USING BEHAVIORAL INTERVENTIONS TO FOSTER RESOURCE SUSTAINABILITY

Inauguraldissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Universität zu Köln

2018

Vorgelegt von
Sebastian Soung-Un Tonke, M.Sc.
Aus Berlin

Referent: Prof. Dr. Bettina Rockenbach

Korreferent: Prof. Dr. Matthias Sutter

Tag der Promotion: 26.03.2019

ACKNOWLEDGEMENTS

I thank Bettina Rockenbach for her excellent supervision during my doctoral studies. I am truly grateful for her support, encouragement, advice, trust and inspiration that led to the completion of this thesis. I am also indebted to the other members of the thesis committee: Matthias Sutter and Oliver Gürtler.

I also thank my other co-authors Timo Vogelsang and Arne Weiß for countless valuable academic discussions that have shaped and contributed to my academic development. It was a particular great pleasure to share the PhD experience in and outside of university with Timo Vogelsang and Sebastian Schneiders. I thank them and Suparee Boonmanunt, Matthias Heinz, Felix Kölle, Thomas Lauer, Anne Schielke, Uta Schier, Marcin Waligora, Lukas Wenner and Jarid Zimmermann for their companionship, generous feedback and numerous insightful conversations over lunch.

I also thank Dagmar Weiler and the Cologne Graduate School of Management, Social Sciences and Economics (CGS) for providing financial support during my PhD studies. I thank Lotte Westermann and Felix Povel from KfW who made the research cooperation in Kosovo possible.

None of the chapters in this thesis would exist without the enormous effort that research teams in Namibia and Kosovo exerted to implement the research projects. I am truly grateful for the excellent field management assistance provided by Benedikt Dechamps, Michail Kortt and Myrielle Gonschor. I also thank all of the involved students in the research projects in Windhoek and Pristina. All of them worked long hours while maintaining rigorous academic standards under tight deadlines. I thank Silvy Lungameni, Selma Nghifindaka, Hendrieka Gorases, Victoria Nekongo, Miryam Amunime, Penny Sheetekela, Maria Nkunkumo, Barbara Scharfbillig, Tjivari Kandjii, Abner Kambonde, Stephanus Kaishungu, Aubrey Goraseb, Kateuru Tjiuana, Kristof Nghidinihamba, Venomukona Tjiseua, Saija Kristof, Aili Thomas, Albertina Johannes, Demetrius Henok, Eillo Shidiwe, Frida Elunde, Johanna Ndengu, Ndapunikwa Moses, Tulinawa Neshuku, Fanni Shikongo, Sesilia Josef, Emma Josef, Linda Kalili, Johanna Nambahu, Menesia Nangolo, Josua Kandume, Kristof Johannes, Daniel Oreseb, Stabby Moyo, Jason Abraham, Erastus Andreas, Dennis Munyingwa, Erkki Iita, Amupolo Junias, Lukas Keinamseb, Fatbardha Muqiqi, Donika Kastarati, Drini Vitija, Egzona Bejtullahu, Erdonita Humolli, Arbenita Gashi, Donika Krasniqi, Rrezarta Bllaca, Ndriqim Bujupi, Linda Berisha, Saranda Berisha, Ninëz Kosumi, Djellzë Bllaca, Kaltrina Tosuni, Veriana Muqiqi, Erza Rexhepi, Ibrahim Murati, Ardita Fazliu, Anduena Kuqi, Gëzim Kurteshi, Martin Günther and Viet A Nguyen.

I would also like to express my gratitude to Job Horohua, Wilma Husselmann, Aino Nsinano, Cornwell Chadya and Vaino Shivute at *NamWater* who supported the research projects throughout all phases. I thank Ilir Abdullahu, Gazmend Behluli, Aferdita Bejta and Fatmir Murtezaj at the *Regional Water Company Pristina* and their welcoming atmosphere and cooperation that has led to the success of the project.

Finally, I thank my family for their moral support and their positive influence on me. This thesis is dedicated to them.

Dedicated to:

The courage of my mother.

The benevolence of my father.

The integrity of my brother.

The love of my wife.

CONTENT

CHAPTER 1.

INTRODUCTION

Nearly a decade ago, Richard Thaler and Cass Sunstein published the international bestseller "Nudge" (2008) which inspired behavioral economics to go beyond the understanding of decision-making and into the design of behavioral interventions as a policy measure. The promise of these behavioral interventions is that even small changes in the decision context can lead to large behavioral effects. This makes behavioral interventions in many cases more cost effective and arguably less restrictive than traditional policy measures like the prohibition of socially undesirable behavior and its enforcement. The effectiveness of behavioral interventions has now been proven across many domains such as retirement savings, health, charitable giving, energy conservation, tax evasion and many more (see for example Dolan et al., 2012).

Since human decision-making is complex and context dependent, policy interventions do not always work as expected. Behavioral interventions are no exception and while the same intervention may work in one context, it may not work or even backfire on its intention in another context (Madrian, 2014). From a policy perspective, understanding what causes these interventions to work or not to work is critical for obvious reasons. Investigating the causes requires a theoretical understanding of how these interventions work. From an ex-post perspective, it is arguably easier to rationalize behavior given the plethora of theoretical models. This approach, however, may not always be informative, since it may suffer from hindsight bias. From an ex-ante perspective, one of the most popular phrases when economists are asked for policy advice is "it depends". Given the nature how most economic models are constructed, this is almost always the correct answer to any question in economics (Rodrik, 2017). This doctoral thesis argues that in order to fully take advantage of economic theory for policy design, one needs to understand the local environment and identify the factors that "it depends" on ex-ante (Datta and Mullainathan, 2012; Duflo, 2017). Economic theory can then be used as a framework to guide survey and intervention design in an increasingly complex world.

The first theme of this thesis is the use of extensive diagnostic pre-intervention surveys to understand the local context *before* designing and testing behavioral interventions. Doing so among highly heterogeneous subject pools in the field requires not only the credible identification of a particular relevant behavioral factor (bottleneck), but also trying to *quantify* the prevalence of that behavioral factor among the population. This is particularly challenging to extract from expert interviews since their views may suffer from availability bias. Therefore, the pre-intervention studies used in this thesis are conducted among hundreds of subjects to respect the heterogeneity of human decision-making in the field. The pre-intervention diagnostic often identify factors that neo-classical economics could have disregarded as seemingly irrelevant factors (Thaler, 2015). Using rigorous experimental methodology, however, the results of this thesis suggest otherwise and show that addressing these factors may lead to highly effective and in particular cost-effective interventions. This approach seems to be very fruitful, since the treatment effects in this thesis are among the largest in the comparable literature. Should we expect the interventions used in this thesis to have similar effects elsewhere? That depends on whether the context is comparable, which may make it necessary to conduct pre-intervention studies. For example, when there is no lack of attention, we should not expect reminder interventions to be particularly effective. When there is no lack of understanding, we should not expect simplifications to be effective. This thesis therefore advocates to proceed from what works to what works when.

The second theme of this thesis is the use of behavioral interventions in settings where traditional policy tools are difficult to enforce due to ethical, legal, technical or financial reasons. In such settings, the complementarity and potential of behavioral interventions can fully unravel. They are attractive here not only because of their cost effectiveness, but also for their unrestrictive character as they do not forbid behavior or restrict access. A prime setting to use behavioral interventions is in the water sector. Water is one of the most important resources on the planet. It is fundamental for the health and wealth of individuals and therefore a key component for sustainable economic development. The lack of access to purified water sources leads to waterborne diseases like diarrhea and typhoid fever (Amrita, Kremer and Zwane, 2007), infant mortality (Gamper-Rabindran, Khan and Timmins, 2010) and inferior educational attainment (Ashraf, Glaeser, Holland and Steinberg, 2017). Affordable and dependable access to water is also crucial input factor for industrial and agricultural productivity. In fact, three out of four of the jobs worldwide are water-dependent for example in agriculture, manufacturing or construction (World Water Assessment Programme, 2016).

Because of its fundamental value to livelihood on the planet, lack of access to purified water has been named the third largest global risk in terms of impact by the World Economic Forum (2017). Two thirds of the world’s population already experience severe water scarcity for at least one month a year (Mekonnen and Hoekstra, 2016) and water demand has been increasing by 1% per year over the past decades (World Water Assessment Programme, 2018).

Water scarcity will likely continue to affect billions of lives due to population growth, economic development, changing consumption patterns, climate change, urbanization and deforestation (World Water Assessment Programme, 2018).

Behavioral interventions are intriguing in the water sector since the usefulness and applicability of traditional policy may be limited. Reducing water demand by increasing water prices has caveats: Meta-studies show that price elasticity can be low in particular in the short term (e.g., Olmstead, Hanemann and Stavins, 2007), implying that price increases need to be large in order to reduce water demand substantially. Such price changes, though, may lack political acceptance, for example because of unequal distributional effects among poorer households and the fact that access to purified water is a basic human right.

Non-payment of water bills is another threat to resource sustainability as it inhibits the maintenance and expansion of infrastructure. The financial health of many water utilities depends crucially on the cost-recovery of high fixed costs infrastructure through its users. Denial of access to non-paying customers is problematic because water is a basic human right. Many countries consequently have legal provisions against, for instance, cutting off the supply of water (Finger, Allouche and Luis-Manso, 2007).

Finding effective interventions to improve water management is therefore a global challenge. In this thesis, I contribute to the endeavor of ensuring resource sustainability by identifying the behavioral bottlenecks obstructing the sustainable use of resources and showing that these bottlenecks can be overcome by behavioral interventions. Behavioral interventions offer a complementary approach to traditional policy measures. They can be employed rather quickly with immediate effects, are cost effective and might be more accepted among the population than traditional measures of water management like pricing strategies or disconnections from the water network. This may be particularly important in less developed countries as both policy makers and customers are often severely budget constrained. In addition, the consequences of water scarcity are often felt the harshest among the poor. The rest of the thesis is structured as follows.

Chapter Two is based on Rockenbach, Tonke and Weiß (2018a).[1] We conduct a large-scale field experiment in cooperation with the national Namibian Water provider *NamWater* to reduce non-payment for piped residential water. We first conduct diagnostic surveys to identify behavioral bottlenecks that hinder payments. The surveys show that around half of customers neither receive nor understand their invoice properly. Moreover, the vast

[1] Chapter two is joint work with Bettina Rockenbach and Arne Weiß. The project was funded under the Institutional Strategy of the University of Cologne within the German Excellence Initiative of the DFG. The experiment was preregistered at the AEA RCT registry (AEARCTR-0000925).

majority seems to be willing to pay for water. Based on this diagnosis, we treat around 10,000 customers of the national water utility by providing simplified invoices via phone and additionally applying different psychological commitment techniques. Initially, payments increase by 30 to 56 percent (depending on treatment) and by about 10 percent over the course of the intervention in comparison to an untreated control group. This chapter provides novel results on both short and long-term effects of commitments to pay and investigates potential backfiring effects. We find that the intervention is highly effective even among high-debt customers and that water consumption is unaffected.

Chapter Three is based on Tonke (2018a).[2] I conduct a large-scale field experiment to encourage water conservation among 15,000 residential customers. At the time of the intervention (2017), parts of Namibia were facing a potential drought due to limited rainfall in past years. The treatments in the experiment vary the content of a text message that customers receive only once on their mobile phone. Messages containing specified conservation tips decrease water consumption by 5.8 percent, while messages encouraging customers to come up with their own ways to save water are entirely ineffective. The treatment effect is driven by high users and does not negatively affect payment behavior. The study shows that customers' lack of knowledge on how to conserve water effectively is a substantial bottleneck to pro-environmental behavior. This bottleneck can be overcome by a text message containing specific saving tips. The intervention is one of the most successful in the mass communication literature and saved 25.6 million liters within six months at a cost of merely 600 USD. The study also provides systematic advice on how to select water savings tips. The specific text message contains three simple, yet in a pre-intervention survey uncommonly mentioned water saving tips in the customers' most salient domains of water usage. The strength of this intervention is its simplicity, even in comparison to other highly cost-effective behavioral interventions: Unlike for example social comparison interventions, the text messages does not require the production of individualized graphs and redesign of invoice letters. This might be crucial factor, when droughts are already imminent or unforeseen.

Chapter Four is based on Tonke (2018b). This experiment shows that identity concerns are an important motivation for the payment of utility bills and that identity of adults is malleable through written statements about one's identity (identity labels), which substantially improve payment behavior. In a large-scale field experiment in Kosovo, customers of the public water provider randomly receive stickers with or without identity label attached to their invoice in order to reduce non-payment for piped residential water. Positively

[2] The research was funded by the Center for Social and Economic Behavior (C-SEB) at the University of Cologne. The experiment was preregistered at the AEA RCT registry (AEARCTR-0002280.).

framed identity labels include the statements “Please be a responsible citizen” or “You are a responsible citizen”. Negatively framed identity labels include the statements “Please don’t be an irresponsible citizen” or “You are not an irresponsible citizen”. Negatively framed identity labels increase collection efficiency (payments divided by billed amount) by 26 percentage points in comparison to an untreated group and are about twice as effective as positively framed labels. Post-intervention, stickers without identity labels yield significantly smaller treatment effects than stickers with identity label. Survey evidence suggests that these effects are caused by changes in customers’ self-perception and rules out alternative mechanisms like social norms, sanctioning, monitoring or reminder effects.

Identity labels are a highly effective policy measure. The best performing sticker type increases annual payments by about 7 Euro per customer in comparison to an untreated group at a cost of only 12 cents. This roughly corresponds to two to three hourly wages or one additional monthly bill paid per year. In addition, survey results suggest that the intervention is well accepted among the population. There are many other domains in which identity labels could be used to encourage civic behavior as for example voter mobilization, avoiding littering, volunteering for public services, driving carefully, energy conservation or paying taxes. Some of these domains might be even more promising than the paying for utility bills since they constitute behaviors that might be more intuitively associated with being a responsible citizen.

Chapter Five is based on Rockenbach, Tonke and Weiß (2018b).[3] We conduct a framed field experiment in an impoverished neighborhood in Namibia to study how charitable donations to disaster management (e.g. to fight floods, droughts, etc.) depend on the sharing behavior of others. While social norms ask the rich to serve society and to share their wealth with the less fortunate, they often fail to comply with this expectation. We ask whether the rich’s self-serving behavior causes contagion effects among the poor. To do so we vary both the reference group (rich and poor) as well as the information of the behavior of others (egoistically and prosocially) in a 2x2 design. The rich’s failure to share is not only unexpected by the poor, it also changes their view on the acceptability of own egoistic behavior and leads to an increase in egoistic behavior. The poor use the rich’s norm violation to justify their own self-serving behavior. The number of egoistic choices increases by 18 percentage points, which translates into a reduction of donations by about 26%. Thus, the rich’s self-serving behavior causes a double damage: Society not only suffers from their low contributions but also from contagion effects among others.

---

[3] Chapter five is joint work with Bettina Rockenbach and Arne Weiß. The project was funded under the Institutional Strategy of the University of Cologne within the German Excellence Initiative of the DFG.

CHAPTER 2.

USING BEHAVIORAL INSIGHTS TO DECREASE NON-PAYMENT FOR PUBLIC UTILITIES

Joint work with Bettina Rockenbach and Arne Weiß

2.1 Introduction

Dependable and affordable access to public utilities such as electricity and water is a key prerequisite for economic and social development. Public utilities provide important input factors for industry and agriculture and are crucial for the development of human capital, for example in schools or hospitals. The maintenance and expansion of infrastructure is constrained when customers do not pay their utility bills. While non-payment is also an issue in rich countries, it is a particular problem in low and middle-income countries (Szabó and Ujhelyi, 2015; Jack and Smith, 2015). The existing literature on reasons for non-payment in these countries are attributed to customers' low income and a low willingness to pay, for instance due to low quality of service or a lack of enforcement of non-paying customers (Aguilar-Benitez and Saphores, 2008; Vásquez, 2015, Vásquez and Alicea-Planas, 2017). Addressing these causes requires investments in the infrastructure to improve service quality or the enforcement of heavy-handed interventions, such as denial of future service. Both options impose serious challenges for policy makers. The first option (investments to improve service) may itself be constrained by non-payment such that the utility and customers remain stuck in a bad equilibrium (Strand, 2012). The second option (denying access to non-paying customers) is constrained by water and electricity being basic needs (see UN sustainable development goals) and fundamental for the welfare of individuals (for micro-level evidence on the negative effects of water outages on health and economic activity see Ashraf, Glaeser, Holland and Steinberg, 2017). Many countries consequently have legal provisions against, for instance, cutting off the supply of water (Finger, Allouche and Luis-Manso, 2007). When heavy-handed interventions are difficult to implement, the toolbox of softer behavioral interventions (e.g., "nudges") may offer a powerful alternative (Madrian, 2014; Chetty, 2015).

In this paper, we investigate the potential of behavioral interventions to address non-payment for water. We present a natural field experiment (Harrison and List, 2004) among 9,833 customers in northern Namibia in cooperation with the national public water provider *Namibia Water Corporation* (*NamWater).* Our administrative dataset covers an extended period of both pre- and post-intervention data, which allows us to study short-term and long-term effects as well as the effectiveness of the intervention among high and low debt customers. We conduct the experiment in an area of 84,610 square kilometers (about the size of South Carolina) over a period of nine months. At the start of our study, 84% of our eventually treated sample had arrears on their account and about 33% had arrears of at least three average monthly bills. The average debt of our treated sample corresponds to about 334 USD (Median 10 USD)[4]. Payment enforcement through legal action is slow and sanctioning non-paying customers by cutting-off supply is difficult due to ethical, technical and logistical constraints.

In the *basic* treatment of our experiment, we call customers to offer a free monthly text message service that contains simplified invoice information. The text message ensures that customers have easy access to the invoice information in a non-technical language. The *basic* treatment serves as a comparison group to two psychological *commitment treatments* (i.e., the *self-concept* and the *plan* treatment) that are implement on top of the *basic* treatment. The comparison of the *basic* and the *commitment* treatments isolates the effects of the *commitment* treatments from the effects of reminders, personal interactions on the phone, reciprocity for being offered a free service, and feeling monitored. In the *self-concept* treatment, we invoke a self-concept of a reliable water payer. This is done by eliciting answers from the customers to questions such as "How important is it to you to be reliable water payer?" The intervention builds on individuals' desire to shape their self-concepts and ties their (desired) self-concept to their payment behavior (Bryan, Walton, Rogers and Dweck, 2011). In the *plan* treatment, customers commit to their own plan for future payments by responding to questions such as "How do you make sure you pay your bills on time?" Plan making creates personal rules and can help customers to develop strategies to overcome logistical obstacles. Further, eliciting plans creates cognitive links between a future cue and behavior. Once a cue becomes available (such as when receiving the invoice), individuals automatically switch to the pre-conceived plan (Gollwitzer, 1999; Gollwitzer and Sheeran,

[4] 10 USD correspond to roughly 9 hourly wages.

2006; for reviews see Beshears, Milkman, and Schwartzstein, 2016; Rogers, Milkman, John and Norton, 2015).

We find that simply offering and sending customers free monthly invoice information by text message raises average payments remarkably, both in the first month (about 30%) and by about 10% over the course of the intervention. On top of these effects, we find substantial short-term effects of the commitment treatments: In the first month, average payments increase - on top of the *basic* treatment - by a further 10% in the *self-concept* treatment and by another 26.5% in the *plan* treatment. The commitment treatments are also highly effective among customers who have paid little or infrequently in the pre-intervention year. The effectiveness wanes over time, yet we find positive treatment effects more than half a year post intervention. Water consumption is unaffected by the treatments. The entire intervention is extremely cost-effective and increased average payments per customer by about 10-11 USD, which corresponds to an estimated return-on-investment of roughly 1,000%.

Why did these interventions work? The effectiveness of an intervention crucially hinges on diagnosing the underlying problem and understanding the local context (Duflo, 2017). Before we implemented our behavioral intervention, we identified and quantified potential behavioral bottlenecks (Datta and Mullainathan, 2014) by conducting extensive pre-intervention interviews among a randomized sample of 329 of customers. This is important because the literature on non-payment for water in Namibia suggests numerous yet unquantified reasons. Among such reasons are poverty and beliefs that water should be free because it is a basic human right and through publicly advertised slogans like “Water is Life”. Further, customers may have a lack of understanding of the cost-covering concept because water was supplied without charge until Namibia’s independence in 1990 (Du Plessis, Neels, Anyim and Matros, 2005; Klintenberg, Mazambani and Natanga, 2007). Moreover, some individuals perceive water as “a gift from God”, which should not be marketed (Mazambani, Schönbrodt-Stitt and Klintenberg, 2006). As these studies were mostly qualitative or relied on small samples, it remained unclear how representative these reasons were.

Our own pre-intervention diagnostics show two surprising observations, which are not commonly mentioned in the previous literature: First, about half of the customers report to have problems receiving their postal invoice and understanding the invoice letter. The *basic* treatment addresses this first bottleneck by providing comprehensible invoice information. Secondly, a remarkably high share of customers seems to be willing to pay for water. For instance, more than 90% of customers state that water should be paid for and name coherent reasons for the necessity of paying. Furthermore, in an incentivized dictator game customers would allocate, on average, 75% of a lottery price (about 50 USD) to paying back arrears on the account. These results indicate a gap between the willingness to pay and the actual payment behavior recorded in the administrative data set. Our commitment interventions address this second behavioral bottleneck by using psychological commitment techniques. Individuals commit either to a desirable self-concept or to a concrete plan and then try to act consistently with their commitment. These interventions are commitment techniques in the sense that acting inconsistently with the elicited self-concept or plan creates cognitive dissonance, making it psychologically costly to renege (Festinger, 1957; Bénabou and Tirole, 2011).

This paper is the first to study both short and long-term, as well as possible counter effects of commitments to pay among heterogeneous types and on the usage of consumption good. Thereby, we address unanswered questions on the impact of behavioral interventions beyond the immediate effect on the targeted behavior (Thaler, 2018). Our paper extents the previous literature in four important and policy-relevant ways: First, we complement the existing literature by showing the effectiveness of commitment mechanisms for a behavior that is monetarily costly to the customer and has a public goods character. Hitherto, self-concept invoking treatments have only been used in the context of voter mobilization (Bryan, Walton, Rogers and Dweck, 2011), to promote pro-social behavior among children (Bryan, Master and Walton, 2014) and to decrease cheating behavior (Bryan, Adams and Monin, 2013). Plan-making interventions have for example been used to increase voter mobilization (Nickerson, Rogers, 2010), vaccination rates (Milkman, Beshears, Choi, Laibson and Madrian, 2011) or preventive colonoscopy screening rates (Milkman, Beshears, Choi, Laibson and Madrian, 2013).

Secondly, we show long-term evidence in a domain where recurrent action is necessary. This is important because behavioral interventions may induce rebound effects in the long-term, resulting in the absence of a net effect (Sunstein, 2016).[5] These concerns are particularly important when interventions target payment behavior (Szabó and Ujhelyi (2015). On the one hand, households might budget over time such that higher payments now induce lower payments in the future. In addition, psychological licensing effects (Merritt, Effron, Monin, 2010) could induce lower future payments. On the other hand, the literature on consistency and habit formation suggests positive spillovers from short-term to long-term effects (Neal, Wood, Labrecque and Lally, 2012, Frey and Rogers, 2014). Evidence on behavioral long-term effects of such commitment treatments is therefore important but hitherto very limited.[6]

[5] For long-term effects of social norms interventions on water and energy conservation see Ferraro, Miranda and Price (2011) and Allcott and Rogers (2014)

[6] Conner and Higgins (2010) measure smoking behavior 48 months after the first elicitation of implementation intentions.

Thirdly, we document the effectiveness of such interventions for high and low debt customers. Taking account of the heterogeneity among customers is not only of scientific importance; it may also be viewed as a fairness imperative to design policy interventions that not only affect the average, but also, or even in particular, those customers that most severely free-ride on the cooperative behavior of others. We show that our commitment interventions are highly effective even among customers that have paid little in the past. This is particularly interesting for the *self-concept* treatment. Who one is or how one behaved in the past seems less important than who one wants to be. Furthermore, we would expect our commitment interventions to be particularly effective among customers with high debt since those are the individuals for which the gap between payment behavior and their stated willingness to pay is largest.

Fourthly, our experiment provides empirical evidence on the question whether behavioral interventions that target payment cause spillovers on a closely related domain: water consumption. Predictions are unclear as higher payments might increase water consumption through psychological entitlement or licensing effects. On the contrary, awareness of the cost of water might lead to reduced usage.

Our study highlights the importance of identifying, quantifying and appropriately addressing behavioral bottlenecks. We show that removing logistical and behavioral bottlenecks can help to bridge the discrepancy between willingness-to-pay and actually payments. This is not only highly cost-effective (we estimate a return-on-investment of around 900%), but is also particularly important in settings where applying incentive-based economic solutions is restricted because of financial, legal or ethical constraints.

2.2 Experimental Setting and Design

Namibia is one of the driest country in Sub-Saharan Africa with limited amount of surface water and low and unpredictable rainfall (Lu et al., 2016). Groundwater is often saline and not potable. Droughts are a recurrent threat in Namibia. About 61% of customers in our study area state that they do not have sources of water other than what is provided by NamWater.

Pre-intervention telephone survey – To diagnose the causes for non-payments for water, we ran a telephone survey in June 2015 with a random sample (N=329) of *NamWater's* customers. In addition, we conducted face-to-face interviews (N=31) among a convenience sample in Windhoek, Namibia's capital, in which we asked interviewees to explain the invoice to us. Our research team of local students carried out the surveys and introduced themselves as part of a research team of the University of Cologne. The telephone questionnaire contains 39 questions and completion lasts on average about 30 minutes. Questions address water management in the household, mode of payment, understanding and perception of water payments, personal and social norms, knowledge about cut-offs and demographics.

Strikingly, about 42% of customers state to have experienced problems with late or no delivery of invoices and about 45% of participants of the Windhoek sample are unable to point out the total amount due when asked to explain the invoice. This finding is in line with older survey results by Du Plessis et al. (2005) who report that interviewees have trouble understanding and receiving their invoice. Secondly, a remarkably high share of customers seems to be willing to pay for water: 98% of customers state that water should be paid for and 92% can name coherent reasons for water payments (e.g. "purification" or "maintenance"). To measure willingness-to-pay with monetary incentives we implement a dictator game as part of the telephone survey. Customers could win a lottery earning of about 50 USD (roughly 45 hourly wages) for participation in the survey. We then asked participants how they would split the potential lottery earning between a direct mobile phone transfer (phone credit) and a repayment of arrears on their *NamWater* account. On average, participants would allocate 75% of the potential lottery earning to reduce arrears. We find that customers with higher debt allocate significantly more money to reduce debt on their account (p=0.008).[7] This finding shows that customers do not only state being willing to pay for water. They also walk the talk, by foregoing money to repay debt. This strongly suggests a gap between customers'

[7] OLS regression with debt measured in average invoices on fraction of money allocated to repay debt. Measured among 254 customers for which we could match account information with survey answers.

willingness to pay and their actual payment behavior. We address this second bottleneck by using psychological commitment techniques.

Experimental Sample – Our study focuses on private customers in the northern *Cuvelai* region. *NamWater* bills these customers directly. In most other regions, households living in cities and towns are typically billed by their municipality, which itself is billed by *NamWater*. We restrict our sample to private customers because the holder of a private account is typically also responsible for the household's payment decisions, unlike community-based customers, government units or municipalities. A large overlap between the legal and the economic agents is an important prerequisite for any commitment-based intervention to work. Our sample is quite heterogeneous and ranges from predominantly small households to a few large businesses. The northern *Cuvelai* region comprises about 79% of *NamWater*'s directly billed private customers (August 2015). Most people in the *Cuvelai* region speak either Oshiwambo, the most widely spoken of Namibia's many officially recognized local languages, or English. Restricting our experiment to this region therefore reduces potential noise that comes from translation into different local languages while at the same time allowing us to reach the vast majority of *NamWater's* private customers.

Experimental Data - NamWater compiles accounting data on a monthly basis, usually at the beginning of each month. The majority of the data is entered manually into the system (about 95% of customers pay in cash at payment points) and then digitally compiled. Clear guidelines for quality checks of the manually entered data and a consistent notation of data correction were missing at the time of the study. This leads to some erratic data points (e.g. negative payments and extreme values) and missing values, which are typically corrected in the ensuing months. We therefore update incomplete payment records and extreme payment records by using the corrected account data from the ensuing month. Similar to field data on charitable giving, health expenditure or income data, the customers' payment data has a large fraction of zeros and is heavily right skewed. The large fraction of zeros results from the fact that customers on average make only about four payments a year and often make bulk payments in multiples of 50 N$.[8] We will provide estimates from several regression models and data transformations typically used with such data types to show robustness of our results.

[8] In October 2015 1 USD was worth 13.5 N$.

2.3 Treatments

Basic – In the *basic* treatment, we call customers and offer a free monthly text message service (SMS) in simplified language to ensure both access and understanding of the invoice information. The message contains the total amount due as well as the water consumption (in N$) of the last month. Table 1 provides the full telephone script and text message content.

Table 1. Basic Script and Text Message

Telephone script (once)	Text message (every month)
Hello, how *are you today?* My name is [caller's name]. I am calling from Namwater. Am I talking to Mr./Mrs. [customer's name]? We would like to get to know our customers. Can I ask you a few questions? Are you the one who is paying for water in your house? We want to inform you about our new SMS service. The SMS service will make it easier for you to pay. The SMS will contain all necessary information for you to make payments, like the amount you have to pay and your NamWater account number. The SMS will be sent to you each month for free. Would you like to receive the SMS in English or Oshiwambo? That's it, I don't have any more questions. Thank you for your time and have a nice day.	Dear [Name], Here is your invoice: In [Month] you used water for N$ [Consumption amount]. In total you have to pay N$ [Amount] (Total amount due). Your NamWater account number is [Account Number]. Please pay via usual payment mode. Ignore this SMS if already paid.

In addition to the *basic* treatment, we conducted two *commitment treatments* – the *self-concept* treatment and the *plan* treatment – to address the diagnosed gap between customers' stated high willingness-to-pay and their low payments.

Self-concept treatment – On top of the *basic* script, the *self-concept* treatment intends to invoke a water-paying self-concept by asking "*How important is it to you to be a reliable water payer?*" and three more comparable questions in which the adjective *responsible* is replaced by *good*, *responsible* and *debt-free.* The first sentence of the text message in the *self-concept* treatment differs in comparison to the *basic* treatment and reads “Here is your invoice to you as a committed water payer” to remind customers of their answers. We use nouns (“water-payer”) rather than verbs (“pay water”) because previous literature has shown that the interventions based on nouns are more effective, as these are more representative of one’s self (Walton and Banaji, 2004) – think of “to lie” vs “being a liar”. The questions are adapted from Bryan et al. (2011), who use this type of intervention to mobilize voters in the US. In order to make sure that the self-concept treatment could be effectively implemented in the local language (Oshiwambo), we conducted a pilot study in which we asked native speakers about their perception about “lying” and “being a liar” in Oshiwambo. As in English, the latter provoked stronger reactions.

Table 2. Self-Concept Script and Text Message

Telephone script (once)	Text message (every month)
[Basic] + We also have a couple of questions to you as a NamWater Customer, is that OK? How important is it to you to be a reliable water payer? How much do you care about being a good water payer? How much do you care about being a responsible water payer? How important is it to you to be a debt-free water payer? That’s it, I don’t have any more questions. Thank you for your time and have a nice day.	Dear [Name], Here is your invoice to you as a committed water payer: In [Month] you used water for N$ [Consumption amount]. In total you have to pay N$ [Amount] (Total amount due). Your NamWater account number is [Account Number]. Please pay via usual payment mode. Ignore this SMS if already paid.

Plan treatment – On top of the text *basic* script, we elicit concrete plans of customers about how, when, where they would make payments and ask them to commit to their plans to pay (see Table 3). The first sentence of the text message reads “As a reminder to your commitment to pay”. Table 1 provides an overview over the details of the telephone script and text message for each treatment.

Table 3. Plan Script and Text Message

Telephone script (once)	Text message (every month)
[Basic] + We also have a couple of questions to you as a NamWater Customer, is that OK? How do you make sure you pay your bills on time? Can we count on your future payments? Are you going to pay next month? When exactly are you going to pay? Where are you going to pay? So Mr./Mrs. <name>, let me repeat: When you are in <city name> in <month> can we expect your payments? That's it, I don't have any more questions. Thank you for your time and have a nice day.	Dear [Name], As a reminder to your commitment to pay for water: In [Month] you used water for N$ [Consumption amount]. In total you have to pay N$ [Amount] (Total amount due). Your NamWater account number is [Account Number]. Please pay via usual payment mode. Ignore this SMS if already paid.

Behavioral mechanism – The *basic* treatment addresses the diagnosed bottleneck of undelivered invoices and the lack of understanding of the invoices. However, the fact that customers interacted with our research team on the phone and repeatedly received invoices via text messages makes it difficult to identify one particular behavioral mechanism behind the *basic* treatment. Besides receiving and understanding their invoice, customers could also feel more scrutinized for having been called by the company and seeing, via text message, that their consumption and payments are monitored. Furthermore, reminder effects (e.g., Cadena and Shoar, 2013, Karlan, McConnel, Mullainathan and Zinman, 2016), or effects caused by personal contact with the phone caller (Karlan, Morten and Zinman, 2015) and reciprocity effects for being offered a free SMS service (Szabó and Ujhelyi, 2015) could drive the effects of the *basic* treatment. By comparing the commitment treatments to the *basic* treatment (instead of an untreated control group), we identify the additional effects of the psychological commitments on top of the potential effects, already present in the *basic* treatment.

The gap between willingness-to-pay and the actual payment behavior is addressed by the commitment treatments. These interventions are commitment techniques in the sense that customers commit to either a plan or a desirable self-concept. Deviating from one' plans and self-concept may create disutility, in form of cognitive dissonance, as humans strive to act consistently (Festinger, 1957; Konow, 2000; Bénabou and Tirole, 2011). This drive for consistency makes it psychologically costly to renege. A necessary though not sufficient condition for these commitment techniques to work is that individuals indeed want to act in line with the targeted behavior (Sheeran, Milne, Webb and Gollwitzer, 2005; Bryan et al., 2011). The two commitment treatments should therefore be ineffective if customers perceive the targeted behavior as undesirable. Commitment interventions are more effective when declared publicly (Rogers, Milkman, John and Norton, 2016). We make use of this, as customers commit publicly to their plan to pay or their self-concept towards the phone caller. To which degree any treatment effect is driven by this public commitment is beyond the scope of this study.

A potential concern is that the commitment treatments change customers' beliefs about being sanctioned differently than in the *basic* treatment. Disconnections, however, are very hard to enforce due to ethical (basic human right), technical and logistical constraints (e.g., limited staff and high cost). Before customers are placed on a cut-off list, they are contacted several times by *NamWater* staff and receive a warning letter about potential disconnections. Being placed on a cut-off list does also not lead to an immediate disconnection. Our own estimates suggest that at most 10% of customers on a cutting-list are eventually cut off. We trained the phone callers to be very polite and friendly during the phone calls. The phone call was framed as a service and "getting to know our customers". Judged by the daily reports of our phone callers on their interaction with customers as well by the almost universal take-up of the SMS service, customers had a positive perception of the intervention. It therefore seems unlikely that customers confounded our phone call with a threat of potential sanctions.

2.4 Conducting the Experiment

Randomization - We randomly create three groups among the 12,719 *NamWater* customers whose account information includes a mobile phone number. We use the min-max t-stat method stratified by geographical location (proxied by pipeline location to which a customers is connected) with 1,000 re-randomizations. We balance between each pair of treatments on the following variables of the pre-intervention year: number of payments made, yearly payment ratio (sum of payments divided by sum of invoices), amount of months being a *NamWater* customer, debt amount and total invoice amount.

Implementation - In order to call the 12,719 phone numbers, we set up a call center in Windhoek with 25 local students in the last two weeks of September 2015 on *NamWater*'s premises. All phone callers took part in a three-day workshop and received in-depth training including mock callings and regular feedback. Treatments are balanced within day and phone caller. This ensures that treatment effects are not confounded by time (e.g. "end of the month effects") or phone-caller idiosyncratic effects (e.g., gender or friendliness). Phone callers received their daily assignments in the morning briefing. Daily briefings in the morning and afternoon ensured that any questions of customers are handled in the same way.

All interactions with customers were fully scripted, practiced and the adherence to the script rigorously monitored. This procedure allows us to provide full transparency and high control about the content of the phone calls. A phone call usually lasted about 3-5 minutes and was limited to this one conversation. The phone callers coded the answers given to the questions as well as the interviewee's gender and the language of the phone call. The phone callers were trained and reminded to be as friendly and helpful as possible. All customers were called up to three times if the customer could not be reached during a previous call attempt. We managed to talk to 9,833 (77.3%) of the assigned customers which we will refer to as the intention-to-treat (ITT) sample. The majority of the customers that we could not reach had inactive or wrong phone numbers or were not answering the phone. Note, that these types of non-responses cannot cause a selection bias since unreachable customers cannot know which treatment they were assigned to.

Based on the phone callers' feedback, the vast majority of customers was delighted about the introduction of the SMS service, which is reflected in near universal take-up (about 98%). Text messages were sent on a monthly basis from October 2015 until June 2016, timed as closely as possible to the mailing of the written invoices. The text message did not substitute the postal invoices.

Table 4 shows summary statistics (mean, standard deviation, 25th percentile, median and 75th percentile) for the *basic, self-concept* and *plan* group for the pre-intervention year as well as balance tests. The sample is well balanced with no statistically significant mean differences between the *basic* treatment and the *self-concept* and *plan* treatment. Note that the standard deviations are relatively large for many of our variables, highlighting the large heterogeneity of customers in our sample.[9]

Table 4. Pre-intervention Summary Statistics (ITT Sample)

	Mean	SD	P25	Median	P75	Mean diff. to Basic	p-value
Basic (N=3,287)							
Payment in N$	111.75	546.39	0	0	100	-	-
Payment in N$ if >0	313.35	879.77	81	158	312	-	-
Consumption in N$	133.01	309.41	21.20	53.40	119.40	-	-
Number of payments	4.15	2.75	2	4	6	-	-
Debt in N$	536.42	2164.79	31.95	138.54	454.3	-	-
Account age in month	44.26	39.178	19	34	53	-	-
Self-Concept (N=3,297)							
Payment in N$	111.11	363.25	0	0	100	-0.642	0.882
Payment in N$ if >0	309.85	553.53	84.50	155.40	300	-3.505	0.770
Consumption in N$	132.26	325.28	26.70	53.70	115.70	-0.741	0.863
Number of payments	4.18	2.68	2	4	6	.0356	0.603
Debt in N$	475.90	1451.69	37.35	125.87	407.69	-60.52	0.198
Account age in month	45.42	39.74	19	34	55	1.152	0.247
Plan (N=3,249)							
Payment in N$	116.02	847.13	0	0	100	4.273	0.535
Payment in N$ if >0	325.13	1393.96	80	156.1	300	11.78	0.541
Consumption in N$	136.83	575.72	21.66	50.75	111.65	3.828	0.562
Number of payments	4.15	2.70	2	4	6	0.005	0.948
Debt in N$	501.11	1594.25	31.86	127.52	417.36	-35.31	0.467
Account age in month	45.13	39.64	19	34	55	0.862	0.388

Notes: The table reports summary statistics of the pre-intervention year for the ITT sample. The table provides mean, standard error, 25th percentile, median and 75th percentile. The last two columns test for pre-treatment differences in means before the intervention using an OLS regression with treatment dummies and standard errors clustered at the customer level. We report the regression coefficients and p-values of the two commitment treatments in comparison to the *basic* treatment.

Summary statistics of the intervention – Summary statistics of the intervention are displayed in Table 5. The take up rate is very high in all three treatments (97-99%). Roughly 13% of text messages each month are on average undeliverable (after automated retries) over the nine-month period. Typical reasons are technological restrictions, for example network errors, deactivated numbers or switched-off phones. Attrition rates are around 1%. There is no

[9] The same table is reproduced with top coded variables at the 99th quantile in the appendix A1. Top coding reduces the influence of outliers on the mean value of the variables.

statistically significant difference in attrition rates between treatments nor among observable characteristics. In most cases, customers that attrite from the data become inactive.

Customers show a high commitment with respect to paying for water, which is what we expected given the answers in our pre-intervention survey. In the *self-concept* treatment about 96% of customers (i.e., even those with very high debt) state that being a reliable water payer is either very important or important to them. In the *plan* treatment, 49% of customers mention more than four concrete steps of the payment process and thus provide relatively detailed plans. About 87% of customers make plans to pay during October. Lastly, about 72% of interviews were conducted in Oshiwambo (the local language).

Table 5. Summary Statistics of the Intervention

Treatment	Take-up rate	Successfully delivered messages	Attrition rate	Most common commitment answer	2nd most common commitment answer	Plans to pay in October	Phone call in Oshiwambo
Basic	0.990	0.873	0.010	-	-	-	0.703
Self-concept	0.972	0.860	0.008	0.627 (V. import.)	0.329 (Important)	-	0.736
Plan	0.973	0.860	0.011	0.493 (4+ steps)	0.441 (2-3 steps)	0.866	0.732

Cost of the intervention – The cost for the intervention was relatively low. A back-of-the-envelope calculation suggests costs of about 1 USD per customer for a 5 minute phone call and the text messages over 9 months. The text messages cost 60 cents per customer and conducting a phone call costs 38 cents, including personnel cost for the phone caller and providing the necessary materials.

2.5 Estimation Strategy and Results

Estimation strategy - This section analyzes how the two commitment treatments affected payment behavior beyond the effect of the *basic* treatment. The additional effect of the *basic* treatment in comparison to an untreated group is evaluated in a separate section.

Table 5 shows the intention-to-treat effects (ITT) in the first month after the intervention (October). Regression 1 shows the marginal effects at means on the extensive margin (probability of making a payment) of a probit regression. Regression 2 shows the effects on the intensive margin (the effect on the natural logarithm of the payment amount conditional on being larger than zero). Regression 3 estimates the combined effects of regression 1 and 2 using a two-part model which multiplies the estimated effects from regression 1 and 2 (Belotti et al., 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. We control for the pre-treatment values of the variables and strata used for randomization in the regression as recommended by Bruhn and McKenzie (2009). All standard errors are clustered at the customer level.

Regressions 4 and 5 provide fixed effects difference-in-difference with clustered standard errors on the customer level (Betrand, Duflo and Mullainathan, 2004). We use the inverse hyperbolic sine to transform our outcome variable. The inverse hyperbolic sine can be interpreted in the same way as the traditional log transformation.[10] Unlike the log transformation, however, the inverse hyperbolic sine is defined for zero and negative values (Burbidge, Magee and Robb, 1988; MacKinnon and Magee, 1990). Such transformations have been recently used for example by McKenzie (2017) for profit data of entrepreneurs in Nigeria. In appendix A2, we present estimations when adding 0.1 and 1 to the log of payment amount and after a cube root transformation, which all yield similar results. Regression 6 shows the effects on the inverse hyperbolic sine of water consumption using the same specification as in regression 5. We use the following difference-in-difference estimator:

$$\begin{aligned} outcome_{it} = {} & \beta \cdot treatment_i \cdot post + month\ fixed\ effect_t \\ & + customer\ fixed\ effect_i + error_{it} \end{aligned}$$

where treatment is an indicator variable for the treatment groups and post is a dummy variable indicating periods after the intervention starts (October 2015). Table 6 reports the initial effects and Table 7 the long-term effects of the two commitment treatments.

[10] The inverse hyperbolic sine is defined as ln $(y+(y^2+1)^{0.5})$.

***Result 1**: Both commitment interventions increase payments compared to the basic treatment in October: average payments increase by about 10% in the self-concept treatment and by about 26.5% in the plan treatment.*

The *self-concept* treatment increases the likelihood of paying by 2.8 percentage points (p=0.014) and the *plan* treatment by 7.5 percentage points (p<0.001). On the intensive margin, both point estimates are positive (approximately 1.3 and 3.5 percent) yet statistically insignificantly different from zero. The estimated combined effect from the two-part model for the *self-concept* treatment is 17.80 N$ (p=0.064) and for the *plan* treatment 47.19 N$ (p<0.001). This corresponds to an estimated increase of 10% and 26.5% respectively in comparison to the mean raw value of the *basic* treatment in October (178 N$).

Table 6. ITT Effects in October in Comparison to Basic Group

	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$	(5) IHS Water usage in N$
Self-concept	0.028**	0.013	17.799*	0.161**	0.042
	(0.011)	(0.031)	(9.608)	(0.072)	(0.042)
Plan	0.075***	0.035	47.187***	0.463***	0.009
	(0.012)	(0.031)	(10.181)	(0.072)	(0.042)
Model	Probit	OLS	Two part model	Diff-in-Diff	Diff-in-Diff
Observations	9,828	5,456	9,828	128,685	128,685
R-Squared	0.115	0.329	-	0.017	0.024

Notes: Table 6 reports ITT effects on payment behavior (regressions 1-4) and water consumption (regression 5) for the first month of the intervention (October 2015). Regression 1 shows the treatment effects of a probit regression on the likelihood of making a payment (marginal effects at means). Regression 2 reports the ITT effect using OLS on the intensive margin. Regression 3 multiplies the effects of regressions 1 and 2 to get an estimate of the combined effect using a two-part model (Belotti et al. 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. Regression 4 and 5 show the estimates from a diff-in-diff regression including time and individual fixed effects. The outcome variables are transformed using the inverse hyperbolic sine (IHS). Regressions adjust for the randomization method as suggested by Bruhn and McKenzie (2009): We include pre-treatment means used for randomization as well as geographical location (proxied by pipeline connection), phone caller fixed effects and date of phone call fixed effects. Control variables are top coded at the 99th percentile. Five customers attrite in October. All reported standard errors are clustered at the customer level to account for serial correlation. ** p<0.05; *** p<0.01

To put this in perspective, an hourly wage corresponds to about 15 N$[11] (Namibian Statistics Agency, 2016). This means that the *self-concept* treatment increases average payments by about one and the *plan* treatment by about three hourly wages. The inverse hyperbolic sine estimates a larger positive and statistically significant effects on the payment amount of 16 (p=0.025) and 46 percent (p<0.001). The coefficient of water consumption has a positive point estimate but is insignificantly different from zero.

***Result 2**: Customers who have paid little in the past react strongly to the treatments*

In Figure 1, we visualize heterogeneous treatments effects across customers with respect to their past payment behavior by depicting the regression estimates of the combined effect of the two-part model by quartiles of the pre-intervention payment ratio (sum of payments divided by sum of invoices) in October. Bars indicate 95% confidence intervals. The full regression tables are presented in appendix Table A3. The 1st quartile for example shows the worst 25% in terms of the pre-intervention payment ratio (i.e. those that paid the least fraction of their invoice). The 4th quartile shows the 25% with the highest fraction. The interventions are highly effective among customers that have paid relatively little in the past. The point estimates for the 1st and 2nd quartile for the *self-concept* treatment are 22.62 N$ (p=0.546) and 43.31 N$ (p=0.041) and therefore larger than the corresponding estimate for the full sample. They are, however, statistically insignificantly different from zero, which may be due to a problem of statistical power. The point estimates for the *plan* treatment are 83.36 N$ (p=0.022) and 57.65 N$ (p=0.041) respectively and therefore also considerably larger than for the full sample. The point estimates for the 3rd and 4th quartile are close to zero and statistically insignificant in both treatments. Appendix A4 and A5 shows the same analysis using the pre-intervention debt quartiles instead of the pre-intervention payment ratio with similar results.

The finding that customers with little payments in the past react strongly to our interventions resonates well with our behavioral diagnosis because these customers showed the largest gap between stated values and actual payment behavior.

[11] 15 N$ is equivalent to about 1.11 USD at the time of the intervention.

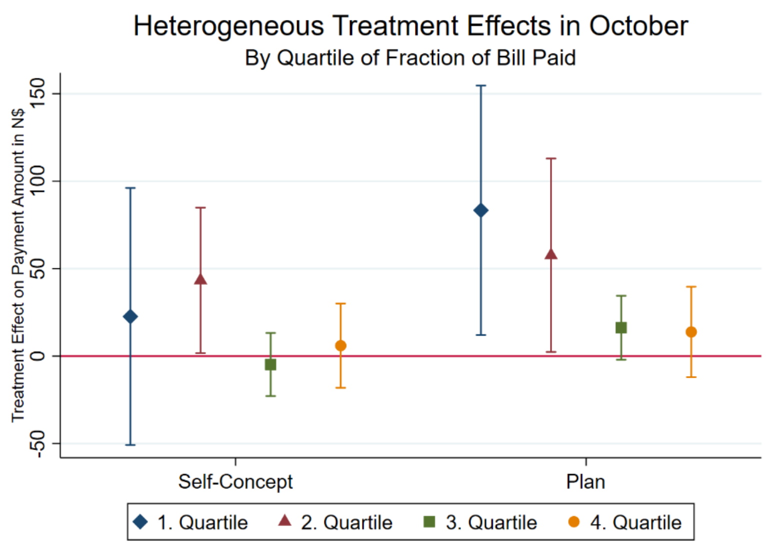


Figure 1: Treatment Effects by Pre-intervention Fraction of Bill Paid

A potential concern of the *self-concept* treatment was that customers with an unfavorable payment history might adopt (or reaffirm) a self-concept of being a non-paying customer (see Bryan, Adams and Monin, 2013). This could happen if customers are uncertain about their types and infer it through their actions (Bénabou and Tirole, 2006; 2011). Our treatment may therefore make salient to them that they do not care about water payments since they have not paid in the past. The treatment could then induce a corresponding self-concept of an unreliable payer, making payments less likely in the future. We find, however, that evoking a desirable self-concept works even for customers that, given their payment history, have little reason to think of themselves as responsible water-payers. Evidence for the effectiveness among such groups is to the best of our knowledge missing. For customers in our *self-concept* treatment it seems to be more important who they want to be than who they are.

This distinguishes the effects of the *self-concept* intervention from priming interventions, which make parts of one's identity salient and therefore are most effective among individuals that already hold a certain identity. For example, priming the identity as a previous donor is more effective for more regular donors (Kessler and Milkman, 2016) and crime-related primes increase dishonest behavior among criminals, but not regular citizens (Cohn, Maréchal and Noll, 2015).

***Result 3**: No backfiring of the commitment treatments in the long term.*

Long-term effects of the intervention are show in Table 7. We use the same specifications and regressions as in previous tables and pool data for the 1st, 2nd and 3rd quarter in the intervention period. The left panel shows effects for the *self-concept* treatment, the right panel shows treatment effects for the *plan* treatment. In particular, we are interested whether rebound effects might offset the short-term effects. We find that both commitment interventions are particularly effective in the first quarter (Oct.-Dec.) of the intervention. The 2nd quarter (Jan.-March) shows the smallest treatment effects overall. The negative, though statistically insignificant point estimate is driven by negative treatment effects on the intensive margin of about -4.2 to -4.5 percent (marginally significant in the *plan* treatment). The payment propensity is unchanged. In the 3rd quarter (April-June) of the intervention, we see a positive effect for the *self-concept* treatment on the likelihood of making a payment of 1.5 percentage points (p=0.036). The point estimate of the two-part model has the same size as in the first quarter (p=0.096). We do not observe such long-term effects for the *plan* treatment.

Table 7. ITT Long Term Effects

	Self-Concept					Plan				
	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$	(5) IHS water usage in N$	(6) Payment propensity (binary)	(7) Log (Payment amount | >0)	(8) Combined Effect on payment amount	(9) IHS Payment amount in N$	(10) IHS water usage in N$
Quarter 1	0.025***	-0.024	5.887	0.128***	0.025	0.037***	-0.021	10.868***	0.206***	-0.011
(Oct-Dec)	(0.007)	(0.023)	(3.992)	(0.041)	(0.034)	(0.007)	(0.024)	(4.159)	(0.041)	(0.034)
Quarter 2	0.005	-0.042	-2.881	0.061	-0.003	0.006	-0.045*	-3.445	0.014	-0.020
(Jan-Mar)	(0.007)	(0.026)	(3.258)	(0.048)	(0.037)	(0.007)	(0.027)	(3.183)	(0.048)	(0.038)
Quarter 3	0.015**	0.013	5.968*	0.078*	-0.029	0.007	-0.002	1.811	0.032	-0.015
(Apr-June)	(0.007)	(0.028)	(3.590)	(0.041)	(0.034)	(0.007)	(0.028)	(3.435)	(0.041)	(0.035)
Model	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff

Notes: Table 7 reports ITT effects on payment behavior (regressions 1-4, 6-9) and water consumption (regression 5, 10) by quarter. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports the ITT effect on the intensive margin. Regression 3 shows the estimated combined effect from regressions 1 and 2 using a two-part model (Belotti et al. 2015). Fitted values from the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. Regression 4 and 5 show the estimates from a diff-in-diff regression including time and individual fixed effects. The outcome variables are transformed using the inverse hyperbolic sine (IHS). Regressions adjust for the randomization method as suggested by Bruhn and McKenzie (2009): We include pre-treatment means used for randomization as well as geographical location (proxied by pipeline connection), phone caller fixed effects and date of phone call fixed effects. Control variables are top coded at the 99th percentile. All reported standard errors are clustered at the customer level to account for serial correlation. * p<0.1; ** p<0.05; *** p<0.01

What could explain the negative point estimates on the intensive margin in the 2nd quarter (regression 2 and 7)? We discuss three possible explanations: First, the high treatment effect in October might lead to budgeting effects in later periods as customers now have less money in their household budget to pay for utility bills. Secondly, some customers might have reduced their debt by paying in October, so that the remaining debt is smaller, causing smaller payments. Lastly, customers might show psychological licensing effects, which induces them to make lower payments in subsequent periods.

The last explanation, licensing effects, seems unlikely to be a driver since they would not only show up in reduced payments but also in a lower propensity to pay. The reason is that making a payment is quite effortful, as customer travel, on average, 25 km to their next payment point. A customer who feels licensed to non-payment would most probably not go at all instead of going and paying less.

To shed more light on the other two explanations, budgeting and reduced debt, we run a regression that controls for (lagged) debt and consumption, which is shown in Table A6 in the appendix. These variables are not included in our main regression since it would distort the overall impact evaluation of the interventions (debt is endogenously affected by our treatments). Controlling for debt and invoice amount yields higher treatment effects throughout all quarters and in particular in the 3rd quarter. This suggests that reduced debt is one of the channels behind waning treatment effects. As the point estimates on the intensive margin in the 2nd quarter are still negative, albeit insignificant, when controlling for debt, both budgeting and reduced debt may explain reduced payments in the 2nd quarter (for similar results see Szabó and Ujhelyi, 2015). Yet, an important observation is that this effect is found only temporarily, as in the 3rd quarter the point estimates on the intensive margin improve by about 5.3 percent in the *self-concept* and 4.3 percent in the *plan* treatment in comparison to the 2nd quarter. The improved payment behavior in the third quarter shows that our intervention did not backfire in the long-term.

Result 4: *Water consumption is unchanged*

Learning about potential spillover effects across domains of behavior is important because the literature on spillovers of behavioral interventions is scarce. One the one hand, customers might increase water consumption due to licensing effects (Merritt, Effron, Monin, 2010). On the other hand, customers could become more aware of the cost of water and thus use water more carefully. From the water utility's cost-profit perspective, it is not straightforward how marginal changes in water consumption affect profits. The marginal cost of producing water is usually low since the main costs of production stem from fixed costs. Hence, large reductions in water usage could hurt the profits of the water utility, because it reduces the billable amount billed unproportionally to production costs. The point estimates for water consumption in October (Table 6, regression 5) are slightly increased, yet insignificantly different from zero (p=0.288 and p=0.743). In the long-run they are unchanged (Table 7, regressions 5 and 10). We conclude that there are no spillovers on water consumption.

The basic treatment - This section estimates the effect of the *basic* treatment in comparison to an *untreated* comparison group. Since the effect of the *basic* treatment was not the main focus of the study, the *untreated* group was not experimentally varied (which allowed higher number of observations and thus higher statistical power for our commitment treatments). Instead, we estimate the treatment effect of the *basic* treatment in comparison to an *untreated* group using matching methods. Note that take-up of the treatment among the *untreated* group is impossible. Therefore, typical concerns with matching methods, for instance that individuals self-select into treatments, do not apply. Furthermore, our panel data allows us to construct groups with equal pre-intervention levels and parallel pre-intervention trends. Therefore, the main concerns about quasi-experimental methods (compared to randomization) should be alleviated when estimating the effects of the *basic* treatment.

We match customers in the *basic* treatment with customers that we never had contact with.[12] We provide estimates from two different matching methods in order to show insensitivity with respect to the matching procedure. As a first matching method, we use entropy balancing (Hainmueller, 2012). Entropy balancing is a procedure that reweights sample units such that the reweighted control group is balanced between treatment and control group with respect to a set of predefined covariates. Entropy balancing ensures high covariate balance even for larger sets of covariates and allows matching beyond the first moment of a variable. As a second approach we use coarsened exact matching (Iacus, King and Porro, 2011). Coarsened exact matching (CEM) creates sets of strata for pre-defined covariates and then finds matches in control and treatment group that share these strata. The advantage of both methods over the more popular propensity score matching is that they guarantee a reduction in covariate imbalance (King and Nielsen, 2018).

We test for differences in levels and trends between our matched comparison groups and the *basic* treatment with respect to payment amount, payment propensity, debt and water

[12] Summary statistics of that group of customers are provided in Appendix A7.

consumption to ensure successful matching. Appendix Table A8 shows that for both matching groups there do not exist statistically significant differences with respect to level or trends over time between the *basic* and the matched *untreated* group.

Table 8 shows the treatment effects using the regression models as in previous tables. Estimates of the entropy match are shown in panel A and estimates using the CEM are displayed in panel B. The treatment effects on payments are largest in October where the payments amount increase by about 38.22 N$ and 44.46 N$ in the basic treatment (regressions 3 and 8), which corresponds to an increase of 30% (Entropy balancing) and 31% (CEM) in comparison to the respective untreated payment mean.[13]

This effect is quite large and corresponds to about 2-3 hourly wages. This means that our two commitment treatment lead to a short term increase on payments of 40% and 56% respectively in comparison to an untreated group. Water consumption is unchanged (regressions 5 and 10). We find that there is a strong and lasting effect on the payment propensity throughout all quarters. Similar to the analysis of the two commitment treatments, we find a negative effect in the basic *treatment* on the intensive margin overall and especially in the second quarter of the intervention, which we interpret as budgeting effects for the same reasons as in the previous analysis.

[13] The weighted mean of the payment amount in October for the entropy balancing sample is 149.12 N$ and for the CEM sample is 123.01 N$. The means differ since CEM matching results in less matched observations than entropy matching.

Table 8. Treatment Effects in Comparison to Untreated Group

	A. Entropy Balancing					B. Coarsened Exact Matching				
	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$	(5) IHS water usage in N$	(6) Payment propensity (binary)	(7) Log (Payment amount | >0)	(8) Combined Effect on payment amount	(9) IHS Payment amount in N$	(10) IHS water usage in N$
October	0.093***	0.033	44.551***	0.292***	0.026	0.089***	0.072*	38.218***	0.541***	0.030
	(0.012)	(0.038)	(8.622)	(0.045)	(0.033)	(0.014)	(0.043)	(7.335)	(0.079)	(0.046)
Quarter 1	0.089***	-0.011	32.907***	0.387***	0.027	0.084***	-0.003	25.950***	0.484***	0.010
(Oct-Dec)	(0.007)	(0.029)	(4.501)	(0.034)	(0.031)	(0.008)	(0.032)	(4.073)	(0.043)	(0.037)
Quarter 2	0.043***	-0.112***	0.287	0.150***	0.035	0.040***	-0.153***	-4.397	0.164***	0.011
(Jan-March)	(0.007)	(0.029)	(3.945)	(0.036)	(0.029)	(0.008)	(0.032)	(3.726)	(0.042)	(0.036)
Quarter 3	0.040***	-0.034	9.227**	0.159***	0.045	0.037***	-0.066*	4.053	0.175***	0.022
(April-June)	(0.007)	(0.031)	(3.903)	(0.036)	(0.029)	(0.009)	(0.035)	(4.016)	(0.044)	(0.037)
Model	Weighted probit	Weighted OLS	Weighted Two-part model	Weighted Diff-in-Diff	Weighted Diff-in-Diff	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff

Notes: Both matching procedures use the following pre-intervention variables to match on, which closely resemble those used for randomization. In addition, we include variables to account for time trends in the payment data. In particular we use: Bi-annual payment propensity, bi-annual payment amount (if payment amount >0), debt in month before intervention, age of account, fraction of bill paid, inverse hyperbolic sine of water consumption, Inactivity (no water consumption) in month prior to intervention, total months of inactivity in pre-intervention year. Entropy uses a weighted sample of untreated customers to match the treated customers, which results in a match of all 3287 basic treatment customers with 8912 weighted untreated customers. CEM uses one-to-one matching with 5 bins for each of matching variables, which results in a sample of 2,626 basic treatment customers and 2,626 untreated control customers. * p<0.1; ** p<0.05; *** p<0.01

2.6 Conclusion

We conduct a large-scale natural field experiment in cooperation with the Namibian public water utility *NamWater* in order to reduce non-payment for water through behavioral interventions. We design our behavioral interventions according to pre-intervention diagnostic surveys. Building on the survey evidence, we design three interventions, treating 9,833 customers in total. The interventions aim to close the discrepancy between customers' stated willingness to pay and their actual payment behavior. In the *basic* treatment, we send customers simplified invoice information via text message. On top of the *basic* treatment, we test two commitment treatments, in which customers commit either to a desirable water paying self-concept or to their own payment plans. In the first month of the intervention, average payments increase by 30% in the *basic* treatment, by 40% in the *self-concept* treatment and by 56% in the *plan* treatment relative to an untreated comparison group. The effects of the commitment treatments wane over time. However, even more than six months after the start of the intervention, we find a significantly higher propensity to pay. On average, payments increased by an equivalent of nine hourly wages per customer in total. The intervention is extremely cost effective: A back-of-the-envelope calculation suggests a return-on-investment of around 900%.

We believe that a crucial success factor of our intervention is the identification of widespread behavioral bottlenecks impeding payments. This allows a broader view on reasons for non-payments beyond technical and economic aspects. In this case, a relatively simple intervention led to relatively large behavioral responses. This is even more remarkable given that we did not reduce the logistical hurdles to paying (customers travel on average 25 km to a payment point and wait on average 45 minutes in line to pay) and only treated one household member out of a median of seven, who are jointly responsible for consumption and possibly payments. Moreover, our sample also includes those customers who did not struggle with reception and understanding of the invoice and those 16% who did not have debt at the start of the intervention. Though difficult to quantify, the treatment-on-the-treated effects might therefore be substantially higher than the intention-to-treat effects reported in the paper. This underlines that tailored behavioral interventions can be a powerful policy measure, in particular in settings where monetary incentives or heavy-handed interventions are difficult or even impossible to implement.

2.7 Appendix to Chapter 2

Table A1. Pre-intervention Summary Statistics (ITT Sample)
Top Coded at 99% Percentile

	Mean	SD	P25	Median	P75	Mean diff. to Basic	p-value
Basic (N= 3,287)							
Payment in N$	100.41	239.78	0	0	100	-	-
Payment in N$ if >0	293.62	383.18	81	158	312	-	-
Consumption in N$	125.17	228.39	21.20	53.40	119.40	-	-
Number of payments	4.14	2.75	2	4	6	-	-
Debt in N$	508.85	1276.79	31.95	138.54	454.30	-	-
Account age in month	44.26	39.15	19	34	53	-	-
Self-Concept (N=3,297)							
Payment in N$	100.44	239.68	0	0	100	0.02	0.991
Payment in N$ if >0	293.63	389.24	84.50	155.40	300	0.01	0.999
Consumption in N$	123.13	224.11	26.70	53.70	115.70	-2.04	0.521
Number of payments	4.18	2.68	2	4	6	0.04	0.603
Debt in N$	466.61	1201.72	37.35	125.87	407.69	-42.23	0.180
Account age in month	45.41	39.70	19	34	55	1.150	0.247
Plan (N=3,249)							
Payment in N$	100.35	240.75	0	0	100	-0.057	0.982
Payment in N$ if >0	293.79	386.97	80	156.10	300	0.166	0.983
Consumption in N$	124.07	227.92	21.66	50.75	111.65	-1.105	0.734
Number of payments	4.15	2.70	2	4	6	0.005	0.948
Debt in N$	490.99	1317.07	31.86	127.52	417.36	-17.86	0.590
Account age in month	45.11	39.61	19	34	55	0.862	0.388

Notes: The table reports summary statistics of the pre-intervention year for the ITT sample. All continuous variables are top coded at the 99th percentile. The table provides mean, standard error, 25th percentile, 50th percentile and 75th percentile. The last two columns test for pre-treatment differences in means before the intervention using an OLS regression with treatment dummies and standard errors clustered at the customer level. We report the regression coefficients and p-values of the two commitment treatments in comparison to the *basic* treatment.

Table A2. Alternative Regression Specifications

	Self-Concept			Plan		
	(1) Log (payment amount in N$ +0.01)	(2) Log (payment amount in N$ +1)	(3) Cube Root (payment amount in N$)	(4) Log (payment amount in N$ +0.01)	(5) Log (payment amount in N$ +1)	(6) Cube Root (payment amount in N$)
October	0.263**	0.143**	0.159*	0.759***	0.411***	0.496***
	(0.116)	(0.064)	(0.083)	(0.117)	(0.065)	(0.085)
Quarter 1 (Oct-Dec)	0.218***	0.112***	0.120***	0.351***	0.181***	0.199***
	(0.067)	(0.036)	(0.046)	(0.067)	(0.037)	(0.047)
Quarter 2 (Jan-March)	0.026	0.007	-0.008	0.040	0.014	0.000
	(0.066)	(0.035)	(0.043)	(0.066)	(0.035)	(0.043)
Quarter 3 (April-June)	0.127*	0.070*	0.079*	0.052	0.029	0.020
	(0.067)	(0.036)	(0.045)	(0.068)	(0.037)	(0.046)
Model	Diff-in-Diff	Diff-in-Diff	Diff-in-Diff	Diff-in-Diff	Diff-in-Diff	Diff-in-Diff

Notes: This table provides ITT estimates of the treatments in comparison to the basic treatment. The precise effects for regressions using the logarithm can be obtained by taking exp (coefficient)-1. The same diff-in-diff framework as in the main specifications is used. * p<0.1; ** p<0.05; *** p<0.01

Table A3. Heterogeneous Treatment for October by Pre-intervention Payment Ratio

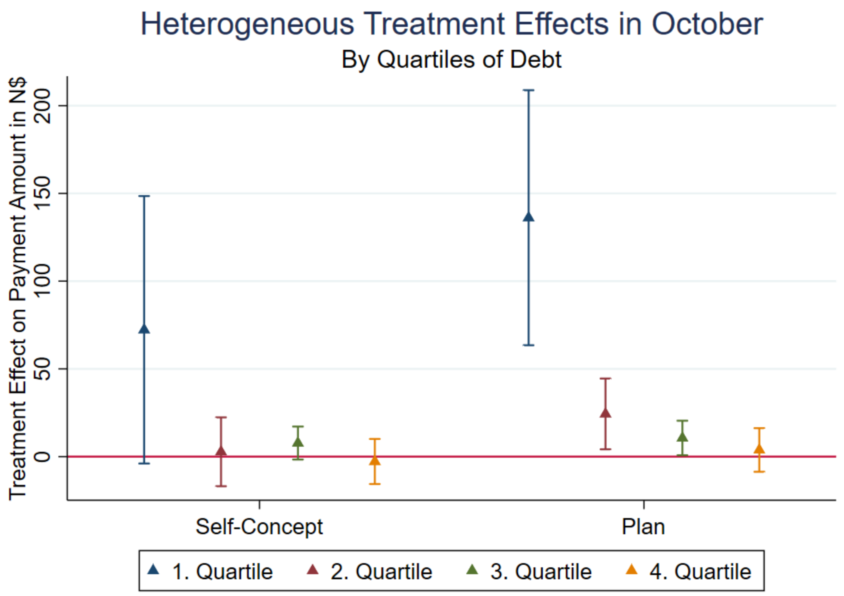
	Self-Concept					Plan				
	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$	(5) IHS water usage in N$	(6) Payment propensity (binary)	(7) Log (Payment amount | >0)	(8) Combined Effect on payment amount	(9) IHS Payment amount in N$	(10) IHS water usage in N$
1st Quartile (worst 25%)	0.036 (0.023)	-0.025 (0.081)	22.622 (37.489)	0.123 (0.105)	-0.064 (0.085)	0.048** (0.023)	0.119 (0.078)	83.363** (36.388)	0.228** (0.107)	-0.054 (0.085)
2nd Quartile	0.042* (0.023)	0.116** (0.053)	43.307** (21.215)	0.134 (0.093)	0.060 (0.069)	0.117*** (0.023)	0.051 (0.053)	57.647** (28.210)	0.324*** (0.095)	0.053 (0.070)
3rd Quartile	0.004 (0.021)	-0.037 (0.054)	-4.8266 (9.204)	0.128 (0.092)	0.152** (0.068)	0.079*** (0.021)	-0.013 (0.054)	16.2141 (9.321)	0.261*** (0.091)	0.050 (0.072)
4th Quartile (best 25%)	0.034 (0.023)	-0.027 (0.057)	5.96963 (12.294)	0.168* (0.095)	-0.007 (0.076)	0.069*** (0.024)	-0.046 (0.061)	13.803 (13.176)	0.223** (0.097)	0.006 (0.073)
Model	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff

Notes: This table reports ITT effects on payment behavior (regressions 1-4, 6-9) and water consumption (regression 5, 10) by pre-intervention fraction of bill paid quartile. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports the ITT effect on the intensive margin. Regression 3 shows the estimated combined effect from regressions 1 and 2 using a two-part model (Belotti et al., 2015). Fitted values from the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. Regression 4 and 5 show the estimates from a diff-in-diff regression including time and individual fixed effects. The outcome variables are transformed using the inverse hyperbolic sine (IHS). Regressions adjust for the randomization method as suggested by Bruhn and McKenzie (2009): We include pre-treatment means used for randomization as well as geographical location (proxied by pipeline connection), phone caller fixed effects and date of phone call fixed effects. In addition, we control for lagged debt amount in N$ and lagged water consumption in N$. Control variables are top coded at the 99th percentile. All reported standard errors are clustered at the customer level to account for serial correlation. * $p<0.1$; ** $p<0.05$; *** $p<0.01$

Table A4. Heterogeneous Treatment Effects for October by Pre-Intervention Debt

	Self-Concept					Plan				
	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$	(5) IHS water usage in N$	(6) Payment propensity (binary)	(7) Log (Payment amount | >0)	(8) Combined Effect on payment amount	(9) IHS Payment amount in N$	(10) IHS water usage in N$
1st Quartile (worst 25%)	0.033 (0.023)	0.092 (0.069)	72.277* (38.878)	0.129 (0.112)	0.004 (0.084)	0.086*** (0.023)	0.121* (0.066)	136.129*** (37.074)	0.408*** (0.114)	0.088 (0.083)
2nd Quartile	0.046** (0.023)	-0.063 (0.047)	2.751 (10.007)	0.142 (0.093)	0.085 (0.071)	0.119*** (0.022)	-0.057 (0.048)	24.335** (10.282)	0.245*** (0.094)	0.011 (0.073)
3rd Quartile	0.038* (0.022)	0.030 (0.041)	7.738 (4.796)	0.095 (0.088)	-0.034 (0.068)	0.075*** (0.023)	0.005 (0.044)	10.6169** (5.020)	0.249*** (0.087)	0.005 (0.070)
4th Quartile (best 25%)	-0.006 (0.023)	-0.021 (0.063)	-2.7748 (6.554)	0.183* (0.093)	0.069 (0.074)	0.018 (0.023)	0.008 (0.062)	3.7908 (6.342)	0.146 (0.093)	-0.055 (0.074)
Model	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff

Notes: This table reports ITT effects on payment behavior (regressions 1-4, 6-9) and water consumption (regression 5, 10) by pre-intervention debt amount. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports the ITT effect on the intensive margin. Regression 3 shows the estimated combined effect from regressions 1 and 2 using a two-part model (Belotti et al., 2015). Fitted values from the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. Regression 4 and 5 show the estimates from a diff-in-diff regression including time and individual fixed effects. The outcome variables are transformed using the inverse hyperbolic sine (IHS). Regressions adjust for the randomization method as suggested by Bruhn and McKenzie (2009): We include pre-treatment means used for randomization as well as geographical location (proxied by pipeline connection), phone caller fixed effects and date of phone call fixed effects. In addition, we control for lagged debt amount in N$ and lagged water consumption in N$. Control variables are top coded at the 99th percentile. All reported standard errors are clustered at the customer level to account for serial correlation.
* p<0.1; ** p<0.05; *** p<0.01



Appendix A5: Treatment effects with 95% confidence intervals from the two-part model in Table A4.

Table A6. ITT Effects Controlling for Lagged Debt and Lagged Consumption

	Self-Concept					Plan				
	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$	(5) IHS water usage in N$	(6) Payment propensity (binary)	(7) Log (Payment amount | >0)	(8) Combined Effect on payment amount	(9) IHS Payment amount in N$	(10) IHS water usage in N$
Quarter 1	0.025***	-0.017	8.832	0.152***	0.010	0.036***	-0.021	14.266**	0.222***	-0.018
(Oct-Dec)	(0.007)	(0.020)	(5.448)	(0.043)	(0.034)	(0.007)	(0.021)	(5.787)	(0.043)	(0.034)
Quarter 2	0.006	-0.022	-1.058	0.093*	-0.021	0.006	-0.033	-2.905	0.038	-0.031
(Jan-March)	(0.007)	(0.022)	(4.284)	(0.050)	(0.038)	(0.007)	(0.022)	(4.235)	(0.050)	(0.038)
Quarter 3	0.014**	0.036	11.716***	0.096**	-0.043	0.007	0.031	7.765*	0.048	-0.025
(April-June)	(0.007)	(0.024)	(4.371)	(0.044)	(0.035)	(0.007)	(0.023)	(4.193)	(0.044)	(0.036)
Model	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff	Probit	OLS	Two-part model	Diff-in-Diff	Diff-in-Diff

Notes: This table reports ITT effects on payment behavior (regressions 1-4, 6-9) and water consumption (regression 5, 10) controlling for lagged debt and water consumption. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports the ITT effect on the intensive margin. Regression 3 shows the estimated combined effect from regressions 1 and 2 using a two-part model (Belotti et al. 2015). Fitted values from the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. Regression 4 and 5 show the estimates from a diff-in-diff regression including time and individual fixed effects. The outcome variables are transformed using the inverse hyperbolic sine (IHS). Regressions adjust for the randomization method as suggested by Bruhn and McKenzie (2009): We include pre-treatment means used for randomization as well as geographical location (proxied by pipeline connection), phone caller fixed effects and date of phone call fixed effects. In addition, we control for lagged debt amount in N$ and lagged water consumption in N$. Control variables are top coded at the 99th percentile. All reported standard errors are clustered at the customer level to account for serial correlation. * p<0.1; ** p<0.05; *** p<0.01

Table A7. Summary Statistics For Not-contacted Sample

	Mean	SD	P25	Median	P75
Untreated (N=8912)					
Payment in N$	98.96	380.35	0	0	70
Payment in N$ if >0	309.07	622.03	82	152	302.55
Consumption in N$	128.20	302.31	20.2	50.5	111.65
Number of payments	3.77	2.79	1.5	3	6
Debt in N$	725.39	2523.84	40.06	168.92	612.4
Age of account in month	48.34	38.42	23	38	61
Fraction of bill paid	1.20	10.49	.545	.887	1.045

Notes: The table reports summary statistics of the pre-intervention year for the not-contacted sample. The table provides mean, standard error, 25th percentile, 50th percentile and 75th percentile.

Table A8. Balance Check for Matching

	Entropy Balancing					Coarsened Exact Matching				
	(1) Payment propensity (binary)	(2) Log (Payment amount in N$ | >0)	(3) IHS Payment amount (in N$)	(4) IHS Water usage (in N$)	(5) Debt in N$ (99% top coded)	(6) Payment propensity (binary)	(7) Log (Payment amount in N$ | >0)	(8) IHS Payment amount in N$	(9) IHS Water usage in N$	(10) Debt in N$ (99% top coded)
Basic	0.002	0.014	0.016	-0.008	7.408	0.004	-0.004	0.023	0.012	-4.403
(Level difference)	(0.007)	(0.032)	(0.040)	(0.033)	(28.937)	(0.008)	(0.034)	(0.044)	(0.039)	(29.954)
Basic * time trend	0.001	-0.002	0.003	0.001	0.904	0.000	0.000	0.000	0.004	0.094
(Trend difference)	(0.001)	(0.004)	(0.005)	(0.004)	(2.341)	(0.001)	(0.004)	(0.005)	(0.004)	(1.963)
Observations	132,265	41,433	132,265	132,265	132,265	59,238	20,284	59,238	59,238	59,238
R-squared	0.000	0.001	0.000	0.002	0.000	0.000	0.001	0.000	0.002	0.001

Notes: This table shows the results of the matching procedures used to estimate the effect of the basic treatment in comparison to an untreated group. If the matching procedure was successful, we should not expect differences between the basic and untreated group with respect to the pre-intervention characteristics. We test for level difference by regressing the outcome variable on the basic treatment dummy and for trend difference by interaction of the *basic* dummy with the time variable (months). Standard errors are clustered at the customer level. As can be seen in the table, both matching procedures result in coefficients close to zero without any statistically significant coefficients.

CHAPTER 3.

THE PIVOTAL ROLE OF SPECIFIC TIPS TO ENCOURAGE WATER CONSERVATION

3.1 Introduction

Two thirds of the world's population experiences severe water scarcity for at least one month a year (Mekonnen and Hoekstra, 2016). Lacking access to purified piped water has serious economic consequences on industrial and agricultural productivity, as well as on the health and wealth of individuals.[14] Finding effective interventions to curb water demand is a key challenge for the sustainable provision of piped water in water-stressed regions.

A standard economic approach would be to address the problem of water scarcity by increasing prices. This, however, has caveats: Meta-studies show that price elasticity can be low in particular in the short term (e.g., Olmstead, Hanemann and Stavins, 2007) implying that price changes need to be large in order to reduce water demand substantially.[15] Such price increases, though, may lack political acceptance for example because of unequal distributional effects among poorer households and the fact that access to clean water is a basic human right.

Non-pecuniary policies are therefore a crucial complementary strategy and are especially important when droughts are imminent or unforeseen (Dietz and Stern, 2008). A standard non-pecuniary policy are mass communication campaigns, which target customers' behavior through leaflets, letters or other mass outlets. Such campaigns are usually quicker and easier to implement than most pecuniary strategies (Katz, Grinstein, Kronrod and Nisan, 2016). Previous experimental literature on mass communication campaigns focuses on interventions reducing water consumption by appealing to social norms, inspired by a study

[14] Lacking access to purified water sources leads to waterborne diseases like diarrhea and typhoid fever (Amrita, Kremer and Zwane, 2007; Ashraf, Glaeser, Holland and Steinberg, 2017), infant mortality (Gamper-Rabindran, Khan, and Timmins, 2010) and inferior educational attainment (Ashraf et al., 2017).

[15] The low elasticity may result from the fact that purified water has limited substitutes and that water payments may only account for a small proportion of a households total expenditures. This might cause the wealthy, which often are the heaviest users, to respond little to price changes (Mansur and Olmstead, 2012; Ferraro and Miranda, 2013). Further, customers often have difficulties understanding invoice information and non-payment is common in many low- and middle-income countries (Szabo and Ujhelyi, 2015). Rode, Gómez-Baggethun and Krause (2014) argue that pricing strategies could even crowd out intrinsic motivation to act pro-environmentally.

of Schultz et al. (2007). These interventions provide information of one's own usage in comparison to other households. The treatments in such large-scale experiments reduce water consumption between 4.8 to 5.6 percent (Ferraro and Price, 2011; Ferraro and Miranda, 2013; Brent, Cook and Olsen, 2015)[16]. They are particularly effective among high users and show substantial long-term effects (Bernedo, Ferraro and Price, 2011; Ferraro, Miranda and Price, 2013).

This paper differs from previous literature by investigating two key features of such campaigns that have received little attention so far. First, I focus on the phrasing of saving tips. Although many interventions include a list of saving tips (e.g., printed on the back of the invoice), there seems to be no evidence on when these tips work and when they do not work (Lu, Deller and Hviid, 2017). In this study, I investigate whether the degree of specificity of the tips matters. Should tips be as specific as possible or should they be less specific and instead encourage individuals to contribute with their own knowledge and ideas on how to reduce water usage? There are two pre-registered and contradicting hypotheses with respect to the effectiveness of specific and less specific tips. While specific tips reduce cognitive effort by giving concretized advice on how to act, they run the risk of being inadequate to a particular household or might even cause reactance by restricting certain behaviors. Less specific tips in contrast foster involvement in the idea-generating process, encouraging individually adequate ways to save water. Moreover, prior research suggests that individuals highly value their own input to products and ideas (Norton, Mochon and Ariely, 2012; Hooshangi and Loewenstein, 2016). However, less specific tips may demand too much cognitive effort to find effective ways to conserve water.

Second, I study potential spillovers on payment behavior. Understanding these spillovers is not only interesting from a behavioral perspective, it is also important from the utility's revenue perspective as reductions in water usage mechanically lower the billable amount. Policy makers might be concerned that a reduction in billable amounts may translate into a one-to-one loss in revenues, which could threaten the cost-recovery of high fixed cost infrastructure. While this concern might be substantial in settings where 100% of the bill is paid, the effects on payment behavior are less clear in settings where many bills are unpaid or only partially paid as typical in many low- and middle-income countries (e.g., Szabo and Ujhelyi, 2015). Therefore, whether and to which degree reductions in billable amounts affect the utility's revenues is yet an unanswered question.

[16] Social norm experiments in the electricity domain are about half the size with treatment effects ranging from 2.0 to 2.1 percent (Allcott, 2011; Alcott and Rogers, 2014; Costa and Kahn, 2013, Ayres, Raseman and Shih, 2013).

To provide answers to these questions, I conduct a natural field experiment (Harrison and List, 2004) with around 15,000 customers in cooperation with the Namibian water utility (*NamWater*). In addition, I present evidence from a pre-intervention survey (N=285) measuring socio-demographics, knowledge and attitudes towards water usage to discuss the intervention and its results.

Namibia is a typical country facing recurrent water scarcity, because of low, seasonal and unpredictable rainfall (Lu et al., 2016). At the time of the intervention (August 2017), several regions were facing water scarcity due to limited rainfall in past years, which called for a swift intervention. The intervention is implemented only once via text message on customers' mobile phones and behavior observed over 6 months until the start of the rain season (as pre-registered). The treatment messages vary the degree of the specificity of saving tips. In the *specific* treatment, customers receive three specific saving tips (e.g. re-use water from cooking to water plants). In the *semi-specific* treatment, customers receive three suggestions in a specific domain (e.g. reuse water from the kitchen), but need to concretize the water saving activities otherwise. The *unspecific* treatment does not provide a specific domain, but instead encourages customers to find three ways to save water by staying alert, observing where too much water is used and by breaking current habits to save water "their own way". It therefore intends to involve customers in the idea-generating process.

This study provides several policy-relevant insights: First, the *specific* treatment is most effective and decreases water usage by 5.8 percent. The point estimate of the *semi-specific* suggests a reduction of 3.1 percent, yet it is statistically insignificantly different from zero. The *unspecific* treatment is ineffective (0.3 percent), which means that merely creating awareness and calling to take action is insufficient to achieve a reduction in water usage. The provision of specific tips is the pivotal factor of the campaign.

Second, the treatment effects are entirely driven by customers with above median usage (high users). Non-pecuniary strategies are therefore a complementary strategy to pecuniary strategies, as high users are often least price sensitive (Mansur and Olmstead, 2012). Survey evidence suggests that high users may have more "slack" to improve on than low users, rather than being different with respect to socio-demographics or conservational attitudes and knowledge.

Third, despite the large reduction in monthly billable amounts (through decreased consumption) the intervention does not significantly affect customers' payments for water. This finding can be explained by the fact that the majority of customers pays less than what they are being billed anyway. Instead, customers rather make bulk payments about four times a year and these payments are uncorrelated with the actually billed amount. If anything, the *semi-specific* and *unspecific* treatment improve the likelihood of making payments.

Fourth, the intervention is extremely simple and can easily be scaled up. That means that even policy makers with limited time or limited resources can employ this intervention because of its simplicity (plain text) and low implementation costs of 0.04 USD per customer. Despite its simplicity, the treatment effect of the *specific* treatment corresponds to the largest ones found in mass communication literature. The campaign saved about 25.6 million liters[17] within 6 months. Had every customer in the experimental sample received the *specific* treatment, water savings would have accumulated to around 50 million liters.

[17] 25.6 million liters corresponds to about 6.76 million gallons.

3.2 Experimental Sample and Data

The experimental sample contains all of *NamWater*'*s* residential customers in Northern Namibia whose accounts provided a phone number at the time of the intervention, which corresponds to 71.2% of the total number of customers in that region. The sample is split into four approximately equally sized groups and balanced via min-max t-stat method stratified by location (with 1,000 redraws) to balance on water consumption, debt and payment behavior. The invoice and payment data is heavily right skewed with some extreme outliers and erratic data. I use Grubb's outlier test (Grubbs, 1969) to exclude 288 households with a 99.9% confidence level that have unusually and extreme water usages of more than 200,000 liters in a single month. This amount approximately corresponds to the volume of a large public swimming pool and is potentially due to extreme leakages, faulty accounting data or commercial usage.[18] A robustness check in the appendix provides several alternative exclusion rules (Table A1 and A2) which do not change the results.

Table 1 provides summary statistics for the experimental sample of 14,943 customers and balance checks between the treatments and the untreated group. I provide estimates on water consumption for the untransformed value and for the inverse hyperbolic sine of water consumption since unlike the log transformation, the inverse hyperbolic sine (IHS) is defined for zero values (Burbidge, Magee and Robb, 1988; MacKinnon and Magee, 1990).[19] The inverse hyperbolic sine can be interpreted in the same way as the traditional log transformation. McKenzie (2017), for example, uses the inverse hyperbolic sine transformation for profit data of entrepreneurs in Nigeria.

The sample is well balanced. The only marginally significant difference is between *control* and *unspecific* treatment group for the age of the account. Such differences are controlled for in a diff-in-diff regression with individual and month fixed effects. The messages were sent out on two consecutive days in August 2017 in English and the local language Oshiwambo. Back and forth translations ensured that the content was identical in both languages. Around 1% of customers become inactive after the intervention, with no statistically significant differences between treatments.

[18] The control group contains more of such cases (30.2%) than any of the other treatments, which would be inconsistent with a selection bias in favor of the treatments.

[19] The inverse hyperbolic sine is defined as $\ln(y+(y^2+1)^{0.5})$.

Table 1. Baseline Summary Statistics and Test of Balance

	Mean	SD	P25	Median	P75	Mean diff. to Control	P-value
Untreated Control (N=3,701)							
Water consumption in N$	125.86	209.89	26.00	65.00	134.55	-	-
IHS water consumption in N$	4.49	1.88	3.95	4.87	5.60	-	-
Debt in N$	764.66	3192.38	64.20	246.59	800.45	-	-
Payment amount in N$	134.80	534.92	0.00	0.00	100.00	-	-
Payment propensity (binary)	0.32	0.47	0.00	0.00	1.00	-	-
Age of account in months	44.97	20.54	29.00	42.00	65.00	-	-
Specific tips (N=3,746)							
Water consumption in N$	124.87	208.72	26.00	65.00	130.50	-0.99	0.730
IHS water consumption in N$	4.49	1.88	3.95	4.87	5.56	-0.01	0.722
Debt in N$	811.35	2304.29	64.69	243.00	789.90	46.69	0.433
Payment amount in N$	139.60	442.98	0.00	0.00	100.00	4.80	0.723
Payment propensity (binary)	0.32	0.47	0.00	0.00	1.00	0.00	0.818
Age of account in months	45.24	20.33	30.00	43.00	64.00	0.27	0.633
Semi-specific tips (N=3,729)							
Water consumption in N$	125.71	208.55	26.10	65.00	130.00	-0.15	0.958
IHS water consumption in N$	4.51	1.85	3.96	4.87	5.56	0.02	0.479
Debt in N$	816.37	2164.34	65.00	236.54	767.36	51.72	0.381
Payment amount in N$	120.36	422.77	0.00	0.00	100.00	-14.44	0.279
Payment propensity (binary)	0.33	0.47	0.00	0.00	0.33	0.00	0.613
Age of account in months	44.99	20.88	28.00	42.00	66.00	0.02	0.979
Unspecified tips (N=3,767)							
Water consumption in N$	128.44	218.35	26.10	65.00	132.50	2.58	0.380
IHS water consumption in N$	4.51	1.87	3.96	4.87	5.58	0.020	0.445
Debt in N$	786.27	2105.45	64.72	241.69	777.62	21.62	0.710
Payment amount in N$	119.55	525.23	0.00	0.00	100.00	-15.25	0.295
Payment propensity (binary)	0.33	0.47	0.00	0.00	1.00	0.00	0.607
Age of account in months	45.93	20.40	30.00	43.00	66.00	0.96	0.088

Notes: The table reports pre-intervention summary statistics of the for the ITT sample. The table provides mean, standard error, 25th percentile, median and 75th percentile. Average price at the time of the intervention is 13.15 N$ (ca. 1 USD) for 1000 liters. The last two columns test for pre-treatment differences in means before the intervention using a pooled OLS regression with treatment dummies and standard errors clustered at the customer level. I report the regression coefficients and p-values of the treatments in comparison to the basic treatment.

3.3 Treatments

Customers randomly either receive specific, semi-specific or unspecific water saving tips on their cell phone. All messages are personalized, inform about the current water scarcity and the necessity to use less water. Table 2 shows the exact content of the three types of messages.

Table 2. Content of Treatment Messages

Specific	Dear [Name], Our nation is low on water supply. You need to start acting today. Please use less water. Please consider the following tips: Shorten the time you spend in the shower by a minute or two. Re-use water from cooking and cleaning food to water plants. Use a broom instead of water to clean floors in and around your house.
Semi-Specific	Dear [Name], Our nation is low on water supply. You need to start acting today. Please use less water. Please consider the following ideas: When can you keep the water tap closed and avoid running water (for example in the bathroom)? How can you re-use water in your household (for example from the kitchen)? How can you avoid unnecessary water usage (for example outside the house)?
Unspecific	Dear [Name], Our nation is low on water supply. You need to start acting today. Please use less water. Please take a couple of minutes to find three ways to use less water in your household. Be alert and observe where you use too much water. What can you personally change to use less water and to break your habits? Save water and do it your own way!

The *specific* treatment contains three concretized ways to save water. The saving tips are selected systematically by targeting the three most commonly self-reported main usages of water from a pre-intervention survey: Cleaning, bathing and cooking (see Table 6 for survey data) and thereby targets behavior that is most salient to the customer. Further, the intervention focuses on tips that are mentioned by few (around 5.9%) of customers in the pre-intervention surveys (see Table 5 for details) to reduce the risk that customers receive tips that they are already aware of.

The *semi-specific* treatment partially involves customers in the idea-generating process. The treatment provides the same specific domain as the *specific* treatment, but the actual activity to reduce water needs to be concretized by the customer.

The *unspecific* treatment intends to involve customers in the idea-generating process. The treatment encourages customers to find three ways to save water in the household without providing a specific domain. The message gives general instructions on how to do so: Staying alert, observing where too much water is used, reconsidering existing habits and saving water "your own way!"

3.4 Hypotheses

There are two pre-registered and contradicting hypotheses with respect to the effectiveness of the treatments.

Hypothesis 1: Specificity of tips increases the effectiveness of the intervention.

Steg (2016) argues that complexity and limited cognitive resources reduce the ability and motivation to act on pro-environmental behavior. Specific water saving tips help to overcome this barrier by providing straightforward suggestions. Further, specific tips may make the desired behavior appear easier and therefore more likely to be acted upon (Ajzen, 1991). Literature on implementation intentions suggests that specifying an implementation plan increases the likelihood of acting as it develops strategies to overcome logistical obstacles (Gollwitzer and Sheeran, 2006). Further, reducing complexity are among the most effective "nudges" (Thaler and Sunstein, 2008). This strand of literature therefore suggests that specificity reduces the cognitive effort to find and act upon effective ways to save water.

Hypothesis 2: Specificity of tips decreases the effectiveness of the intervention.

Customers have better knowledge of their household characteristics and current water saving activities. Too specific tips might propose inapplicable actions (e.g. "don't wash your car with a water hose" when one does not own a car) or behaviors that are already acted upon. Moreover, a strand of literature argues that individuals overvalue their own input to ideas and products. Hooshangi and Loewenstein (2016) find that individuals are overconfident about the value of, and overly likely to invest in their own entrepreneurial idea. On the other hand, when investing in another person's idea, they are underconfident about the value of, and insufficiently likely to invest in it. Along these lines, the "IKEA-effect" (Norton, Mochon and Ariely, 2012) describes the phenomenon that individuals have a higher valuation of self-made products. Franke et al., (2010) show that involving customers in the specification and design process of products increases their willingness-to-pay for these products. Further, involvement of individuals in pro-environmental campaigns improves the acceptability of pro-environmental strategies (Steg, 2016). Attempts to restrict specific behaviors could also lead to (psychological) reactance (Brehm, 1966; Kronrod, Grinstein and Wathieu, 2012).

3.5 Estimation Strategy and Results

3.5.1 Water Consumption

Table 3 provides difference-in-difference regressions with month and customer fixed effects and clustered standard errors on the customer level (Betrand, Duflo and Mullainathan, 2004). I use the following difference-in-difference estimator:

$$usage_{it} = \beta \cdot treatment_i \cdot post + month\ fixed\ effect_t + customer\ fixed\ effect_i + error_{it}$$

The *post* dummy variable indicates the six experimental months from August 2017 until January 2018, the pre-registered end date.[20] Regression 1 and 2 in Table 3 show the intention-to-treat effects (ITT) on water usage among the full sample. Regression 3 and 4 show the ITT effects among the below median water consumers (low-users). Regression 5 and 6 show the ITT effects among the above median consumers (high users). Note, that the treatment-on-the-treated (ToT) effects are larger, as not all text messages were successfully delivered. An upper bound for the compliance rate is 91% as the remaining text messages were sent to inactive numbers. Other reasons why customers may have remained untreated are for example network errors or full memory on a customer's phone. Thus, a lower bound for the ToT effects can be estimated by dividing the ITT effects reported below by the compliance rate.

Result 1: *The specific treatment decreases water consumption by 5.8 percent. Unspecific tips are ineffective.*

The specific treatment reduces water consumption by about 5.8 percent (p=0.045) per customer per month in comparison to the untreated group. The effect on the untransformed water usage is 7.86 N$ (p=0.029) which translates into a reduction of about 600 liters per month per customer. This amount corresponds for example to a reduction of showering time by a minute per day. The point estimate of the *semi-specific* suggests a reduction of 3.1 percent,

[20] The pre-registered end date is the 31.12.2017, which falls into the *NamWaters* billing period of January (The accounting month of December ends in mid-December due to holidays). The end date was chosen since the first substantial rainfalls arrive around January and thereby alleviate the water scarcity. In 2018, the heavy rain started in mid-January. In appendix table A3, I provide the same regressions using the end of the rain season (March 2018) as the end date (which is when available water resources reach their peak). The results do not change.

yet is statistically insignificantly different from zero (*p*=0.269). The *unspecific* treatment has a point estimate of 0.3 percent (*p*=0.931) and is therefore ineffective.[21]

Result 2: *The treatment effect is driven by high-users.*

Regressions 3 and 4 show that for low users all coefficients are statistically insignificantly different from zero. Regressions 5 and 6 show large effects for high users in the *specific* treatment. The *specific* treatment reduces water consumption by 12.7 percent (*p*=0.004). The *semi-specific* treatment reduces water consumption among the high users by 6.9 percent (*p*=0.107). The point estimate of the *unspecific* treatment suggests a reduction of 2.1 percent, yet is statistically insignificantly different from zero.

Table 3. ITT Effects on Water Usage

	Full Sample		Low Users		High Users	
	(1) IHS Water usage in N$	(2) Water usage in N$	(3) IHS Water usage in N$	(4) Water usage in N$	(5) IHS Water usage in N$	(6) Water usage in N$
Specific	-0.058**	-7.862**	0.032	-2.643	-0.127***	-11.253*
	(0.029)	(3.607)	(0.035)	(2.180)	(0.044)	(6.773)
Semi-Specific	-0.031	-3.483	0.006	-0.723	-0.069	-6.170
	(0.028)	(3.526)	(0.036)	(2.307)	(0.043)	(6.661)
Unspecific	0.003	-3.668	0.030	-1.431	-0.021	-5.527
	(0.029)	(3.625)	(0.036)	(2.289)	(0.044)	(6.867)
Constant	4.139***	115.490***	3.419***	32.956***	4.867***	198.769***
	(0.016)	(1.786)	(0.020)	(0.612)	(0.024)	(3.545)
Observations	146,878	146,878	73,647	73,647	73,231	73,231
R-squared	0.045	0.030	0.083	0.097	0.032	0.030
N	14,943	14,943	7,470	7,470	7,473	7,473

Notes: Regressions show the ITT estimates from a diff-in-diff regression including time and individual fixed effects on water consumption. The outcome variables in regression 1, 3 and 5 are transformed using the inverse hyperbolic sine (IHS) which can be interpreted as the log transformation. Regressions 2, 4 and 6 show the effects on the untransformed value. The average price of water is 13.5 N$ for 1000 liters. Standard errors (in parentheses) are clustered on the customer level. * p<0.1; ** p<0.05; *** p<0.01

3.5.2 Payment Behavior

Table 4 shows the treatment effects on payment behavior. Regression 7 shows the marginal effects at means on the extensive margin (probability of making a payment) of a probit regression. Regression 8 shows the effects on the intensive margin using OLS (the effect on the natural logarithm of the payment amount conditional on being larger than zero).

[21] The *specific* treatment is also significantly more effective than the *unspecific* treatment (table A4 in appendix).

Regression 9 estimates the combined effects of regression 7 and 8 using a two-part model which multiplies the estimated effects from regression 7 and 8 (Belotti et al., 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment values of strata used for randomization in the regression as recommended by Bruhn and McKenzie (2009). All standard errors are clustered at the customer level. Regression 10-12 show the estimates for low users and regressions 13-15 show estimates for high users.

Result 3: *The campaign does not harm payment behavior.*

Regression 7 shows that the *semi-specific* treatment increases the likelihood of making a payment by 1 percentage point (p=0.043). The *unspecific* treatment increased the probability of making a payment by 1.2 percentage points (p=0.018). On the intensive margin (conditional on paying), the treatments have negative point estimates (marginally significant for the *semi-specific* treatment). None of these effects however accumulates to an economically meaningful increase or decrease in payment behavior: The estimates for the combined effect (regression 3) range from to -0.41 N$ to 2.87 N$[22] and are statistically insignificant from zero. Interestingly, among low users, the intensive margin is significantly smaller (5.5 to 5.7 percent) although the billable amount among this group is unchanged. The *unspecific* treatment shows a positive treatment effect among the high users of 10.73 N$ (p=0.027). I find no effects on extensive or intensive margin of paying for the high users in the *specific* treatment, whose monthly billable amount was reduced by about 12.7 percent.

[22] At the time of the intervention, 1 USD was worth about 13 N$.

Table 4. Effects on Payment Behavior

	Full Sample			Low Users			High users		
	(7) Payment propensity (binary)	(8) Log Payment amount in N$ (if >0)	(9) Combined Effect	(10) Payment propensity (binary)	(11) Log Payment amount in N$ (if >0)	(12) Combined Effect	(13) Payment propensity (binary)	(14) Log Payment amount in N$ (if >0)	(15) Combined Effect
Specific	0.005	-0.011	0.447	0.002	-0.031	-1.562	0.008	0.000	4.219
	(0.005)	(0.018)	(2.487)	(0.006)	(0.023)	(1.667)	(0.007)	(0.025)	(4.965)
Semi-Specific	0.010**	-0.035*	-0.409	0.009	-0.057**	-1.499	0.010	-0.021	1.785
	(0.005)	(0.018)	(2.449)	(0.006)	(0.023)	(1.777)	(0.007)	(0.025)	(5.007)
Unspecific	0.012**	-0.015	2.871	0.012*	-0.055**	-0.759	0.011	0.024	10.733**
	(0.005)	(0.018)	(2.452)	(0.006)	(0.023)	(1.728)	(0.007)	(0.025)	(4.847)
Observations	89,228	30,089	89,228	45,333	13,598	45,333	43,895	16,491	43,895
R-squared	0.0989	0.369	-	0.0881	0.316	-	0.1015	0.308	-

Notes: This table reports ITT effects on payment behavior (regressions 7-9) for the six months post intervention for the full sample. Regression 7 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 8 reports the ITT effect on the intensive margin. Regression 9 multiplies the effects of regressions 7 and 8 to get an estimate of the combined effect using a two-part model (Belotti et al. 2015). Regressions 10-12 provide equivalent estimations for low users and regressions 14-15 for high users. Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. Regression control for strata used for randomization as recommended by Bruhn and McKenzie (2009). Control variables are top coded at the 99th percentile. Four customers attrite in August. All reported standard errors are clustered at the customer level to account for serial correlation.* p<0.1; ** p<0.05

3.6 Discussion of Results

Result 1 shows that the *specific* treatment is most effective. As discussed in the hypotheses section, the main reason why specific tips might be more effective is that they reduce cognitive effort, which is one of the key barriers to act pro-environmentally when cognitive resources are limited (Steg, 2016). The *specific* treatment provides three concrete tips in plain language that are tailored to the specific context and thereby reduce the cognitive effort costs of finding effective ways to save water.

Finding effective ways to save water is not easy for customers in the sample. Table 5 shows the categorized answers from the pre-intervention survey on whether a customer plans to reduce water and if so, to name concrete steps.

Table 5. Plans to reduce water and concrete ideas from pre-intervention survey

	All	Low Users	High Users	P-value
Plans to reduce consumption in the future	0.703	0.732	0.682	0.371
	(0.458)	(0.444)	(0.467)	
i) Has suggestion on how to reduce water use	0.709	0.671	0.738	0.311
	(0.455)	(0.473)	(0.442)	
a) Store water in containers and lock tap[a]	0.314	0.305	0.321	0.782
	(0.465)	(0.463)	(0.469)	
b) Tell others to save water[b]	0.165	0.134	0.189	0.742
	(0.372)	(0.343)	(0.393)	
c) Use alternative water sources	0.106	0.134	0.085	0.169
	(0.309)	(0.343)	(0.280)	
d) Tip included in specific treatment	0.059	0.049	0.066	0.473
	(0.235)	(0.217)	(0.250)	
e) Other ideas	0.170	0.122	0.208	0.197
	(0.377)	(0.329)	(0.407)	
ii) Doesn't know how to reduce water use	0.291	0.329	0.262	0.311
	(0.455)	(0.473)	(0.442)	
f) "Don't know"	0.296	0.333	0.259	0.551
	(0.461)	(0.480)	(0.447)	
g) Vague answer (e.g. "use water wisely")	0.704	0.667	0.741	0.551
	(0.461)	(0.480)	(0.447)	

Notes: This table shows means and standard errors in parentheses. Multiple answers are possible for a) - e).
[a] This water saving strategy is meant to avoid running water by storing water in containers, which allows them to visibly monitor their consumption. [b] Others refers to people in the household, kids or neighbors . The last column shows the p-value of a two-sided Chi-square test testing for a difference between low and high users.

The first row shows the percentage of interviewees planning to reduce water consumption for all and for low and high users. Row *i)* indicates the fraction of customers proposing a way to save water conditional on planning to reduce water consumption and row *ii)* displays the remaining fraction of interviewees that is unable to propose a way to save water. Rows *a) - e)* show the fractions of answer types conditional on planning to reduce water consumption. Rows *f)* and *g)* display fractions of answer types among those that are unable to propose a way. The last column provides p-values of testing for differences between high and low users.

Intentions to reduce water consumptions are high since around 70.3% plan to reduce water consumption. Yet, around 29.6% of users are unable to mention any concrete steps to save water or make vague statements like "use water wisely". While the other interviewees make suggestions, it is unclear to which degree these steps are effective to reduce water consumption. The fact that the *unspecific* treatment is completely ineffective, casts doubt on their effectiveness. Table 5 also shows that the tips used in the *specific* treatment are not commonly mentioned, as only 5.9% of interviewees propose ways that are similar to those in the treatment. Overall, I do not find any statistically significant differences between low and high users for the measured variables.

If customers indeed find it difficult to come up with effective ways to save water, the fact that the *unspecific* and *semi-specific* treatment show an increase in the likelihood of making payments could be interpreted as a response to their struggle. Customers in the *semi-specific* and *unspecific* treatment may then make these additional payments to fight water scarcity in their "own way", while customers in the specific treatment might feel that they already do their part by using less water. Such behaviors in which individuals try to compensate their past shortcomings for example by acting more prosocially are known as moral cleansing (e.g., Jordan, Mullen and Murnighan, 2011).

Result 2 shows that the treatment effect is driven by high users. Such heterogeneous effect have also been documented for social norm interventions and real-time smart-meter feedback.[23] Tiefenbeck et al. (2018) interpret this as higher levels of "slack" among high users, meaning that high user might just be more wasteful with water and therefore might have more room to improve on. However, it could also be that high users are structurally different with respect to their demographics or attitudes toward water conservation. To provide suggestive answers on this question, Table 6 provides demographics for low and high users from the pre-intervention survey. The first five rows show the fractions of mentioned main purposes of water usage. The table also displays whether interviewees have a job, their gender, household earnings, people living in the household and availability of alternative water sources (besides piped water from *NamWater*). Lastly, the table reports whether the customer owns plants for gardening or animals. The table also reports, conditional on owning animals, the absolute

[23] Heterogeneous effects in social norm interventions have been found by Allcott (2011), Ayres, Raseman and Shih (2013), Ferraro and Price, (2011) and Brent, Cook, and Olsen (2015).

number and the types of animals. The last column provides p-values of a two-sided Chi-square test testing for a difference between low and high users

Table 6. Structural differences between high and low users

	All	Low Users	High Users	P-Value
Main purpose of water usage				
Cleaning	0.712	0.694	0.726	0.563
	(0.453)	(0.463)	(0.448)	
Bathing	0.488	0.504	0.476	0.634
	(0.501)	(0.502)	(0.501)	
Cooking	0.393	0.413	0.378	0.548
	(0.489)	(0.494)	(0.486)	
Gardening	0.249	0.231	0.262	0.552
	(0.433)	(0.423)	(0.441)	
Animals	0.253	0.215	0.280	0.208
	(0.435)	(0.412)	(0.451)	
Has a job	0.484	0.438	0.518	0.180
	(0.501)	(0.498)	(0.501)	
Female	0.505	0.561	0.466	0.118
	(0.501)	(0.498)	(0.500)	
Household earnings in N$	4312.6	3412	4965.20 [b]	0.328
	(6465.6)	(4893.10)	(7348.10)	
People living in household	7.449	7.009	7.761	0.191
	(4.346)	(3.974)	(4.579)	
Has alternative water sources	0.376	0.383	0.371	0.846
	(0.485)	(0.488)	(0.485)	
Owns plants	0.263	0.248	0.274	0.616
	(0.441)	(0.434)	(0.448)	
Owns animals	0.211	0.190	0.226	0.467
	(0.408)	(0.394)	(0.419)	
Number total animals	7.968	4.711	10.37	0.659
	(52.58)	(17.55)	(67.65)	
Number of goats	5.895	3.215	7.872 [a]	0.675
	(48.39)	(11.88)	(62.99)	
Number of cattle	1.811	1.248	2.226	0.629
	(8.098)	(6.281)	(9.209)	
Number of donkeys	0.263	0.248	0.274	0.985
	(1.252)	(1.135)	(1.335)	
N	285	121	164	

Notes: The table shows means and standard errors in parentheses. Multiple answers are possible for main usage. [a] Includes a customer with 800 goats (mean when dropping this customer is 5.662). [b] Includes a customer who reported a household income of 280 000 N$ (mean when dropping this customer is 4665.46 N$). Binary variables are tested using a two-sided Chi-square test. Other variables are tested using a two-sided Wilcoxon rank-sum test.

For none of the presented covariates, I find statistically significant differences between high and low users. While this does not mean that there are no differences, it suggests that the presented covariates cannot explain much of the difference in consumption. I interpret this as evidence in favor of the “slack” interpretation.

Result 3 shows that the text message campaign does not negatively affect payment behavior, although on average the billable amounts is substantially reduced. An explanation for this is that the majority of customers pays less than what they are being billed anyway (the median fraction of the bill paid is 69%). Payments correlate surprisingly little with the monthly-billed amounts. Figure 1 shows a scatter plot of payments and invoices for the whole dataset. Dots on the 45-degree line would mark customers that exactly pay their current bill. However, only 1.4% of the values lie on that line. Instead, many customers make bulk payments in multiples of 100 on average about four times a year. It therefore seems that the actually billed amount is rather irrelevant to the payment behavior of the customers.

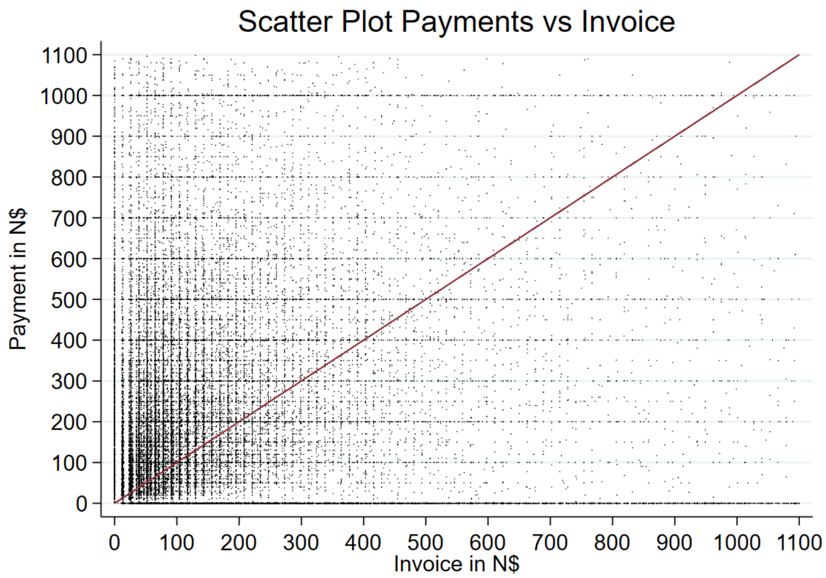


Figure 2: Payments correlate little with invoices.

Also note that a mechanical reduction in payments through lower billable amounts is inconsistent with the finding that the reduction on the intensive margin of payments in Table 4 are smallest in the *specific* treatment whose billable amount was reduced the largest. The results do also not change when extending the time horizon until June 2018 (Table A5). Since payment behavior is largely constant, the reduction in water usage in the *specific* treatment translates into an *increase* of the fraction of the bill paid of 8.3% (p=0.031; Table A6).[24] If the marginal cost of water is larger than zero, this means that the intervention might even improve profits of the utility.

[24] The fraction of the bill paid is also called the collection efficiency and is calculated as the sum of payments divided by sum of invoices for a given time period.

3.7 Conclusion

The *specific* treatment yields treatment effects that are among the largest treatment effects found in the mass communication literature and shows that specific advice on how to conserve water is a crucial factor for the effectiveness of the intervention. Since customers in the *unspecific* treatment show no reduction in water usage, the experiment shows that merely creating awareness is insufficient to reduce water consumption. This suggests that customers' lack of knowledge on how to save water effectively is a severe bottleneck hindering conservational behavior, which can be overcome by a simple intervention.

This study also proposes a systematic way to select specific saving tips to fit the local context (Duflo, 2017): The *specific* treatment targets the average consumers' most salient domains of water use and provides tips that the average customer is unaware of. The *specific* treatment thereby reduces the effort cost of finding effective ways to save water, which seems to be the main driver of the results. The difficulty to find effective ways to save water might be a particularly limiting factor among the poor whose cognitive capacity has been shown to be impeded by having less resources per se (Shah, Mullainathan and Shafir, 2012).

There are implications for policy makers beyond the insight that saving tips should be as specific as possible: The treatment effects are driven by customers with above median usage, which are typically least price-sensitive and therefore the study emphasizes the complementarity of non-pecuniary to pecuniary strategies as found in previous studies (e.g., Ferraro and Price, 2011). Targeting such customers would make non-pecuniary strategies even more cost-effective. In addition, the intervention does not negatively influence overall payment behavior, which can be explained by the fact that many customers pay less than what they are billed for anyway and that the correlation of payments with the amount billed is low. This finding alleviates potential conflicts of interest between cost-recovery and curbing water demand. If the marginal cost of providing water is larger than zero, than the intervention might even have increased profits. Whether and to which degree these results translate to settings in which payment morale is substantially higher, however, remains unclear.

This experiment also contributes to the literature beyond the mass communication campaigns. In the real-time smart-metering literature, Tiefenbeck et al. (2018) suggest that mixed results with respect to their effectiveness (see e.g., Buchanan, Russo and Anderson, 2015) could stem from the fact that providing feedback is ineffective, if the adequate behavior to reduce consumption is not sufficiently specified. This paper provides valuable apples-to-apples evidence in support of this claim.

Given that billions of people face water scarcity each year across the globe, the simplicity of the intervention is the key feature for policy makers that are often time and budget constrained. The text message campaign does not require any computational programming to provide individualized information as for example necessary in social norm interventions. Instead, it only costs around 0.04 USD per customer to implement and in return reduces water consumption on average by 5.8%, which corresponds to about 25.6 million liters saved within 6 months.

3.8 Appendix to Chapter 3

Table A1. ITT Effects with Different Cut-offs

	(1) Water Usage (in N$) Cut-off at 1000 N$	(2) Water Usage (in N$) Cut-off at 1500 N$	(3) Water Usage (in N$) Cut-off at 2000 N$	(4) Water Usage (in N$) Cut-off at 2500 N$
Specific	-6.173***	-6.723**	-7.475**	-7.549**
	(2.005)	(2.694)	(3.121)	(3.410)
Semi-specific	-4.111**	-4.599*	-5.319*	-3.980
	(2.037)	(2.707)	(3.162)	(3.420)
Unspecific	-2.071	-1.489	-2.173	-1.982
	(1.992)	(2.667)	(3.054)	(3.342)
Constant	80.256***	96.354***	105.817***	112.326***
	(0.956)	(1.313)	(1.521)	(1.700)
Observations	131,900	140,904	144,457	146,159
R-squared	0.075	0.045	0.037	0.033
Number of customer	13,379	14,337	14,697	14,869

Notes: Regressions show the ITT estimates from a diff-in-diff regression including time and individual fixed effects on water consumption. Regression 1, 2, 3 and 4 exclude customers whose highest invoice exceeded 1000N$, 1500N$, 2000N$ and 2500N$ respectively. The average price of water is 13.5 N$ for 1000 liters. All standard errors (in parentheses) are clustered on the customer level.

Table A2. ITT Effects Trimming Top 1% of Values

	Full Sample		Low Users		High Users	
	(1) IHS Water usage in N$	(2) Water usage in N$	(3) IHS Water usage in N$	(4) Water usage in N$	(5) IHS Water usage in N$	(6) Water usage in N$
Specific	-0.052*	-7.170***	0.034	-1.881	-0.124***	-11.598**
	(0.029)	(2.694)	(0.035)	(1.843)	(0.043)	(5.053)
Semi-Specific	-0.040	-5.797**	0.007	-0.094	-0.083*	-11.401**
	(0.028)	(2.749)	(0.036)	(1.955)	(0.043)	(5.167)
Unspecific	0.001	-4.618*	0.027	-1.578	-0.022	-7.519
	(0.028)	(2.687)	(0.035)	(1.887)	(0.044)	(5.074)
Constant	4.101***	101.312***	3.429***	33.153***	4.793***	171.140***
	(0.016)	(1.282)	(0.020)	(0.551)	(0.024)	(2.561)
Observations	148,130	148,130	74,803	74,803	73,327	73,327
R-squared	0.045	0.043	0.082	0.129	0.031	0.038
N	15,228	15,228	7,596	7,596	7,632	7,632

Notes: Regressions show the ITT estimates from a diff-in-diff regression including time and individual fixed effects on water consumption. Regressions in this table trim the top 1% of values in the data. The outcome variables in regression 1, 3 and 5 are transformed using the inverse hyperbolic sine (IHS) which can be interpreted as the log transformation. Regressions 2, 4 and 6 show the effects on the untransformed value. The average price of water is 13.5 N$ for 1000 liters. All standard errors are clustered on the customer level.

Table A3. ITT Effects until End of Rain Season (March)

	Full Sample		Low Users		High Users	
	(1) IHS Water usage in N$	(2) Water usage in N$	(3) IHS Water usage in N$	(4) Water usage in N$	(5) IHS Water usage in N$	(6) Water usage in N$
Specific	-0.061**	-7.558**	0.010	-2.728	-0.114***	-10.520
	(0.028)	(3.484)	(0.034)	(2.006)	(0.043)	(6.638)
Semi-Specific	-0.029	-3.151	0.007	-1.194	-0.061	-4.632
	(0.028)	(3.406)	(0.035)	(2.097)	(0.042)	(6.522)
Unspecific	0.018	-2.900	0.040	-0.533	0.002	-4.739
	(0.028)	(3.499)	(0.034)	(2.107)	(0.043)	(6.716)
Constant	4.133***	114.943***	3.426***	33.167***	4.870***	199.942***
	(0.016)	(1.814)	(0.020)	(0.625)	(0.025)	(3.646)
Observations	176,500	176,500	89,763	89,763	86,737	86,737
R-squared	0.058	0.028	0.077	0.076	0.061	0.033
N	14,943	14,943	7,577	7,577	7,366	7,366

Notes: Regressions show the ITT estimates from a diff-in-diff regression including time and individual fixed effects on water consumption. The outcome variables in regression 1, 3 and 5 are transformed using the inverse hyperbolic sine (IHS) which can be interpreted as the log transformation. Regressions 2, 4 and 6 show the effects on the untransformed value. The average price of water is 13.5 N$ for 1000 liters. All standard errors (in parentheses) are clustered on the customer level. ** p<0.05; *** p<0.01

Table A4. ITT Effects in Comparison To Unspecific Treatment

	Full Sample		Low Users		High Users	
	(1)	(2)	(3)	(4)	(5)	(6)
	IHS Water usage in N$	Water usage in N$	IHS Water usage in N$	Water usage in N$	IHS Water usage in N$	Water usage in N$
Semi-Specific	-0.034	0.185	-0.024	0.707	-0.047	-0.645
	(0.029)	(3.588)	(0.036)	(2.297)	(0.043)	(6.765)
Specific	-0.061**	-4.194	0.002	-1.213	-0.106**	-5.727
	(0.029)	(3.668)	(0.036)	(2.169)	(0.044)	(6.875)
Constant	4.146***	116.251***	3.414***	32.776***	4.880***	199.860***
	(0.018)	(2.051)	(0.023)	(0.695)	(0.028)	(4.062)
Observations	110,523	110,523	55,198	55,198	55,325	55,325
R-squared	0.045	0.029	0.084	0.098	0.033	0.029
N	11,242	11,242	5,596	5,596	5,646	5,646

Notes: Regressions show the ITT estimates from a diff-in-diff regression including time and individual fixed effects on water consumption. The outcome variables in regression 1, 3 and 5 are transformed using the inverse hyperbolic sine (IHS) which can be interpreted as the log transformation. Regressions 2, 4 and 6 show the effects on the untransformed value. The average price of water is 13.5 N$ for 1000 liters. All standard errors (in parentheses) are clustered on the customer level.** $p<0.05$; *** $p<0.01$

Table A5. Effects on Payment Behavior until June 2018

	Full Sample			Low Users			High Users		
	(1) Payment propensity (binary)	(2) Log Payment amount in N$ (if >0)	(3) Combined Effect	(4) Payment propensity (binary)	(5) Log Payment amount in N$ (if >0)	(6) Combined Effect	(7) Payment propensity (binary)	(8) Log Payment amount in N$ (if >0)	(9) Combined Effect
Specific	0.005	-0.019	-0.301	0.005	-0.038*	-1.255	0.005	-0.014	0.343
	(0.004)	(0.017)	(1.662)	(0.005)	(0.022)	(1.295)	(0.006)	(0.024)	(3.148)
Semi-Specific	0.010**	-0.026	0.896	0.011**	-0.046**	-0.150	0.009	-0.019	1.252
	(0.004)	(0.017)	(1.670)	(0.005)	(0.021)	(1.303)	(0.006)	(0.023)	(3.201)
Unspecific	0.008*	-0.003	2.582	0.009*	-0.043**	-0.570	0.007	0.030	8.269**
	(0.004)	(0.017)	(1.624)	(0.005)	(0.021)	(1.348)	(0.006)	(0.023)	(3.255)
Observations	176,033	56,557	176,033	89,687	25,848	89,687	86,346	30,709	86,346
R-squared	0.0969	0.294		0.0878	0.249		0.0995	0.237	

Notes: This table reports ITT effects on payment behavior (regressions 1-3) for the full sample until June 2018. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports the ITT effect on the intensive margin. Regression 3 multiplies the effects of regressions 1 and 2 to get an estimate of the combined effect using a two-part model (Belotti et al. 2015). Regressions 4-6 provide equivalent estimations for low users and regressions 7-9 for high users. Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. Regression control for strata used for randomization as recommended by Bruhn and McKenzie (2009). Control variables are top coded at the 99th percentile. Four customers attrite in August. All reported standard errors are clustered at the customer level to account for serial correlation.* p<0.1; ** p<0.05

Table A6. Treatment Effects on the Collection Efficiency until June 2018

	(1) Collection Efficiency	(2) Collection Efficiency
Specific	0.083**	0.083**
	(0.039)	(0.039)
Semi-Specific	0.026	0.026
	(0.039)	(0.039)
Unspecific	0.019	0.018
	(0.039)	(0.039)
Pre collection efficiency		0.006
		(0.006)
Observations	14,808	14,808
R-squared	0.000	0.000

Notes: The table shows the ITT effects on the collection efficiency, i.e. the fraction of the overall bill paid from the start of the intervention until June 2018. Collection efficiency is top coded at the 99.9^{th} percentile to reduce influence of extreme outliers. Regression 1 shows a simple OLS regression. Regression 2 controls for the pre-intervention mean of the collection efficiency. ** $p<0.05$

CHAPTER 4.

USING IDENTITY LABELS TO DECREASE NON-PAYMENT FOR WATER: EVIDENCE FROM KOSOVO

4.1 Introduction

Non-payment of public utility bills is a common problem in low and middle-income countries and constrains the maintenance and expansion of infrastructure (Szabó and Ujhelyi, 2015; Jack and Smith, 2015). Sanctioning of non-paying customers, for example trough disconnection from the network, is often difficult to enforce due to ethical, legal and technical reasons. Why do customers pay at all in such a setting? In this experiment, I study the role of identity (one's sense of self or self-perception) for the payment of public utility bills. Identity concerns are one of the main drivers of human behavior because individuals strive to act consistently with their identities (Festinger, 1957; Akerlof and Kranton, 2000; Bénabou and Tirole, 2006; 2011). Individuals maintain multiple identities and each identity carries distinct prescriptive norms, informing which behavior is appropriate in a given situation (Akerlof and Kranton, 2000). This suggest a straightforward intervention to change undesirable behavior: exhort people to take on a new identity (Shah and Ludwig, 2016). In this study, I test whether adults' identity is malleable through written statements about one's identity (identity labels) and whether identity labels can be used to reduce non-payment for water.

To investigate these questions, I conduct a large-scale field experiment among around 11,800 customers in cooperation with the Regional Water Company Prishtina (*RWCP*) in Kosovo in order to reduce non-payment for water. Before designing the intervention, we (as a research team) conduct pre-intervention surveys to identify the behavioral factors influencing the payment decision (Datta and Mullainathan, 2015; Duflo, 2017). The survey provides three key findings which motivate the intervention: *(i)* the current invoice placement does not attract attention, since the invoices are typically folded and put in the door crack for lack of owning a mailbox, *(ii)* people have difficulties understanding the information on the invoice and most importantly, *(iii)* people state that being a "responsible citizen" is important to them. The intervention builds upon these findings: To increase attention, we place a sticker along with the invoice at eye-level on a customer's door. To improve understanding of the invoice, we highlight the most important figures on the invoice using a text marker. Lastly, and in the focus of this experiment, we print different identity labels on the stickers. Identity labels provide individuals with a statement about a desirable identity in an attempt to provoke actions that are consistent with that label (Cornelissen, Dewitte, Warlop and Yzerbyt, 2007). In total, there are six different sticker treatments and an untreated control group. Four stickers include an identity label, while two stickers do not (*basic* stickers). Positively framed *basic* stickers inform customers that payment ensures 24h water supply, ask customers to pay and to use the stickers as a reminder to pay. Negatively framed *basic* stickers are identical to the positive ones, except that they inform customers that payment risks 24h water supply. On top of the content of the *basic* stickers, stickers with positively framed identity labels include the *persuasion* statement "Please be a responsible citizen" and the *attribution* statement "You are a responsible citizen". Stickers with negatively framed identity labels additionally include the *persuasion* statement "Please don't be an irresponsible citizen" and the *attribution* statement "You are not an irresponsible citizen" on top of the content of the *basic* sticker. The experiment is implemented over three consecutive months in June to August 2017.

The experiment disentangles the potential effectiveness of *persuasion* ("Please be" and "Please don't be") and *attribution* ("You are" and "You are not") labels as well as the role of *positively* and *negatively* framed identity labels ("responsible citizen" and "irresponsible citizen"). The identity labels in this experiment are inspired by a classic psychological study by Miller, Brickman and Bolen (1975). They argue that it is easier to influence behavior by telling individuals that they *already possess* a certain identity instead of persuading them to adopt a new identity. In *their attribution* treatment, a teacher repeatedly tells fifth-graders that they already are a "litter-conscious class". In *their persuasion* treatment, a teacher tells the children for example "Don't be a litter-bug". The *attribution* treatment leads to less littering behavior than the other treatments and therefore becomes a self-fulfilling prophecy. Such labeling techniques have been suggested outside of the classroom for policy making recently (Walton and Wilson, 2018). Identity labels are intriguing from a policy perspective, since changing individuals' identity through more elaborate programs like *behavioral cognitive therapy* might require several weeks of training (Blattman, Jamison and Sheridan, 2017).

Yet, identity labels are not commonly studied in economics and therefore their influence among adults not well understood (Shah and Ludwig, 2016). Whether and which identity labels are an effective measure to change behavior among adults is hitherto insufficiently answered. The experimental design of Miller, Brickman and Bolen, 1975) has a series of shortcomings, which make it difficult to assess the true effectiveness of such techniques. First, it is conducted among children. Second, there is a multitude of other treatment specific communication (from the school principal and janitor) which might confound the mechanism. Third, the experiment has a very small sample size (only one class per treatment), which leads to confounds like teacher specific effects and experimental contamination effects. Lastly, they cannot disentangle the effects of positive identity label framings (e.g., "litter-conscious class") and negative identity label framings (e.g., "don't be a litter-bug").[25]

According to deviance regulation theory (Blanton, Stuart and Van den Eijenden, 2001), however, the positive or negative framing of identity labels might matter because individuals self-regulate more on perceived consequences of deviating from behavioral norms than on the basis of conforming to behavioral norms. In other words, identity is more defined by one's distinctness rather than by one's commonalities with others. Therefore, individuals identity is more influenced by actions that would make them "stick out from the crowd" (Blanton and Christie, 2003). This idea is also formalized by Bénabou and Tirole (2006) who describe that reputational payoffs from actions depend on social norms. The reputational rewards or the stigma attached to an action depends on how many others engage in that certain behavior. For example, if prosocial behavior is believed to be relatively scarce (e.g. becoming an organ donor) the derived utility from being the "lone hero" is large. On the contrary, if prosocial behavior is common (e.g., not having a criminal record) then the stigma attached to deviating (committing crime) is large.

Moreover and as pre-registered, I study heterogeneous effects after the intervention ends. Post-intervention treatment effects may dramatically influence the overall effectiveness of interventions, yet little is known about what influences their persistence (Madrian, 2012; Sunstein, 2016; Brandon et al., 2017; Thaler, 2018). Frey and Rogers (2014) argue that changes in self-perception (identity) could be a pivotal mechanism. If self-perception influences behavior and is

[25] There are additional other reasons why Miller, Brickman and Bolen (1975) themselves state to be careful about considering these results as the basis for a solution to any social problem such as the long-term effects of such interventions.

malleable, then interventions that change self-perception produce persistent behavioral change as long as the change in self-perception is retained.

Overall, the sticker intervention is highly successful: Sticker types improve the collection efficiency (payments divided by billed amount) during the intervention between 9.9 percentage points and 26 percentage points. The most effective sticker ("Please don't be an irresponsible citizen") translates into a 4.5 Euro increase in payments per customer over the three months of the intervention in total. This amount corresponds to about 62% of an average monthly bill. Negatively framed identity labels are the most effective and on average twice as effective as positively framed identity labels. However, I do not find that *persuasion* or *attribution* stickers yield differential treatment effects. After the intervention ends, I find that customers receiving the *basic* stickers (without identity label) pay significantly less in comparison to customers receiving stickers with an identity label.

To shed light on the behavioral mechanism, I provide evidence from two post-intervention surveys. In a telephone survey (N=808), most importantly, we measure self-perceptions of being a responsible citizen and the amount of other responsible citizens in the city. Further, we measure beliefs on sanctioning and social norms. We also asked which utility bills were paid in the last three months to measure unintended spillover effects on other utility bills and demographics such as income, education and gender. In a street survey (N=638), we show interviewees randomly one of the sticker types and ask whether they have seen the sticker before, what phrases from the sticker they remember, reasons for making water payments, their impressions of the sticker and their intentions to comply with the requests on the sticker.

I find direct survey evidence that individuals link payments for water to being a responsible citizen. This is not only the most commonly mentioned reason for making payments when customers are asked why the stickers would encourage them to pay, but also found in a telephone survey: Negatively framed identity labels significantly change customers' self-perception with respect to being a responsible citizen in comparison to the *untreated* group. A similar, yet smaller, effect is found for the positively framed identity labels. Such an effect is, however, not found for customers receiving the *basic* sticker. Survey evidence rules out that the larger treatment effects of the negatively framed identity labels are driven by a shift in beliefs on sanctioning, monitoring, reminders or social norms. Instead, the large effect of the negatively framed stickers is consistent with theories suggesting that customers regulate their identity more on what makes them distinct rather than what makes them similar to others (Blanton, Stuart and Van den Eijenden, 2001; Bénabou and Tirole, 2006).

This experiment provides several novel insights: First, I show that customers' identity concerns are a strong motivation to pay utility bills, which explains why customers might pay at all in settings where non-payment is difficult to sanction.

Second, I show that identity among adults is malleable trough identity labels in written form. The positive or negative framing of the label is crucial. Self-perception changes particularly strongly for negatively framed identity labels. This change in self-perception coincides with the largest improvement in payment behavior among all stickers in line with deviance regulation theory (Blanton, Stuart and Van den Eijenden, 2001) and shows that identity labels can be an effective measure to change behavior. Yet, I do not find that *attribution* labels are more effective than *persuasion labels* as found by Miller, Brickman and Bolen (1975). Potentially, because their study is conducted among children who might be more prone to such identity manipulation attempts (Yeager, Dahl and Dweck, 2018). Furthermore, in contrast to their study, the majority of stickers in this experiment are placed without personal interaction.

Third, I find that stickers with identity labels improve payment behavior post-intervention in comparison to the *basic* stickers that do not contain an identity label. This provides apples-to-apples evidence on how interventions that change self-perception might cause persistent treatment effects, which is important to understand for settings in which policy makers are unable to treat individuals repeatedly.

Lastly, this experiment documents that behavioral interventions can be a highly attractive policy measure when traditional policy tools, as for example disconnections from the network, are difficult to enforce and ethically objectionable. The intervention in this study are tailored to the specific context and extremely cost effective. Over the whole year (intervention and post-intervention) payments increase by up to 7 Euro at the cost of 12 cent per customer. In addition, the intervention is perceived positively and is well accepted among the population according to survey answers.

4.2 Research Setting and Behavioral Diagnosis

Research Setting - The field experiment is conducted in Pristina, Kosovo's capital and largest city. The experimental sample includes eight neighborhoods containing 11,808 customers, which corresponds to about 17.5% of the private accounts in the city. The experimental sample is chosen in cooperation with *RWCP* to be representative of the rest of the city. At the time of the intervention, about 75% of customers in the sample show debt of more than 8.90 Euro. The median customers has accumulated debt beyond 36 Euro, which corresponds on average to about 6.7 unpaid invoices. Given the average monthly salary is around 400 Euro these amounts are quite substantial.

To illustrate the extent of non-payment, Figure 1 shows a histogram of customers' individual collection efficiency for the pre-intervention year. The collection efficiency is defined as the sum of all payments divided by the sum of all invoices for a given time period. About 22.6% of customers paid nothing in the pre-intervention year. The rest of the customers show a hump shaped distribution peaking at 100% (i.e., paying exactly what is billed). Customers to the left of the 100% paid less than what they used in the pre-intervention year and customers to the right of 100% paid more than what they used (for example to repay outstanding debt or to pay in advance).

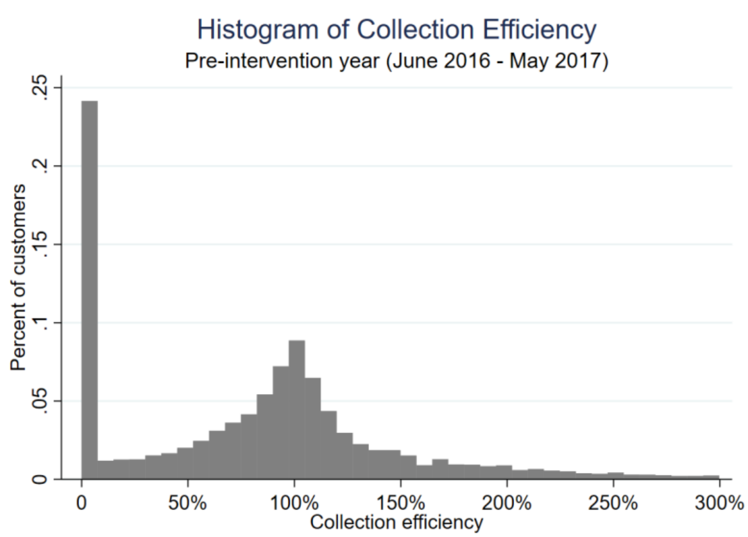


Figure 1: Histogram of the collection efficiency in the pre-intervention year among the experimental sample. Graph excludes 4.6% of customers with a collection efficiency larger than 300%.

The median collection efficiency in the experimental sample is 90.8%. The figure excludes customers whose collection efficiency exceeds 300% which is the case for 4.6% of customers. Note, that the highly right-skewed data causes the mean collection efficiency to be relatively high (102.3%).

Behavioral Diagnosis – Effective interventions depend on understanding the local context and identifying the relevant behavioral bottlenecks (Datta and Mullainathan, 2015; Duflo, 2017). To get a better understanding on customers' perceptions, a research team of local students conducted 934 structured pre-intervention surveys among citizens in public spaces throughout the city. The interviewers introduced themselves as part of a research team of the *University of Cologne* and interviewees answered anonymously. First, interviewees were asked to rank utility bills for water, electricity, internet and waste by their relative importance. This question was asked before interviewees knew that the survey elicited perceptions towards water payments to avoid interviewer demand effects. To assess potential difficulties in understanding, interviewees were then handed an exemplary invoice and asked to identify the total amount due. Finally, the survey contained several questions on payments and usage for water and whether being a responsible citizen was important to them.

In addition, the invoice delivery process was identified as an important element influencing the customers' payment decision. To get an in-depth understanding of the bill delivery process, we accompanied the *RWCP* staff (*invoicers*) over several days and conducted unstructured interviews with customers that were met in person. The *invoicers* read customers' water meters each month and deliver customers' invoices in person to the customers' door. Each *invoicer* typically visits around 50-200 customers a day depending of accessibility of a water meter in the given area. When at site, the *invoicer* reads the customers' water meter and then prints the corresponding water bill using a handheld printing device. Since 84% of customers in the sample don't own a mailbox, the bill is typically folded and stuck in the door crack or hung on the door handle. Pictures of the bill delivery process can be found in Figure A1 in the appendix. If a customer is present during the reading of the water meter, the invoice is handed directly to that person (e.g., 14.7% in June). There are three key insights from the pre-intervention surveys:

i) *Current invoice placement does not create much attention* – When the bill is folded and then stuck in the door crack or hung on the doorknob, the customer's attention is not drawn to the bill. This problem is confirmed in the unstructured interviews as well as when customers are observed when coming home. The invoice is often hardly visible which can be seen in Figure A1 in the appendix. The invoice placement does not create a feeling of importance or urgency.

ii) *Complexity of the invoice* – About 49.7% of interviewees are unable to correctly identify the total amount due when shown the invoice. This fraction does not substantially improve when analyzing the subsample that states to personally be responsible for utility payments (53.4%). The complexity of the bill arises from the multitude of numbers that are printed on small slip of paper (see appendix Figure A2). The most common mistake (70.5% conditional on making a guess) is that interviewees in the survey refer to the monthly fee as the amount that needs to get paid in total.

iii) *Importance of paying for water and being a responsible citizen* – About 87.9% of interviewees rank the importance of water bills either on first or second place. Water is the overall highest ranked bill and ranks statistically significantly higher than electricity (p=0.032), internet (p<0.001) and waste (p<0.001, two-sided Mann-Whitney-U (MWU) test). Given the large fraction of customers having unpaid water bills in the administrative data (about 90.5%), only a relatively low fraction (34.7%) of interviewees states to have unpaid water bills.[26] I interpret this as a social desirability response bias caused by the stigma attached to non-payment. Since 100% of customers state that being a responsible citizen is important to them,[27] customers try to avoid being perceived as irresponsible and therefore don't admit to have debt when asked.

[26] Of those stating to have debt, 88% intend to repay their outstanding bills in the future. The most common stated reason for non-payment is insufficient financial means.

[27] Elicited among a subsample of 80 citizens, after which we decided to drop the question from the questionnaire due to the uniform consent.

4.3 Experimental Design

Treatments – The treatments build upon the findings in the behavioral diagnosis. To increase the attention that customers pay to the invoice, the research team uses stickers that places the invoice at eye-level on a customer's door instead of placing the invoice in the door crack or the door handle.[28] To improve understanding of the invoice, the research team uses markers to highlight the most important numbers on the invoice if the customer's invoice shows debt. In addition, we print identity labels on the sticker attempting to provoke label consistent behavior of an identity that is important to customers. Figure A2 in the appendix shows pictures of the sticker intervention. In total, there are six sticker treatments which are shown in Figure 2 and an *untreated* control group.

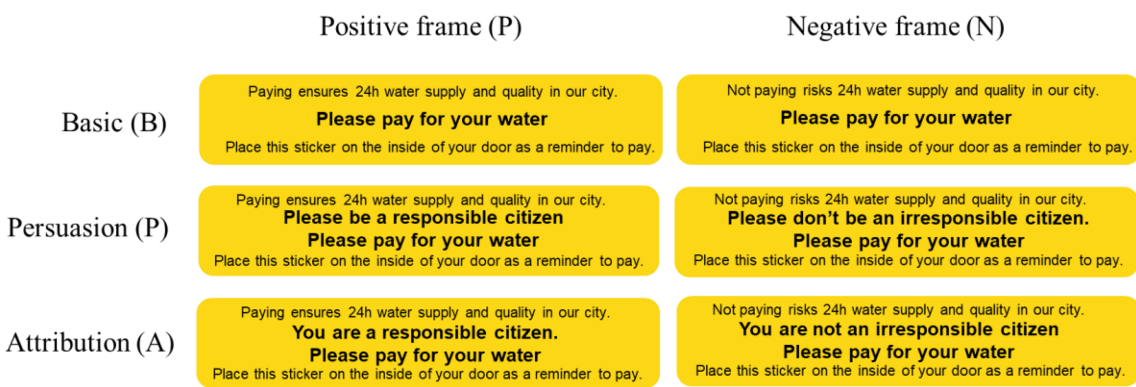


Figure 2: Treatment messages printed on the stickers.

Stickers in the left column are positively framed (P for positive). Positively framed stickers inform customers about the positive consequences of paying ("Paying ensures 24h water supply and quality in our city. Please pay for your water."). In addition, the sticker reads "Place this sticker on the inside of your door as a reminder to pay", which is intended to work as a permanent reminder.[29] Sticker PB (Positive Basic) differs from sticker PP (Positive Persuasion)

[28] Customers with a mailbox received the sticker attached to the back of the invoice. When there was personal contact between the research team and a customer, the research team followed a standard script in which the research team informed the customer that stickers were given to everyone in the street to remind them to pay for water. Then the exact wording on the sticker was read to the customer and the amount to pay was highlighted and explained. I control for these variables in the appendix table A1, which does not substantially change the results.

[29] Garner (2005) is to the best of my knowledge the only comparable study that uses post-it notes as reminders to increase response rates to a survey.

and PA (Positive Attribution) only with respect to one single sentence, but is otherwise identical. PP stickers additionally include the sentence "*Please be a responsible citizen*". PA stickers include the sentence "*You are a responsible citizen*". Stickers in the right column are negatively framed (N for negative). They state the negative consequences of not paying ("Not paying risks 24h water supply and quality in our city. Please pay for your water"). Stickers NP (Negative Persuasion) and NA (Negative Attribution) include one additional sentence, but are otherwise identical to NB (Negative Basic). NP additionally includes the sentence "*Please don't be an irresponsible citizen*" and NA includes "*You are not an irresponsible citizen*". Throughout this paper, I will refer to PB and NB as the basic stickers. PP and PA are called positively framed identity labels. NP and NA are referred to as negatively framed identity labels.

Randomization – Treatment assignment is clustered and randomized by date and *invoicer*. This minimizes any potential date and *invoicer* specific effects. Therefore all neighbors on a given day receive the same sticker type to avoid treatment contamination effects. Table 1 shows summary statistics for the untreated comparison group and the mean difference and p-value of each of the treatments in comparison to the untreated group to test for imbalances. Overall, the balance table shows no statistically significantly different variables at the 5% level for any of the measured covariates. Since the data is highly right skewed and noisy, I provide the same regressions as in the main text with an extensive set of control variables in the appendix (Table A1) which does not substantially change the results.

Implementation – The research team placed the stickers according to the assigned treatment over three months (June to August 2017). The research team usually learned their treatment assignment in the morning each day. We did not treat customers who were inactive (no water usage) during the intervention. Each customer was assigned to receive three stickers of the same type throughout the experiment.[30] The invoices were folded in a way that third parties could not see the invoice information. The team documented a range of covariates for example how visible the placement of the stickers is to third parties (e.g., on door or mailbox), whether there was personal contact (including acceptance and rejection of the sticker), whether the water consumption was estimated

[30] Roughly 10.8% received only one sticker during the three months because *invoicers* became sick, the customer became inactive (no consumption) or the handheld printing devices had malfunctions. The sample excludes six customers who received different types of stickers during the intervention due to human error.

(e.g., not possible when water meter is flooded) or read exactly, building types (apartment block or individual house) and on which floor level an apartment is located (-3 to 15) and building quality (5-point scale) as an approximation variable for wealth. The research team also accompanied the *invoicer* and measured these covariates on days when no stickers were placed (untreated group) to control for any potential effects caused by having more people than usually reading the meter or observing people taking notes in front of their house. Figure 3 shows the timeline of the experiment.

Post-experimental sticker perception survey (N=638) – After the end of the intervention in August, we conducted a street survey in public spaces throughout the city. The interviews were conducted by local students who introduced themselves as part of a research team of the *University of Cologne*. We showed interviewees randomly one of the sticker types and asked whether they had seen the sticker before. If yes, we asked them to recall the content of the sticker. This was done to measure potential experimental contamination effects. If they had not seen the sticker before, they were given the sticker for sufficient time to read, before being asked to hand it back and to recall the content. This was done to measure which phrases on the sticker were most memorable. Afterwards the interviewer handed the sticker back yet again to ask whether and why the sticker would encourage them to pay. We then measured their overall impressions of the sticker, asked explicitly whether they would feel more monitored than usually if the saw the sticker placed on their door and whether they would use the sticker as a reminder.

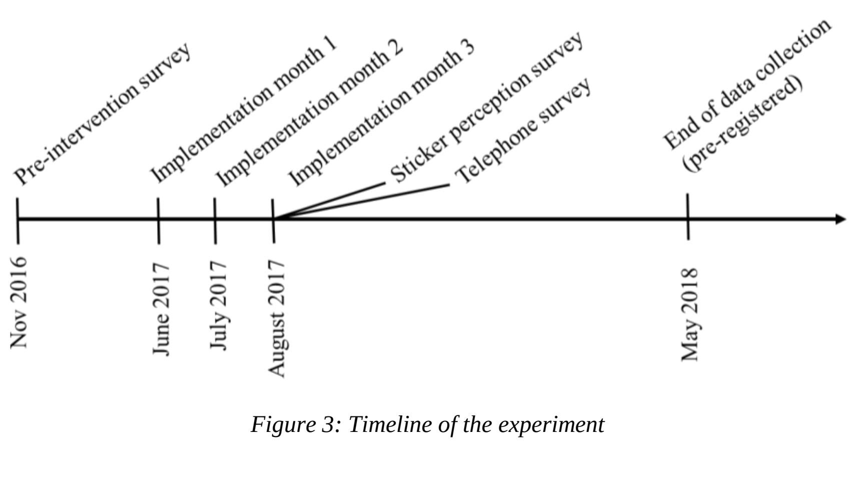


Figure 3: Timeline of the experiment

Post-experimental telephone survey (N=808) – Also in August, the research team of local students conducted a telephone survey. The research team introduced themselves as part of a research team of the *University of Cologne.* They called all customers whose *RWCP* account provided a phone number and who were part of the experimental sample. Interviewers were unaware of a customer's identity and payment behavior at all times. Customers were called in random order and up to three times if a customer was not reached during a previous call attempt. In total, the survey sample contains 808 customers who confirmed to be responsible for payments in the household.[31] The questionnaire focuses on self-perceptions of being a responsible citizen, prevalence of other responsible citizens in *Pristina*, social norms and perceived consequences and thresholds of non-payment at which potential sanctioning would be enforced. Additionally, the questionnaire asked which utility bills were paid in the last three months in order to measure unintended spillover effects on other utility bills and demographics such as income, education and gender.

[31] We managed to talk to 37.36% of customers whose account provided a phone number. Completion rate is overall high with 91.1% on average. I do not find treatment differences in the likelihood to interrupt the interview once started i.e., answering yes to "are you the one responsible for water payments in your household?". The largest pairwise difference is found between the *untreated* group and the *P1* treatment and is statistically insignificantly different from zero (p=0.360, χ^2 - Test). These findings speak against a selection bias into the survey based on the treatment assignment.

Table 1. Descriptive Statistics and Balance Table

	Untreated (N=1,854)		PB (N=1,589)		NB (N=1,671)		PP (N=1,731)		NP (N=1,579)		PA (N=1,699)		NA (N=1,685)	
	Mean (1)	SD (2)	Mean Diff. (3)	P-value (4)	Mean Diff. (5)	P-value (6)	Mean Diff. (7)	P-value (8)	Mean Diff. (9)	P-value (10)	Mean Diff. (11)	P-value (12)	Mean Diff. (13)	P-value (14)
Collection efficiency	0.988	1.420	0.104	.153	-0.031	.724	0.024	.743	0.078	.451	0.052	.604	-0.015	.861
Payment amount in EUR	6.321	18.92	0.782	.359	0.505	.534	1.342	.136	0.715	.473	1.307	.228	0.165	.854
Payment propensity (binary)	0.282	0.450	0.044	.065	0.012	.623	0.051	.053	0.025	.397	0.037	.179	-0.008	.741
Payment amount (if >0)	22.45	30.15	-0.607	.722	0.775	.671	0.622	.733	0.527	.776	1.458	.451	1.283	.475
Water consumption in EUR	7.281	6.236	0.091	.924	0.340	.719	0.905	.359	-0.007	.994	0.960	.371	-0.073	.941
Debt amount in EUR	265.6	556.0	-78.354	.332	-58.438	.471	-71.871	.376	-82.122	.294	-62.854	.430	-73.927	.381
Has debt (binary)	0.912	0.283	0.007	.708	0.011	.565	0.011	.573	0.020	.305	0.014	.406	0.010	.637
Shared meter	0.238	0.426	-0.096	.346	-0.070	.493	-0.071	.453	-0.022	.841	0.108	.367	-0.082	.466

Notes: The first two columns report the mean and standard deviation for the untreated group in the pre-intervention year. The remaining columns show the coefficients and the p-value testing for a difference between the untreated group and the respective treatment. Differences are tested using an OLS regression in which the outcome variable is regressed on the treatment dummies. Standard errors are clustered on the invoicer-date level (unit of randomization). Continuous variables are winsorized at the 99.9th percentile to reduce influence of extreme outliers.

4.4 Estimation Strategy and Results

The main outcome variable is the collection efficiency (as displayed in Figure 1). The collection efficiency is defined as the sum of all payments divided by the sum of all invoices for a given time period. This aggregate measure reduces noise in the outcome variable by averaging payment behavior over several months, which is useful for data with high variability and low autocorrelation (McKenzie, 2012). Regression 1 shows the marginal effects at means on the extensive margin of a probit regression. It measures the probability of making any payment during the three month intervention. Regression 2 shows the effects of an OLS regression on the intensive margin (the effect on the natural logarithm of the collection efficiency conditional on being larger than zero). Regression 3 estimates the combined effects of regression 1 and 2 using a two-part model (Belotti et al., 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment means of the outcome variable (McKenzie, 2012), which is the collection efficiency and whether a customer has made any payments in the pre-intervention year. In addition, I control for the strata used for randomization (*invoicer* fixed effects and date fixed effects) in the regression as recommended by Bruhn and McKenzie (2009). Regressions controlling for a wide range of other covariates are presented in the appendix A1. They do not substantially change the results. Standard errors are clustered at the level of randomization (*invoicer* date level). Table 2 shows the average treatment effects in comparison to the *untreated* group for the actual intervention period and for the post-experimental periods. Figure 4 plots the treatment effects with whiskers indicating 95% confidence intervals for the intervention periods.

Result 1*: Large positive treatment effects of stickers on collection efficiency.*

Regression 3 in Table 2 shows that PB (Positive Basic) yields a combined treatment effect of 16.3 percentage points (p=0.010) on the collection efficiency during the intervention period in comparison to the untreated group. NB (Negative Basic) yields a treatment effect of 18.6 percentage points (p=0.005). The point estimates of the positively framed identity labels are smaller. PP (Positive Persuasion) increases collection efficiency by 14.8 percentage points

(p=0.027) and PA (Positive Attribution) by 9.9 percentage points, yet the latter effect is statistically insignificantly different from zero (p=0.140). The negatively framed labels yield the largest treatment effects. The NP (Negative Persuasion) sticker shows a treatment effect of 25.0 percentage points (p<0.001) and the NA (Negative Attribution) sticker an effect of 26.2 percentage points (p<0.001). The increase for the NP and NA sticker corresponds to about 4.20 Euro and 4.50 Euro, respectively, in total per customer over the three experimental months.

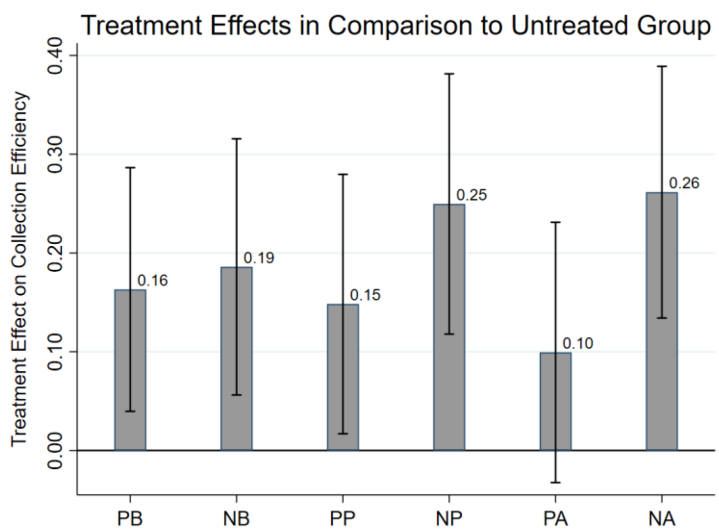


Figure 4: Average treatment effects on collection efficiency with 95% confidence intervals by sticker type in comparison to untreated control group during the intervention (June to August 2017).

***Result 2**: No evidence that attribution labels are more effective than persuasion labels.*

The regression analysis suggests that *attribution* stickers are not more effective than *persuasion* stickers when holding the positive or negative identity frame constant (i.e. comparing PP vs. PA and NP vs. NA). The differences are tested in Table 3. Regression 9 shows that the combined treatment effect of PP is statistically insignificantly different from PA (p=0.403). The point estimate of PA is even 4.9 percentage points smaller than the one of PP indicating that even with larger statistical power, the *attribution* treatment would not be more effective than the *persuasion* treatment. The treatment differences between decreases to 2.5 percentage points when controlling for covariates (see Table A1). The treatment difference of 1.3 percentage points between NP and NA is also statistically insignificantly different from each other (p=0.836).

Result 3: *Negative identity labels are twice as effective as positively framed identity labels.*

Negatively framed identity labels (NP and NA) are on average about twice as effective as positively framed identity labels (PP and PA). Differences are tested in the right panel (Panel B) of Table 3. Regression 12 shows that the difference of 10.1 percentage points is marginally statistically significant (p=0.095) for *persuasion* (PP vs. NP). The difference between positive and negative *attribution* stickers (PA vs. NA) is 16.2 percentage points and statistically significantly different from zero (p=0.009).[32] The pooled difference between positive and negative identity labels is 13.1 percentage points (p=0.005) as tested in regression Table A2 in the appendix.

Result 4*: Stickers with identity label are more effective than basic sticker after intervention ends.*

As pre-registered, I now analyze how the different sticker types affect payment behavior for the nine months after the intervention ends. The right panel (regressions 4-6) in Table 2 shows the regression results for the post-intervention data (September 2017 until May 2018). I use the same regression model as in the previous analysis. I find that both basic stickers show negative effects. The post-intervention point estimates of the combined effect of the PB sticker is negative 5.9 percentage points (p=0.131) and negative 4.2 percentage points in the NB treatment (p=0.268) in comparison to the untreated group. The difference becomes statistically significant after controlling for covariates in Table A1. The point estimates of the stickers containing a positively framed identity label are in the range of 3.1 to 3.4 percentage points, yet statistically insignificantly different from the untreated group. While the NP treatment is marginally statistically significant on the extensive margin (p=0.074), the overall combined point estimate of 4.4 percentage points is statistically insignificant as well. The NA shows has the smallest point estimate among the stickers with identity label and is close to zero.

[32] For the basic stickers the positive framing is not statistically different from the negative framing (p=0.709)

Table 2. Average Treatment Effects on Collection Efficiency

	Intervention (June `17 – Aug. `17)			Post-Intervention (Sept. `17 – May `18)		
	(1) Payment Propensity	(2) Log Collection efficiency>0	(3) Combined Effect	(4) Payment Propensity	(5) Log Collection efficiency>0	(6) Combined Effect
Positive Basic (PB)	0.050**	0.031	0.163***	0.006	-0.006	-0.059
	(0.021)	(0.054)	(0.063)	(0.018)	(0.027)	(0.039)
Negative Basic (NB)	0.064***	0.022	0.186***	0.004	-0.030	-0.042
	(0.018)	(0.051)	(0.066)	(0.020)	(0.029)	(0.038)
Pos. Persuasion (PP)	0.036*	0.044	0.148**	0.026	0.007	0.031
	(0.022)	(0.047)	(0.067)	(0.020)	(0.030)	(0.038)
Neg. Persuasion (NP)	0.074***	0.050	0.250***	0.036*	-0.021	0.044
	(0.021)	(0.053)	(0.067)	(0.020)	(0.025)	(0.040)
Pos. Attribution (PA)	0.035	0.010	0.099	0.016	0.022	0.034
	(0.022)	(0.053)	(0.067)	(0.021)	(0.035)	(0.041)
Neg. Attribution (NA)	0.072***	0.065	0.262***	0.019	-0.005	-0.006
	(0.022)	(0.058)	(0.065)	(0.018)	(0.026)	(0.039)
Observations	11,808	7,028	11,808	11,794	9,182	11,794
R-squared	0.133	0.115	-	0.1574	0.0460	-

Notes: The first three columns of Table 2 report average treatment effects on collection efficiency for the treatment periods (June-August 2017) in comparison to the untreated group. The last three columns show treatment effects for the post-intervention periods (September 2017-May 2018). Regression 1 and 4 show the treatment effects of a probit regression on the likelihood of making a payment (marginal effects at means). Regression 2 and 5 report average treatment effect of an OLS regression on the intensive margin. Regressions 3 combines the effects of regressions 1 and 2 using a two-part model (Belotti et al. 2015). Similarly, regression 6 combines the effects of regression 4 and 5. Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment values of collection efficiency and whether any payments were made (main outcome variables), as well as the strata used for randomization (invoicer and date dummies) in the regression as recommended by Bruhn and McKenzie (2009). Control variables are winsorized at the 99.9th percentile. All reported standard errors are clustered at the invoicer-date level (unit of randomization). 14 customers attrite from the sample after having been treated during the intervention. * p<0.1; ** p<0.05; *** p<0.01

In comparison to the *basic* sticker, however, stickers with identity labels yield statistically significantly larger treatment effects. Positively framed identity labels treatment effects are 8.4 percentage points larger (p=0.002) than the *basic* stickers and negatively framed identity labels 6.9 percentage points larger (p=0.010). Individually tested, all but NA are yield statistically larger post-intervention treatment effects than the corresponding basic stickers (Table A4).[33]

[33] The post-intervention treatment effects of NB is statistically significantly different from NP (p=0.027), but not from NA (p=0.370). PB is statistically different from PP (p=0.013) and so is PA (p=0.017).

Table 3. Testing for Differences Between Sticker Treatments

Panel A	Persuasion vs. Attribution		
	(7) Payment Propensity	(8) Log Collection efficiency>0	(9) Combined Effect
Basic (PB+NB)	0.057***	0.027	0.175***
	(0.018)	(0.048)	(0.057)
Pos. Identity (PP+PA)	0.037*	0.045	0.149**
	(0.022)	(0.047)	(0.067)
Pos. Identity*Attribution (PA)	-0.002	-0.035	-0.049
	(0.022)	(0.040)	(0.059)
Neg. Identity (NP+NA)	0.075***	0.051	0.250***
	(0.021)	(0.054)	(0.067)
Neg. Identity*Attribution (NA)	-0.002	0.014	0.013
	(0.021)	(0.048)	(0.061)
Observations	11,808	7,028	11,808
R-squared	0.133	0.115	-

Panel B	Positive vs. Negative		
	(10) Payment Propensity	(11) Log Collection efficiency>0	(12) Combined Effect
Basic (PB+NB)	0.064***	0.022	0.163***
	(0.018)	(0.051)	(0.063)
Basic*Positive (PB)	-0.014	0.009	0.023
	(0.017)	(0.044)	(0.061)
Persuasion (PP+NP)	0.074***	0.051	0.250***
	(0.021)	(0.054)	(0.067)
Persuasion*Positive (PP)	-0.038*	-0.006	-0.101*
	(0.022)	(0.038)	(0.061)
Attribution (PA+NA)	0.072***	0.066	0.262***
	(0.022)	(0.058)	(0.065)
Attribution*Positive (PA)	-0.037	-0.056	-0.162***
	(0.023)	(0.051)	(0.062)
Observations	11,808	7,028	11,808
R-squared	0.1330	0.115	-

Notes: This table reports heterogeneous treatment effects with respect to persuasion or attribution framing (left panel) and with respect to positive or negative framing (right panel) on collection efficiency for the treatment periods (June-August 2017) in comparison to the untreated control group. Regression 7 and 10 show the treatment effects of a probit regression on the likelihood of making a payment (marginal effects at means). Regression 8 and 9 report average treatment effect of an OLS regression on the intensive margin. Regression 9 and 12 combine the effects using a two-part model (Belotti et al. 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment values of collection efficiency and whether any payments were made, as well as the strata used for randomization (invoicer and date dummies) in the regression as recommended by Bruhn and McKenzie (2009). Control variables are winsorized at the 99.9^{th} percentile. All reported standard errors are clustered at the invoicer-date level (unit of randomization). * p<0.1; ** p<0.05; *** p<0.01

4.5 Post-Experimental Survey Results

4.5.1 Behavioral Mechanism

This section provides answers from both the street and the telephone survey to shed light on the behavioral mechanism that might have caused the negatively framed identity labels to be so effective. Table 4 displays descriptive statistics of both surveys. Panel A shows the average answers sorted by the sticker type that was shown to interviewees in the street survey.

Table 4. Main Survey Results

	Untreated	Basic		Positive Identity		Negative Identity	
		PB	NB	PP	PA	NP	NA
Panel A: Street Survey							
Encouraging to pay		0.544	0.468	0.667	0.620	0.574	0.628
Be a responsible citizen	-	0.510	0.364	0.500	0.649	0.455	0.327
Sticker grabs attention	-	0.388	0.568	0.448	0.414	0.389	0.449
Otherwise cut-offs	-	0.041	0.205	0.224	0.053	0.167	0.286
Would use as reminder	-	0.467	0.340	0.314	0.462	0.394	0.397
Feels more monitored	-	0.567	0.538	0.682	0.527	0.462	0.436
Useful and good		0.878	0.915	0.906	0.879	0.935	0.899
N=540	-	91	94	87	95	94	79
Panel B: Telephone Survey							
Self-perception (“always”)	45.87	46.53	41.44	38.68	37.61	34.58	31.68
Social norms (more than half)	0.615	0.564	0.628	0.664	0.578	0.660	0.529
Aware of any consequences	0.894	0.951	0.878	0.899	0.903	0.937	0.952
Aware of cut-offs	0.491	0.558	0.496	0.449	0.573	0.519	0.563
How much debt (median)	100	100	90	70	100	100	100
How much debt in bills (md.)	3	2.75	2.5	3	3	3	3
Other utility bills paid	2.109	2.236	2.174	2.079	2.052	2.137	2.063
N=808	120	106	115	117	121	117	112

Notes: The table shows selected answers from both post experimental surveys. Panel A shows results from the street survey excluding customers stating to have seen the sticker at home. The first row shows the fraction of customers stating that the sticker would encourage them to pay. The following rows show the categorized reasons for why so, conditional on being encouraged to pay. The next rows show the fraction of customers that would use the sticker as a reminder to pay and the fraction of customers that would feel more monitored than usually. Panel B shows results from the telephone survey among those stating to be responsible for payments of the water bill. The first row shows the fraction of people always thinking of themselves as responsible citizens. The next row displays the fraction of customers believing that more than half of customers in Pristina pay their bills regularly. The following rows measure whether customers are aware of any consequences of non-payment (and in particular cut-offs) and the median thresholds at which these sanctions would be enforced. The last fraction shows the sum of other utility bills that customers state to have paid during the last three months.

The sample in Panel A excludes customers stating to have seen the sticker at home since their opinion might be distorted due to being part of the actual experiment. Panel B displays answers from the telephone survey. The table shows answers sorted by the treatment group that the interviewed customers were actually part of in the experiment.

Being a responsible citizen is main motivation for making payments – In the post-experimental street survey, we asked interviewees why the stickers would encourage them to pay. Answers were open-ended and categorized by the interviewer. The descriptive statistics are shown in the indented rows in Panel A. The most commonly stated reason is being or wanting to be a responsible citizen (47.4%).[34] The second most common answer is that the sticker would grab their attention (41.1%). Note, that these are precisely the two channels (identity and attention) that the intervention was designed to address (since we identified these mechanisms to be relevant in the pre-intervention survey). This shows that identity concerns are a crucial motivation to make payments in the experiment.[35]

Negative identity labels change customers' self-perception – If the stickers establish or strengthen a link between the payment decision and one's identity, we should observe a reflection of this in customers' self-perception. In the telephone survey, we measured to which degree customers identified as responsible citizen by asking "How often do you think of yourself as responsible citizen?" Figure 5 shows interviewees' answers pooled across all treatments.

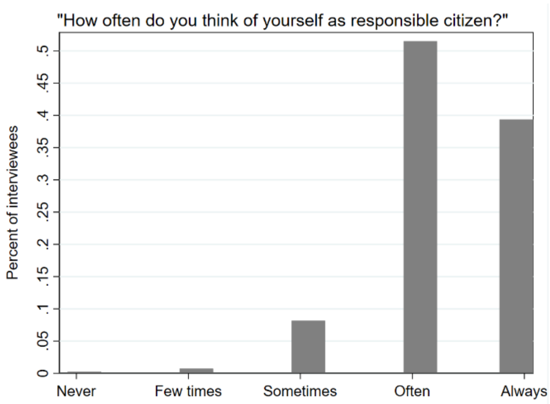


Figure 5: Self-perception of being a responsible citizen.

[34] This answer is also the most common reason for the basic sticker, which speaks against an interviewer demand effect since these stickers do not contain any "responsible citizen" wording.
[35] The answer categories were defined after running pilot surveys. Alternative reasons why the sticker would encourage them to pay are significantly less often mentioned, for example through fearing cut-offs (16.08%) or to ensure 24h water supply in the city (18.33%) or because they would use the sticker as a reminder (10.93%).

Around 90% of customers always or often think of themselves as responsible citizen. In both the NP and NA treatment, the self-perception is different: In comparison to the *untreated* group, customers in the *NP* and *NA* treatment on average think *less often* of themselves as responsible citizen. In the untreated group 45.9% state to be always responsible. This fraction statistically significantly decreases to 34.6% (p=0.040, χ^2-Test) in the NP and 31.7% (p=0.017, χ^2 -Test) in the NA treatment. A similar yet smaller effect is found for positively framed identity labels but not for the *basic* stickers. [36] Customers who made below median payments during the three months of the intervention drive this downward shift in self-perception.[37]

Higher effectiveness of negatively framed identity labels is in line with deviance regulation theory – According to deviance regulation theory, individuals self-regulate more based on what would make them "stick out". In order to check whether this could explain the larger effectiveness of the negatively framed identity labels, I provide survey answers with respect to social norms and the prevalence of other responsible citizen in Pristina. The left panel of Figure 6 shows beliefs on social norms ("How many in your street/building pay their water bill regularly?") and the right panel beliefs on prevalence of other responsible citizens in the city ("How many people in Pristina are responsible citizen?").

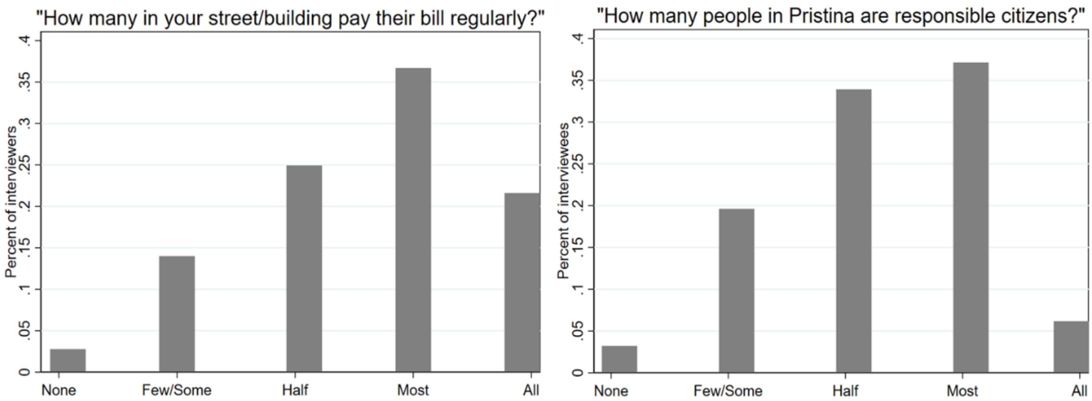


Figure 6: The left panel shows beliefs on social norms and the right panel beliefs on the prevalence of other responsible citizens in the city.

[36] The pooled difference in self-perception between positively framed identity stickers and the untreated group is marginally significant (p=0.065, MWU-test). The pooled difference between basic stickers and the untreated group is statistically insignificant (p=0.612, MWU-test). The pooled difference in self-perception between negatively framed identity stickers is statistically significant in comparison to the basic stickers (p=0.014, MWU-test) as well as in comparison to the untreated group (p=0.010, MWU-test).

[37] Customers with below median payments think less often of themselves as responsible citizen in comparison to the untreated group, but only if they received an identity label. Table A5 shows that self-perception is lower for below median payers receiving positively framed identity labels (p=0.043) or negatively framed identity labels (p=0.055), but not for customers receiving the basic stickers (p=0.831). This finding suggests that customers might not simply adopt the labeled identity. Instead, the identity label may causes customers to regulate (self-signal and shape) their identity through observation of their own actions (Bem, 1967, Bénabou and Tirole, 2006; 2011).

Both measures show that the majority of citizens in Pristina is believed to pay the bill regularly and that customers believe that the majority of citizens in Pristina are responsible citizens. Therefore, non-payment and being an irresponsible citizen would make them "stick out from the crowd", which might explain why the negatively framed identity labels outperform the positively framed identity labels.[38]

4.5.2 Ruling Out Alternative Channels

The results so far show that the customers' payment decision is linked to their identity and that customers' self-perception changes particularly strongly among customers receiving the negatively framed stickers, which coincides with the largest increase in payment behavior. There are however several alternative mechanisms that could potentially confound the mechanism. I present a number of alternative channels and argue that these can be ruled out.

No treatment differences in social norms – The sticker intervention could have changed beliefs on other's payment behavior (social norms). For example, customers might potentially believe that everyone else's payment behavior was similar to theirs, since all stickers in a building contained the same message. Alternatively, the intervention could cause neighbors to discuss their payment habits spreading information about others' payment behavior. One can also imagine that the identity labels implicitly convey information about the behavior of others. For example, an irregular payer receiving the "You are a responsible citizen" sticker might shift his beliefs about social norms downwards. All these channels would reflect in changes in the beliefs about others' payment behavior. The intervention does not cause a significant shift in beliefs on social norms when comparing the negatively framed identity labels (or any of the treatments) with the *untreated* group. The direction of the effect size is also not uniform, for example beliefs shift insignificantly upwards in the NP treatment (p=0.213, *MWU-test*) and insignificantly downwards in the NA treatment (p=0.808, *MWU-test*). A change in social norms causing the effectiveness of the negatively framed identity labels is therefore inconsistent with the survey evidence.

[38] Alternative explanation might be that the positive framings have little room to improve on, since most customers already think of themselves as responsible citizens (ceiling effects). Alternatively, individuals might be loss-averse in their identity "stock".

No treatment differences in sanctioning beliefs – As a result of receiving a sticker on their door, customers might believe that they could potentially face sanctions (cut-offs, law suits) or might be at the threshold of getting sanctioned. To measure such effects, we elicited awareness of sanctions and the debt thresholds at which customers would expect sanctions in the telephone survey. Customers are not statistically significantly more aware of any sort of sanctions in general or specifically for disconnections. Neither, do I find that customers in the treatments have different beliefs regarding the threshold at which one is considered for sanctions in comparison to the untreated group. The median beliefs for the threshold are identical for the untreated group and the negatively framed identity labels. Lastly, I find that if anything, the intervention is more effective among customers sharing a collective meter (Table A6). Sanctions among this group are particularly difficult to enforce, since the whole building would need to be disconnected from water supply, even if individual households are paying. I conclude, that it is unlikely that fear of getting sanctioned explains the treatment effects.

No treatment differences in perceptions of being monitored - The street survey asked customers explicitly about whether customers would feel more monitored than usually if they saw the sticker hanging on their door. Around half of interviewees (53.6%) agree that they would feel more monitored than usually. While monitoring might play a role why the stickers worked in general, they cannot explain why the negatively framed identity labels were so effective. If anything, customers seeing the negatively framed labels would feel less monitored in comparison to the basic stickers (p=0.040, MWU-test). Therefore, an increase in the perception of being monitored is also inconsistent with the survey results.

No treatment differences in intentions of placing the sticker as reminder – Another channel that can be ruled out is that customers were more likely to place certain sticker type on their door as a reminder to pay. However, only relatively few interviewees (39.6%) state that they would place the sticker on their door. Moreover, there are no differences in the intentions of placing the negatively framed stickers on the door that could explain the larger effectiveness of the negatively framed stickers.

4.5.3 Further Survey Insights

High acceptance of the intervention - The overall impression of the intervention is very positive. On average 90.2% agree or partially agree that the stickers are “useful and good”. The intervention has an even higher approval rate among customers that report to have seen a sticker at home in comparison to those seeing the sticker for the first time (p=0.019, χ^2 - Test). The high approval rate is also reflected in the acceptance rate of the sticker when customers are met in person as less than 1% reject the sticker. Moreover, the majority of customers state that the stickers would encourage them to pay (58.9%). In comparison to the basic stickers, positive labels are statistically significant more encouraging (p=0.008, χ^2 - Test) and so are negative identity labels (p=0.070, χ^2 - Test). Among customers stating that they would not be encouraged, the most common answer is that they are already paying (66.9%) and that they don’t need a reminder or help (24.6%).

Negligible contagion effects – The experiment was carefully designed to minimize experimental contagion effects. First, treatments are clustered by *invoicer* and day such that neighbors receive the same sticker type. Second, we did not treat neighborhoods too close to the city center to avoid visibility to pedestrians going window shopping or drinking coffee. We also did not treat businesses and inactive customers to avoid that stickers would be left hanging on the door when no one was home. The vast majority was placed inside of apartment blocks where a key was necessary to enter the building. In addition, the company did not mention the intervention publicly. In total, only about 9% of stickers were placed in spots where they could potentially be seen from the street level.

Nevertheless, we measured experimental contagion effects in our street survey by asking whether customers had seen the sticker before. Around 15.7% of interviewees state to have seen the sticker at home. This is in line with what we would expect given that we treated about 17.5% of the city. Only about 4.1% state to have seen the sticker in the street and 6.2% report to have seen the sticker at friends or their family’s house. The remaining interviewees state to see a sticker for the first time. Among those that saw stickers in the street or at a friend’s house, the recall of the content is low. They only have a good recall (as coded by interviewers) for the phrase “Please pay for your water“ which is mentioned by 70.3% of interviewees. This phrase is common to all stickers and therefore should not cause contagion effects between the treatment arms. For all other phrases on the sticker, these interviewees are more likely to recall nothing at all rather than being able to at least partially recall any of the other phrases. Contagion effects therefore do not seem to confound the effects found in this study.

Manipulation check of identity label – Survey evidence suggests that customers being part of the actual experiment read the sticker content, and in particular, the relevant identity label. About 46% of customers who report having seen the sticker at home recall the identity phrase. However, customers that received a basic sticker are technically unable to recall an identity label because it is not content of the sticker. We could adjust the recall rate for the fact only 66.5% of all customers in the experiment received a sticker with identity label. Dividing the recall rate of 46% by the share of customers receiving an identity labeled sticker yields a rough estimate of about 69.1%.[39] Interviewees stating to have seen the sticker at home have a better recall of the identity label than all other groups (p=0.009, MWU-test) and also compared to those having seen it in the street (p=0.019, MWU-test).

No spillovers on other utility bills – A question that has received little attention in the literature is to which degree such interventions may influence the payment of other utility bills. Negative spillovers might stem from budgeting effects or from increased attention to payment of the water bills at the cost of other utility bills. On the other hand, customers' overall willingness to pay towards utility payment may increase due to the changes in the self-perception of being a responsible citizen. The survey evidence here does not provide hard facts. Judging from the self-reported data however, it seems that there are no significant spillovers on the reported amount of other utility bills paid. The largest difference is found between the PA sticker with 2.05 bills and the PB sticker with 2.24 bills, yet this difference is statistically insignificantly different from zero (p=0.228, χ^2-test). This suggests that there were no spillovers on other utility bills, at least not on the extensive margin.

[39] This is a rough estimate since interviewees might recall what they saw when we briefly showed them the sticker in the beginning of the survey and not what the saw at home. In addition, by chance, we could have interviewed a larger proportion of a certain sticker type in the interview in comparison to the actual experiment, which would also bias the estimate.

4.6 Conclusion

This paper investigates the role of identity concerns for public utility payments using identity labels. It shows that identity concerns are a crucial motivation to pay for water utility and that adults' identity is malleable through simple written identity labels on a sticker. The identity labels are a highly effective measure to improve collection efficiency. The positive or negative framing of identity labels matters. Negatively framed identity labels are about twice as effective as positively framed identity labels in improving payment behavior. After the interventions ends, the *basic* stickers pay significantly less than stickers with identity label, which provides novel evidence on the persistence of treatment effects. Understanding such effects is crucial in settings where policy makers are unable to treat individuals repeatedly for example for logistical or financial reasons since the long-term effects may dramatically influence the overall effectiveness (Sunstein, 2016). Frey and Rogers (2014) argue that if identity affects behavior and identity is malleable, then treatment effects may persist as long as the change in identity is retained. This study thereby contributes to this literature by providing apples-to-apples evidence that self-perception changing interventions may indeed change behavior persistently.

The post-experimental surveys provide evidence that identity concerns are an important motivation to pay for water. This is not only the most commonly mentioned reason for making payments in open-ended answer questions, but also found in changes of customers' self-perception in a telephone survey. The survey also rules out several alternative mechanisms that could explain the high effectiveness of the negatively framed identity labels as for example changed beliefs on sanctioning, monitoring, social norms or reminders. Instead, the large effectiveness of the negatively framed stickers correlates with the largest change in self-perception among all sticker types which is consistent with behavioral theories suggesting that individuals identity is regulated more based on what makes them distinct rather than what they have in common with others (Blanton, Stuart and Van den Eijenden, 2001; Bénabou and Tirole, 2006).

There are many other domains in which identity labels could be used to encourage civic behavior as for example voter mobilization, avoiding littering, volunteering for public services, driving carefully or paying taxes. Some of these domains might be even more promising than payment of utility bills since they constitute behaviors that might be more intuitively associated with being a responsible citizen. The results of this study could also be interesting beyond policymaking as identity labels are frequently used in everyday communication without the intention to influence behavior. Individuals are for example labeled as nerd, millennial or techie which might affect how they see themselves and thereby their educational, career and health choices.

Lastly, traditional tools like disconnecting non-paying customers from the supply network are not only technically and legally difficult to enforce, they may also be politically unwanted and ethically objectionable. In contrast, the use of appropriate behavioral interventions is appealing. In this study, the sticker intervention is not only well accepted among the population, it is also highly cost efficient. The best performing sticker type ("Please don't be an irresponsible citizen") increases payments by about 7 Euro per customer over the span of a year at a cost of merely 12 cents. This suggest a back-of-the-envelope return-on-investment of around 5800%.

4.7 Appendix to Chapter 4

Figure A3: Typical bill delivery process. Step 1: Reading the water meter. Step 2: Entering meter reading and printing the invoice using handheld device. Step 3: Placing the invoice on door handle or in door crack. The placement does not draw attention to invoice.

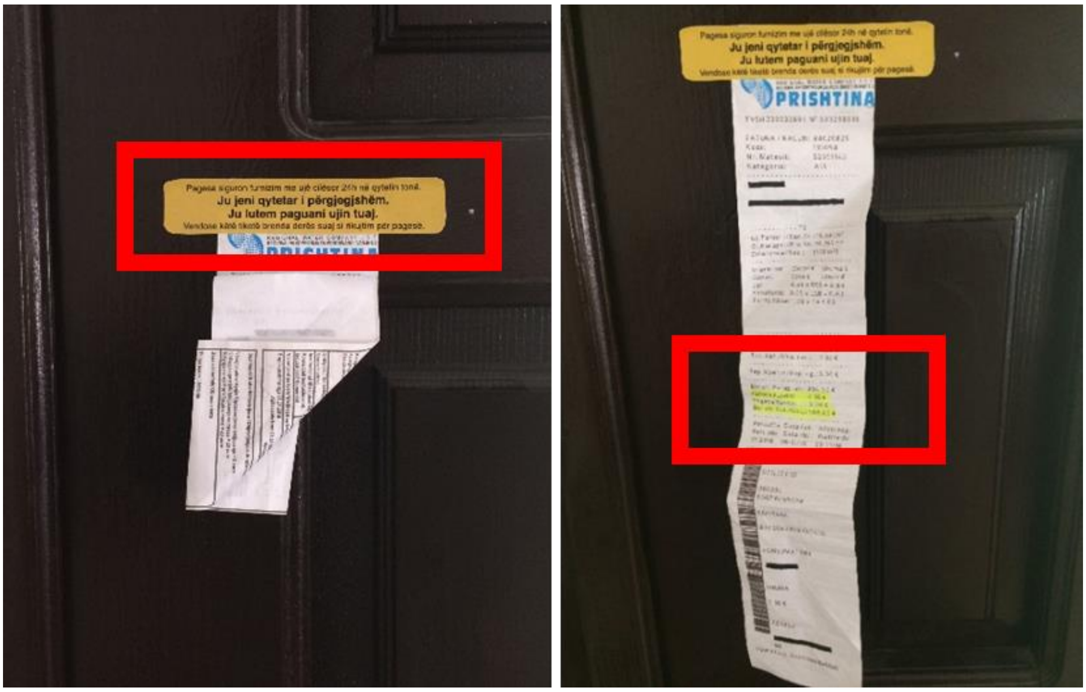


Figure A4: Treatment implementation. Left picture shows the sticker along with the folded invoice placed on a door. The content cannot be seen by third parties. Right picture shows the highlighted numbers on the invoice if the invoice was unraveled.

Table A1. Treatment Effects with Controls

	Intervention			Post-Intervention		
	(1) Payment Propensity	(2) Log Collection Efficiency>0	(3) Combined Effect	(4) Payment Propensity	(5) Log Collection Efficiency>0	(6) Combined Effect
PB	0.037**	0.042	0.146**	-0.002	-0.068**	-0.078**
	(0.017)	(0.052)	(0.062)	(0.014)	(0.029)	(0.038)
NB	0.039**	0.036	0.142**	-0.015	-0.049*	-0.078**
	(0.015)	(0.048)	(0.066)	(0.018)	(0.028)	(0.038)
PP	0.016	0.059	0.114*	0.014	-0.019	0.001
	(0.019)	(0.044)	(0.066)	(0.017)	(0.030)	(0.037)
NP	0.055***	0.061	0.213***	0.027	-0.026	0.012
	(0.020)	(0.050)	(0.067)	(0.019)	(0.033)	(0.040)
PA	0.020	0.031	0.089	0.006	0.004	0.014
	(0.020)	(0.050)	(0.068)	(0.018)	(0.034)	(0.041)
NA	0.058***	0.071	0.234***	0.004	-0.035	-0.033
	(0.019)	(0.055)	(0.064)	(0.015)	(0.032)	(0.038)
Obs.	11,808	7,028	11,808	11,794	9,182	11,794
R-squared	0.162	0.126	-	0.212	0.057	-

Notes: Table A1 reports average treatment effects on collection efficiency for the treatment periods (June-August 2017) in comparison to the untreated control group. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports average treatment effect on the intensive margin. Regression 3 combines the effects of regressions 1 and 2 using a two-part model (Belotti et al. 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment values of collection efficiency, whether any payments were made, average invoice, average debt, average payments, dummies whether the customer has a shared water meter, mailbox or was met personally during the intervention as well as the strata used for randomization (invoicer and date dummies) in the regression as recommended by Bruhn and McKenzie (2009). Control variables are winsorized at the 99.9th percentile. All reported standard errors are clustered at the invoicer-date level (level of randomization). * p<0.1; ** p<0.05; *** p<0.01

Table A2. Differences in Treatment Effects Between Positive and Negative Stickers

	(1) Payment Propensity	(2) Log Collection Efficiency>0	(3) Combined Effect
Sticker	0.057***	0.027	0.175***
	(0.018)	(0.048)	(0.057)
Sticker * identity label	-0.021	0.001	-0.050
	(0.015)	(0.029)	(0.042)
Sticker * identity label * negative	0.037**	0.030	0.131***
	(0.016)	(0.033)	(0.047)
Observations	11,808	7,028	11,808
R-squared	0.133	0.115	-

Notes: Table A2 test how treatment effects differ between positive and negatively framed identity stickers during the treatment period during the treatment periods (June-August 2017). To test for differences, I use a regression with *sticker* indicating all stickers. *Identity label* indicates stickers containing an identity label and *negative* indicates whether the identity label is negative. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports average treatment effect on the intensive margin. Regression 3 combines the effects of regressions 1 and 2 using a two-part model (Belotti et al. 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment values of collection efficiency and whether any payments were made in the pre-intervention year as well as the strata used for randomization (invoicer and date dummies) in the regression as recommended by Bruhn and McKenzie (2009). Control variables are winsorized at the 99.9th percentile. All reported standard errors are clustered at the invoicer-date level (unit of randomization). ** p<0.05; *** p<0.01

Table A3. Pooled Differences in Post-intervention Treatment Effects

	(1) Payment Propensity	(2) Log Collection Efficiency>0	(3) Combined Effect
Sticker	0.005	-0.052**	-0.051
	(0.017)	(0.025)	(0.033)
Sticker*Positive identity label	0.016	0.051**	0.084***
	(0.013)	(0.025)	(0.027)
Sticker*Negative identity label	0.022*	0.031	0.069***
	(0.012)	(0.024)	(0.027)
Observations	11,794	9,182	11,794
R-squared	0.157	0.046	-

Notes: Table A3 test how treatment effects differ between positive and negatively framed identity stickers for the post-intervention period (September-May 2018). To test for differences, I use a regression with *sticker* indicating all stickers. *Identity label* indicates stickers containing an identity label and *negative* indicates whether the identity label is negative. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports average treatment effect on the intensive margin. Regression 3 combines the effects of regressions 1 and 2 using a two-part model (Belotti et al. 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment values of collection efficiency and whether any payments were made in the pre-intervention year as well as the strata used for randomization (invoicer and date dummies) in the regression as recommended by Bruhn and McKenzie (2009). Control variables are winsorized at the 99.9th percentile. All reported standard errors are clustered at the invoicer-date level (unit of randomization). * p<0.1; ** p<0.05; *** p<0.01

Table A4. Post-intervention Sticker Differences

	(1) Payment Propensity	(2) Log Collection Efficiency>0	(3) Combined Effect
NB+NP+NA	0.004	-0.042	-0.042
	(0.020)	(0.029)	(0.038)
NP	0.032	0.031	0.085**
	(0.020)	(0.034)	(0.039)
NA	0.015	0.011	0.036
	(0.018)	(0.037)	(0.040)
PB+PP+PA	0.006	-0.061**	-0.059
	(0.018)	(0.030)	(0.039)
PP	0.020	0.052	0.090**
	(0.016)	(0.033)	(0.036)
PA	0.010	0.068*	0.092**
	(0.016)	(0.035)	(0.039)
Observations	11,794	9,182	11,794
R-squared	0.1574	0.046	-

Notes: Table A4 test heterogeneous effects with respect to stickers with or without identity label for the post-intervention period (September-May 2018). N1+N2+N3 pools all negative framed stickers. P1+P2+P3 pools all positively framed stickers. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports average treatment effect on the intensive margin. Regression 3 combines the effects of regressions 1 and 2 using a two-part model (Belotti et al. 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment values of collection efficiency and whether any payments were made in the pre-intervention year as well as the strata used for randomization (invoicer and date dummies) in the regression as recommended by Bruhn and McKenzie (2009). Control variables are winsorized at the 99.9th percentile. All reported standard errors are clustered at the invoicer-date level (unit of randomization). * p<0.1; ** p<0.05

Table A5. Changes in Self-Perception in Comparison to Untreated Group

	(1) Self-Perception "How often do you think of yourself as responsible citizen?"
Basic	0.066
	(0.310)
Positive Identity Label	-0.660**
	(0.326)
Negative identity Label	-0.619*
	(0.323)
Collection efficiency>100% (ca. median)	0.299
	(0.353)
Basic*Collection efficiency>100%	-0.332
	(0.440)
Pos. Identity Label*Collection efficiency>100%	0.519
	(0.450)
Neg. identity Label*Collection efficiency>100%	0.089
	(0.445)
Observations	744
Pseudo R2	0.0150

Notes: This table shows heterogeneous effects with respect to customers' self-perception depending on whether they paid fully during the intervention or not, which coincides with a median split. Ordered logit regression model with robust standard errors in parentheses and no additional controls ** p<0.05, * p<0.1

Table A6. Heterogeneous Treatment Effects by Water Meter Type

	(1) Payment Propensity	(2) Log Collection Efficiency>0	(3) Combined Effect
PB	0.093	0.170	0.449***
	(0.069)	(0.188)	(0.155)
NB	0.093**	-0.001	0.228
	(0.042)	(0.186)	(0.158)
PP	0.193***	0.043	0.531***
	(0.046)	(0.193)	(0.161)
NP	0.088**	0.229	0.513***
	(0.041)	(0.182)	(0.159)
PA	0.121***	0.124	0.458***
	(0.043)	(0.183)	(0.146)
NA	0.155**	0.133	0.554***
	(0.068)	(0.194)	(0.174)
Individual meter	0.150***	0.050	0.434***
	(0.038)	(0.206)	(0.136)
PB* Individual meter	-0.062	-0.157	-0.355**
	(0.072)	(0.207)	(0.170)
NB* Individual meter	-0.046	0.028	-0.077
	(0.048)	(0.207)	(0.172)
PP* Individual meter	-0.197***	0.008	-0.476***
	(0.054)	(0.221)	(0.176)
NP* Individual meter	-0.025	-0.211	-0.333*
	(0.049)	(0.206)	(0.174)
PA* Individual meter	-0.105**	-0.146	-0.447***
	(0.047)	(0.208)	(0.163)
NA* Individual meter	-0.106	-0.072	-0.355*
	(0.074)	(0.215)	(0.182)
Observations	11,808	7,028	11,808
R-squared	0.137	0.116	-

Notes: This table shows heterogeneous treatment effects during the intervention period (June-August 2017) with respect to owning a shared versus an individual water meter. Regression 1 shows the effect of the treatments on the likelihood of making a payment (marginal effects at means). Regression 2 reports average treatment effect on the intensive margin. Regression 3 combines the effects of regressions 1 and 2 using a two-part model (Belotti et al. 2015). Fitted values from the log transformation of the two-part model are obtained using Duan's (1983) smearing retransformation and standard errors are obtained by bootstrapping. I control for the pre-treatment values of collection efficiency and whether any payments were made in the pre-intervention year as well as the strata used for randomization (invoicer and date dummies) in the regression as recommended by Bruhn and McKenzie (2009). Control variables are winsorized at the 99.9th percentile. All reported standard errors are clustered at the invoicer-date level (unit of randomization). * p<0.1; ** p<0.05; *** p<0.01

CHAPTER 5.

THE SELF-SERVING BEHAVIOR OF THE RICH CAUSES CONTAGION EFFECTS AMONG THE POOR

Joint work with Bettina Rockenbach and Arne Weiß

5.1 Introduction

In many societies wealth and economic power is concentrated among a small group. As manifested in a wide range of religious, philosophical and political thought, social norms ask this rich group to serve society and to share their wealth with the less fortunate.[40] However, the rich often fail to comply with this expectation. This is, arguably, a particular concern in many developing countries in which the rich are held less accountable and constrained by institutional checks and balances (e.g., Acemoglu and Robinson, 2012; Besley and Perrson, 2014). This allows the rich to engage in multiple ways of self-serving behavior, for example in form of corruption (Bardhan, 2006), elite capture (e.g., Platteau and Gaspart, 2003; Bardhan and Mookherjee, 2006) and tax evasion (e.g., "Paradise Papers").[41] With their disproportionate influence on economic and political outcomes, the rich's self-serving behavior has obvious direct consequences for economic development. In this paper, we ask whether there is an additional indirect effect when other members of society learn about self-serving behavior of the rich. The answer to this question is largely empirical because the theoretical literature (which we review in section 3) yields conflicting predictions.

We report the results of a framed field experiment in Namibia designed to provide causal evidence on potential behavioral contagion effects from the rich's self-serving behavior to the behavior of the poor. Moreover, by experimental design as well as through

[40] An early manifestation of this expectation can be found in the bible "From everyone who has been given much, much will be demanded" (Luke 12:48). A moral obligation of the rich to share with society can also be found in Zakat in the Quran, in the writings of many great philosophers such as Aristotle, Bentham, Marx, Rawls and Smith and in the widely used progressive income taxation.

[41] More information can be found under https://www.icij.org/investigations/paradise-papers.

complementary qualitative evidence, we disentangle them from mere conformity effects. Such conformity effects could stem from the poor's expectation that the rich possess better information (Vesterlund, 2003) or the poor's desire to imitate the rich in order to signal a higher status (Kumru and Vesterlund, 2010). Reliably ruling out such confounds requires varying both the reference group and the content of the information about the behavior of others. The existing empirical literature, however, does not provide such variations and thus does not allow a conclusive answer to our question. It either predominantly focuses on reference groups that are of the same socioeconomic status (Bicchieri and Xiao, 2010; Cason and Mui 1998; Croson and Shang, 2009; Frey and Meyer, 2004; Iriberri and Rey-Biel, 2013; Krupka and Weber, 2009; Shang et al., 2008; Shang and Croson, 2009) or does not vary the content of the information provided (Ebeling, Feldhaus and Fendrich, 2017; Kumru and Vesterlund, 2010), i.e., whether the rich behaved self-servingly or pro-socially. Furthermore, these studies are conducted exclusively in developed countries, typically in student subject pools. To close this gap, we measure the behavior of participants from a poor neighborhood after they received information about other's self-serving or pro-social behavior, either from a rich neighborhood or their poor neighborhood. Thus, in a controlled way we vary both the socioeconomic status of the comparison group (rich and poor) and their reported behavior (self-serving and pro-social).

Our framed field experiment (Harrison and List, 2004) is conducted in Namibia, one of the countries with the highest income inequality worldwide (World Bank, 2014). Differences in wealth are therefore pronounced and naturally existing. Thus, our experiment combines strict experimental control with a high degree of realism (Levitt and List, 2009). We find that information about the egoistic behavior of the rich induces the poor to act significantly more egoistically. By contrast, neither the rich's pro-social behavior nor information on how other poor individuals behaved affects the poor. The spread of egoism is therefore not a simple conformity effect. We show that it is instead caused by a change in injunctive norms: the rich's (unexpected) failure to share leads other members of society to view their own egoistic behavior as more acceptable.[42] Thus, the rich's egoistic behavior causes a double damage: Society not only suffers from their low contributions but also indirectly from contagion effects.

[42] Other self-serving exploitations of the decision context have been found in situations of uncertainty: for example about the consequences of one's decisions (e.g., Dana, Weber, Krupka, 2006; Haisley and Weber, 2010; Exley, 2015), the responsibility for an outcome (Bartling and Fischbacher, 2012) or the relevant norms applying to a decision situation (Spiekermann and Weiss, 2016; Charness, Naef, and Sontuoso, 2018).

5.2 Experimental Design

5.2.1 Experimental Set-up

Our experimental study took place in the poor Ombili neighborhood in Windhoek, Namibia's capital and largest city. The participants received information about the sharing behavior from the rich Klein-Windhoek neighborhood or their own neighborhood. Wealth differences are not only apparent in administrative statistics (Table A1) but also dominated how our experimental participants perceived the rich neighborhood (Table A2).

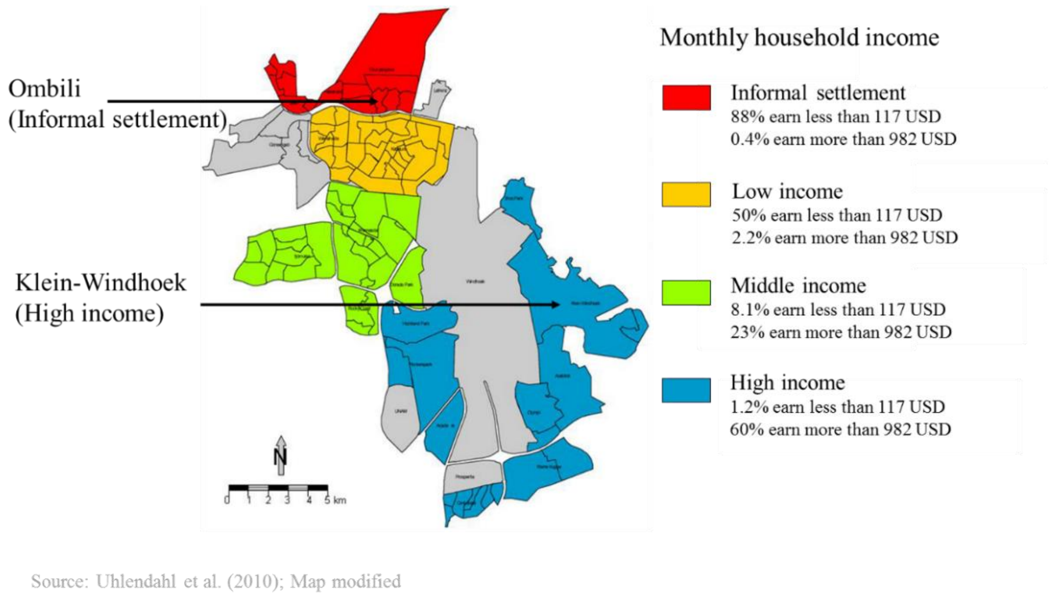


Figure 5. Map of Windhoek by income level (Uhlendahl et al., 2010). The poor receive social information from either the impoverished Ombili neighborhood or the rich Klein-Windhoek neighborhood.

In order to conduct the experiment, we set up a new laboratory at the Ombili community center. Before running the experimental sessions, we informed the community leaders as well as the district counsellor about our experimental protocol without disclosing any research hypotheses. The Ministry of Home Affairs issued a research visa. We employed a team of nine Namibian research assistants, all fluent in English. Six were also fluent in Oshiwambo, the most common language in the Ombili neighborhood. Most assistants were native speakers of several other languages spoken in Windhoek, which helped overcome potential problems in understanding. The research assistants were not informed about the hypotheses underlying our experimental study either.

To recruit participants, the research assistants approached potential participants on the streets and asked them to participate in an experiment. If someone agreed, the recruiter would put her or his signature on a “ticket” allowing participation in the experiment within the next 40 minutes. The research assistants signed up 349 participants within a week in November 2014. We did not allow persons without a ticket to take part in this experiment in order to avoid contaminaton effects between experimental sessions. Furthermore, the research assistants never issued tickets to persons that asked for a ticket themselves. They also changed their recruiting locations within the Ombili neighborhood regularly.

To avoid security concerns regarding cash handling in the field laboratory, participants were paid in air-time for their mobile phones. The requirement that only owners of a cell phone could participate is not particularly restrictive, as most people own a cell phone even in poor neighborhoods (see Table A1). Air-time is a credit for pre-paid mobile phones, which can be used for texting, phone calls, internet data packages as well as small purchases. Air-time can easily be transferred for free from one user to the next. All our experimental participants knew and used air-time. We transferred all transactions in the afternoon of the same day that the experimental session took place. To strengthen participants’ trust in the transaction, they were informed that they could come the following day in case they would not receive their air-time. No participant complained about the payment process throughout the whole experiment.

We required participants to be able to read and write on their own in order to allow for full anonymity. The instructions and questions were pre-tested to check and improve understandability of the experiment. We offered sessions in English and Oshiwambo, which were run simultaneously by our Namibian research assistants in two separate rooms. When preparing the instructions, the English version were translated into Oshiwambo and back to English several times to ensure that instructions were as identical as possible in both languages. We showed participants how to mark their decisions on an exemplary sheet to ensure understanding. Participants made their choices privately in cubicles made out of cardboard (see picture in A10) and turned in all completed paper sheets into urns by themselves. We never recorded any names of participants.

5.2.2 Treatments

Participants played a modified version of the dictator game (Forsythe, Horowitz, Savin and Sefton, 1994), in which they could choose between two options to share an amount of 60N$ (~5US$) between themselves and the Disaster Management division of the widely known Namibian Red Cross (NRC). Option A gave participants 55N$ and 5N$ to the NRC. For readability purposes, we will label this option EGO ("egoistic") throughout the rest of the paper. Option B gave 30N$ to participants and 30N$ to the NRC. We will label this option PS ("pro-social"). The endowment of 60N$ corresponded to roughly four hourly wages to participants (Namibian Statistics Agency, 2016). By choosing the Disaster Management division as a recipient, we avoided that neither subject group directly benefited from the contributions. Prior to the experiment, we signed a contract with the Namibian Red Cross (NRC) that stated our obligation to give money to the Namibian Red Cross Disaster Management Division according to the decisions of the participants. This contract was shown to all participants before the experiment to reaffirm participants that their decisions were not hypothetical. No participant expressed unfamiliarity with the Red Cross. We studied five treatments with a total of 349 participants (see Table 1) to identify the underlying mechanisms behind any changes in behavior.

Table 1. Treatment Overview (N=349)

Reference Group	No Information	Majority EGO	Majority PS
No Reference	Baseline (N=81)		
Rich Neighborhood (Klein-Windhoek)		Rich EGO Treatment (N=72)	Rich PS Treatment (N=50)
Poor Neighborhood (Ombili)		Poor EGO Treatment (N = 86)	Poor PS Treatment (N=60)

In the *Baseline* treatment, participants did not receive any information about the behavior of others. In the other four treatments, participants received, prior to taking their own decision, information about the majority behavior of participants either from the rich neighborhood (Klein-Windhoek) or from their own neighborhood (Ombili). In the Rich EGO treatment, we informed participants that in "a session we recently conducted [in Klein-Windhoek] most participants chose option A." In the Poor EGO treatment, the bracket term was substituted by "here at Ombili Community Center." In the Rich PS treatment, we informed that in "a session we conducted recently [in Klein-Windhoek], the majority of participants decided for option B." In the Poor PS treatment, the bracket term was substituted by "here at Ombili Community Center."

This 2x2 experimental design leads to the four information treatments Rich EGO, Rich PS, Poor EGO and Poor PS. The social information was not deceptive. As in previous literature (e.g., Biccieri and Xiao, 2009), it referred to a selectively chosen experimental session, which we conducted in Ombili and also in Klein-Windhoek for the purpose of gathering the social information for the experiment.

Table 2. Demographics Across Treatments

	Baseline	Rich EGO	Rich PS	Poor EGO	Poor PS
Age	25.98	26.76	25.38	25.32	23.9
	(8.52)	(7.28)	(9.41)	(7.26)	(6.57)
Female	0.4	0.35	0.54	0.38	0.45
	(0.49)	(0.48)	(0.50)	(0.49)	(0.50)
No Income	0.62	0.61	0.75	0.65	0.67
Corrugated Iron	0.63	0.67	0.67	0.71	0.57
Secondary School	0.71	0.65	0.8	0.72	0.65
People in HH	6.1	5.99	6.18	6.01	5.9
	(3.26)	(2.6)	(2.7)	(3.57)	(2.4)
Session in Oshiwambo	0.54	0.47	0.58	0.51	0.35**
	(0.50)	(0.53)	(0.50)	(0.50)	(0.48)
Observations	81	72	50	86	60

Notes: No Income indicates whether the participant has regular income or not. Secondary school indicates completion of secondary school. We report the percentage share of the modus for ordinally scaled variables income level, materials of which house is constructed and education. Interval variables show means with standard errors in parenthesis. For interval variables, we use t-tests (allowing for unequal variances). For binary categorical variables, we use the Chi-Square tests. For ordinally scaled variables, we use a Mann-Whiney-U-test. All tests are two-sided. ** indicates a statistically significant difference (p<0.05) in comparison to the Baseline.

The experimental sessions ended with a questionnaire in which we elicited beliefs about social norms, participants' perceptions of both comparison groups (see Tables A2-A6), as well as demographics (Table 2). Most importantly, we measured injunctive norms (Cialdini, Reno and Callgreen, 1990) by asking "Which option do you think SHOULD be chosen by participants here?" In order to clarify the concept of injunctive norms to the participants, we gave the examples that one should not litter and that one should not drink and drive. Further, we asked whether and, if so, why participants were surprised by the information given to them. The questionnaire data allows us to better understand the motives guiding behavior and to test whether participants indeed expected the rich to share with society. Further details can be found in the appendix sections *experimental design* and *instructions*.

Participants' demographics in our treatments are displayed in Table 2. About two thirds of our participants live in shacks built from corrugated iron and stated to have no monthly income. In contrast, the monthly median income in the rich district is between 1100-2000 USD (Uhlendahl et al., 2010). The sample is overall well balanced between treatments. The only variable that was statistically significant different in comparison to the Baseline was the share of participants who participated in Oshiwambo in the Poor PS treatment.[43]

[43] We control for this and other covariates in the regression analysis (Table 3) which does not change our main result.

5.3 Predictions

The literature on social comparisons leads to inconclusive predictions for our research question: A first strand predicts conformity to the ingroup and divergence from the outgroup based upon the assumption that conforming and diverging behavior is driven by psychological and socioeconomic closeness to comparison groups (Mussweiler, 2001; Mussweiler, 2003). A social identity perspective yields the same prediction as participants might aim to maintain a distinct positive social identity: “Whatever they are, we are not” (Tajfel and Turner, 1986; Hogg, 2006). A second strand of literature, by contrast, argues for conformity to the rich group because of status seeking and upward assimilation: Conforming to the rich allows the poor to conclude that they are among the “better ones” (Collins, 2000; Kumru and Vesterlund, 2010). Importantly, both strands predict conformity to either group regardless of the content of the social information. For example, if behavior is driven by a status-seeking motivation, we should observe conformity to the rich in both the Rich EGO *and* the Rich PS treatment. However, if it is learning about a social norm violation that influences the poor’s behavior, we should observe an effect specific to a certain comparison group *and* a certain behavior, i.e., only when the rich behaved egoistically. Our 2x2 experimental design allows separating between mere conformity effects and effects driven by a social norm violation and thus identifying the behavioral consequences of the rich’s self-serving behavior for the rest of society.

5.4 Results

In the baseline treatment, absent any comparison information, 59% of the participants in the poor neighborhood Ombili choose the egoistic option (see Figure 2A). When subjects are informed about the egoistic behavior of the rich, the share of egoistic choices significantly increases to 75% (Baseline vs. Rich EGO treatment $\chi^2(1,153)=4.25$, p=0.039). This means that the number of pro-social choices decreases by almost 40% and the total amount donated by about 26%. However, when the poor receive the information that the rich behaved predominantly pro-socially, their behavior does not change compared to the baseline treatment (Baseline vs. Rich PS $\chi^2(1,131)=0.02$, p=0.887). The share of pro-social choices is neither significantly different when subjects are informed that the poor behaved egoistically (Baseline vs. Poor EGO treatment $\chi^2(1,167)=1.20$, p=0.272) nor when subjects receive information that the poor behaved pro-socially (Baseline vs. Poor PS treatment $\chi^2(1,141)=0.08$, p=0.773).

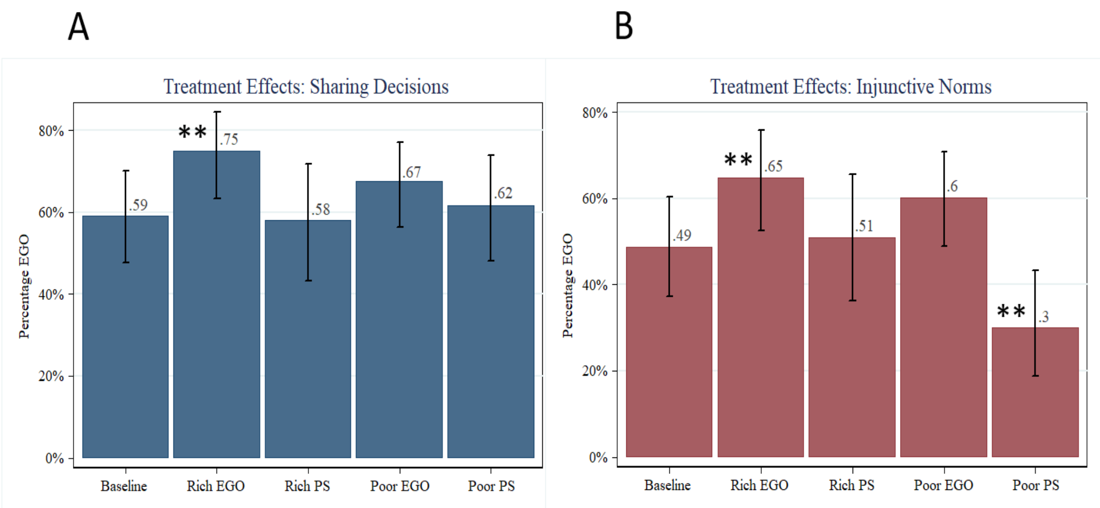


Figure 2. Bars in **Panel A** show the share of egoistic choices across treatments and binomial 95% confidence intervals (Clopper-Pearson). Bars in **Panel B** indicate the share of participants stating that the egoistic option should be chosen by participants (injunctive norms). ** indicates a statistically significant difference (p<0.05) of a two-sided chi-squared-test in comparison to the Baseline.

Table 3 reports parametric regression results. The regressions allow us to control for possible between-treatment differences in participants' demographics and, by estimating robust standard errors clustered at the session level, to account for possible session effects (Frechette, 2012). In models 1 and 2, we run a probit regression with dummy variables for the four social information treatments. The regression table shows marginal effects at means. Model 1 shows that without control variables the probability of choosing the egoistic option is 16 percentage points higher (p=0.02) in the Rich EGO treatment than in the baseline treatment. This difference increases to 18 percentage points when adding controls (p=0.004). The coefficients for the other treatments always remain insignificantly different from zero. Note that a Wald test rejects equality of coefficients between Rich EGO and Poor EGO (p=0.19 without controls and p=0.003 when adding controls), suggesting that the participants are more likely to choose the egoistic option when the information about egoistic behavior comes from the rich instead of the poor reference group.

Since participants' behavior is only significantly changed by information about the egoistic behavior of the rich, our results can neither be explained by stronger conformity to groups that are similar (Fatas, Heap and Arjona, 2018; Gino, Ayal and Ariely, 2009; Jetten, Spears, and Manstead, 1996; Shang, Reed and Croson, 2008) nor by a status-seeking motivation (Ebeling, Feldhaus and Fendrich, 2017; Kumru and Vesterlund, 2010). The absence of a visible status-seeking motivation in our experimental results corroborates earlier findings that status concerns are less important among the very poor (Akay, Martinsson and Medhin, 2012). Instead, our data suggest that only the self-serving behavior by the rich spreads among other members of society.

How does the social information change participants' beliefs about what *should* be done? The bars in Figure 2B correspond to participants' personal injunctive norms. In the baseline treatment, 47% participants state that EGO should be chosen. When being informed that the rich behaved egoistically (Rich EGO treatment), a statistically significantly larger share of participants (65%) state that the egoistic option should be chosen ($\chi^2(1,149)=3.90$, p=0.048). The injunctive norm is unaffected in the Rich PS treatment ($\chi^2(1,127)=0.063$, p=0.801). Note that these results are consistent with the existence of a social norm that asks the rich to share with society. Learning that the rich share merely confirms what the poor expect of this group anyway. By contrast, learning about the egoistic behavior of the rich allows the poor to justify their own egoistic behavior: "If even the rich do not share, why should I?" Interestingly, in the Poor PS treatment injunctive norms also change ($\chi^2(1,138)=4.93$, p=0.026), but, as noted above, actual behavior is unaffected.

Table 3. Probability to Choose Egoistic Option

	EGO Choice (1)	EGO Choice (2)
Rich EGO	0.163**	0.180***
	(0.070)	(0.066)
Rich PS	-0.012	-0.037
	(0.123)	(0.111)
Poor Ego	0.081	0.088
	(0.067)	(0.069)
Poor PS	0.023	0.014
	(0.106)	(0.097)
Observations	349	344
Pseudo R-Squared	0.013	0.033
Controls	No	Yes
Number of Session	25	25

Notes: The table reports marginal effects at means from a probit regression on the probability of choosing the egoistic option in comparison to the baseline. Controls include age, gender, a dummy controlling for participants reporting no income, a dummy controlling for the materials of which a house is build, e.g., corrugated iron or brick, a dummy variable controlling for the type of toilet, a dummy controlling for education and a dummy controlling whether the session was in English or Oshiwambo. Standard errors are clustered by experimental session.** p<0.05; *** p<0.01

The questionnaire data supports the interpretation of the main result in light of a social norm violation. First, participants overwhelmingly associate people in Klein-Windhoek with wealth (Table A2). Second and more importantly, the poor expect the rich to decide pro-socially, as 60% of participants in the Rich EGO treatment are surprised by that information. This exceeds the respective shares in the other treatments, where the level of surprised participants is between 42% and 44% (Rich EGO vs. Poor EGO: (χ^2 (1,152)=4.56, p=0.033); Rich EGO vs. Poor PS: (χ^2 (1,130)=3.52, p=0.061); Rich EGO vs. Rich PS: (χ^2(1,119)=4.10, p=0.043). In a qualitative content analysis (Mayring, 2000), three independent raters classified participants' statements on why they were surprised. Krippendorff's Alpha (Hayes and Krippendorf, 2007) indicates satisfactory inter-coder reliability. Roughly 25% of the statements were classified as "The rich should share." Another 18% of participants explicitly relate their surprise to the wealth of the people in Klein-Windhoek (e.g., "yes, because they are rich"). Another 23% of participants merely reiterate that they were surprised (Krippendorff's Alpha: 0.74). The statements of participants in the Rich PS treatment who were not surprised paint a similar picture: 50% state that the rich have enough money to donate and 19% declare that the rich have the responsibility to donate. 19% state that PS is the right option (Krippendorff's Alpha: 0.8). In contrast, statements from the Poor EGO and Poor PS treatment do not indicate a role-specific norm for the poor to share.

5.5 Conclusion

We show that the responsibility of the rich goes beyond material contributions: They set moral benchmarks. The failure of the rich to share with society causes a double damage: Society not only directly suffers from low contributions by the rich but also from a contagion effect. In our experiment, this effect was strong enough to change the allocation by an equivalent of about two hourly wages among a substantial share of participants. Further, we show that the spread of egoism is not a simple conformity effect, but caused by a change in injunctive norms: the rich's (unexpected) norm violation triggers a change in what the others' think they should do towards egoism. Understanding this mechanism is particularly important for developing countries, where limited state capacity makes society even more depend on the voluntary contributions of all members of society.

We believe that our findings may be relevant to a wide range of settings, such as tax compliance, environmental protection or performing civic duties. Taking the example of tax compliance, Traxler (2010) models the spillover effects of the tax evasion of moral reference groups on the tax compliance of others, which can lead to a large reduction in overall tax revenue. Our experimental results show that the rich may be a particularly influential reference group because their norm violation allows others to feel legitimized for their own egoistic behavior. In times when information travels fast and the behavior of the rich and powerful is at the center of (social) media attention, the detrimental consequences of their self-serving behavior can quickly amplify.

5.6 Appendix for Chapter 5

Table A1. Wealth and Infrastructure Differences Between the Comparison Groups

Household owns / has access to	Improvised Housing Unit (Shack)	Cement Block/ Brick Houses	Internet at Home	Piped Water Inside	Electricity for Lighting	Mobile Telephone
Ombili (Tobias Hainyeko)	65.8 %	14.8 %	2.4 %	13.2 %	27 %	68.8 %
Klein-Windhoek (Windhoek East)	0.4 %	97.1 %	57.6 %	97.5 %	99 %	87.2 %

Source: Namibian Statistics Agency (2011).

Table A2. First associations with Klein-Windhoek (rich neighborhood)

Category Classification	Percentage Share
People in Klein-Windhoek are rich and have high living standards	55.9%
Positive Characteristics (good, helpful, respect)	15.7%
Difference not related to wealth	8.3%
Expressing questions / missing knowledge	9.3%
Other type of association	7.9%
Incomprehensible statement	4.2%

Notes: We classify first associations with Klein-Windhoek given by participants in the baseline treatments, who did not receive social information.

Table A3. First associations with own neighborhood.

Category Classification	Percentage Share
People live in poverty. Bad standard of living. Help/change is needed	53.7%
Some people misbehave	9%
Good people that respect each other	22.4%
Other reasons	9%
No idea / Not understandable	6%

Notes: We classify first associations with Klein-Windhoek given by participants in the baseline treatments, who do not receive social information

Table A4. Amount of "good people" in Klein-Windhoek vs Ombili

Amount of Good People (N= 81)	Own Neighborhood	Klein-Windhoek
Most	17	28
Many	17	12
Some	29	29
Few	16	9
None	2	3

Notes: We find no difference in the share of "good people" in Klein-Windhoek in comparison to their own neighborhood according to the views of participant in the Baseline (two-sided Friedman test, p=0.5785).

Table A5. Perceived similarity with people from either neighborhood

Perceived Similarity (N= 81, Baseline sample)	Own Neighborhood	Klein-Windhoek
Very Similar	17	6
Similar	23	14
Different	13	15
Very Different	22	25
None of the above	6	21

Notes: Participants perceive themselves to be more similar to people from own neighborhood (two-sided Friedman test, p=0.0027).

Table A6. Average empirical and normative expectations for different neighborhoods across treatments.

	Empirical expectations for the participant's own session	2nd order normative expectations for own session	Empirical expectations for Klein-Windhoek	2nd order normative expectations for Klein-Windhoek
Baseline A	70.8%	72.5%	82.3%	72.5%
	(30.5%)	(27.1%)	(24%)	(31.2%)
Baseline B	64.9%	54.4%	47.8%	45.6%
	(34.8%)	(38.3%)	(35.7%)	(38.4%)
Poor EGO	81.4%	78.2%	69.5%	68.2%
	(23%)	(30%)	(31.9%)	(33.3%)
Poor PS	50.7%	40.8%	41.3%	37.2%
	(33.7%)	(34.2%)	(31.3%)	(30.0%)
Rich EGO	87.9%	86.9%		85.6%
	(18.6%)	(20.5%)		(20.3%)
Rich PS	46.5%	40%		22.6%
	(30.4%)	(33.4%)		(22.4%)

Notes: Standard errors in parentheses. Empirical expectations for Klein-Windhoek were not measured in the Rich EGO and Rich PS treatment because empirical information was already given to them as a treatment manipulation. Empirical Expectations: "Please guess: What share of participants here in this session decided for Option A? You will receive an additional N$ 1 in air time for a right guess." 2nd Order Normative Expectations in this session: "Please guess: What share of participants here in this session thinks that option A SHOULD be chosen. You will receive an additional N$ 1 in air time for a right guess. "Empirical Expectations for Klein Windhoek: "What share of participants in Klein-Windhoek decided for option A?"2nd Order Normative Expectations for Klein Windhoek: "Please guess: What share of participants in Klein-Windhoek in the mentioned session thinks that option A SHOULD be chosen? You will receive an additional N$ 1 in air time for a right guess."

Table A7. Probit Regression Only for “Strong” Beliefs

	EGO Choice (1)	EGO Choice (2)
Rich EGO	0.165**	0.184***
	(0.071)	(0.068)
Rich PS	- 0.038	-0.067
	(0.117)	(0.108)
Poor EGO	0.082	0.089
	(0.068)	(0.07)
Poor PS	- 0.038	-0.067
	(0.117)	(0.108)
Zero income		-0.012
		(0.06)
No brick house		0.065
		(0.069)
No flush toilet		-0.118
		(0.075)
No secondary school		-0.089
		(0.113)
Oshiwambo language		0.062
		(0.067)
Age		0.002
		(0.003)
Gender		-0.002
		(0.053)
Neighborhood Dummies		Yes
Observations	313	310
Pseudo R-squared	0.023	0.045
Number of Clusters	25	25

Notes: The table reports marginal effects at means from a probit regression on the probability of choosing the egoistic option in comparison to the baseline. Controls include age, gender, a dummy controlling for participants reporting no income (Zero income), a dummy controlling for the materials of which a house is build, e.g., corrugated iron or brick (No brick house), a dummy variable controlling for the type of toilet (No flush toilet), a dummy controlling for education (No secondary school) and a dummy controlling whether the session was in English or Oshiwambo (Session language). Standard errors are clustered by experimental session.** p<0.05; *** p<0.01

A8. Details of the Experimental Design

At the beginning of the experiment, we inform participants in all treatments that the sessions are conducted in both Klein-Windhoek (the rich neighborhood) and in the Ombili community center (the poor neighborhood). We then ask participants to write down whether they are taking part in Ombili or in Klein-Windhoek. This prepares the participants for the forthcoming social information. The fact that 98% of participants correctly answered that they are taking part in Ombili also serves as a successful check that participants were able to follow the instructions and could communicate in writing.

In the information treatments (Rich EGO, Poor Ego, Rich PS and Poor PS), participants then receive information about the behavior of participants from a previous session from their own poor neighborhood (Ombili) or of participants from the rich neighborhood (Klein-Windhoek). In the Poor EGO treatment, we inform participants that in "a session we recently conducted [here at Ombili Community Center] most participants chose option A." In the Rich EGO treatment, the term in brackets was substituted by "in Klein-Windhoek." In the Poor PS treatment, we informed that in "a session we conducted recently [here at Ombili Community Center], the majority of participants decided for option B." In the Rich PS treatment, the term in brackets was substituted by "in Klein-Windhoek." To provide the social information without deception, we ran sessions consisting of ten participants on the street in Klein-Windhoek and selected those in which the majority behaved either prosocially or egoistically. For the Poor EGO and Poor PS treatments, we used selected sessions from the baseline treatment.

After receiving information about the majority behavior of another group, participants were asked to guess how many participants in the mentioned session behaved prosocially or egoistically. Participants in the EGO treatments answered to the question "What share of participants in that session in Klein-Windhoek [Ombili Community Center] decided for Option A?" by marking their guess on a scale from 8 out of 10 to 10 out of 10. In the PS treatments, participants were asked to guess "What share of participants in that session in Klein-Windhoek [Ombili Community Center] decided for Option B?" on a scale from 6 out of 10 to 10 out of 10. The believe range elicitation is kept equal between comparison groups, the crucial element in the social comparison theories, but varies slightly between the PS and EGO treatments. We perform a robustness check of our main results by running the regression of Table 3 only with participants that had similarly strong beliefs, i.e., by dropping those observations in the PS treatments that believed that less than 80% decided for option B. Table A7 shows that the results stay qualitatively the same. The coefficient for Rich Ego is statistically different from zero (p=0.02 and p<0.01 when adding controls). In the baseline treatment, participants are neither provided with social information nor asked for their own estimates about the behavior of another group.

All participants then play a modified pen-and-paper version of the dictator game. They can choose between two options, which determine how a total of 60N$ (~5US$) will be split between themselves and the Namibian Red Cross Disaster Management, which is an established charity and well known among participants. Option A gives participants 55N$ and 5N$ to the Namibian Red Cross Disaster Management. Option B gives 30N$ to themselves and 30N$ to the Namibian Red Cross Disaster Management. The Disaster Management division of the Red Cross was chosen to avoid that participants might directly benefit themselves when giving to the charity.

The second part of the experiment measures beliefs on social norms. In line with the experimental manipulation, we elicited expectations in the two EGO treatments by asking what share of participants decided for option A and in the two PS treatments by asking for the share who decided for option B. In order to compare expectation in the social information treatments to the baseline treatment, we ran two different versions of the latter. One randomly selected half of the participants in the baseline treatment (baseline A) were asked what share decided for option A and the remaining half (baseline B) were asked for the share that decided for option B. The belief elicitations were incentivized by paying an additional 1 N$ in air time for every correct guess.

We measure empirical expectations by asking “Please guess: What share of participants here in this session decided for Option A?” We then ask for personal injunctive norms by asking “Which option do you think SHOULD be chosen by participants here?“ Further, we measure 2^{nd}-order normative expectations by asking “Please guess: What share of participants here in this session thinks that option A SHOULD be chosen?” Afterwards we measure empirical expectations for Klein-Windhoek by asking “Please guess: What share of participants in Klein-Windhoek decided for option A?” and 2^{nd} order normative expectations by asking “Please guess: What share of participants in Klein-Windhoek in the mentioned session thinks that option A SHOULD be chosen?”

Participants of pilot sessions struggled with the belief elicitation by percentage shares. A debriefing revealed that several participants gave answers that were inconsistent with what they wanted to express. Many had difficulties in understanding the concept of percentage shares. Before each of the main sessions, we tried to enhance understanding of percentage shares by giving brief visual examples with apples and tomatoes. We are nevertheless still skeptical about the quality of the belief elicitation data by percentage shares and therefore focus our analysis on expectations and norms that required only binary answers. Note that difficulties of study participants in developing countries with complex psychological scales have already been thoroughly documented elsewhere (see Laajaj and Marcours, 2017). The data of the belief elicitation by percentage shares is given in Table A6.

The third part of the experiment asks participants in which neighborhood they live as well as for their first associations with people from their own neighborhood and from Klein-Windhoek. Furthermore, we measure the perceived similarity with either neighborhood and the amount of "good people" in either neighborhood. In the information treatments, we additionally asked whether and why participants were surprised by the information given to them (e.g., "Did it surprise you that in the mentioned session the majority of people in Klein-Windhoek decided for option A?")

The questionnaire ended with demographics containing questions on the materials from which their house is built, type of toilet, income per month, languages spoken, age, gender, occupation, education, how many people live in their household and whether and if so they knew details about the experiment prior to participating. Lastly, we ask whether they could be contacted for future studies and for their phone number for air-time transfers.

Below are the instructions for all treatment variations. Part one and two of the instructions vary by treatment. Part three is identical in all treatments except that in the information treatments, we additionally asked whether and why participants were surprised by the information given to them (e.g., "Did it surprise you that in the mentioned session the majority of people in Klein-Windhoek decided for option A?")

As explained before, there exist two variations for the Baseline for the second part of the instructions. In version Baseline A we elicit beliefs about norms by asking "What share of participants […] decided for option A?", whereas in Baseline B we ask "What share of participants […] decided for option B?"

[Treatment Instructions Baseline A]

Survey – First Part

This survey is being conducted in Klein-Windhoek and at Ombili Community center.
Are you taking part in Klein-Windhoek or at Ombili Community Center?

Information:
Your decision in this part of the survey will decide how a total of 60 N$ will be split between yourself and the Namibian Red Cross Disaster Management, which is a charitable organization. There are two options: option A and option B. Option A and option B are described in the following:

Option A
You receive N$ 55 and the Namibian Red Cross Disaster Management receives N$ 5
Option B
You receive N$ 30 and the Namibian Red Cross Disaster Management receives N$ 30

Now split the N$ 60 between yourself and the Namibian Red Cross Disaster Management. **Only mark one option**, otherwise you will not receive air-time. You cannot change your decision later in this survey.

☐ **Option A** You receive N$ 55
The Namibian Red Cross Disaster Management receives N$ 5

OR

☐ **Option B** You receive N$ 30
The Namibian Red Cross Disaster Management receives N$ 30

Survey – Second Part

1. Please guess: What share of participants here in **this** session decided for Option A? You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	BBBBBBBBBB
☐	1 out of 10	A BBBBBBBBB
☐	2 out of 10	AA BBBBBBBB
☐	3 out of 10	AAA BBBBBBB
☐	4 out of 10	AAAA BBBBBB
☐	5 out of 10	AAAAA BBBBB
☐	6 out of 10	AAAAAA BBBB
☐	7 out of 10	AAAAAAA BBB
☐	8 out of 10	AAAAAAAA BB
☐	9 out of 10	AAAAAAAAA B
☐	10 out of 10	AAAAAAAAAA

2. Which option do you think **SHOULD** be chosen by participants here?

☐ Option A ☐ Option B

3. Please guess: What share of participants here in **<u>this</u>** session thinks that option A SHOULD be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	BBBBBBBBBB
☐	1 out of 10	A BBBBBBBBB
☐	2 out of 10	AA BBBBBBBB
☐	3 out of 10	AAA BBBBBBB
☐	4 out of 10	AAAA BBBBBB
☐	5 out of 10	AAAAA BBBBB
☐	6 out of 10	AAAAAA BBBB
☐	7 out of 10	AAAAAAA BBB
☐	8 out of 10	AAAAAAAA BB
☐	9 out of 10	AAAAAAAAA B
☐	10 out of 10	AAAAAAAAAA

4. Please guess: What share of participants in Klein-Windhoek decided for option A? You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	BBBBBBBBBB
☐	1 out of 10	A BBBBBBBBB
☐	2 out of 10	AA BBBBBBBB
☐	3 out of 10	AAA BBBBBBB
☐	4 out of 10	AAAA BBBBBB
☐	5 out of 10	AAAAA BBBBB
☐	6 out of 10	AAAAAA BBBB
☐	7 out of 10	AAAAAAA BBB
☐	8 out of 10	AAAAAAAA BB
☐	9 out of 10	AAAAAAAAA B
☐	10 out of 10	AAAAAAAAAA

5. Please guess: What share of participants in Klein-Windhoek thinks that option A **SHOULD** be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	BBBBBBBBBB
☐	1 out of 10	A BBBBBBBBB
☐	2 out of 10	AA BBBBBBBB
☐	3 out of 10	AAA BBBBBBB
☐	4 out of 10	AAAA BBBBBB
☐	5 out of 10	AAAAA BBBBB
☐	6 out of 10	AAAAAA BBBB
☐	7 out of 10	AAAAAAA BBB
☐	8 out of 10	AAAAAAAA BB
☐	9 out of 10	AAAAAAAAA B
☐	10 out of 10	AAAAAAAAAA

Survey – Third Part

1. In which neighborhood do you live?

2. What comes to your mind first when you think about the people in your neighborhood?

3. Please mark the statement to which you agree the most:

   ☐ I am very similar to the people in the neighborhood I live in
   ☐ I am similar to the people in the neighborhood I live in
   ☐ I am different to the people in the neighborhood I live in
   ☐ I am very different to the people in the neighborhood I live in
   ☐ none of the above

4. Please mark the statement to which you agree the most:

   ☐ Most people in in the neighborhood I live in are good people
   ☐ Many people in in the neighborhood I live in are good people
   ☐ Some people in the neighborhood I live in are a good people
   ☐ few people in in the neighborhood I live in are good people
   ☐ There are no good people in my neighborhood

5. What comes to your mind first if you think about people in Klein-Windhoek?

6. Please mark the statement to which you agree the most:

   ☐ I am very similar to people in Klein-Windhoek
   ☐ I am similar to people in Klein-Windhoek
   ☐ I am different to people in Klein-Windhoek
   ☐ I am very different to people in Klein-Windhoek
   ☐ none of the above applies

7. Please mark the statement to which you agree the most:

   ☐ Most people in Klein Windhoek are good people
   ☐ Many people in Klein Windhoek are good people
   ☐ Some people in Klein Windhoek are a good people
   ☐ Few people in Klein Windhoek are a good people
   ☐ There are no good people in Klein-Windhoek

8. From which materials is the place you live in mostly build:

   ☐ I do not live in a house or flat ☐ Plastic bags and nature materials

   ☐ Corrugated iron ☐ Bricks

   ☐ Bricks with more than three rooms

9. What kind of toilet do the members of your household use

   ☐ Flush-toilette ☐ Dry-toilette / Longdrop

   ☐ Flush-by-hand toilette ☐ none

10. What is your income per month?

    ☐ no income ☐N$ 1 to N$ 300

    ☐N$ 301 to N$ 600 ☐N$ 601 to N$ 1800

    ☐N$ 1801 to N$ 5000 ☐more than N$ 5000

11. What is your home language?

12. Which other language or languages do you speak fluently?

    ☐ No ☐ Yes, I speak ____________________

13. How old are you?

14. What is your gender?

    ☐ Female ☐ Male

15. What is your current occupation?

    ☐ I do not have a job ☐ My current occupation is ____________________

16. If you work, in which district do you work?

17. How many people live in your household (including yourself)?

18. Which is your highest completed degree of education?

    ☐ No School ☐ Primary ☐ Secondary ☐ University/College

19. **Before** coming to this survey today, have you heard details about specific questions in this survey?

    ☐ No ☐ Yes, I have heard that ______________________

20. Can we contact you for further surveys in the future?

    ☐ No ☐ Yes

21. **Please state your cell phone number for payments**

____

[Treatment Instructions Baseline B]

Survey – First Part

1. This survey is being conducted in Klein-Windhoek and at Ombili Community center. Are you taking part in Klein-Windhoek or at Ombili Community Center?

Information:

Your decision in this part of the survey will decide how a total of 60 N$ will be split between yourself and the Namibian Red Cross Disaster Management, which is a charitable organization. There are two options: option A and option B. Option A and option B are described in the following:

Option A

You receive N$ 55 and the Namibian Red Cross Disaster Management receives N$ 5

Option B

You receive N$ 30 and the Namibian Red Cross Disaster Management receives N$ 30

2. Now split the N$ 60 between yourself and the Namibian Red Cross Disaster Management. **Only mark one option**, otherwise you will not receive air-time. You cannot change your decision later in this survey.

☐ **Option A** You receive N$ 55
The Namibian Red Cross Disaster Management receives N$ 5

OR

☐ **Option B** You receive N$ 30
The Namibian Red Cross Disaster Management receives 30 N$

Survey – Second Part

1. Please guess: What share of participants here in **<u>this</u>** session decided for Option B? You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	AAAAAAAAAA
…		
☐	10 out of 10	BBBBBBBBBB

2. Which option do you think **SHOULD** be chosen by participants here?

☐ Option A ☐ Option B

3. Please guess: What share of participants here in **<u>this</u>** session thinks that option B SHOULD be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	AAAAAAAAAA
…		
☐	10 out of 10	BBBBBBBBBB

4. Please guess: What share of participants in Klein-Windhoek decided for option B? You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	AAAAAAAAAA
…		
☐	10 out of 10	BBBBBBBBBB

5. Please guess: What share of participants in Klein-Windhoek thinks that option B **SHOULD** be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	AAAAAAAAAA
…		
☐	10 out of 10	BBBBBBBBBB

Survey – Third Part

(identical in all treatments)

[Treatment Instructions Rich Ego]

Survey – First Part

1. This survey is being conducted in Klein-Windhoek and at Ombili Community center. Are you taking part in Klein-Windhoek or at Ombili Community Center?

Information:

Your decision in this part of the survey will decide how a total of 60 N$ will be split between yourself and the Namibian Red Cross Disaster Management, which is a charitable organization. There are two options: option A and option B. Option A and option B are described in the following:

Option A

You receive N$ 55 and the Namibian Red Cross Disaster Management receives N$ 5

Option B

You receive N$ 30 and the Namibian Red Cross Disaster Management receives N$ 30

2. In a survey session we conducted recently in **Klein-Windhoek, most of the participants decided for option A.**
   Please guess: What share of participants in that session in Klein-Windhoek decided for Option A? You will receive N$ 1 in air time for a right guess.

☐	8 out of 10	AAAAAAAA BB
☐	9 out of 10	AAAAAAAAA B
☐	10 out of 10	AAAAAAAAAA

3. Now split the N$ 60 between yourself and the Namibian Red Cross Disaster Management. **Only mark one option**, otherwise you will not receive air-time. You cannot change your decision later in this survey.

☐ **Option A** You receive N$ 55

The Namibian Red Cross Disaster Management receives N$ 5

OR

☐ **Option B** You receive N$ 30

The Namibian Red Cross Disaster Management receives N$ 30

Survey – Second Part

4. Please guess: What share of participants here in **<u>this</u>** session decided for Option A? You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	BBBBBBBBBB
…		
☐	10 out of 10	AAAAAAAAAA

5. Which option do you think **SHOULD** be chosen by participants here?

☐ Option A ☐ Option B

6. Please guess: What share of participants here in **<u>this</u>** session thinks that option A SHOULD be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	BBBBBBBBBB
…		
☐	10 out of 10	AAAAAAAAAA

7. Please guess: What share of participants in Klein-Windhoek in the mentioned session thinks that option A **SHOULD** be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	BBBBBBBBBB
…		
☐	10 out of 10	AAAAAAAAAA

Survey – Third Part

(Identical in all treatments)

[Treatment Instructions Rich PS]
Survey – First Part

1. This survey is being conducted in Klein-Windhoek and at Ombili Community center. Are you taking part in Klein-Windhoek or at Ombili Community Center?

Information:

Your decision in this part of the survey will decide how a total of 60 N$ will be split between yourself and the Namibian Red Cross Disaster Management, which is a charitable organization. There are two options: option A and option B. Option A and option B are described in the following:

Option A

You receive N$ 55 and the Namibian Red Cross Disaster Management receives N$ 5

Option B

You receive N$ 30 and the Namibian Red Cross Disaster Management receives N$ 30

2. In a survey session we conducted recently in **Klein-Windhoek, the majority of participants decided for option B.**
Please guess: What share of participants in that session in Klein-Windhoek decided for **Option B?** You will receive N$ 1 in air time for a right guess.

☐	6 out of 10	BBBBBB AAAA
☐	7 out of 10	BBBBBBB AAA
☐	8 out of 10	BBBBBBBB AA
☐	9 out of 10	BBBBBBBBB A
☐	10 out of 10	BBBBBBBBBB

3. Now split the N$ 60 between yourself and the Namibian Red Cross Disaster Management. **Only mark one option**, otherwise you will not receive air-time. You cannot change your decision later in this survey.

☐ **Option A** You receive N$ 55

The Namibian Red Cross Disaster Management receives N$ 5

OR

☐ **Option B** You receive N$ 30

The Namibian Red Cross Disaster Management receives N$ 30

Survey – Second Part

4. Please guess: What share of participants here in **<u>this</u>** session decided for Option B? You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	AAAAAAAAAA
…		
☐	10 out of 10	BBBBBBBBBB

6. Which option do you think **SHOULD** be chosen by participants here?

☐ Option A ☐ Option B

7. Please guess: What share of participants here in **<u>this</u>** session thinks that option B SHOULD be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	AAAAAAAAAA
…		
☐	10 out of 10	BBBBBBBBBB

8. Please guess: What share of participants in Klein-Windhoek in the mentioned session thinks that option B **SHOULD** be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐	0 out of 10	AAAAAAAAAA
…		
☐	10 out of 10	BBBBBBBBBB

Survey – Third Part

(identical in all treatments)

[Treatment Instructions Poor EGO}

Survey – First Part

1. This survey is being conducted in Klein-Windhoek and at Ombili Community center. Are you taking part in Klein-Windhoek or at Ombili Community Center?

Information:

Your decision in this part of the survey will decide how a total of 60 N$ will be split between yourself and the Namibian Red Cross Disaster Management, which is a charitable organization. There are two options: option A and option B. Option A and option B are described in the following:

Option A

You receive N$ 55 and the Namibian Red Cross Disaster Management receives N$ 5

Option B

You receive N$ 30 and the Namibian Red Cross Disaster Management receives N$ 30

2. In a survey session we conducted recently here at **Ombili Community Center, most of the participants decided for option A.** Please guess: What share of participants in that session at Ombili Community Center decided for Option A? You will receive N$ 1 in air time for a right guess.

☐	8 out of 10	AAAAAAAA BB
☐	9 out of 10	AAAAAAAAA B
☐	10 out of 10	AAAAAAAAAA

3. Now split the N$ 60 between yourself and the Namibian Red Cross Disaster Management. **Only mark one option**, otherwise you will not receive air-time. You cannot change your decision later in this survey.

☐ **Option A** You receive N$ 55

The Namibian Red Cross Disaster Management receives N$ 5

OR

☐ **Option B** You receive N$ 30

The Namibian Red Cross Disaster Management receives N$ 30

Survey – Second Part

4. Please guess: What share of participants here in **<u>this</u>** session today decided for Option A? You will receive an additional N$ 1 in air time for a right guess.

☐ 0 out of 10 BBBBBBBBBB

…

☐ 10 out of 10 AAAAAAAAAA

5. Which option do you think **SHOULD** be chosen by participants here?

☐ Option A ☐ Option B

6. Please guess: What share of participants here in **<u>this</u>** session thinks that option A SHOULD be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐ 0 out of 10 BBBBBBBBBB

…

☐ 10 out of 10 AAAAAAAAAA

5. Please guess: What share of participants in Klein-Windhoek decided for option A? You will receive an additional N$ 1 in air time for a right guess.

☐ 0 out of 10 BBBBBBBBBB

…

☐ 10 out of 10 AAAAAAAAAA

6. Please guess: What share of participants in Klein-Windhoek thinks that option A **SHOULD** be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐ 0 out of 10 BBBBBBBBBB

…

☐ 10 out of 10 AAAAAAAAAA

Survey – Third Part

(identical in all treatments)

[Treatment Instructions Poor PS]

Survey – First Part

1. This survey is being conducted in Klein-Windhoek and at Ombili Community center. Are you taking part in Klein-Windhoek or at Ombili Community Center?

Information:

Your decision in this part of the survey will decide how a total of 60 N$ will be split between yourself and the Namibian Red Cross Disaster Management, which is a charitable organization. There are two options: option A and option B. Option A and option B are described in the following:

Option A

You receive N$ 55 and the Namibian Red Cross Disaster Management receives N$ 5

Option B

You receive N$ 30 and the Namibian Red Cross Disaster Management receives N$ 30

2. In a survey session we conducted recently here **at Ombili Community Center, the majority of participants decided for option B.** Please guess: What share of participants in that session at Ombili Community Center decided for Option B? You will receive N$ 1 in air time for a right guess.

☐	6 out of 10	BBBBBB AAAA
☐	7 out of 10	BBBBBBB AAA
☐	8 out of 10	BBBBBBBB AA
☐	9 out of 10	BBBBBBBBB A
☐	10 out of 10	BBBBBBBBBB

3. Now split the N$ 60 between yourself and the Namibian Red Cross Disaster Management. **Only mark one option**, otherwise you will not receive air-time. You cannot change your decision later in this survey.

☐ **Option A** You receive N$ 55

The Namibian Red Cross Disaster Management receives N$ 5

OR

☐ **Option B** You receive N$ 30

The Namibian Red Cross Disaster Management receives N$ 30

Survey – Second Part

4. Please guess: What share of participants here in **this** session today decided for Option B? You will receive an additional N$ 1 in air time for a right guess.

☐ 0 out of 10 AAAAAAAAAA
…
☐ 10 out of 10 BBBBBBBBBB

5. Which option do you think **SHOULD** be chosen by participants here?

☐ Option A ☐ Option B

6. Please guess: What share of participants here in **this** session thinks that option B SHOULD be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐ 0 out of 10 AAAAAAAAAA
…
☐ 10 out of 10 BBBBBBBBBB

7. Please guess: What share of participants in Klein-Windhoek decided for option B? You will receive an additional N$ 1 in air time for a right guess.

☐ 0 out of 10 AAAAAAAAAA
..
☐ 10 out of 10 BBBBBBBBBB

8. Please guess: What share of participants in Klein-Windhoek thinks that option B **SHOULD** be chosen. You will receive an additional N$ 1 in air time for a right guess.

☐ 0 out of 10 AAAAAAAAAA
…
☐ 10 out of 10 BBBBBBBBBB

Survey – Third Part

(identical in all treatments)

Figure A10: Picture of Laboratory Session

BIBLIOGRAPHY

Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: the origins of power, prosperity and poverty* (1st ed). New York: Crown Publishers.

Aguilar-Benitez, I., & Saphores, J.-D. (2008). An analysis of nonpayment of water services in Nuevo Laredo, Mexico: Nonpayment of water services. *Water Resources Research*, *44*.

Ahuja, A., Kremer, M., & Zwane, A. P. (2010). Providing Safe Water: Evidence from Randomized Evaluations. *Annual Review of Resource Economics*, *2*, 237–256.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*, 179–211.

Akay, A., Martinsson, P., & Medhin, H. (2012). Does Positional Concern Matter in Poor Societies? Evidence from a Survey Experiment in Rural Ethiopia. *World Development*, *40*, 428–435.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and Identity. *Quarterly Journal of Economics*, *115*, 715–753.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*, 1082–1095.

Allcott, H., & Rogers, T. (2014). The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, *104*, 3003–3037.

Ashraf, N., Glaeser, E., Holland, A., & Steinberg, B. M. (2017). Water, Health and Wealth (*Working Paper No. w23807*). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w23807.pdf

Ayres, I., Raseman, S., & Shih, A. (2013). Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. *Journal of Law, Economics, and Organization*, *29*, 992–1022.

Bardhan, P. (2006). The economist's approach to the problem of corruption. *World Development*, *34*, 341–348.

Bardhan, P., & Mookherjee, D. (2006). Pro-poor targeting and accountability of local governments in West Bengal. *Journal of Development Economics*, *79*, 303–327.

Bartling, B., & Fischbacher, U. (2012). Shifting the Blame: On Delegation and Responsibility. *The Review of Economic Studies*, *79*, 67–87.

Bayliss, K., & Fine, B. (Eds.). (2008). *Privatization and Alternative Public Sector Reform in Sub-Saharan Africa*. London: Palgrave Macmillan UK.

Belotti, F., Deb, P., Manning, W. G., & Norton, E. C. (2015). twopm: Two-part models. *Stata Journal*, *15*, 3–20.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74,* 183–200.

Bénabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review, 96,* 1652–1678.

Bénabou, R., & Tirole, J. (2011). Identity, Morals, and Taboos: Beliefs as Assets. *The Quarterly Journal of Economics*, *126,* 805–855.

Berliner, J. S. (2002). Jean-Philippe Platteau, Institutions, Social Norms, and Economic Development. *Economic Development and Cultural Change*, *50,* 758–759.

Bernedo, M., Ferraro, P. J., & Price, M. (2014). The Persistent Impacts of Norm-Based Messaging and Their Implications for Water Conservation. *Journal of Consumer Policy*, *37,* 437–452.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates? *The Quarterly Journal of Economics*, *119*, 249–275.

Beshears, J., Milkman, K. L., & Schwartzstein, J. (2016). Beyond Beta-Delta: The Emerging Economics of Personal Plans. *American Economic Review: P&P*, *106,* 430–434.

Besley, T., & Persson, T. (2014). Why Do Developing Countries Tax So Little? *Journal of Economic Perspectives*, *28,* 99–120.

Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology*, *62*, 155–172.

Bicchieri, C., & Xiao, E. (2009). Do the Right Thing. *Journal of Behavioral Decision Making*, *22,* 191–208.

Blanton, H., & Christie, C. (2003). Deviance regulation: A theory of action and identity. *Review of General Psychology*, *7,* 115–149.

Blanton, H., Stuart, A. E., & Van den Eijnden, R. J. J. M. (2001). An Introduction to Deviance-Regulation Theory: The Effect of Behavioral Norms on Message Framing. *Personality and Social Psychology Bulletin*, *27,* 848–858.

Blattman, C., Jamison, J. C., & Sheridan, M. (2017). Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia. *American Economic Review*, *107,* 1165–1206.

Brandon, A., Ferraro, P., List, J., Metcalfe, R., Price, M., & Rundhammer, F. (2017). *Do The Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments* (No. w23277). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w23277.pdf

Brehm, J. W. (1966). *A theory of psychological reactance*. Oxford, England: Academic Press.

Brent, D. A., Cook, J. H., & Olsen, S. (2015). Social Comparisons, Household Water Use, and Participation in Utility Conservation Programs: Evidence from Three Randomized Trials. *Journal of the Association of Environmental and Resource Economists*, *2*, 597–627.

Bruhn, M., & McKenzie, D. (2009). In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, *1*, 200–232.

Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences*, *108*, 12653–12656.

Bryan, Christopher J., Adams, G. S., & Monin, B. (2013). When cheating would make you a cheater: Implicating the self prevents unethical behavior. *Journal of Experimental Psychology: General*, *142*, 1001–1005.

Bryan, Christopher J., Master, A., & Walton, G. M. (2014). “Helping” Versus “Being a Helper”: Invoking the Self to Increase Helping in Young Children. *Child Development*, *85*, 1836-1842

Buchanan, K., Russo, R., & Anderson, B. (2015). The question of energy reduction: The problem(s) with feedback. *Energy Policy*, *77*, 89–96.

Burbidge, J. B., Magee, L., & Robb, A. L. (1988). Alternative Transformations to Handle Extreme Values of the Dependent Variable. *Journal of the American Statistical Association*, *83*, 123–127.

Cadena, X., & Schoar, A. (2011). *Remembering to Pay? Reminders vs. Financial Incentives for Loan Payments* (*Working paper No. w17020*). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w17020.pdf

Cason, T., & Mui, V. (1998). Social Influence in the Sequential Dictator Game. *Journal of Mathematical Psychology*, *42*, 248–65.

Charness, G., Naef, M., & Sontuoso, A. (2018). Opportunistic Conformism. *Journal of Economic Theory*, accepted.

Chetty, R. (2015). Behavioral Economics and Public Policy: A Pragmatic Perspective. *American Economic Review*, *105*, 1–33.

Cialdini, R. B., Reno, R. R., & Kaligren, C. A. (1990). A Focus Theory of Normative Conduct. *Journal of Personality & Social Psychology*, *58*, 1015–1026.

Cohn, A., Maréchal, M. A., & Noll, T. (2015). Bad Boys: How Criminal Identity Salience Affects Rule Violation. *The Review of Economic Studies*, *82*, 1289–1308.

Collins, R.L. (2000). Among the better ones: upward assimilation in social comparison. In J. Suls & L. Wheeler (Eds.) *Handbook of social comparison: Theory and research* (pp.159-173). New York, NY: Springer Science+Business Media.

Conner, M., & Higgins, A. R. (2010). Long-term effects of implementation intentions on prevention of smoking uptake among adolescents: A cluster randomized controlled trial. *Health Psychology*, *29*, 529–538.

Conner, M., & Norman, P. (Eds.). (2007). *Predicting health behaviour: research and practice with social cognition models* (2. ed.). Maidenhead: Open Univ. Press.

Cornelissen, G., Dewitte, S., Warlop, L., & Yzerbyt, V. (2007). Whatever people say I am, that's what I am: Social labeling as a social marketing tool. *International Journal of Research in Marketing*, *24*, 278–288.

Costa, D. L., & Kahn, M. E. (2013). Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment. *Journal of the European Economic Association, 11*, 680–702.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*, 67–80.

Datta, S., & Mullainathan, S. (2014). Behavioral Design: A New Approach to Development Policy. *Review of Income and Wealth, 60*, 7–35.

Dietz, T., Stern, P. C., National Research Council (U.S.), & National Research Council (U.S.) (Eds.). (2008). *Public participation in environmental assessment and decision making*. Washington, DC: National Academies Press.

Dolan, P., Hallsworth, M., Halpern, D., King, D., Metcalfe, R., & Vlaev, I. (2012). Influencing behaviour: The mindspace way. *Journal of Economic Psychology*, *33*, 264–277.

Duan, N. (1983). Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association*, *78,* 605.

Duflo, E. (2017). The Economist as Plumber. *American Economic Review: Papers & Proceedings*, *107*, 1–26.

Du Plessis, N., Neels, C., Anyim, I., & Matros, A. (2005). *Perceptions on affordability and willingness to pay for water services*. Research Document.

Ebeling, F., Feldhaus, C., & Fendrich, J. (2017). A field experiment on the impact of a prior donor's social status on subsequent charitable giving. *Journal of Economic Psychology, 61, 124–133*

Exley, C. L. (2016). Excusing Selfishness in Charitable Giving: The Role of Risk. *The Review of Economic Studies*, *83*, 587–628.

Fatas, E., Hargreaves Heap, S. P., & Rojo Arjona, D. (2018). Preference conformism: An experiment. *European Economic Review*, *105,* 71–82.

Ferraro, P. J., & Miranda, J. J. (2013). Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment. *Resource and Energy Economics*, *35*, 356–379.

Ferraro, P. J., Miranda, J. J., & Price, M. K. (2011). The Persistence of Treatment Effects with Norm-Based Policy Instruments: Evidence from a Randomized Environmental Policy Experiment. *American Economic Review: P&P*, *101*, 318–322.

Ferraro, P. J., & Price, M. K. (2013). Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment. *Review of Economics and Statistics*, *95*, 64–73.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Finger, M., Allouche, J., & Luis-Manso, P. (Eds.). (2007). *Water and liberalisation: European water scenarios*. London: IWA Publ.

Forsythe, R., Horowitz, J. L., Savin, N. E., & Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior, 6,* 347.

Franke, N., Schreier, M., & Kaiser, U. (2010). The “I Designed It Myself” Effect in Mass Customization. *Management Science*, *56,* 125–140.

Fréchette, G. (2012). Session-effects in the laboratory. *Experimental Economics*, *15,* 485-498.

Frey, B. S., & Meier, S. (2004). Social Comparisons and Pro-social Behavior: Testing “Conditional Cooperation” in a Field Experiment. *American Economic Review*, *94,* 1717–1722.

Frey, E., & Rogers, T. (2014). Persistence: How Treatment Effects Persist After Interventions Stop. *Policy Insights from the Behavioral and Brain Sciences*, *1,* 172–179.

Gamper-Rabindran, S., Khan, S., & Timmins, C. (2010). The impact of piped water provision on infant mortality in Brazil: A quantile panel data approach. *Journal of Development Economics*, *92,* 188–200.

Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and Differentiation in Unethical Behavior. *Psychological Science*, *3,* 393.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54,* 493–503.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. In *Advances in Experimental Social Psychology* (Vol. 38, pp. 69–119). Elsevier.

Grant, A. M., & Gino, F. (2010). A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of Personality and Social Psychology*, *98*, 946–955.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, *11*, 1–21.

Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis*, *20*, 25–46.

Haisley, E. C., & Weber, R. A. (2010). Self-serving interpretations of ambiguity in other-regarding behavior. *Games and Economic Behavior*, *68,* 614–625.

Harrison, G. W., & List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, *42,* 1009–1055.

Hayes, A. F., & Krippendorff, K. (2007). Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures*, *1*, 77–89.

Hogg, M.A. (2006). Social identity theory. In P.J. Burke (Ed.) *Contemporary Social Psychology Theories* (pp.111-136). Stanford, CA: Stanford University Press.

Hooshangi, S., & Loewenstein, G. (2018). The Impact of Idea Generation and Potential Appropriation on Entrepreneurship: An Experimental Study. *Management Science*, *64*, 64–82.

Iriberri, N., & Rey-Biel, P. (2013). Elicited beliefs and social information in modified dictator games: What do dictators believe other dictators do?: Beliefs and information in dictator games. *Quantitative Economics*, *4*, 515–547.

Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, *20*, 1–24.

Jack, B. K., & Smith, G. (2015). Pay as You Go: Prepaid Metering and Electricity Expenditures in South Africa. *American Economic Review: P&P*, *105*, 237–241.

Jetten, J., Spears, R., & Manstead, A. S. R. (1996). Intergroup Norms and Intergroup Discrimination. *Journal of Personality & Social Psychology*, *71*, 1222–1233.

Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the Moral Self: The Effects of Recalling Past Moral Actions on Future Moral Behavior. *Personality and Social Psychology Bulletin*, *37*, 701–713.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the Top of Mind: How Reminders Increase Saving. *Management Science*, *62*, 3393–3411.

Karlan, D., Morten, M., & Zinman, J. (2016). A personal touch in text messaging can improve microloan repayment. *Behavioral Science & Policy*, *1*, 25–31.

Katz, D., Grinstein, A., Kronrod, A., & Nisan, U. (2016). Evaluating the effectiveness of a water conservation campaign: Combining experimental and field methods. *Journal of Environmental Management*, *180*, 335–343.

Keefer, P., & Knack, S. (2008). Social Capital, Social Norms and the New Institutional Economics. In C. Ménard & M. M. Shirley (Eds.), *Handbook of New Institutional Economics* (pp. 701–725). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kessler, J. B., & Milkman, K. L. (2018). Identity in Charitable Giving. *Management Science*, *64*, 845–859.

King, G. and Nielsen, R. (2018). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis* (forthcoming). Retrieved from: http://j.mp/2ovYGsW

Klintenberg, P., Mazambani, C., & Nantanga, K. (2007). Integrated water resources management in the Namibian part of the Cuvelai Basin, central northern Namibia. *CuveWaters Papers*, *2*. Retrieved from https://d-nb.info/1031489843/34

Konow, J. (2000). Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review*, *90*, 1072–1092.

Kronrod, A., Grinstein, A., & Wathieu, L. (2012). Go Green! Should Environmental Messages Be So Assertive? *Journal of Marketing*, *76*, 95–102.

Krupka, E., & Weber, R. A. (2009). The focusing and Informational Effects of Norms on Pro-social Behavior. *Journal of Economic Psychology*, *30*, 307–320.

Kumru, C. S., & Vesterlund, L. (2010). The Effect of Status on Charitable Giving. *Journal of Public Economic Theory*, *12*, 709–735.

Levitt, S. D., & List, J. A. (2009). Review paper: Field experiments in economics: The past, the present, and the future. *European Economic Review, 53,* 1–18.

Lu, L., Deller, D. & Hviid, M. (2017). Price and Behavioural Signals to Encourage Household Water Conservation in Temperate Climates. *Working paper*. Retrieved from https://www.anglianwater.co.uk/_assets/media/Price_and_Behavioural_Signals_to_Encourage_Water_Conservation.pdf

Lu, X., Wang, L., Pan, M., Kaseke, K. F., & Li, B. (2016). A multi-scale analysis of Namibian rainfall over the recent decade – comparing TMPA satellite estimates and ground observations. *Journal of Hydrology: Regional Studies, 8*, 59–68.

MacKinnon, J. G., & Magee, L. (1990). Transforming the Dependent Variable in Regression Models. *International Economic Review, 31,* 315-339.

Madrian, B. C. (2014). Applying Insights from Behavioral Economics to Policy Design. *Annual Review of Economics*, *6*, 663–688.

Mansur, E. T., & Olmstead, S. M. (2012). The value of scarce water: Measuring the inefficiency of municipal regulations. *Journal of Urban Economics*, *71*, 332–346.

Mayring, P. (2000). Qualitative Content Analysis. *Forum: Qualitative Social Research*, *1*, 105–114.

Mazambani, C., Schönbrodt-Stitt, S., & Klintenberg, P. (2006). Water supply: a gift from God or does it come with a cost? *World Water Week 2006 report*.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more T in experiments. *Journal of Development Economics*, *99*, 210–221.

McKenzie, D. (2017). Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition. *American Economic Review*, *107*, 2278–2307.

McKenzie-Mohr, D. (2011). *Fostering sustainable behavior: an introduction to community-based social marketing* (Third edition). Gabriola Island: New Society Publishers.

Mekonnen, M. M., & Hoekstra, A. Y. (2016). Four billion people facing severe water scarcity. *Science Advances, 2,* e1500323

Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral Self-Licensing: When Being Good Frees Us to Be Bad: Moral Self-Licensing. *Social and Personality Psychology Compass*, *4*, 344–357.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, *108*, 10415–10420.

Milkman, Katherine L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2013). Planning prompts as a means of increasing preventive screening rates. *Preventive Medicine*, *56*, 92–93.

Miller, R. L., Brickman, P., & Bolen, D. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology*, *31*, 430–441.

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. Psychological Review, 110, 472-489.

Mussweiler, Thomas. (2001). 'Seek and ye shall find.' *European Journal of Social Psychology*, *31*, 499–509.

Namibian Statistics Agency. (2011). *Khomas 2011: Census Regional Profile*. Retrieved from http://cms.my.na/assets/documents/p19dptss1ri5r1f2kt6i5931i31g.pdf.

Namibian Statistics Agency (2016). *Key Highlights of the Namibian Labour Force Survey 2016 Report.* Retrieved from https://cms.my.na/assets/documents/Key_Highlights_of_the_Namibia_Labour_Force_Survey_2016_Report2.pdf

Neal, D. T., Wood, W., Labrecque, J. S., & Lally, P. (2012). How do habits guide behavior? Perceived and actual triggers of habits in daily life. *Journal of Experimental Social Psychology*, *48*, 492–498.

Nickerson, D. W., & Rogers, T. (2010). Do You Have a Voting Plan?: Implementation Intentions, Voter Turnout, and Organic Plan Making. *Psychological Science*, *21*, 194–199.

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, *22*, 453–460.

Olmstead, S. M., Michael Hanemann, W., & Stavins, R. N. (2007). Water demand under alternative price structures. *Journal of Environmental Economics and Management*, *54*, 181–198.

Platteau, J. P. (2000). *Institutions, social norms, and economic development*. Amsterdam, The Netherlands: Harwood Academic Publishers.

Rockenbach, B., Tonke, S., & Weiß, A. (2018a). Using Behavioral Insights to Decrease Non-Payment for Public Utilities. *Working paper*.

Rockenbach, B., Tonke, S., & Weiß, A. (2018b). The Self-serving Behavior of the Rich Causes Contagion Effects Among the Poor. *Working paper*.

Rode, J., Gómez-Baggethun, E., & Krause, T. (2015). Motivation crowding by economic incentives in conservation policy: A review of the empirical evidence. *Ecological Economics*, *117*, 270–282.

Rodrik, D. (2017). *Economics rules: why economics works, when it fails, and how to tell the difference* (First published in paperback). Oxford New York, NY: Oxford University Press.

Rogers, T., Milkman, K. L., John, L. K., & Norton, M. I. (2015). Beyond good intentions: Prompting people to make plans improves follow-through on important tasks. *Behavioral Science & Policy*, *1*, 33–41.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science*, *18*, 429–434.

Shah, A. K., Mullainathan, S., & Shafir, E. (2012). Some Consequences of Having Too Little. *Science*, *338*, 682–685.

Shah, Anju K., & Ludwig, J. (2016). Option Awareness: The Psychology of What We Consider. *American Economic Review P&P*, *106*, 425–429.

Shang, J., Reed, A., & Croson, R. (2008). Identity Congruency Effects on Donations. *Journal of Marketing Research*, *45*, 351–361.

Sheeran, P., Webb, T. L., & Gollwitzer, P. M. (2005). The Interplay Between Goal Intentions and Implementation Intentions. *Personality and Social Psychology Bulletin*, *31*, 87–98.

Spiekermann, K., & Weiss, A. (2016). Objective and subjective compliance: A norm-based explanation of 'moral wiggle room.' *Games and Economic Behavior*, *96*, 170–183.

Steg, L. (2016). Values, Norms, and Intrinsic Motivation to Act Proenvironmentally. *Annual Review of Environment and Resources*, *41*, 277–292.

Strand, J. (2012). Low-level versus high-level equilibrium in public utility services. *Journal of Public Economics*, *96*, 163–172.

Sunstein, C. R. (2017). Nudges that fail. *Behavioural Public Policy*, *1*, 4–25.

Szabó, A., & Ujhelyi, G. (2015). Reducing nonpayment for public utilities: Experimental evidence from South Africa. *Journal of Development Economics*, *117*, 20–31.

Tajfel, H. & Turner, J.C. (1986). Theoretical Perspectives on Intergroup Relations. In S. Worchel & G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7-24). Chicago, IL: Nelson-Hall.

Thaler, R. H. (2016). *Misbehaving: the making of behavioural economics*. New York, London: W. W. Norton & Company.

Thaler, R. H. (2018). From Cashews to Nudges: The Evolution of Behavioral Economics. *American Economic Review*, *108*, 1265–1287.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: improving decisions about health, wealth, and happiness*. New York: Penguin Books.

Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., & Staake, T. (2018). Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation. *Management Science*, *64*, 1458–1476.

Traxler, C. (2010). Social norms and conditional cooperative taxpayers. *European Journal of Political Economy*, *26*, 89–103.

Tonke, S. (2018a). The Pivotal Role of Specific Tips to Encourage Water Conservation. *Working paper*.

Tonke, S. (2018b). Using Identity Labels to Decrease Non-payment for Water: Evidence from Kosovo. *Working paper*.

Uhlendahl, T., Ziegelmayer, D., Wienecke, A., Mawisa, M.L. & du Pisani, P. (2010). *Water Consumption at Household Level in Windhoek, Namibia*. *Survey about Water Consumption at*

Household Level in different Areas of Windhoek depending on Income Level and Water Access in 2010 (Final Project Report). Retrieved from http://www.freidok.uni-freiburg.de/volltexte/7937/pdf/water_cosumption_windhoek_2010_final_report.pdf

Vásquez, W. F. (2015). Nonpayment of water bills in Guatemala: Dissatisfaction or inability to pay? *Water Resources Research, 51,* 8806–8816.

Vásquez, W. F., & Alicea-Planas, J. (2017). Factors associated with nonpayment behavior in the water sector of Nicaragua. *Utilities Policy*, *47,* 50–57.

Vesterlund, L. (2003). The informational value of sequential fundraising. *Journal of Public Economics*, *87*, 627–657.

Walton, G. M., & Banaji, M. R. (2004). Being What You Say: The Effect of Essentialist Linguistic Labels on Preferences. *Social Cognition*, *22*, 193–213.

Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, *125*, 617–655.

World Bank. (2014). *World development indicators 2013.* Retrieved from http://data.worldbank.org/data-catalog/world-development-indicators

World Economic Forum. (2017). *The Global Risks Report 2017.* Retrieved from http://www.deslibris.ca/ID/10090180

World Water Assessment Programme (2016). *Water and jobs*. Retrieved from http://www.unesco.org/new/en/natural-sciences/environment/water/wwap/wwdr/2016-water-and-jobs/

World Water Assessment Programme. (2018). *Nature-based solutions for water*. Retrieved from http://www.unesco.org/new/en/natural-sciences/environment/water/wwap/wwdr/2018-nature-based-solutions/

Yeager, D. S., Dahl, R. E., & Dweck, C. S. (2018). Why Interventions to Influence Adolescent Behavior Often Fail but Could Succeed. *Perspectives on Psychological Science*, *13*, 101–122.

CURRICULUM VITAE

Personal Details

Name	Sebastian Soung-Un Tonke
Place of birth	Berlin

Education

Since 2013	Ph.D. Student in Economics *University of Cologne*
2013	Master of Science in Economics *Barcelona Graduate School of Economics*
2012	Bachelor of Science in Economics *University of Cologne*
2007	Abitur *Evangelisches Gymnasium zum Grauen Kloster*