia’s mobile website compared to the desktop website. Several trials run on the mobile website suggest that offering fewer default values can increase the average donation amount. In contrast, most trials run on the desktop website suggest that offering *more* default values increases the average donation amount.

Overall, our results document that the number of defaults and the default values interact in a complex way. Characterizing what constitutes an ‘optimal’ set of defaults is therefore a challenging exercise with the answer also depending on the user interface. For the moment, our analysis focuses on a mere ‘what works’ approach. In future work, we also plan to give a theoretical explanation for different drivers of default effects. In doing so, we hope to have the opportunity to run further trials in cooperation with WMDE. However, we anticipate that it is going to be challenging to precisely pin down specific mechanisms (such as anchoring effects, social reference points, etc.; see Altmann et al. 2019).

The evidence contributes to earlier work that examines the impact of default values on donations.¹ Three of the most closely related papers are Edwards and List (2014), Adena et al. (2014) and Altmann et al. (2019). The first two papers test the impact of offering a ‘suggested’ donation amount. The latter studies random exposure of online donors to different default donation amounts and also varies the so-called ‘co-donation’, which is a donation directed to the platform. All three studies find large extensive and intensive margin effects and document, similar to our research, a strong impact of default amounts.² Adena et al. (2014) and Altmann et al. (2019) also show that intensive margin increases in average donations might come at the cost of an extensive margin decline in the share of donors with an ambiguous overall impact.

We differ from these publications in two ways. First, our context is characterized by multiple rather than one default donation value. This means, we can study the impact of different numbers and values of defaults both separately and jointly in a consistent setting.³ Second, we seem to be the first to explore and document differential effects of defaults on online donations with respect to the device used to access the internet.

The remainder of this chapter is structured as follows. The next section introduces our data and summarizes the different trials. In section 4.3, we present the distribution of donations in a setting without defaults as a benchmark. Then in section 4.4, we present the results from the trials. We document the impact of defaults on donations in trials run on Wikipedia’s desktop separately from trials run on the mobile website. Section 4.5 concludes.

4.2 Data

Our study focuses on Wikipedia users’ donation decisions in response to exposure to what we call a ‘donation banner’ (see, e.g., Appendix Figure 4.A1). In Wikipedia’s online fundraising campaigns, these banners may pop-up in a desktop computer or mobile phone user’s browser window. The banners typically allow to choose the donation amount as well as other donation modalities (e.g., payment method). From these banners, potential donors are forwarded to another website that asks them to enter further payment details and to confirm the donation.

1. For a study on an extensive margin default setting see, e.g., Zarghamee et al. (2017).

2. For related evidence from a lab experiment, see Fiala and Noussair (2017).

3. The usage of multiple default options is shared with earlier studies in social psychology, that were based on randomized mailing experiments (Doob and S.McLaughlin 1989; Weyant 1984). These studies report mixed effects of defaults, with negative overall effects arising from extensive margin responses to ‘overly aggressive’ suggestions of donation amounts.

WMDE provided us with data that track banner impressions and information on completed donations.

Within this setting, we study the effect of default values for donation amounts presented in these banners. Our data cover 12 trials conducted by WMDE between 2014 and 2018. Table 4.1 provides an overview over these trials. The first two ‘trials’ (IDs `hk18_wpde_1` and `hk18_mob_4`) were non-experimental and simply displayed given donation banners. These banners only offered an entry field to manually type any given donation amount, meaning there were no default options with respect to the donation amount. Hence, the two ‘trials’ serve as a benchmark in our analysis. Trial 1 was conducted via the desktop version of the German Wikipedia website, and trial 2 ran on the mobile version.

All remaining trials were designed as ‘A/B tests’ that randomly assigned users to banners that differed regarding the default values for the donation amount. Trials 3 to 9 were run on the desktop websites (Wikipedia.org as well as Wikipedia.de). Trials 10 to 12 were run on the mobile version of these sites. Our data contain information about individual donations and the number of impressions by treatment status in each trial. Overall, we observe 58,079 donations with a total volume of slightly more than 1 million Euro.

Table 4.1: Overview Campaigns

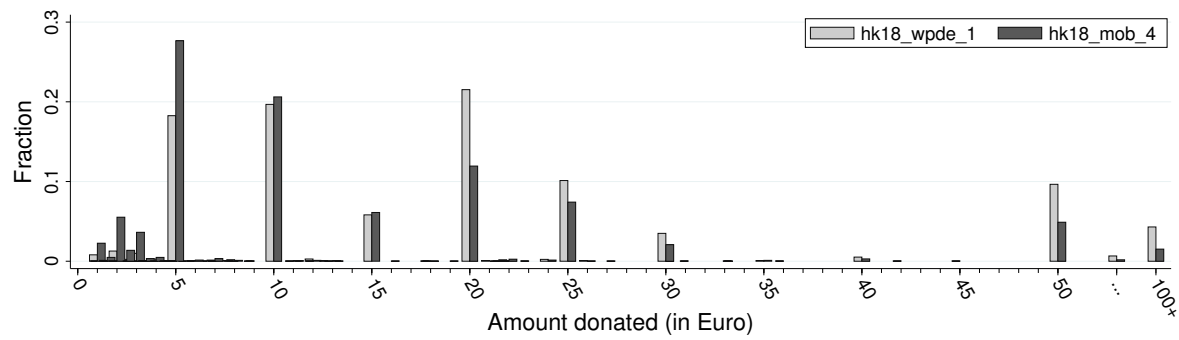
Number	Trial-ID	Group	Defaults	Number of impressions	Number of donations	Total revenue (in EUR)	Average donation amount (in EUR)	Year
Desktop devices, Non-experimental tests (no variation in banner)								
1	hk18_wpde_1		<input type="checkbox"/>	415,744	2,114	52,317	24.75	2018
2	hk18_mob_4		<input type="checkbox"/>	22,260,380	2,697	40,244	14.92	2018
Desktop devices, A/B tests (run only on desktop websites)								
3	hk18_org_1	control	5/15/25/50/75/100/250/ <input type="checkbox"/>	898,440	2,252	39,708	17.63	2018
		treated	<input type="checkbox"/>	898,050	2,209	34,866	15.78	
4	hk16_org_27	control	5/15/25/50/75/100/250/ <input type="checkbox"/>	3,203,570	4,240	82,942	19.56	2016
		treated	5/25/50/100/ <input type="checkbox"/>	3,207,930	4,213	76,217	18.09	
5	hk17_org_23	control	5/15/25/50/75/100/ <input type="checkbox"/>	3,573,310	3,675	76,385	20.78	2017
		treated	5/15/25/50/75/100/250/ <input type="checkbox"/>	3,582,550	3,721	74,737	20.09	
6	hk18_wpde_2	control	5/15/25/50/75/100/ <input type="checkbox"/>	695,525	2,207	53,539	24.26	2018
		treated	5/10/20/25/50/100/200/ <input type="checkbox"/>	695,896	2,178	60,960	27.99	
7	hk18_org_4	control	5/15/25/50/75/100/ <input type="checkbox"/>	984,770	1,755	33,091	18.86	2018
		treated	5/10/20/25/50/100/ <input type="checkbox"/>	982,260	1,877	33,927	18.08	
8	hk14_org_2	control	5/15/25/50/75/100/250/ <input type="checkbox"/>	830,600	1,416	26,513	18.72	2014
		treated	5/20/25/50/75/100/250/ <input type="checkbox"/>	830,500	1,468	27,080	18.45	
9	hk18_org_16	control	5/10/20/25/50/100/ <input type="checkbox"/>	3,894,560	5,867	110,555	18.84	2018
		treated	5/10/21,48/25/50/100/ <input type="checkbox"/>	3,901,880	5,819	109,667	18.85	
Mobile devices, A/B tests (run on mobile websites only)								
10	hk18_mob_6	control	5/15/25/50/100/ <input type="checkbox"/>	11,979,690	2,075	30,712	14.80	2018
		treated	5/10/15/20/25/50/100/ <input type="checkbox"/>	11,976,220	2,048	26,077	12.73	
11	hk18_mob_5	control	5/15/25/50/100/ <input type="checkbox"/>	11,661,380	2,429	35,939	14.80	2018
		treated	2/5/10/15/20/25/50/ <input type="checkbox"/>	11,657,530	2,581	33,795	13.09	
12	hk14_mob_3	control	5/15/25/50/75/ <input type="checkbox"/>	5,191,200	587	6,472	11.02	2014
		treated	5/15/25/50/75/100	5,292,400	651	7,215	11.08	

Notes: indicates the presence of an entry field in the banner to manually specify a donation amount.

4.3 Donations without Defaults

Figure 4.1 displays the distributions of donated amounts in the non-experimental ‘trials’ 1 and 2.⁴ Trial 1 (with the ID `hk18_wpde_1`) was run on Wikipedia’s desktop website. In contrast, trial 2 (with the ID `hk18_mob_4`) was run on Wikipedia’s mobile website. Otherwise, both trials are very similar with respect to the date of the campaign and banner texts.⁵ Rather than offering any default value(s) for a donation amount, the banners contained an entry field to indicate any given amount. Users could type in the desired amount before proceeding to the web form to finalize their donation.⁶

Figure 4.1: Distributions of donations without defaults



Obs.: 2,114 (`hk18_wpde_1`) ; 2,697 (`hk18_mob_4`)
 Dates of campaign: Nov. 13-20, 2018 (`hk18_wpde_1`) ; Nov. 19-26, 2018 (`hk18_mob_4`)

Figure 4.1 shows that a large majority of donors donates multiples of 5 EUR. In trial 1 (desktop), 94% of donations are exact multiples of 5 EUR, and 77% of all donations are in the range between 5 EUR and 25 EUR. Similarly, in trial 2 (mobile users), 82% of donations are exact multiples of 5 EUR, and 76% of all donations range between 5 EUR and 25 EUR.

The two distributions differ in the average amount donated. In the desktop trial, donors are more likely to donate amounts of 20 EUR and above, whereas donations on mobile devices resulted in donation values that were more likely 5 EUR or less. This difference may be due to the campaign being run on (i) potentially different types of users. However, it

4. Note that the following exposition focuses on the intensive margin. In future work on this chapter, we will more carefully examine extensive margin choices (i.e., the decision to donate), too.

5. Both trials were run in November 2018. The only difference in texts are the claimed average donation of 20 EUR (`hk18_wpde_1`) or 21.48 EUR (`hk18_mob_4`) and one sentence. The sentence is “Millions of people use Wikipedia, but only a small fraction donates” (`hk18_wpde_1`) vs. “Our Donation Banner is viewed more than 6 Million times a day, but only 337.589 people have donate so far.” (`hk18_mob_4`).

6. The banner of `hk18_mob_4` on Wikipedia’s mobile website also contained a button labeled “5 € per SMS”, which opened a pre-formulated SMS on a mobile phone and allowed for a direct transaction through the telecommunication provider. This button, however, was not at the regular position of the donation amount defaults, but at the very bottom of the banner just before the “Proceed with donation” button.

could also be due to (ii) difference in the banners, (iii) differences in their visual display or (iv) differences in the usability on different types of devices.

4.4 Donations with Defaults

This section now turns to the results from trials that test the effect of defaults on donations. The first subsection 4.4.1 presents the results from the seven trials that were conducted via the desktop website. Subsection 4.4.2 presents the results from the three trials that were run via the mobile website.

4.4.1 Trials on Desktop Website

The Effects of Defaults

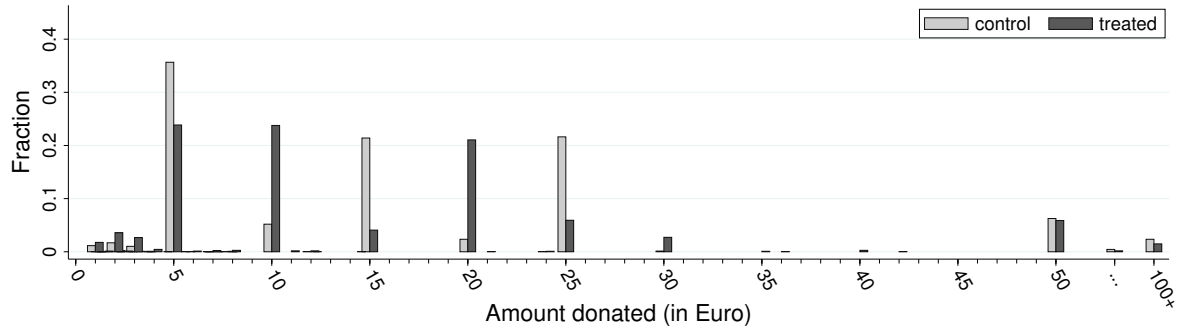
Figure 4.2 displays the distributions of donated amounts in trials 3 (hk18_org_1) and 4 (hk16_org_27) for each treatment status separately. Both trials share the identical set of default values in the control banner (which is a standard set widely used in many of WMDE's campaigns). The set contains 7 default buttons plus an entry field (5/15/25/50/75/100/250/□). Trial 3 (hk18_org_1) tests this default set only against an entry field (□), i.e. a banner without any default(s) for the donation. Trial 4 (hk16_org_27) tests the default set against a subsets of 4 default values plus the entry field (5/25/50/100/□). Note that these 4 values are part of the set of defaults used in the control banner. It should be further noted that – despite the large size of the two trials – the data do not show any significant extensive margin effects (i.e. differences in the donation rate, defined as the number of donations per 1,000 impressions). For this reason, the distributions of donations are presented as fractions (relative to all donations within the experimental groups of the trials).

Figure 4.2 Panel (A) presents the distributions in the control and treatment group of trial 3 (hk18_org_1). In the control group, 87% of donations take exactly one of the values of the 7 default values (5/15/25/50/75/100/250). In contrast, under the treatment condition (i.e. without any default value), only 41% of donations take one of these 7 values.⁷ In the control group, a large fraction of donations (43%) is at 15 or 25 EUR, which are default amounts. At the same time, only a small fraction of donations (8%) take the prominent, round numbers 10 or 20 EUR. In the treatment condition, these numbers are reversed: only 10% of donations

7. Note that the distribution of donations in this treatment group of trial 3 (hk18_org_1) is extremely similar to the one observed above, for the non-experimental trial 1 (hk18_wpde_1).

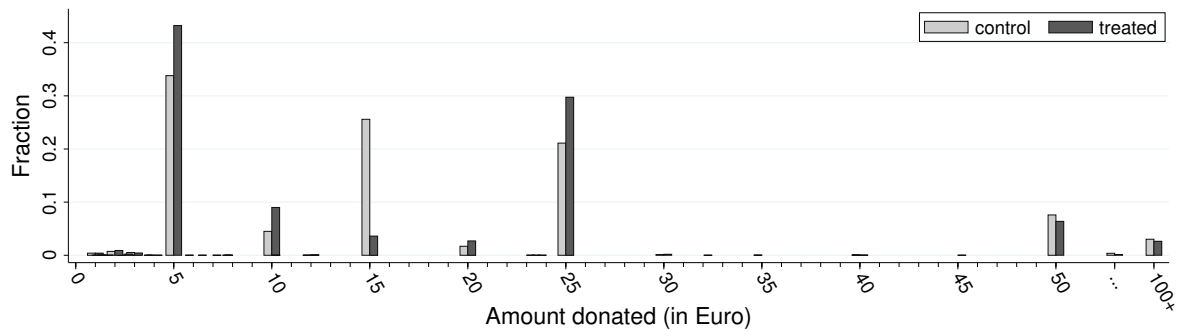
Figure 4.2: Distributions of donations with defaults

(A) Trial 3 (hk18_org_1)



Control – Obs.: 2,252 ; Defaults: 5/15/25/50/75/100/250/□
 Treated – Obs.: 2,209 ; Defaults: □
 Donation rate: 2.51 (control) ; 2.46 (treated) ; P-value 0.530

(B) Trial 4 (hk16_org_27)



Control – Obs.: 4,240 ; Defaults: 5/15/25/50/75/100/250/□
 Treated – Obs.: 4,213 ; Defaults: 5/25/50/100/□
 Donation rate: 1.32 (control) ; 1.31 (treated) ; P-value 0.722

take values of 15 or 25 EUR, whereas 45% of donations have a value of 10 or 20 EUR. Hence, the presented default values have a strong impact on the actual amount donated.

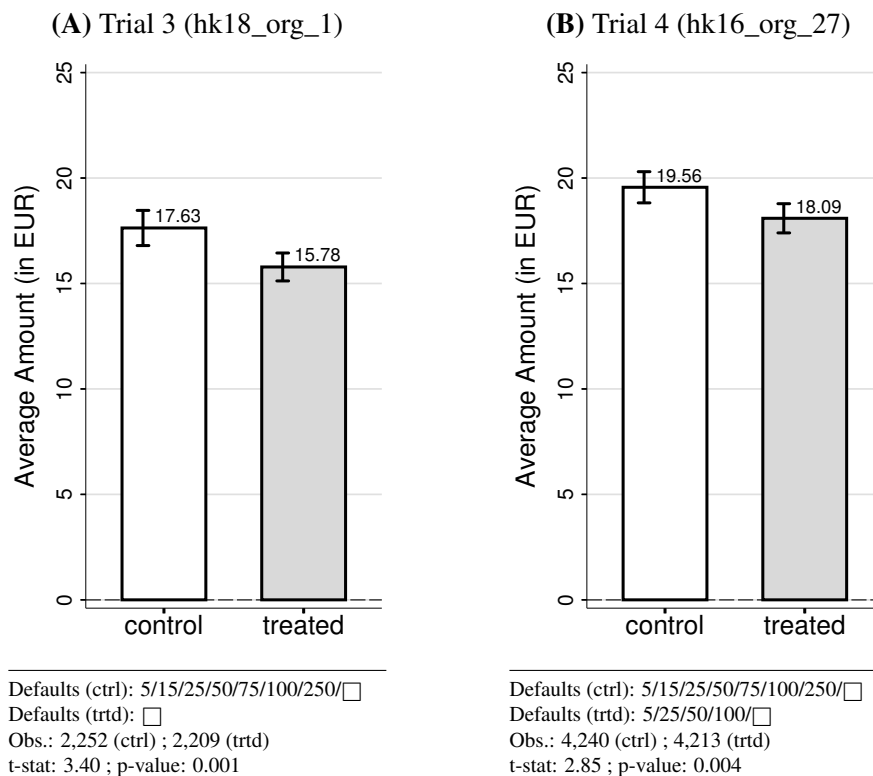
Panel (B) presents the distributions of donated amounts for the control and treatment condition in trial 4 (hk16_org_27). As mentioned above, the treatment banner differs from the control banner in that it ‘drops’ the default values of 15, 75, and 250 EUR. This smaller default set in the treatment banner reduces the fraction of donations with a default value from 91% to 82%. Compared against the control distribution, the treatment distribution counts relatively more donations at 5 and 25 EUR, but far less donations of 15 EUR. The pattern again shows that defaults strongly affect the distribution of donated amounts, and that a large fraction of donors stick to one of the default options – rather than entering another amount.

Beyond these shifts in the distribution, fundraisers are typically most interested in whether defaults can increase revenues. The effect of defaults on revenue may either be

through raising the propensity to donate (extensive margin effect), which is reflected in the donation rate, or through increasing the average donation amount (intensive margin effect). As noted above, we find neither economically nor statistically significant extensive margin effects in both trials. However, defaults clearly impacted the intensive margin. How did this affect donations overall?

Figure 4.3 presents the treatment effects on the average donations in the two trials. The average donation amounts significantly differ at the 99% confidence level in both trials. Trial 3 (hk18_org_1), which tested WMDE’s standard set of defaults against an entry field () documents a 10% lower average donation amount for the treatment banner with only the entry field. This corresponds to an almost 2 EUR decline. A somewhat smaller drop is observed in trial 4 (hk16_org_27). The data indicate a 7.5% (roughly 1.5 EUR) decline in the average donation when the standard set of defaults is reduced to the smaller subset of default values (5/25/50/100/).⁸

Figure 4.3: Average donation amounts with defaults



Overall, the two trials document that (a) default values have a strong effect on the distribution of donated amounts and that (b) the defaults can increase the average amount donated.

8. Both of these results are insensitive to unusually high donation values: restricting the samples to ‘smaller donations’ that are equal to or less than 100 EUR yields almost identical findings.

However, it remains unclear whether the number of defaults, the default values or some combination of the two drive the intensive margin effects.

Number of Defaults or Default Values

Figure 4.4 displays the distributions among the control and treatment groups of trials 5 (hk17_org_23) and 6 (hk18_wpde_2). Both trials share the same set of default values in the control banner. The set of defaults corresponds to WMDE's standard set of defaults (discussed above) but excludes the highest value, i.e. the 250 EUR default option. The 'constrained' default set thus contains 6 default buttons plus the entry field (i.e. 5/15/25/50/75/100/□). Trial 5 tests this reduced set against the 'unconstrained' default set (i.e. including the 250 EUR default). In contrast, trial 6 varies two things: its treatment banner also expands the set of default options from 6 to 7 and, at the same time, replaces some default values. Motivated by the distribution observed in the trials without donation amount defaults⁹, it replaces the default values of 15, 75, and 250 EUR with defaults of 10, 20, and 200 EUR.

Figure 4.4 Panel (A) compares the distributions of the donated amounts observed in the treatment and control group of trial 5 (hk17_org_23).¹⁰ For donations below 100 EUR, the control and treatment distributions are almost identical. However, they differ in the small fraction (3.1%) of donations of 100 EUR and above. Among this latter group, 93% of donors in the control group give exactly 100 or 200 EUR, only 0.8% donate 250 EUR. In the treatment condition (where the banner included a 250 EUR default), 76% of donations in this high-donation group are either 100 or 200 EUR, but also 20.3% are exactly 250 EUR. Hence, one can summarize the impact from the addition of the 250-default as follows: it did affect the distribution of large donations, but not the distribution of small donations.

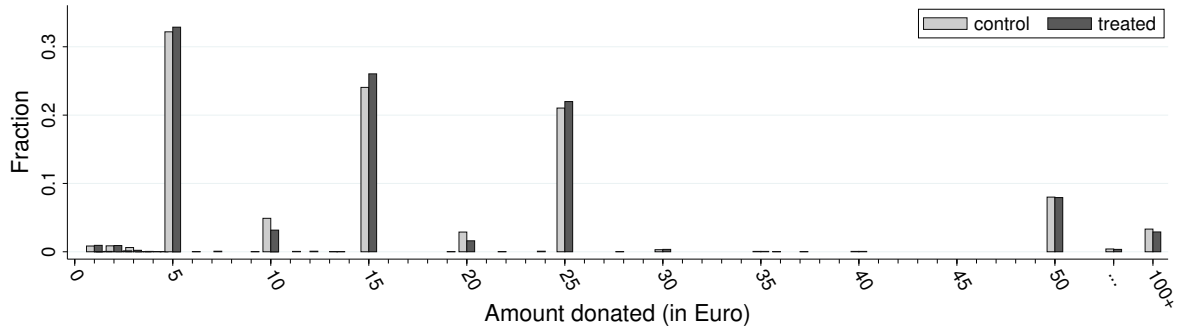
Panel (B) presents the distributions of donations in the control and the treatment group of trial 6 (hk18_wpde_2). As compared to the control group (84%), slightly more donors stick to default values in the treatment group (93%). As mentioned above, the latter uses default values that are motivated by the most prominent donation amounts observed in trials without default amounts. The two distributions observed for the control and the treatment group corroborate the findings from above: most donors stick to defaults (compare, e.g., 10 vs. 15

9. For the distribution without donation amount defaults see trial 1 (hk18_wpde_1) in Figure 4.1 and the treatment distribution of trial 3 (hk18_org_1) in Figure 4.2.

10. As before, none of the two trials shows any significant differences in donations rates (extensive margin effects). Hence, we will continue to present the distribution in fractions (relative to all donations) rather than in absolute numbers.

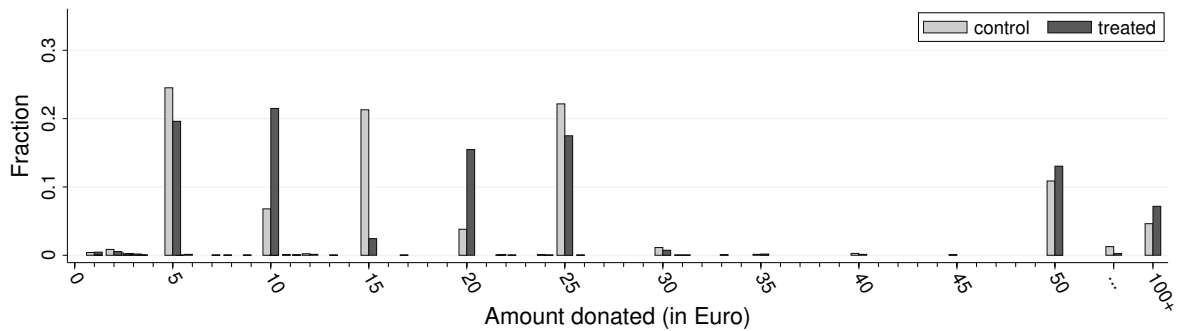
Figure 4.4: Distributions of donations with different defaults

(A) Trial 5 (hk17_org_23)



Control – Obs.: 3,675 ; Defaults: 5/15/25/50/75/100/□
 Treated – Obs.: 3,721 ; Defaults: 5/15/25/50/75/100/250/□
 Donation rate: 1.03 (control) ; 1.04 (treated) ; P-value 0.672

(B) Trial 6 (hk18_wpde_2)



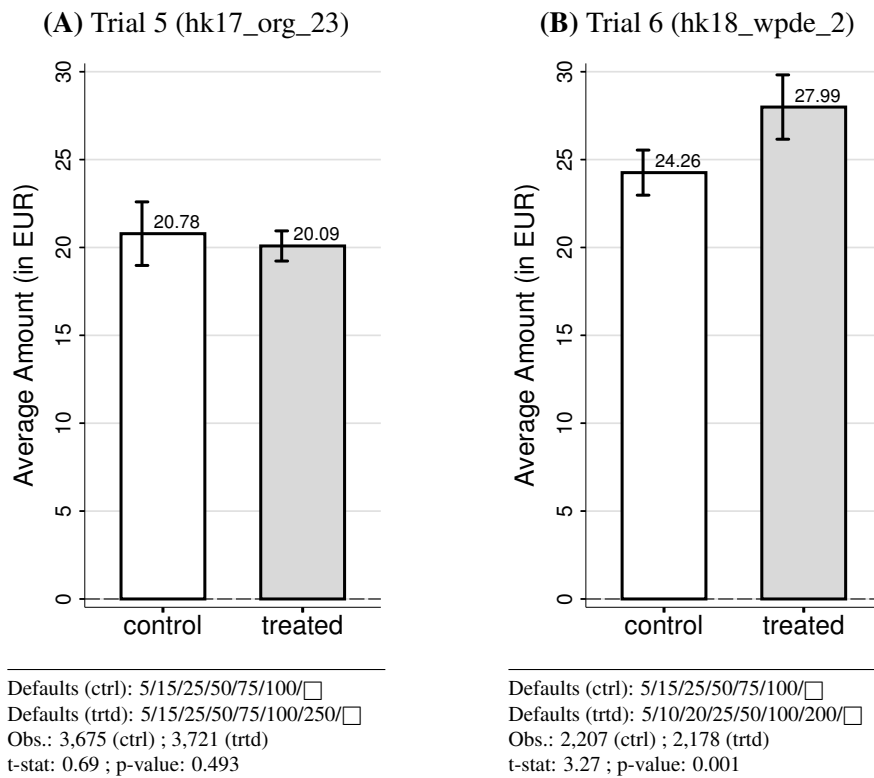
Control – Obs.: 2,207 ; Defaults: 5/15/25/50/75/100/□
 Treated – Obs.: 2,178 ; Defaults: 5/10/20/25/50/100/200/□
 Donation rate: 3.17 (control) ; 3.13 (treated) ; P-value 0.649

EUR and 20 vs. 25 EUR) and that different defaults produce major shifts in the distribution of donated amounts.

Figure 4.5 presents the average donations for trials 5 (hk17_org_23) and 6 (hk18_wpde_2). In trial 5, there is no statistically significant difference in average donations between the control and the treatment banner; the impact of adding the large default value (250 EUR) is therefore insignificant on average. This mainly reflects the observation from above: including this high default value seems to impact, at most, the small fraction of donors that consider giving amounts of 100 EUR or more. The finding further suggests that increasing the number of default values per se does not necessarily assure an increase in the donation amount.

For trial 6, one obtains a different result. Here we observe a significant treatment effect: adding one additional default option and using more ‘prominent’ default values (10, 20 and

Figure 4.5: Average donation amounts with different defaults



200 EUR instead of 15 and 75 EUR) triggers a sizable increase in the average donation amount. The result from this trial again highlights the potential benefits from optimizing the set of defaults. Overall, the findings from both trials suggest that default values might be more important than the number of defaults.

Default Values

Three trials aim at evaluating the effect of amount default values. The trials are trial 7 (hk18_org_4), trial 8 (hk14_org_2), and trial 9 (hk18_org_16). Trial 7 is a replication of trial 6 (see above) that keeps the number of default options constant. Specifically, it tests the same, constrained set of donation defaults as trial 6 (5/15/25/50/75/100/□) against a set in which the defaults of 15 and 75 EUR are replaced with 10 and 20 EUR.¹¹ Trial 8 is minor variation of trial 7, which uses the standard set of defaults in the control (5/15/25/50/75/100/250/□). The treatment banner simply replaces the 15 EUR default with a 20 EUR value. Trial 9 also builds on trial 7. It uses the treatment default set from trial 7 in the control banner and tests

11. Trial 6 additionally contains a 200 EUR default in the treatment banner and therefore jointly varies the number of defaults and the default values.

it against a set where the 20 EUR default is replaced by a default of 21.48 EUR. The latter is the (long-run) average donation amount that is also communicated in the texts of both banners (treatment and control). Hence, this number might be attractive as it is made salient in the text. However, it might also appear unattractive (as compared to 20 EUR) because it is not a round number.

Trial 7 finds a marginally significant extensive effect (p-value 0.036) with a larger donation rate under the treatment banner. However, the almost identical trial 6 from above (to which trial 7 is a replication) yields no statistically significant extensive margin effect – despite the fact that the latter trial 6 had a larger sample size. In addition, trial 8, which tests a similar (partial) variation in donation amount defaults as trial 7, and also trial 9 show no extensive margin effects. Hence, the significant extensive margin effect in trial 7 might therefore have emerged by chance. We therefore maintain our focus on presenting distributions relative to the total number of donations (within a trial and for a given banner).

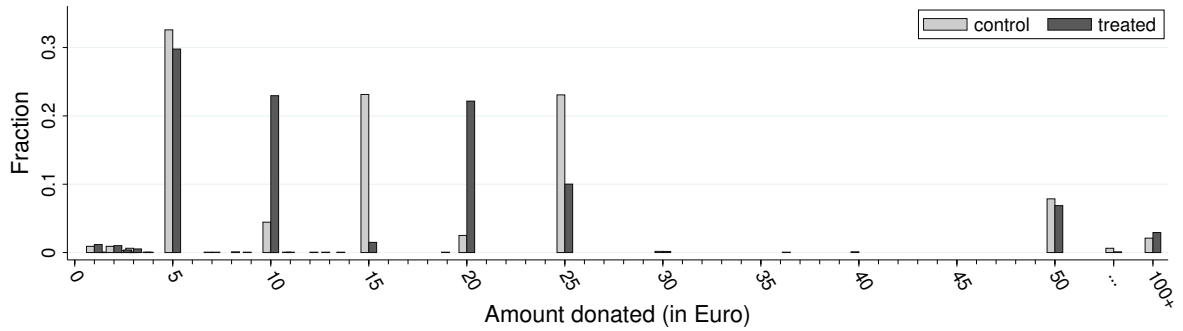
Figure 4.6 shows the distributions of donated amounts of trial 7 (hk18_org_4), trial 8 (hk14_org_2), and trial 9 (hk18_org_16). As before, the distributions document that most donors give amounts that correspond to default values. Trial 9 is a particularly interesting case as its treatment banner contains a default of 21.48 EUR. Despite the high salience from this number being contained in the banner texts, only a small fraction made use of this default value (4%). This finding indicates that donors' preference to donate round number values dominates potential 'salience gains' from stressing numbers in the banner texts.

Figure 4.7 presents the average donation amounts observed in the three trials. Average donations do not statistically differ between control and treatment banner in any of the three trials. This finding is surprising. Trial 6 from above – which is very similar to trial 7 – revealed a statistically significant intensive margin effect. It is therefore surprising that this effect is not observed in trial 7. Considering the distributions, we conclude that any particular set of amount defaults induces some people to donate more and others to donate less than they would donate under a different set of defaults. These fractions of people must be such that the average donation amount remains unchanged.

In sum, all of the trials discussed above show that default values have strong effects on the distribution of donations. While there clearly seems to be scope to increase the overall donation volume, this objective seems to be more complicated to achieve. In particular, the last three trials make evident that variation in default values by themselves do not necessarily help to raise average donations. Optimizing the number of defaults and the default values to maximize overall donations clearly is a challenging task.

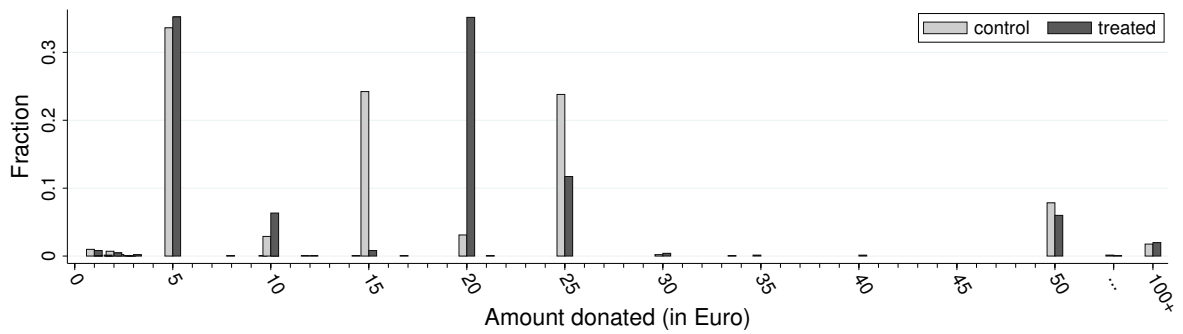
Figure 4.6: Distributions of donations with different default values

(A) Trial 7 (hk18_org_4)



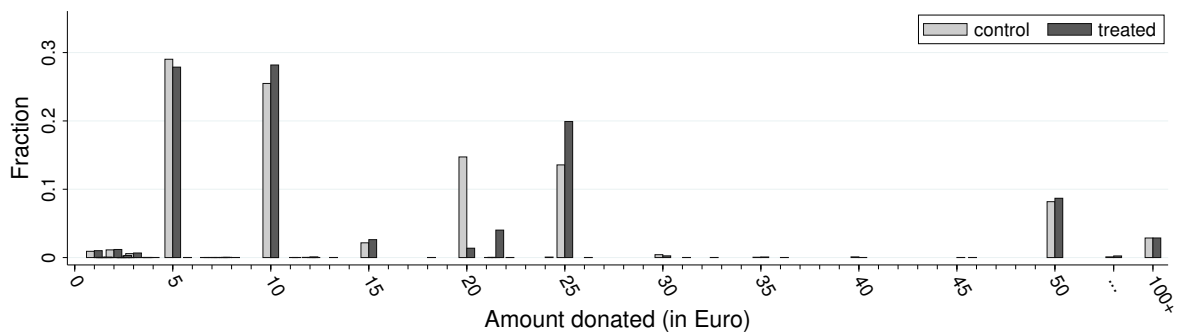
Control – Obs.: 1,755 ; Defaults: 5/15/25/50/75/100/□
 Treated – Obs.: 1,877 ; Defaults: 5/10/20/25/50/100/□
 Donation rate: 1.78 (control) ; 1.91 (treated) ; P-value 0.036

(B) Trial 8 (hk14_org_2)



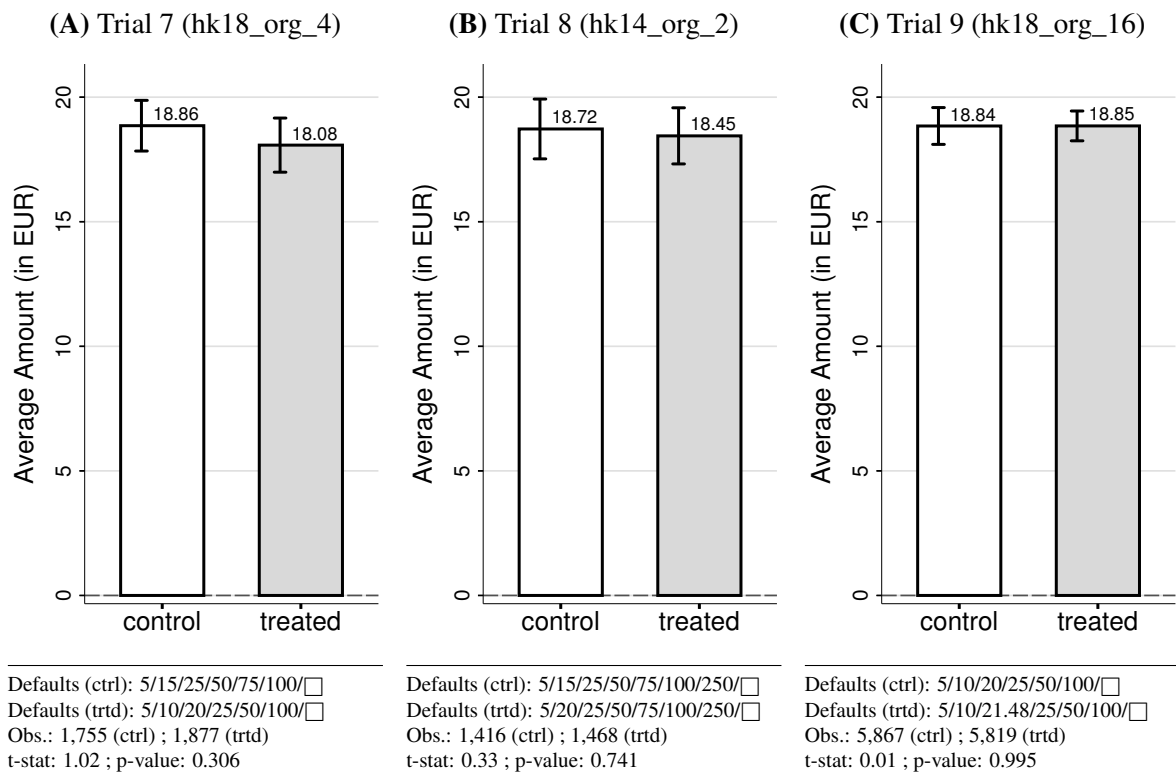
Control – Obs.: 1,416 ; Defaults: 5/15/25/50/75/100/250/□
 Treated – Obs.: 1,468 ; Defaults: 5/20/25/50/75/100/250/□
 Donation rate: 1.70 (control) ; 1.76 (treated) ; P-value 0.332

(C) Trial 9 (hk18_org_16)



Control – Obs.: 5,867 ; Defaults: 5/10/20/25/50/100/□
 Treated – Obs.: 5,819 ; Defaults: 5/10/21.48/25/50/100/□
 Donation rate: 1.51 (control) ; 1.49 (treated) ; P-value 0.586

Figure 4.7: Average donation amounts with different default values



4.4.2 Trials on Mobile Website

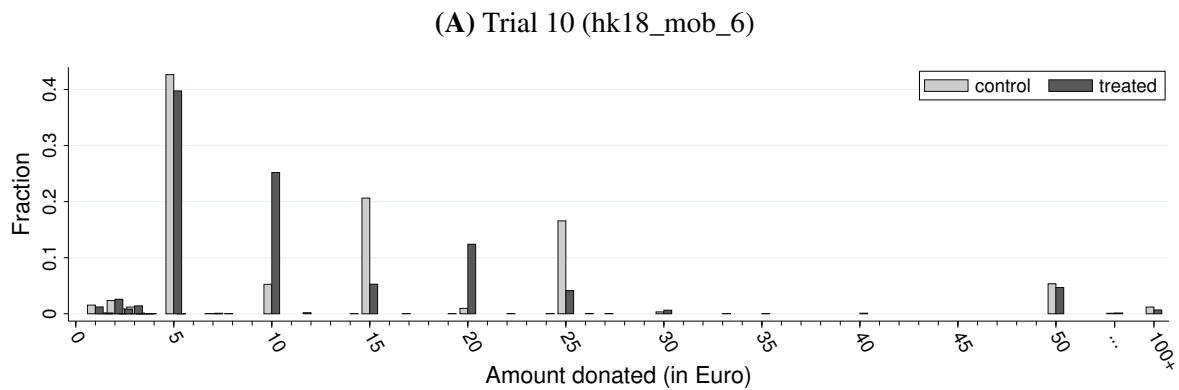
Number of Defaults

Figure 4.8 displays separate distributions of donated amounts by treatment status from two trials that were run on Wikipedia's mobile website. The two trials are trial 10 (hk18_mob_6) and trial 11 (hk18_mob_5). Both trials use a variation of WMDE's standard default set in the control banner that is adapted to the smaller screen of mobile devices. Specifically, the set contains 5 defaults plus the entry field (i.e. 5/15/25/50/100/□), dropping the two defaults of 75 and 250 EUR that pertain to the full standard set. Trial 10 tests this set of defaults against a larger set that additionally includes defaults of 10 and 20 EUR. This variation increases the number of defaults with low values. Trial 11 tests the same default set in the control banner against a similar treatment default set that only differs from the treatment banner default set of trial 10 by exchanging the high value default of 100 EUR against a low value default of 2 EUR, which even further increases the number of defaults with low values.

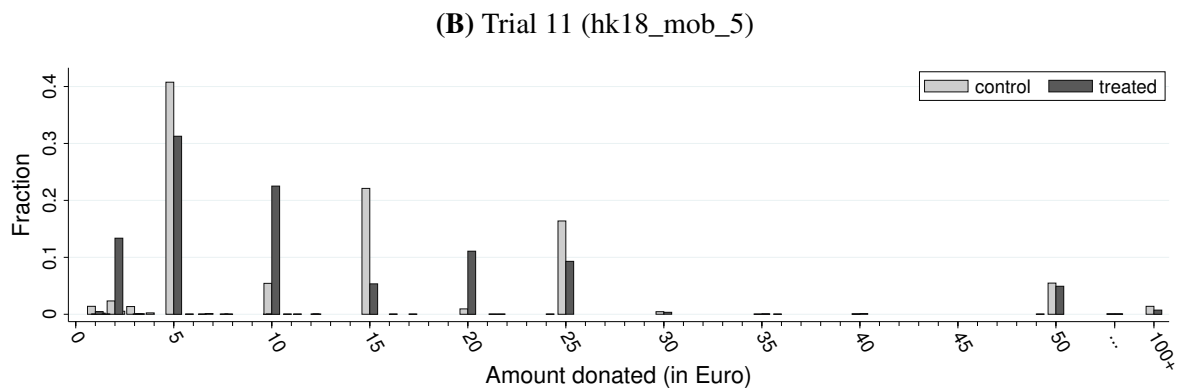
As before, the distributions document that most donors donate amounts that were defaults. In trial 10 (hk18_mob_6), 86% of donors that saw the control banner donated a default amount, compared to 92% of donors that saw the treatment banner. This finding reconfirms that more defaults increase the percentage of donors that donates exactly a default value. In trial 11 (hk18_mob_5), the same fraction of donors that saw the control banner donated a default amount (86%). However, a much larger fraction of donors that saw the treatment banner donated a default amount (98%). This finding suggests that offering amount defaults that are close in value to the amounts that donors would donate without defaults increases the fraction of donors that stick to defaults and donate a default value. Comparing the distributions in Figure 4.8 to those generated from trials run on Wikipedia's desktop website shows that a larger fraction of donors donate small amounts.

Finally, it should be noted that the donation rates of treatment and control banners are insignificantly different in trial 10, but marginally significant in trial 11 with a higher rate generated by the treatment banner. It is unclear whether this finding is reliable or spurious. One conceivable hypothesis could be that an additional default below 5 EUR is inducing more people to donate. However, this hypothesis is unsupported by a mobile trial with a similar setup, which is not presented here because we only have aggregate data for it. This trial (hk13_mob_5) tests a similar control default set of 5, 15, 25, 50, 75, and 100 EUR against a treatment default set of 3, 5, 10, and 20 EUR and finds no statistically significant difference in donation rates. With its extremely low donation rates of 0.073 donations per 1,000 impressions in treatment and control, this trial may lack power to detect a potentially

Figure 4.8: Distributions of donations with different defaults (mobile)



Control – Obs.: 2,075 ; Defaults: 5/15/25/50/100/□
 Treated – Obs.: 2,048 ; Defaults: 5/10/15/20/25/50/100/□
 Donation rate: 0.173 (control) ; 0.171 (treated) ; P-value 0.681

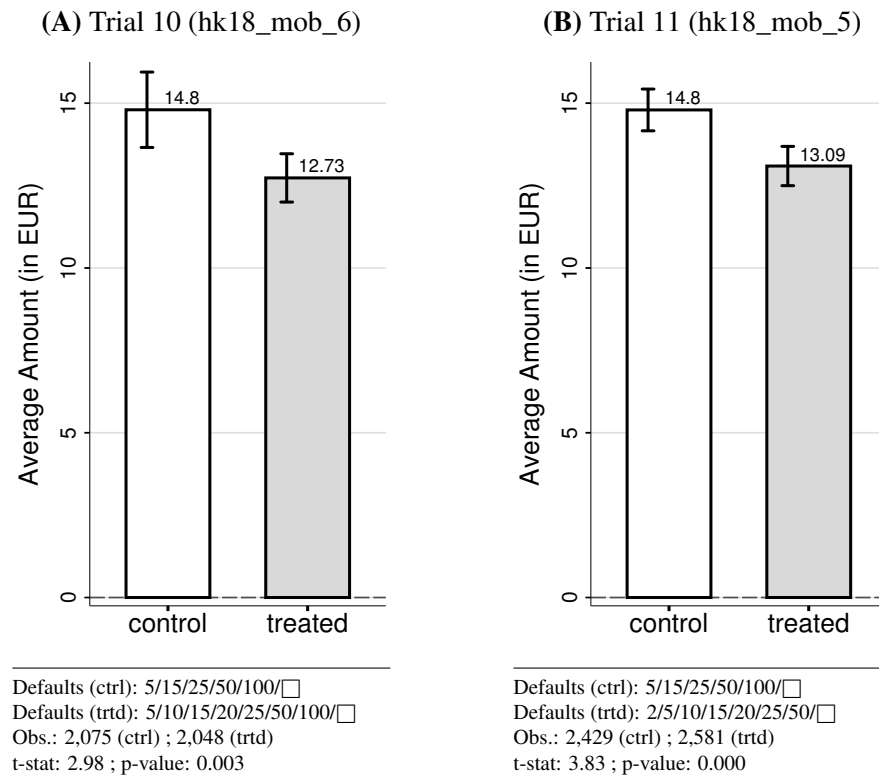


Control – Obs.: 2,429 ; Defaults: 5/15/25/50/100/□
 Treated – Obs.: 2,581 ; Defaults: 2/5/10/15/20/25/50/□
 Donation rate: 0.208 (control) ; 0.221 (treated) ; P-value 0.031

very small extensive margin effect. On the other hand, it may indicate that the extensive margin effect of trial 10 is spurious.

Figure 4.9 presents the average donation amounts generated by treatment and control banners in trial 10 (hk18_mob_6) and trial 11 (hk18_mob_5). In both trials, the average donation amount in the treatment condition is significantly smaller than the average amount in the control condition. This result contradicts the findings from similar trials run on Wikipedia’s desktop website (see, e.g., trial 4 (hk16_org_27) in Figure 4.3 and trial 6 (hk18_wpde_2) in Figure 4.5), in which more default amounts increase the average donation amount. This difference in findings shows that defaults seem to affect donors on the mobile website differently than donors on the desktop website. While there may be fundamental differences in the user groups, different response to defaults may also arise from differences in the visual display on a desktop PC versus a mobile device.

Figure 4.9: Average donation amounts with different defaults (mobile)

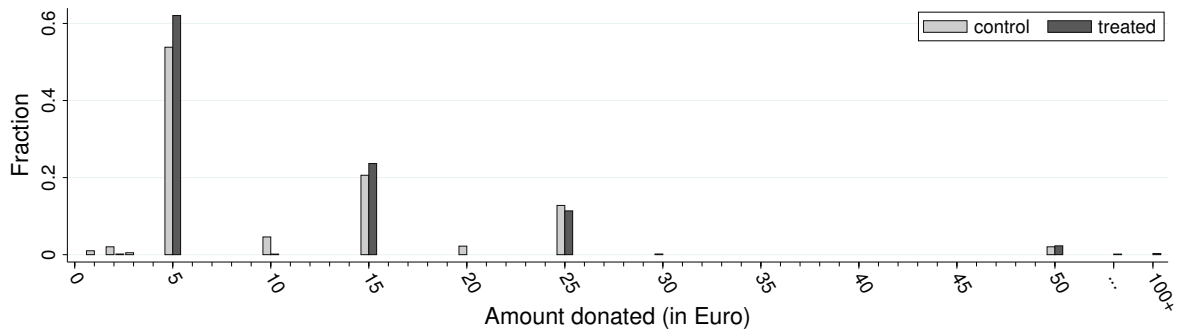


The Entry Field

Figure 4.10 displays the distributions of donated amounts by treatment status for trial 12 (hk14_mob_3), which was run on Wikipedia’s mobile website. This trial tests a variation of WMDE’s standard default set in the control banner plus the usual entry field (5/15/25/50/75/□) against a default set in the treatment banner in which the entry field is replaced by a default of 100 EUR.

Figure 4.10 shows that by excluding the free entry field, the fraction of donations with a default amount increases from 89% to 99.8%. This finding indicates that the entry field is used by a large fraction of donors. The difference in donation rates generated by treatment and control banner is not statistically significant. The p-value is 0.139 with the larger rate from the treatment banner. However, considering that donation rates are very low compared to the other trials, it could be the case that trial 12 lacks power to detect a potential extensive margin effect. In this case, a larger rate from the treatment banner may indicate that the inclusion of the entry field reduces the donation rate.

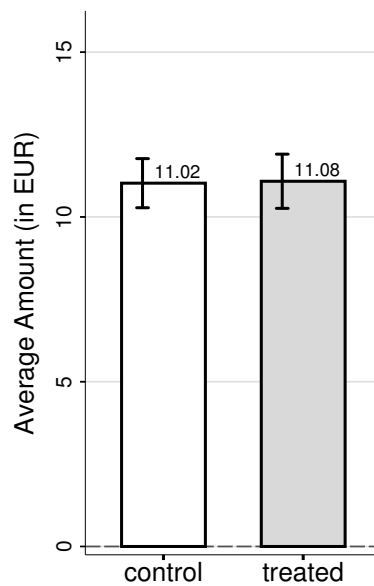
Figure 4.10: Distributions of donations: Trial 12 (hk14_mob_3)



Control – Obs.: 587 ; Defaults: 5/15/25/50/75/□
 Treated – Obs.: 651 ; Defaults: 5/15/25/50/75/100
 Donation rate: 0.113 (control) ; 0.123 (treated) ; P-value 0.139

Figure 4.11 presents the average donation amounts generated by treatment and control banner in trial 12 (hk14_mob_3). These are statistically indifferent suggesting that the inclusion or omission of the free entry field among the donation amount defaults has no effect on the average donation amount.

Figure 4.11: Average donation amounts: Trial 12 (hk14_mob_3)



Defaults (ctrl): 5/15/25/50/75/□
 Defaults (trtd): 5/15/25/50/75/100
 Obs.: 587 (ctrl) ; 651 (trtd)
 t-stat: 0.10 ; p-value: 0.918

4.5 Conclusion

In this chapter, we studied the effects of amount defaults on the decision to donate and on donated amounts, using data from trials of WMDE. The trials tested different numbers of default options, different default values and combinations of both in donation banners on the German Wikipedia website. In our analysis, we explored which of the modifications may contribute to increase the sum of donations relative to the number of banner impressions. To do so, we studied extensive and intensive margin effects.

While our results document strong effects on the distribution of donations and on the average donation amount (intensive margin effects), we find only scarce and unreliable evidence of effects on the decision whether to donate or not to donate (extensive margin effects). This finding suggests that amount defaults increase the sum of donations primarily through their effect on the average donation amount. Our analysis further reveals that there exists a complex interaction between the number of defaults and default values that may also depend on the visual display of donation banners on different devices used to access Wikipedia. Further research is needed to advance on this complex interaction. In future work, we plan to theoretically explore default effects and hope to test the resulting hypotheses through further trials in cooperation with WMDE.

Appendix

Figure 4.A1: Example donation banner



Note: Control banner from campaign hk15_enorg_2. Campaign hk15_enorg_2 was run by WMDE on the english version of Wikipedia. All campaigns analyzed in this chapter are from the German website with donation banners in German language.

Bibliography

- Abera, Wuletawu, Luca Brocca, and Riccardo Rigon. 2016. “Comparative Evaluation of Different Satellite Rainfall Estimation Products and Bias Correction in the Upper Blue Nile (UBN) Basin.” *Atmospheric research* 178:471–483.
- Acemoglu, Daron, Tristan Reed, and James A Robinson. 2014. “Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone.” *Journal of Political Economy* 122 (2): 319–368.
- Adena, Maja, Steffen Huck, and Imran Rasul. 2014. “Charitable Giving and Nonbinding Contribution-Level Suggestions - Evidence from a Field Experiment.” *Review of Behavioral Economics* 1, no. 3 (May): 275–293. doi:10.1561/105.00000010. <https://ideas.repec.org/a/now/jnlrbe/105.00000010.html>.
- Akee, Randall KQ, William E Copeland, Gordon Keeler, Adrian Angold, and E Jane Costello. 2010. “Parents’ Incomes and Children’s Outcomes: A Quasi-Experiment Using Transfer Payments from Casino Profits.” *American Economic Journal: Applied Economics* 2 (1): 86–115.
- Alpizar, Francisco, Fredrik Carlsson, and Olof Johansson-Stenman. 2008. “Anonymity, Reciprocity, and Conformity: Evidence from Voluntary Contributions to a National Park in Costa Rica.” Publisher: Elsevier, *Journal of Public Economics* 92 (5-6): 1047–1060.
- Altmann, Steffen, Armin Falk, Paul Heidhues, Rajshri Jayaraman, and Marrit Teirlinck. 2019. “Defaults and Donations: Evidence from a Field Experiment.” *Review of Economics and Statistics* 101, no. 5 (December): 808–826. <https://ideas.repec.org/a/tpr/restat/v101y2019i5p808-826.html>.
- AMS. 2012. *Rain*. American Meteorological Society Glossary, April. Accessed March 6, 2020. <http://glossary.ametsoc.org/wiki/Rain>.

- Anderson, Siwan, Patrick Francois, and Ashok Kotwal. 2015. "Clientelism in Indian Villages." *American Economic Review* 105 (6): 1780–1816.
- Ashraf, Nava, Oriana Bandiera, and B Kelsey Jack. 2014. "No Margin, No Mission? A Field Experiment on Incentives for Public Service Delivery." *Journal of Public Economics* 120:1–17.
- Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2010. "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India." *American Economic Journal: Economic Policy* 2 (1): 1–30.
- Bénabou, Roland, and Jean Tirole. 2006. "Incentives and Prosocial Behavior." *American Economic Review* 96 (5): 1652–1678.
- . 2011. "Identity, Morals, and Taboos: Beliefs as Assets." Publisher: MIT Press, *Quarterly Journal of Economics* 126 (2): 805–855.
- Bernheim, Bert Douglas. 1994. "A Theory of Conformity." Publisher: The University of Chicago Press, *Journal of Political Economy* 102 (5): 841–877.
- Berry, Marie E. 2015. "When "Bright Futures" Fade: Paradoxes of Women's Empowerment in Rwanda." *Signs: Journal of Women in Culture and Society* 41 (1): 1–27.
- Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do People Mean What They Say? Implications for Subjective Survey Data." *American Economic Review* 91 (2): 67–72.
- Bicchieri, Cristina, and Eugen Dimant. 2019. "Nudging with Care: The Risks and Benefits of Social Information." Publisher: Springer, *Public Choice*: 1–22. doi:<https://doi.org/10.1007/s11127-019-00684-6>.
- Bicchieri, Cristina, and Erte Xiao. 2009. "Do the Right Thing: But Only if Others Do So." Publisher: Wiley Online Library, *Journal of Behavioral Decision Making* 22 (2): 191–208.
- Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment on Community-based Monitoring in Uganda." *The Quarterly Journal of Economics* 124 (2): 735–769.
- Björkman Nyqvist, Martina, Damien de Walque, and Jakob Svensson. 2017. "Experimental Evidence on the Long-run Impact of Community-based Monitoring." *American Economic Journal: Applied Economics* 9 (1): 33–69.

- Bodker, Rene, J Akida, D Shayo, W Kisinza, HA Msangeni, EM Pedersen, and SW Lindsay. 2003. "Relationship Between Altitude and Intensity of Malaria Transmission in the Usambara Mountains, Tanzania." *Journal of Medical Entomology* 40 (5): 706–717.
- Bonnier, Evelina, Jonas Poulsen, Thorsten Rogall, and Miri Stryjan. 2020. "Preparing for Genocide: Quasi-Experimental Evidence from Rwanda." *Journal of Development Economics* (forthcoming).
- Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics*, 5:3705–3843. Elsevier.
- Bugingo, Irene, and Reverien Interayamahanga. 2010. *A Study On The Development And Use Of Governance Indicators in Rwanda*. Technical report. Kigali: Institute of Research and Dialogue for Peace (IRDPA).
- Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *The Quarterly Journal of Economics* 127 (4): 1755–1812.
- Celhay, Pablo, Paul Gertler, Paula Giovagnoli, and Christel Vermeersch. 2018. "Long Run Effects of Temporary Incentives on Medical Care Productivity." *American Economic Journal: Applied Economics*, no. forthcoming.
- Cesarini, David, Erik Lindqvist, Matthew J Notowidigdo, and Robert Östling. 2017. "The Effect of Wealth on Individual and Household Labor Supply: Evidence from Swedish Lotteries." *American Economic Review* 107 (12): 3917–46.
- Collins, William J, and Robert A Margo. 2007. "The Economic Aftermath of the 1960s Riots in American Cities: Evidence From Property Values." *The Journal of Economic History* 67 (4): 849–883.
- Croson, Rachel, and Jen Shang. 2013. "Limits of the Effect of Social Information on the Voluntary Provision of Public Goods: Evidence from Field Experiments." Publisher: Wiley Online Library, *Economic Inquiry* 51 (1): 473–477.
- Dal Bó, Pedro, Andrew Foster, and Louis Putterman. 2010. "Institutions and Behavior: Experimental Evidence on the Effects of Democracy." *American Economic Review* 100 (5): 2205–29.
- DellaVigna, Stefano, and Devin Pope. 2018. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy* 126 (6): 2410–2456.

- Dhaliwal, Iqbal, and Rema Hanna. 2017. "The Devil Is in the Details: The Successes and Limitations of Bureaucratic Reform in India." *Journal of Development Economics* 124:1–21.
- Doob, Anthony N., and Danielle S. McLaughlin. 1989. "Ask and You Shall be Given: Request Size and Donations to a Good Cause." *Journal of Applied Social Psychology* 19 (12): 1049–1056.
- Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "School Governance, Teacher Incentives, and Pupil–teacher Ratios: Experimental Evidence from Kenyan Primary Schools." Publisher: Elsevier, *Journal of Public Economics* 123:92–110.
- Duflo, Esther, Rema Hanna, and Stephen P. Ryan. 2012. "Incentives Work: Getting Teachers to Come to School." *American Economic Review* 102 (4): 1241–78.
- Edwards, James T., and John A. List. 2014. "Toward an Understanding of Why Suggestions Work in Charitable Fundraising: Theory and Evidence from a Natural Field Experiment." *Journal of Public Economics* 114 (C): 1–13. doi:10.1016/j.jpubeco.2014.02. https://ideas.repec.org/a/eee/pubeco/v114y2014icp1-13.html.
- Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael W. Walker. 2019. *General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya*. Technical report. National Bureau of Economic Research.
- Enikolopov, Ruben, and Ekaterina Zhuravskaya. 2007. "Decentralization and Political Institutions." *Journal of Public Economics* 91 (11–12): 2261–2290.
- Faillo, Marco, Daniela Grieco, and Luca Zarri. 2013. "Legitimate Punishment, Feedback, and the Enforcement of Cooperation." *Games and Economic Behavior* 77 (1): 271–283.
- Farmer, Didi Bertrand, Leslie Berman, Grace Ryan, Lameck Habumugisha, Paulin Basinga, Cameron Nutt, Francois Kamali, et al. 2015. "Motivations and Constraints to Family Planning: A Qualitative Study in Rwanda's Southern Kayonza District." *Global Health: Science and Practice* 3 (2): 242–254.
- Fiala, Lenka, and Charles N. Noussair. 2017. "Charitable Giving, Emotions, And The Default Effect." *Economic Inquiry* 55, no. 4 (October): 1792–1812. https://ideas.repec.org/a/bla/ecinqu/v55y2017i4p1792-1812.html.

- Finan, Frederico, Benjamin A Olken, and Rohini Pande. 2017. "The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments*, 2:467–514. Elsevier.
- Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." Publisher: Elsevier, *Economics Letters* 71 (3): 397–404.
- Fraga, Bernard, and Eitan Hersh. 2010. "Voting Costs and Voter Turnout in Competitive Elections." *Quarterly Journal of Political Science* 5:339–356.
- Frey, Bruno S., and Stephan Meier. 2004. "Social Comparisons and Pro-Social Behavior: Testing "Conditional Cooperation" in a Field Experiment." *American Economic Review* 94 (5): 1717–1722.
- Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Economic Journal: Applied Economics* 8 (4): 160–188.
- Gächter, Simon. 2007. "Conditional Cooperation: Behavioral Regularities from the Lab and the Field and Their Policy Implications." In *Economics and Psychology: A Promising New Cross-Disciplinary Field*, edited by Bruno S. Frey and Alois Stutzer, 19–50. CE-Sifo Seminar Series. Cambridge, MA: MIT Press.
- Gomez, Brad T, Thomas G Hansford, and George A Krause. 2007. "The Republicans Should Pray for Rain: Weather, Turnout, and Voting in US Presidential Elections." *Journal of Politics* 69 (3): 649–663.
- GoR. 2008. *Rwanda Joint Governance Assessment Report*. Technical report. Kigali: Government of Rwanda, October. Accessed March 20, 2020. <https://www.afdb.org/fileadmin/uploads/afdb/Documents/Project-and-Operations/ADF-BD-IF-2008-220-EN-RWANDA-JOINT-GOVERNANCE-ASSESSMENT-JGA.PDF>.
- Groh, Matthew, Nandini Krishnan, David McKenzie, and Tara Vishwanath. 2016. "The Impact of Soft Skills Training on Female Youth Employment: Evidence From a Randomized Experiment in Jordan." *IZA Journal of Labor & Development* 5 (1): 9.
- Hansford, Thomas G, and Brad T Gomez. 2010. "Estimating the Electoral Effects of Voter Turnout." *American Political Science Review* 104 (2): 268–288.

- Haushofer, Johannes, and Jeremy Shapiro. 2016. "The Short-Term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya." *The Quarterly Journal of Economics* 131 (4): 1973–2042.
- Heldt, Tobias. 2005. "Conditional Cooperation in the Field: Cross-Country Skiers' Behavior in Sweden." Uppsala University, Sweden. Accessed September 28, 2020. <https://s3.amazonaws.com/fieldexperiments-papers2/papers/00268.pdf>.
- Hirshleifer, Sarojini, David McKenzie, Rita Almeida, and Cristobal Ridao-Cano. 2016. "The Impact of Vocational Training for the Unemployed: Experimental Evidence from Turkey." *The Economic Journal* 126 (597): 2115–2146.
- Huggins, Chris. 2009. "Agricultural Policies and Local Grievances in Rural Rwanda." *Peace Review* 21 (3): 296–303.
- Imbens, Guido W, Donald B Rubin, and Bruce I Sacerdote. 2001. "Estimating the Effect of Unearned Income on Labor Earnings, Savings, and Consumption: Evidence from a Survey of Lottery Players." *American Economic Review* 91 (4): 778–794.
- Jachimowicz, Jon M., Shannon Duncan, Elke U. Weber, and Eric J. Johnson. 2019. "When and Why Defaults Influence Decisions: A Meta-analysis of Default Effects." *Behavioural Public Policy* 3 (2): 159–186.
- Johnson, Eric J., and Daniel Goldstein. 2003. "Do Defaults Save Lives?" Publisher: American Association for the Advancement of Science, *Science* 302 (5649): 1338–1339. doi:10.1126/science.1091721.
- Jones, Damon, and Ioana Marinescu. 2018. *The Labor Market Impacts of Universal and Permanent Cash Transfers: Evidence from the Alaska Permanent Fund*. Technical report. National Bureau of Economic Research.
- Joyce, Robert J, John E Janowiak, Phillip A Arkin, and Pingping Xie. 2004. "CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution." *Journal of Hydrometeorology* 5 (3): 487–503.
- Kangas, Olli, Signe Jauhiainen, Miska Simanainen, and Minna Ylikännö (toim.) 2020. *Suomen Perustulokeilun Arviointi (Evaluation of the Finnish Basic Income Experiment)* [in Finish]. Technical report 2020:15. Helsinki: Finnish Ministry of Social Affairs and Health, May. Accessed September 9, 2020. <http://urn.fi/URN:ISBN:978-952-00-9890-2>.

- Karlan, Dean S. 2007. "Social Connections and Group Banking." *The Economic Journal* 117 (517): F52–F84.
- Kasy, Maximilian. 2018. "Why a Universal Basic Income Is Better Than Subsidies of Low-Wage Work." *Data for Progress* (November): 1–22.
- Kearney, Melissa S., and Magne Mogstad. 2019. "Universal Basic Income (UBI) as a Policy Response to Current Challenges." *The Aspen Institute - Economic Strategy Group* (August): 1–19.
- Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken. 2015. "Tax farming redux: Experimental evidence on performance pay for tax collectors." *The Quarterly Journal of Economics* 131 (1): 219–271.
- Kraehnert, Kati, Tilman Brück, Michele Di Maio, and Roberto Nisticò. 2019. "The Effects of Conflict on Fertility: Evidence from the Genocide in Rwanda." *Demography*: 1–34.
- La Ferrara, Eliana. 2003. "Kin Groups and Reciprocity: A Model of Credit Transactions in Ghana." *American Economic Review* 93 (5): 1730–1751.
- Lind, Jo Thori. 2019. "Spurious Weather Effects." *Journal of Regional Science* 59 (2): 322–354.
- Linek, Maximilian. 2018. "Time Use in Hypothetical Basic Income Scenarios." October. <https://doi.org/10.1257/rct.3434-1.0>.
- Madestam, Andreas, Daniel Shoag, Stan Veuger, and David Yanagizawa-Drott. 2013. "Do Political Protests Matter? Evidence from the Tea Party Movement." *The Quarterly Journal of Economics* 128 (4): 1633–1685.
- Manski, Charles F. 2004. "Measuring Expectations." *Econometrica* 72 (5): 1329–1376.
- Mansuri, Ghazala, and Vijayendra Rao. 2013. *Localizing Development: Does Participation Work?* Policy Research Report. Washington, DC: World Bank. Accessed March 25, 2020. <https://openknowledge.worldbank.org/handle/10986/11859>.
- Martin, Richard, and John Randal. 2008. "How is Donation Behaviour Affected by the Donations of Others?" Publisher: Elsevier, *Journal of Economic Behavior & Organization* 67 (1): 228–238.
- McKenzie, David. 2017. "Can Business Owners Form Accurate Counterfactuals? Eliciting Treatment and Control Beliefs About Their Outcomes in the Alternative Treatment Status." *Journal of Business & Economic Statistics*: 1–9.

- Miguel, Edward, and Mary Kay Gugerty. 2005. "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya." *Journal of Public Economics* 89 (11): 2325–2368.
- MINALOC. 2007. *Law Establishing Community Works in Rwanda* [in English], November. Accessed September 26, 2017. <http://lip.alfa-xp.com/lip/Amategekodb.aspx?Mode=r&pid=8729&iid=2289&rid=30700136>.
- . 2008. *Community Development Policy*. Revised version. Ministry of Local Government, Rwanda, April. Accessed September 17, 2019. <https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/98538/117303/F187363017/RWA-98538.pdf>.
- . 2010. *Concept Paper on Imihigo Planning and Evaluation*. Technical report. Kigali: Ministry of Local Government, Rwanda, February. Accessed September 27, 2017. https://www.minaloc.gov.rw/IMG/pdf/Imihigo_Concept_Note_22_feb_2010_1.pdf.
- . 2011. *Concept Note on Umuganda Current Status*. Technical report. Ministry of Local Government, Rwanda. Accessed October 2, 2017. <http://www.rwandapedia.rw/cmism/views/workspace%253A%252F%252FspacesStore%252Ff8d00bc6-4b7b-4d92-bf64-c02373e61f25>.
- MoH. 2004. *Health Sector Strategic Plan 2005-2009*. Technical report. Kigali: Ministry of Health, Rwanda, April. Accessed April 23, 2020. <http://www.healthyfutures.eu/images/healthy/deliverables/D5.1/Rwanda/rwanda%20health%20sector%20strategic%20plan%202005%20-%202009%202004.pdf>.
- . 2006. *National Family Planning Policy and Its Five-Year Strategies (2006-2010)*. Technical report. Ministry of Health, Rwanda, March. Accessed March 16, 2020. <https://www.gfmer.ch/SRH-Course-2010/national-guidelines/pdf/National-FP-policy-MOH-Rwanda-2006.pdf>.
- . 2008. *Annual Report 2007*. Technical report. Ministry of Health, Rwanda. Accessed May 20, 2016. <http://www.moh.gov.rw/fileadmin/templates/MOH-Reports/RAPPORT-ANNUEL-2007.pdf>.
- . 2009. *Annual Report 2008*. Technical report. Ministry of Health, Rwanda, April. Accessed May 20, 2016. <http://www.moh.gov.rw/fileadmin/templates/MOH-Reports/Final-MoH-annual-report-2008.pdf>.

- Mukarubuga, Cecile. 2004. *Social Movements in Rwanda: Experience of Social Forums Against Poverty*. Technical report. Agency for Co-operation and Research in Development.
- Munyandamutsa, Jean Paul. 2011. *2010 Citizen Report Card Survey: Major Findings*. Kigali, December. Accessed March 23, 2020. <https://opendocs.ids.ac.uk/opendocs/bitstream/handle/20.500.12413/10076/2010%20Citizen%20report%20card%20survey.pdf?sequence=1>.
- Murray-Zmijewski, Alexandra, and Charles Gasana. 2010. *Monitoring and Evaluation in Rwanda Country Case Study*, October.
- NAR and Interpeace. 2016. *Governing With and For Citizens*. Technical report. Never Again Rwanda, Interpeace, August.
- News, BBC. 2007. *Rwanda Moves to Limit Family Size*, February. Accessed May 13, 2020. <http://news.bbc.co.uk/2/hi/africa/6358381.stm>.
- NISR and ORC Macro. 2006. *Rwanda Demographic and Health Survey 2005*. Technical report. Calverton, Maryland, U.S.A.: National Institute of Statistics of Rwanda (NISR) and ORC Macro.
- NISR, MoH, and Macro International Inc. 2008. *Rwanda Service Provision Assessment Survey 2007*. Technical report. Calverton, Maryland, U.S.A.: National Institute of Statistics of Rwanda (NISR), Ministry of Health of Rwanda (MoH), and Macro International Inc.
- NISR, MoH, and ICF International. 2012. *Rwanda Demographic and Health Survey 2010*. Technical report. Calverton, Maryland, USA: National Institute of Statistics of Rwanda (NISR), Ministry of Health of Rwanda (MoH), and ICF International.
- Olken, Benjamin A. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–249.
- . 2010. "Direct Democracy and Local Public Goods: Evidence from a Field Experiment in Indonesia." *American Political Science Review* 104 (2): 243–267.
- Olken, Benjamin A, Junko Onishi, and Susan Wong. 2014. "Should Aid Reward Performance? Evidence From a Field Experiment on Health and Education in Indonesia." *American Economic Journal: Applied Economics* 6 (4): 1–34.

- OSSREA. 2007. *Rapid and Extensive Assessment of Performance Management Contracts - Imihigo*. Technical report. Organisation for Social Science Research in Eastern and Southern Africa, July.
- Pencavel, John. 1986. "Labor Supply of Men: A Survey." *Handbook of Labor Economics* 1:3–102.
- Potters, Jan, Martin Sefton, and Lise Vesterlund. 2007. "Leading-By-Example and Signaling in Voluntary Contribution Games: An Experimental Study." Publisher: Springer, *Economic Theory* 33 (1): 169–182.
- Pradhan, Menno, Daniel Suryadarma, Amanda Beatty, Maisy Wong, Arya Gaduh, Armida Alisjahbana, and Rima Prama Artha. 2014. "Improving Educational Quality through Enhancing Community Participation: Results from a Randomized Field Experiment in Indonesia." *American Economic Journal: Applied Economics* 6 (2): 105.
- Purdeková, Andrea. 2011. "'Even if I am not here, there are so many eyes': Surveillance and State Reach in Rwanda." *The Journal of Modern African Studies* 49 (3): 475–497.
- Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.
- Rasul, Imran, and Daniel Rogger. 2018. "Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service." *The Economic Journal* 128 (608): 413–446.
- Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *The Quarterly Journal of Economics* 119 (2): 679–705.
- RGB. 2014. *The Assessment of the Impact of the Home Grown Initiatives*. Special Issue 4. Kigali: Rwanda Governance Board.
- . 2020. *Umuganda*, March. <http://www.rgb.rw/index.php?id=37>.
- Satyanath, Shanker, Nico Voigtländer, and Hans-Joachim Voth. 2017. "Bowling for Fascism: Social Capital and the Rise of the Nazi Party." *Journal of Political Economy* 125 (2): 478–526.
- Scher, Daniel, and Christine MacAulay. 2010. "The Promise of Imihigo: Decentralized Service Delivery in Rwanda, 2006-2010." *Innovations for Successful Societies: Policy Note* 133.

- Shang, Jen, and Rachel Croson. 2009. "A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods." Publisher: Oxford University Press Oxford, UK, *The Economic Journal* 119 (540): 1422–1439.
- Sommers, Marc. 2012. *Stuck: Rwandan Youth and the Struggle for Adulthood*. Athens, Georgia: University of Georgia Press.
- Thaler, Richard, and Cass Sunstein. 2008. *Nudge – Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
- Thomson, Susan M. 2008. "Rwanda." In *Africa Yearbook Online*, edited by Andreas Mehler and Klaas Van Walraven, vol. 5.
- Ueberschär, Nicole. 2018. *Access to and Utilization of Health Care: Spatial Disparities of Health Center Utilization in Huye District (Rwanda)*. 1st ed. Vol. 3. Geographische Gesundheitsforschung. Düren, Germany: Shaker Verlag.
- UNDP. 2014. *Millenium Development Goals Rwanda: Final Progress Report 2013*. Technical report. Kigali: United Nations Development Program, December. Accessed April 8, 2020. https://www.undp.org/content/dam/rwanda/docs/Research%20and%20publications/Millenium%20development%20goals/UNDP_RW_MDGR%20Rwanda_31_03_2015.pdf.
- USAID, MCHIP, and TRP. 2011. *Assessment of Rwanda's National Family Planning Policy And Its Five-Year Strategies (2005-2010)*. Technical report. USAID, Maternal and Child Health Integrated Program, The Respond Project, June. Accessed February 14, 2020. <http://www.respond-project.org/archive/files/4/4.3/4.3.2/Report-2011-Rwanda-Assessment.pdf>.
- USAID and MoH. 2002. *Report of a Qualitative Assessment of Family Planning in Rwanda*. Technical report. USAID and Ministry of Health of Rwanda.
- Uwimbabazi, Penine. 2012. "An Analysis of Umuganda: the Policy and Practice of Community Work in Rwanda." PhD diss., University of KwaZulu-Natal.
- Vesterlund, Lise. 2003. "The Informational Value of Sequential Fundraising." Publisher: Elsevier, *Journal of Public Economics* 87 (3-4): 627–657.
- Weyant, James. 1984. "Applying Social Psychology to Induce Charitable Donations." *Journal of Applied Social Psychology* 14 (5): 441–447.

- Zarghamee, Homa S., Kent D. Messer, Jacob R. Fooks, William D. Schulze, Shang Wu, and Jubo Yan. 2017. "Nudging Charitable Giving: Three Field Experiments." *Journal of Behavioral and Experimental Economics* 66 (C): 137–149.
- Zhang, Junsen. 2017. "The Evolution of China's One-Child Policy and its Effects on Family Outcomes." *The Journal of Economic Perspectives* 31 (1): 141–159.

List of Papers

List of papers, authors in alphabetical order, and location in the dissertation.

Title: “Dictating Development? The Effects of Local Institutions under Upward Accountability”

Authors: Maximilian Linek

Pages: 6-39

Title: “Time Use with Basic Income: Evidence from a Large-Scale Survey Experiment”

Authors: Maximilian Linek

Pages: 40-72

Title: “Are Wikipedia Users Conditionally Cooperative? Evidence from Fundraising Trials”

Authors: Maximilian Linek and Christian Traxler

Pages: 73-84

Title: “Defaults in Online Charity: Evidence from Wikipedia Germany”

Authors: Maximilian Linek and Christian Traxler

Pages: 85-105