Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

A Large-Scale Field Experiment to Reduce Non-Payments for Water: From Diagnosis to Treatment

Bettina Rockenbach[1,2], Sebastian Tonke[2,*] and Arne R. Weiss[3,4,5]

In a field experiment among 9,823 customers of the Namibian water utility, we implement interventions to reduce non-payments. The interventions are based on diagnostic surveys to identify key obstacles to payments. They address informational frictions and apply psychological commitment techniques to narrow the gap between customers' willingness to pay and actual payments. Initially, payments increase by 29% to 55%, making the interventions highly cost-effective. While removing informational frictions has a lasting impact, the commitment techniques produce only short-term effects. We demonstrate the effectiveness and limitations of behavioral interventions in settings where heavy-handed tools, e.g., disconnecting non-payers, are difficult to implement.

Keywords: Field experiment, non-payment, behavioral intervention, public utility

JEL Codes: C93, D91, Q25

[1] University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany

[2] Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, 53113 Bonn, Germany

[3] University of Alicante, Fundamentos del Análisis Económico (FAE), 03071 Alicante, Spain

[4] University of Mannheim, DFG SFB 884, 68131 Mannheim, Germany

[5] University of Bern, Wyss Academy for Nature and Department of Economics, 3004 Bern, Switzerland

* Corresponding author. Email address: tonke@coll.mpg.de

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Acknowledgments: We thank Chris Bryan, Gary Charness, Lola Collado, Paul Ferraro, Markus Frölich, Uri Gneezy, Matthias Heinz, Kelsey Jack, Steven Levitt, Axel Ockenfels, Dirk Sliwka, Matthias Sutter, Andrea Szabó, Bertil Tungodden, Gergely Ujhelyi, Timo Vogelsang, Björn Vollan and participants of several seminars and conferences for helpful comments. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1– 390838866. We thank Myrielle Gonschor and Martin Günther for their excellent research assistance. We are grateful to Benedikt Dechamps as well as students from the Namibia University of Science and Technology for their outstanding implementation in the field. We thank Job Horohua, Wilma Husselmann, Aino Nsinano, and Vaino Shivute for NamWater's generous support during all phases of the research project. We declare no conflict of interest.

1. Introduction

Economics provides an ever-growing toolbox of behavioral interventions (Thaler, 2018). Yet, designing a suitable intervention to address a specific real-world problem is challenging (Duflo, 2017). Recent evidence suggests that behavioral interventions have heterogeneous effects across contexts, typically wane over time and do not easily scale up (Bryan et al., 2021; DellaVigna and Linos, 2022). The effectiveness of an intervention hinges on addressing the relevant obstacles to behavioral change in the target population (Rodrik, 2010; Datta and Mullainathan, 2014). In this paper, we present evidence from a large-scale natural field experiment to reduce non-payment for water. We address the aforementioned challenges by first diagnosing and quantifying relevant obstacles in the target population. On this basis, we design interventions that tackle these obstacles and test their effectiveness both in the short term and over time.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

In order to identify the most common reasons for non-payment, we use a simple diagnostic tool in the form of surveys. We then design interventions to address the identified obstacles and evaluate their short-term and medium-term effectiveness. The field experiment is conducted in Namibia in cooperation with the Namibia Water Corporation (*NamWater*), where non-payment among customers is a wide-spread problem.

Non-payment for public utilities is a complex global problem in both low- and high-income countries (Bridges and Disney, 2004; Aguilar-Benitez and Saphores, 2008; Vásquez and Alicea-Planas, 2017; Jensen and Chindarkar, 2019; Tonke, 2023), with cost-recovery rates below 50% not being uncommon (Kayaga et al., 2004; Mugabi et al., 2010). Non-payment constrains the maintenance and expansion of the infrastructure necessary to provide access to water and electricity (Jack and Smith, 2015; McRae, 2015; Szabó and Ujhelyi, 2015). These play a vital role for economic and social development (Duflo and Pande, 2007; Dinkelman, 2011), but are not adequately accessed by billions of people (UNICEF and WHO, 2019). Practical solutions to reduce non-payment are not obvious. The standard policy tools like denying non-paying customers access and legal enforcement are difficult to implement. Many countries have legal provisions against, for instance, cutting off the supply of water because it is a basic human need (Finger et al., 2007). Enforcement of payments through overburdened courts is often not effective either (World Bank, 2017). This makes interventions from the toolbox of softer behavioral interventions an attractive alternative.

There are various reasons why individuals do not pay their water bills. Yet, knowledge about the extent to which these reasons were common among the targeted large-scale population was missing. Previous studies highlight unsatisfactory quality of service, lack of trust in the utility, and insufficient enforcement of sanctions (Aguilar-Benitez and Saphores, 2008; McRae, 2015; Vásquez, 2015; Vásquez and Alicea-Planas, 2017; Jensen and Chindarkar,

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

2019).[1] Additionally, in the Namibian context, individuals are reported to believe that water, as a "gift from God", should be free and not marketed (Mazambani et al., 2006). Further, some individuals may not understand the cost-covering concept because water was supplied without charge until Namibia's independence in 1990 (Du Plessis et al., 2005; Klintenberg et al., 2007).

To diagnose and quantify which obstacles to payment are most relevant among the targeted population, we conduct pre-intervention interviews among a randomly drawn sample of customers via phone. Our pre-intervention diagnostics reveal two main obstacles that may be addressed by low-cost behavioral interventions. First, many customers suffer from informational frictions: 43% of customers report receiving their invoice, which is sent by postal mail, either late or not at all.[2] Secondly, the vast majority seem to be willing to pay for water but do not act on their intention. For instance, more than 90% of customers state that water should be paid for and come up with coherent reasons (e.g., "purification" or "maintenance") for the necessity of paying. Nevertheless, most customers do not pay consistently and have amassed on average about 3.4 unpaid monthly water bills. This suggests a gap between the willingness to make payments and the actual payment behavior recorded in the administrative data – a so-called intention-to-action gap. Our diagnosis provides surprisingly little support for reasons assumed in the literature so far. For example, there seem to be strong social norms supporting payments, which runs counter to beliefs that water is a "gift of God" and that households do not understand the cost-recovery concept.

In the second step of our study, we set up a call center and treat around 10,000 private customers over a time span of nine months to study treatment effects in the first month after

---

[1] Investments to improve service may themselves be constrained by non-payment, such that the utility and customers remain stuck in a bad equilibrium (Strand, 2012).

[2] Most customers do not own a mailbox at their residence and instead have a post box in the nearest village or city.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

the intervention (short term) as well as over the ensuing eight months for which we have data (medium term). We implement three treatments via phone call and subsequent text messages. In the *basic* treatment, we offer a free monthly text-message service that contains simplified invoice information. The text-message service ensures that customers have access to the invoice information in a non-technical language and addresses the diagnosed problem of undelivered invoices. The *basic* treatment serves as a comparison group to two psychological commitment treatments that are implemented on top of the *basic* treatment. The psychological commitment treatments target the intention-to-action gap. In addition to receiving the text message, the phone call elicits a commitment from the treated customers either to a desirable self-concept (in the *self-concept* treatment) or to a concrete action plan (in the *plan* treatment). After the phone call, the treated customers are subsequently reminded of their commitment via the monthly text message. We label these two interventions as psychological commitment treatments because acting inconsistently with the elicited self-concept or action plan creates cognitive dissonance, which makes it psychologically costly to renege (Festinger, 1957; Konow, 2000; Bénabou and Tirole, 2011).

In the *self-concept* treatment, we invoke a self-concept of a reliable water payer. This is implemented by eliciting answers from the customers to questions such as "How important is it to you to be a reliable water payer?" The intervention builds on the idea that individuals are uncertain about their self-concept and want to shape their view of themselves through their actions (Bénabou and Tirole, 2011; Bryan et al., 2011; Tonke, 2023). 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?" In addition to a commitment to pay, making plans helps customers to develop specific strategies to overcome logistical obstacles and might also reduce forgetfulness (Gollwitzer, 1999; Gollwitzer and Sheeran, 2006; Rogers et al., 2015; Beshears et al., 2016). A necessary condition for the commitment treatments to

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

work is that individuals indeed want to act in line with the targeted behavior (Sheeran et al., 2005; Bryan et al., 2011). This motivation may stem from different sources, including social norms that prescribe this behavior.

We find that customers highly appreciate the text-message service, which is reflected by an almost universal take-up of our treatments (98%). Simply offering and sending customers free monthly invoice information by text message raises average payments remarkably. Payments increase by about 29% in the first month and by about 8% over the ensuing eight months 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 – compared to the *basic* treatment – by a further 10% in the *self-concept* treatment and by another 26% in the *plan* treatment. The additional effects of the commitment treatments, however, wane subsequently.

This paper makes several contributions: First, our diagnostic approach and experimental findings add to our understanding of non-payments for public utilities. Customers seem to have the intention to pay but often fail to act. The informational frictions of not consistently receiving the invoice can be resolved through low-cost information provision that results in a persistent treatment effect over the course of our experiment. The intention-to-action gap in payment behavior can be tackled through psychological commitment (e.g., Laibson, 1997; Ariely and Wertenbroch, 2002). We discuss various alternative mechanisms underlying the treatment effects, such as scrutiny, the threat of being punished, or reminders and argue that these are unlikely to drive the results. Our study thus adds novel reasons and policy options to the literature on non-payment and non-compliance.

Second, we address unanswered questions on the impact of the studied behavioral interventions over time (Frey and Rogers, 2014; Thaler, 2018). This is important because they may induce rebound effects (e.g., due to budgeting over time or moral licensing), resulting in

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

the absence of a net effect.[3] While the effectiveness of the *basic* treatment is long-lasting, the effectiveness of the psychological commitment techniques wanes over time. This important insight complements previous literature that has not studied their effectiveness over extended time horizons. Self-concept interventions have been used to mobilize voters, to promote pro-social behavior among children, and to decrease cheating behavior (Bryan et al., 2011; Bryan et al., 2013; Bryan et al., 2014). Plan-making interventions have been used to increase voter mobilization, vaccination rates, preventive colonoscopy-screening rates, and job search (Nickerson and Rogers, 2010; Milkman et al., 2011; Milkman et al., 2013; Abel et al., 2019). Our findings suggest that our psychological commitment techniques can be highly effective, but they seem best suited to address one-off decision-making.

Third, our paper attempts to open the black box of how to choose and design interventions that address large-scale, real-world problems. While we cannot claim that we identified the most effective among all possible interventions, our study nevertheless suggests high potential benefits from diagnosing which behavioral obstacles are most common at scale before designing and implementing interventions. Our interventions are a clear departure from previous practices of the utility company and neither featured in local experts' assessment[4] nor in the literature on non-payment for water. Yet, our interventions are highly cost-effective: They increased average payments by about 10-11 USD per customer, which corresponds to an estimated return on investment of roughly 1,000%. This leads to one of the highest months in

[3] Bursztyn et al. (2019), for example, study credit card debt repayment and find that moral messages affect short-term delinquency rates, but they do not find a long-term effect on defaults among their full sample.

[4] We had extensive discussions with local experts, both at the utility and in Namibia's water sector, during exploratory scoping missions, while designing the diagnostic survey and during the implementation of the interventions.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

terms of average payments by private customers in the entire history of the public utility. This suggests that a simple diagnostic survey can help identify seemingly overlooked reasons for non-payment, which can be addressed through scalable and highly cost-effective interventions.

The remainder of the paper is as follows: Section 2 describes the research setting and dataset. Section 3 presents the treatments, their implementation, and the estimation strategy. In Section 4, we provide the experimental results of the two psychological commitment treatments and discuss mechanisms in Section 5. In Section 6, we estimate the effects of the *basic* treatment. Section 7 concludes.

2. Research Setting and Behavioral Diagnosis

2.1 Research Setting and Dataset

Namibia is a middle-income country in Sub-Saharan Africa with a limited amount of surface water and low and unpredictable rainfall (Lu et al., 2016). Groundwater is often saline and not drinkable. Our study focuses on private customers who are directly billed by *NamWater* and reside in Namibia's North across an area of more than 85,000 square kilometers (about the size of Austria or the United Arab Emirates). This represents the vast majority of *NamWater's* directly billed customers (approximately 79% as of August 2015).[5] Our sample is quite heterogeneous, ranging from predominantly small households to a few large customers, which are likely commercial customers.[6]

NamWater compiles administrative data on a monthly basis, usually at the beginning of each month. The majority of the data is entered manually into the system and then digitally

[5] In most other regions, households living in cities and towns are typically billed by their municipality, which itself is billed by *NamWater*.

[6] While our data does not specify which customers are businesses, our results are robust to the exclusion of very large payments, indicative of a commercial customers.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

compiled. Clear guidelines for quality checks of the 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. Whenever possible, we updated incomplete payment records and extreme payment records by using the corrected account data from the ensuing month. Subsequently, we excluded five customers with negative payment values as well as the 0.1% of customers with the most extreme payments. These high payments either strongly suggest commercial usage or large accounting mistakes. These exclusions are less restrictive than applying Grubbs' outlier test (Grubbs, 1969). At an alpha level of 0.001, this test suggests the exclusion of an additional 142 households, which have payments larger than 4,050 N$ in a single month (see Table A1 of the online appendix for a robustness check of our main results when excluding these households).

Customers pay on average only about four times a year and then typically make bulk payments in multiples of 50 N$. These payments are, however, typically insufficient to cover the amount that is charged by *NamWater,* such that many customers accumulate debt over time.[7] Figure A2 shows how unpaid invoices accumulate over the course of the study for each of the treatment groups as well as the untreated comparison group. Customers can pay by cash or bank transfer, but not through mobile phone credits. In practice, almost all customers pay in cash.

2.2 Behavioral Diagnosis

We implemented a diagnostic pre-intervention survey via telephone in June 2015 with a randomly selected sample (N=329) of *NamWater's* customers. Our research team of local

[7] *NamWater* charges an interest rate of 9.75% per year. According to our data set, these interest rates are, however, not consistently charged.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

students carried out these surveys and introduced themselves as part of a research team of the *University of Cologne*. Our diagnostic approach has two components: Identifying potential obstacles through behavioral mapping (based on Datta and Mullainathan, 2014) and subsequently quantifying which of these obstacles are common in a randomly drawn sample of the target population.

Behavioral Mapping – This component is based on the behavioral-design framework by Datta and Mullainathan (2014).[8] We first elicit the process of behaviors that customers have to go through to settle their bills. The main steps of the payment process typically entail collecting the bill from the post office, reading and understanding its content, traveling to payment points, and making the actual payments. In addition, we include candidate obstacles into the questionnaire suggested by the literature and by local experts. Questions address, for instance, perception of water payments, personal and social norms, knowledge about sanctions for non-payments, and demographics. This approach is an iterative and dynamic process in which one set of questions may lead to the next set of questions and irrelevant questions get dropped (following Datta and Mullainathan (2014)).

The survey also includes a dictator game. The purpose of the dictator game is to measure the willingness to make payments in an incentive-compatible manner when customers are freed of any implementation costs. Customers can win a lottery earning of about 50 USD (roughly 45 hourly wages) for participation in the survey. We then ask participants how they

[8] There are several alternative diagnostic approaches in psychology, in particular related to health behavior, notably the Theoretical Domains Framework (Atkins et al., 2017) and the closely linked Behaviour Change Wheel (Michie et al., 2011). While these are conceptually similar to Datta and Mullainathan (2014), they are more complex and use different terminologies and distinctions, as they speak to a different academic field.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

would split the potential lottery earning between a direct mobile-phone credit transfer ("Airtime")[9] and a repayment of arrears on their *NamWater* account. The exact question is: "In case you get drawn as winner in our lottery, we can send you airtime or pay outstanding debts on your *NamWater* account. Of the 500 N$, please state the amount you want to receive in airtime and the amount you want to repay debts of *NamWater* with".

Quantifying Obstacles – We go beyond Datta and Mullainathan (2014) with our second component: Since there were likely many obstacles, we need to identify the obstacles that were common among a large fraction of customers and obstacles that could be addressed through cost-effective and scalable solutions. We do so by running a phone-based survey with 329 randomly drawn customers out of all customers of which we had phone numbers. The final questionnaire and summary statistics are provided in online appendices A3 and A4.

Findings – First, many customers suffer from informational frictions: 43% of customers report receiving their postal invoice either late or not at all. This delivery problem has to be seen in light of the fact that most customers do not own a mailbox at their residence and instead have a post box in the nearest village or city.

Secondly, the vast majority seem to be willing to pay for water. For instance, 93% of customers state that it is fair to pay for water, 86% state that others should pay for water, and 92% can name coherent reasons for this normative belief (e.g., "purification" or "maintenance"). Since self-reported survey responses may suffer from a social-desirability bias, we next turn to the incentivized dictator game. Customers would allocate, on average,

[9] Airtime is a credit on mobile phones, which can be used for texting, phone calls, internet data packages as well as small purchases. Airtime can easily be transferred for free from one user to the next.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

75% of a lottery prize (about 50 USD) to pay off arrears on their accounts.[10] While acknowledging the possible influences of a house-money and an experimenter-demand effect, this nevertheless suggests that without any implementation obstacles, most customers are willing to forego money to repay debt. This finding indicates a gap between the customers' willingness to make payments and their actual payment behavior (intention-to-action gap). A factor contributing to this gap are high transaction costs to make payments: About 95% of respondents make cash payments. In order to pay, they need to travel, on average, 25 km to the next payment point and spend about 47 minutes waiting in line. These logistical obstacles are also a plausible reason why customers make infrequent payments (every three months, on average). Our results provide surprisingly little support for reasons put forth by experts and in the literature. For instance, there seem to be strong social norms supporting payments, which runs counter to presumed beliefs that water is a "gift of God" and that households do not understand the cost-recovery concept. For example, only 7% (N=24) of interviewees state that it is unfair to pay for water. Our survey data also does not provide clear support that financial inability is a reason for non-payment. We find no significant correlation between self-reported household income and the self-reported fraction of the invoice being paid (p=0.564).

3. The Experimental Setup

The findings of the behavioral diagnosis inform our interventions. We use the following criteria to guide the intervention design: First, the interventions implement established behavioral mechanisms to tackle the diagnosed barriers to payments. Second, they can be delivered at scale in the local context and in partnership with *NamWater*. Third, in order to generate

[10] The distribution of choices is presented in Figure A5 of the online appendix. The winner of the lottery was called and asked to confirm the choice.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

generalizable knowledge, the interventions ought to be transferable to similar contexts; and fourth, they promise high cost-effectiveness based on existing empirical evidence. In the following, we describe how our treatments meet these criteria.

We address informational frictions in the *basic* treatment by improving the presentation and delivery of information: We offer a free monthly text-message service (SMS) that contains simplified invoice information. The text-message service ensures that customers have easy access to the invoice information in a non-technical language.[11] We target the intention-to-action gap through psychological commitment mechanisms (e.g., Laibson, 1997; Ariely and Wertenbroch, 2002). These elicit, through a phone call and a subsequent reminder by SMS, a commitment either to a desirable self-concept (in the *self-concept* treatment) or to a concrete action plan (in the *plan* treatment). Other obvious remedies to the high transaction cost of payments, such as direct debiting and technological innovations like mobile payments, were not possible at the time of the study.[12]

3.1 Treatments

The Basic Treatment – In the *basic* treatment, we call customers and offer a free monthly text-message service 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

[11] The short text format of the SMS also gave us the opportunity to simplify the wording of the invoice. We decided to use simplified language because, in a face-to-face survey conducted with a convenience sample (N=31) in Windhoek, Namibia's capital, about 45% of the participants were unable to indicate the total amount due when asked to explain a postal invoice.

[12] *NamWater* introduced some automated payment machines and a mobile payment bus after our intervention.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

consumption (in N$) of the last month. The text-message service did not substitute the regular postal invoice. Text messages were sent every month from October 2015 until June 2016, timed as closely as possible to the mailing of the written invoices. The *basic* treatment addresses the diagnosed obstacle of undelivered invoices. Tables A6-A8 of the online appendix provide the full telephone scripts and text message contents of all three treatments.

The Self-Concept Treatment – The *self-concept* treatment includes the content of the *basic* treatment. On top, it intends to invoke a water-paying self-concept by asking customers four similar questions, such as "How important is it to you to be a reliable water payer?". In the other three questions, the adjective *reliable* is replaced by either *good, responsible* or *debt-free.* The first sentence of the text message in the *self-concept* treatment reminds customers of their responses. It reads "Here is your invoice to you as a committed water payer". We use nouns ("water payer") rather than verbs ("pay water") because previous literature has shown that 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.[13]

[13] In order to make sure that the self-concept treatment could be implemented in the local language (Oshiwambo), we conducted a pilot study in which we asked native speakers about their perception of "lying" and "being a liar" in Oshiwambo. As in English, the latter provoked stronger reactions. To ensure that the framings in the *self-concept* treatment had the same meaning in both languages, we used an extensive version of back-translation: We asked our 25 research assistants individually, who are fluent in both languages and did not translate from English to Oshiwambo, to back-translate the Oshiwambo wording to English. The back-

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

The *self-concept* treatment may work through two channels. First, it can make already existing identities with respect to water payments salient. Second, it can induce customers to commit to a new water-paying self-concept. The first channel is related to interventions that prime identities and hence activate associated norms (Akerlof and Kranton, 2000; Benjamin et al., 2010). For example, crime-related primes increase dishonest behavior among criminals, but not regular citizens (Cohn et al., 2015). Priming the identity as a previous donor is more effective for more regular donors (Kessler and Milkman, 2018). Priming bank employees of their professional identity causes them to become more dishonest, whereas other professions do not become more dishonest (Cohn et al., 2014). Low-caste boys in India solve mazes as well as high-caste boys when caste is not revealed (Hoff and Pandey, 2014). The second channel builds on the notion that individuals' identities are not fixed, but constantly maintained and shaped through one's actions (behavior-identity link). According to this literature, individuals have uncertainty about their identity and inform themselves about their types by observing their actions (Bem, 1967; Bénabou and Tirole, 2011; Bryan et al., 2011). Individuals, therefore, have an incentive to choose actions that signal that they are a "good" type.

Importantly, the two channels make different predictions about the effectiveness of the intervention among customers with bad payment history. Priming high-debt customers of their non-paying identity might lead to an activation of non-paying norms. This could cause them to make even fewer payments in the future. Given that many customers have arrears, the intervention may in this case backfire. The second channel, by contrast, predicts increased water payments even for high-debt customers, because the *self-concept* intervention frames future behavior as a way to claim a desired identity. Since our pre-intervention survey

translation showed that the wording was very similar in both languages. We investigate the relevance of language in Section 5.6.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

suggested a high social desirability of paying regardless of respondents' debt status, backfiring among high-debt customers seems unlikely.

The Plan Treatment – The *plan* treatment is also implemented on top of the *basic* treatment. The *plan* treatment elicits specific action plans of customers about how, when, and where they would make payments, and asks them to commit to their plans to pay. As in the *self-concept* treatment, the first sentence of the text message intends to remind customers of their responses. Its wording is unique to the *plan* treatment and reads "As a reminder to your commitment to pay". There are several reasons how plan-making interventions might help individuals to follow through on their intentions (Rogers et al., 2015; Beshears et al., 2016; Abel et al., 2019). First, plan-making serves as a commitment and is arguably more effective when the commitment is made in front of others. Second, plan-making helps to overcome logistical obstacles, as individuals are prompted to think about specific implementation steps. Lastly, plan-making works as a reminder, as individuals are less likely to forget to act upon their intentions when they have specific plans to pay. We will discuss the role of these factors in our setting in Section 5.

3.2 Conducting the Experiment

We preregistered the experiment at the AEA RCT registry (AEARCTR-0000925). To randomize, we create three groups among the 12,719 *NamWater* customers whose account information includes a mobile phone number. Compared to private customers without a registered mobile phone number in our data set, these show a similar payment as well as consumption behavior in the pre-intervention year and accumulate a similar level of debt per month. As they represent more recently opened accounts, the overall level of debt is lower. That accounts with an associated phone number are, on average, older may be explained by the

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

fact that mobile-phone ownership rose exponentially from almost non-existent in the 2000s.[14] Table A9 provides summary statistics of the non-experimental sample.

We use the min-max t-stat method (see Bruhn and McKenzie, 2009) stratified by geographical location (proxied by pipeline location to which a customer 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), number of months as a *NamWater* customer, debt amount, and total invoice amount.

To call the 12,719 phone numbers, we set up a call center with 25 local students on *NamWater*'s premises in *Windhoek* in the last two weeks of September 2015. All phone callers took part in a three-day workshop and received in-depth training, including mock calls 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 from customers were handled in the same way.

All interactions with customers were fully scripted, practiced and the adherence to the script rigorously monitored. This procedure ensured high control of 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

[14] Mobile cellular subscription for Namibia can be found in the International Telecommunication Union (ITU) World Telecommunication/ICT Indicators Database (https://data.worldbank.org/indicator/IT.CEL.SETS.P2?locations=NA).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

could not be reached during a previous call attempt. We managed to talk to 9,823 (77%) of the assigned customers, which we will refer to as the intention-to-treat (ITT) sample. The majority of the customers 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 self-selection bias in the estimation of the treatment effects since unreachable customers cannot know that they were assigned to a treatment.

Table 1 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 considerable heterogeneity of customers in our sample.[15]

[Table 1 about here]

Summary statistics on the implementation of the intervention are presented in Table 2. Based on the phone callers' feedback, the vast majority of customers were delighted about the introduction of the SMS, which is reflected in near-universal take-up (about 98%). Roughly 12% of text messages each month are on average undeliverable (after automated retries) over the nine-month period. Typical reasons are technological restrictions such as network errors, deactivated numbers, or switched-off phones. Attrition rates are around 1.4%. There is no statistically significant difference in attrition rates between treatments, nor among observable

[15] The same table without top-coding variables is reproduced in Table A10. Top-coding reduces the influence of outliers on the mean value of the variables.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

characteristics. In most cases, customers attrite from the data because they have become inactive.[16]

[Table 2 about here]

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, about 87% of customers make plans to pay during October, the month after the phone calls. 49% of customers mention more than four concrete steps of the payment process and thus provide relatively detailed plans. Lastly, about 72% of interviews were conducted in Oshiwambo (the local language in our study region).

The cost for the intervention itself was 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.

3.3 Estimation Strategy

The infrequent payments of customers result in a large fraction of zeros in the dataset (66%). Furthermore, as is common for expenditure data, the distribution has a large variation (the

[16] Inactivity means that the account shows no account activity anymore (no payments or consumption). Accounts become inactive, for example, due to moving out, passing away, or being terminally disconnected from the water network.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

standard deviation is about twice the mean), is heavily right-skewed, and includes outliers. We, therefore, need a statistical model that can handle the large fraction of zeros and the skewness in the data.

In our main specification, we model the data-generating process by separately estimating the extensive margin (whether a customer pays or not) and the intensive margin (the payment amount condition on making a payment) through a two-part model. At the intensive margin, we apply the log-transformation to bring the payment data closer towards normality, without losing any observations. The combined effects of the two parts of the model are obtained by multiplying the predictions from these (Belotti et al., 2015). A further advantage of a two-part model is that the estimated effects on the extensive and the intensive margin are more informative than one aggregated absolute effect.

In addition, we also report the result from a typical transformation that is recommended instead of the log-transformation when the data includes a substantial fraction of zero values (Bellemare and Wichmann, 2020), the inverse hyperbolic sine (IHS). Online appendices A11 and A12 provide a more detailed discussion of our data as well as estimation strategy and compare different estimation approaches via simulations. Table A13 shows the robustness of our results to alternative outcome variable transformations.

4. Experimental Results

This section analyzes how the two commitment treatments affect 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 Section 6.

The left panel of Table 3 shows the intention-to-treat effects (ITT) in the first month after the intervention (October). Column 1 shows the marginal effects at means on the extensive margin (probability of making a payment) through a probit regression. The outcome variable of this model is dummy-coded, with 1 representing a payment in the respective month. Column

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

2 shows the effects on the intensive margin (the effect on the natural logarithm of the payment amount conditional on it being larger than zero). Column 3 estimates the combined effects of columns 1 and 2, using a two-part model which multiplies the predictions from columns 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). Columns 5-8 run the same statistical models as in columns 1-4 on the pooled monthly data from November to June, i.e., there are now eight observations per customer. All standard errors are clustered at the customer level.

[Table 3 about here]

Result 1: In the first month after the intervention, both commitment interventions significantly increase payments compared to the *basic* treatment: average payments increase by about 10% in the *self-concept* treatment and by about 26% in the *plan* treatment.

The *self-concept* treatment increases the likelihood of paying by 2.8 percentage points (p=0.015) and the *plan* treatment by 7.5 percentage points (p<0.001). The results are robust to multiple hypothesis adjustment using the conservative Bonferroni adjustment (p^{adj}=0.030 and p^{adj}<0.001) for two treatment arms (Savin, 1980). On the intensive margin, i.e., conditional on paying, 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 17.40 N$ (p=0.071) and for the *plan* treatment 46.58 N$ (p<0.001). This corresponds to an estimated increase of 10% and 26%, respectively, in comparison to the mean value of the *basic* treatment in October (178.89 N$). To put this into

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

perspective, an hourly wage at the time of the study corresponds to about 15 N$ or about 1.11 USD (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 show even larger and statistically significant overall effects on the payment amount: 19.7% (p=0.012) in the *self-concept* treatment and 59.5% (p<0.001) in the *plan* treatment. The suggested effect sizes of the IHS and other logarithmic transformations, however, should be interpreted with caution. The results of our simulation exercise (Table A12) show that these tend to overestimate the true effects in the presence of an excess of zeros.

Result 2: The effectiveness of the commitment treatments wanes over time, but the treatments show no rebound effect.

Treatment effects over time are shown in columns (5-8) in Table 3. The behavioral expectations were unclear ex-ante. On the one hand, the literature on consistency and habit formation suggests positive spillovers from short-term to long-term effects. On the other hand, negative rebound effects might offset the short-term effects because households might budget over time such that higher payments now induce lower payments in the future (see e.g., Szabó and Ujhelyi, 2015) or because of psychological licensing effects (Merritt et al., 2010). We find no evidence of such negative effects. If anything, we find that the *self-concept* treatment increases the likelihood of paying by about 1.3 percentage points per month (p=0.009) in comparison to the *basic* treatment. However, the estimated average combined treatment effect from the two-part model is insignificantly different from zero (p=0.457 and p=0.614, respectively).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

5. Discussion of Mechanisms

So far, we have argued that our psychological commitment treatments affect payment behavior through a commitment to either a specific payment plan or to a desirable self-concept, which narrows the gap between intentions and actual payment behavior. However, there are other mechanisms that could potentially drive the treatment effects. In the following, we discuss a number of alternative explanations and argue that they are unlikely to account for our treatment effects.

5.1 Scrutiny

One might be concerned that the commitment treatments change customers' perceptions of being scrutinized by *NamWater*, similar to the effects of scrutiny on reducing tax evasion through more truthful reporting by the taxpayers (e.g., Bott et al., 2020). In our context, however, the amount of money owed is not self-reported. Instead, it is calculated from the amount of water consumption that is read from the water meters by the water utility itself. Thus, in contrast to the literature on tax evasion, the utility already knows what needs to be paid. In addition, the *basic* treatment ensures that current account balances are common knowledge, in particular for those customers who do not receive or understand their mailed invoice.

5.2 Enforcement of Sanctions

Another concern could be that the commitment treatments serve as a signal about the probability of being sanctioned by the utility company. While we cannot rule this out entirely, there are several arguments against such an interpretation. If the company would like to enforce sanctions against customers, they are required to state so explicitly. Before customers are placed on a "cut-off list", they are contacted several times and receive a warning letter that informs customers which steps they have to take to avoid disconnections, for example by agreeing to a debt repayment plan. Disconnections are, in any case, very difficult to implement

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

due to ethical (water as a basic human right), technical and logistical constraints (e.g., due to limited staff) and are thus rarely enforced.

It is not obvious why customers could have confused our intervention with a threat of being sanctioned. Our phone call was framed as a service and "getting to know customers". The phone callers were very polite and friendly during the phone calls. Judging from the daily reports of our phone callers on their interaction with customers as well by the almost universal take-up of the SMS, customers had a very positive perception of the intervention. We are unaware of any callbacks to the customer care center asking whether the phone calls implied potential sanctions. Furthermore, *NamWater* implemented debt-forgiveness programs in the past, which may signal leniency rather than sanctions. Note also that any effect that comes merely from being contacted by phone calls of the water utility and by receiving short messages would already be captured by the *basic* treatment.

Nevertheless, to investigate this potential channel further, our dataset allows us to estimate whether customers that qualify for disconnections react differently to the treatment than customers that do not qualify for disconnections. According to the statement on the invoice, the water supply may be cut off if an invoice is not paid within 45 days. While this does not exactly match *NamWater's* internal threshold at which sanctions are enforced, it may be the information that customers use, when wondering whether our interventions imply that sanctions are now more likely. To analyze heterogeneous effects, we split our sample into three groups of customers based on pre-treatment payment behavior. The first group comprises customers that pay in advance and thus have paid more than they owe (18% of the sample). The second group contains customers that have arrears, but paid their invoices before 45 days

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

are over and thus are not eligible for sanctions (41%).[17] The last group comprises customers that have arrears that would make them eligible for disconnections (41%), but – as mentioned earlier – are rarely disconnected in practice. If our treatments affected behavior by changing beliefs about the enforcement of sanctions, we would expect the latter group to react most strongly. Heterogeneous treatment effects for the three groups are displayed in Table 4.

[Table 4 about here]

We do not find treatment differences between customers of the second and the third group (i.e., between customers with debt above and below the disconnection "eligibility" threshold). The effect on the propensity to pay is very similar between customers in both groups. The combined effects of the two-part model are also statistically insignificantly different between the two groups in the *self-concept* and the *plan* treatment (p=0.670 and p=0.457, Wald test). On a different note, this also suggests that the intervention is similarly effective for customers with high relative debt.[18]

---

[17] Since *NamWater* uses monthly data, we classify customers in that group if their billed debt does not exceed their last two invoices (i.e., they have no debt that is older than 60 days).

[18] In a second test, we mirror NamWater's internal rule to decide against which customers to initiate sanctions, as we do not have comprehensive access to the sanctioning lists of *NamWater*. *NamWater's* approach to decide whom to sanction is to divide customers' debt by their invoice (a rough measure of how many bills are unpaid) and to initiate enforcement action against the most severe debtors. In Table A14, we median-split our sample into two groups based on that metric, with the high-debt group being much more likely to face or have faced sanctions. The treatment effects among both groups remain similar, suggesting that there are

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

5.3 Reminders and Personal Contact

Since we compare the commitment treatments to the *basic* treatment (instead of an untreated control group), we can rule out another set of alternative explanations. The phone call and the free text-message service serve as a reminder to pay (e.g., Cadena and Schoar, 2011; Karlan et al., 2016a) in all three treatments. Hence, any additional effect from the two commitment treatments cannot be explained by being reminded. Similar arguments hold true for having unexpected personal contact with the phone caller (Karlan et al., 2016b) and for showing reciprocity for the water company reaching out to its customers (Szabó and Ujhelyi, 2015).

5.4 Logistical Hurdles (Plan Treatment)

Besides its function as a commitment device, planning prompts might help individuals follow up on their intentions by overcoming logistical obstacles. Logistical obstacles should be larger for customers who live further away from their postboxes. If the *plan* treatment helped to overcome logistical obstacles, we would expect the *plan* treatment to be more effective among customers that have to cover longer distances to make payments (more time and more logistical steps). To estimate heterogeneous treatment effects with respect to travel distance to the post box, we collect geo-location proxies for a subsample of customers for which these are available. Estimates are provided in Table A15. We do not find statistically significant heterogeneous treatment differences.[19]

---

no differences between customers against whom *NamWater* is more likely to have already initiated enforcement action.

[19] We proxied a customer's travel distance to the post office by estimating the street distance between the midpoint of a customer's pipeline scheme (our most precise geographical location identifier) and the location of the post office (village or city) where the postbox was located.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

5.5 No Evidence of Priming (Self-concept Treatment)

A potential concern of the *self-concept* treatment is that non-paying customers might be primed with a non-paying self-concept and thus pay even less in the future. The intervention would then backfire among customers with high debt. However, we find that the *self-concept* treatment increases payments even for customers who, given their payment history, have little reason to think of themselves as responsible or debt-free water payers (see Table 4). Evidence on the effectiveness among such groups is, to the best of our knowledge, missing. This distinguishes the effects of our *self-concept* treatment from typical findings in the priming literature (see Section 3). Our results, therefore, suggest that the intervention does not prime a fixed identity. Instead, the commitment to being a responsible water payer allows them to claim a desirable identity.

5.6 Treatment Differences with Respect to Language

Another concern, in particular for the *self-concept* treatment, is whether the psychological commitment depends on the language in which it is expressed. Oshiwambo is the native language of around 80% of the population where the experiment was conducted. In addition to Oshiwambo, we also offered to communicate in English. Most likely, English was chosen by the respondents (in about a fourth of the phone calls, see Table 2) as a language of the conversation if their mother tongue was one of the many other languages spoken in Namibia and they felt more fluent in English than in Oshiwambo. We find that phone calls conducted in Oshiwambo are associated with much stronger treatment effects than phone calls in English

---

We further restricted our sample to customers who had their postbox within the experimental region, as some customers had their invoices sent to Windhoek or other parts of Namibia (relatives or second home).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

(see Table A16). This suggests that commitment works better when expressed in one's mother tongue. Note, however, that this is only tentative evidence as the language of the phone calls was chosen by the respondents. Therefore, we cannot control for potential and unobservable confounding factors that might drive these heterogeneous treatment effects.

6. Effectiveness of the Basic Treatment

This section estimates the effect of the *basic* treatment in comparison to an *untreated* comparison group. We estimate the treatment effect of the *basic* treatment in comparison to an *untreated* group, using matching methods. We match customers in the *basic* treatment with customers whom we never had contact with but are from the same region as the customers from our experimental sample. The matching group comprises customers who did not provide a phone number when they set up their account as well as customers whom we tried unsuccessfully to reach three times and consequently never actually had contact with, for example, because of network errors, outdated phone numbers, or unanswered calls (N=8,902). The *untreated* group is therefore not part of the randomization process. This allowed a higher number of observations and thus higher statistical power for our commitment treatments. Note that take-up of the treatment among the *untreated* group is impossible. Therefore, the most important concerns of non-randomized designs (self-selection and targeting of program participants based on their pre-intervention payment history) do not apply. Another concern could be that customers without a registered phone are more liquidity-constrained than customers with a registered phone and hence less able to pay. However, not having a registered phone number seems to correlate with the age of the account, rather than with the customer's ability to pay the bills (see Table A9 and Figure A2). Furthermore, our rich panel dataset allows us to construct groups with similar pre-intervention levels and parallel pre-intervention trends. The age of the account is included in our set of matching variables (see Table 5 for a full list).

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

For our main estimation, we use entropy balancing (Hainmueller, 2012) to create a comparison group for the *basic* treatment. 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 robustness check, we provide estimates from a second matching method, coarsened exact matching (Iacus et al., 2012), to show insensitivity with respect to the matching procedure (see Table A17). Coarsened exact matching (CEM) creates sets of strata for pre-defined covariates and then finds matches (in the 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, 2019).

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 consumption to ensure successful matching. Table A18 shows that for both matching methods there are no statistically significant differences between the *basic* and the matched *untreated* groups, with respect to levels or trends over time. Table 5 shows the treatment effects from entropy balancing, using the regression models as in previous tables.

[Table 5 about here]

Result 3: The *basic* treatment increases average payments by about 29% in the first month of the intervention in comparison to an untreated group. Over the course of the intervention, the *basic* treatment increases average payments by about 8%.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

The treatment effects on payments are largest in the first month of the intervention (October) when the payment amounts increase by about 43.33 N$ in the *basic* treatment (p<0.001). This corresponds to an increase of 29% (to the respective untreated payment mean). This effect is quite large and corresponds to about 2-3 hourly wages. This means that our two commitment treatments lead to a short-term increase in payments of 39% and 55%, respectively, in comparison to an untreated group.

For the *basic* treatment, we find that there is a strong and lasting effect on the payment propensity over the course of the study (p<0.001). This is particularly interesting given the substantial transaction costs of making payments. There is a negative effect on the intensive margin (p=0.003), which may be due to budgeting effects or lower amounts of outstanding debt. The treatment effect of the two-part model shows an increase in payments of 8.93 N$ (p<0.001), which corresponds to a persistent increase of about 8% in comparison to the untreated payment mean.

7. Conclusion

Designing effective interventions hinges on understanding the local context and the relevant obstacles to behavioral change (e.g., List, 2011; Duflo, 2017). Yet, Rodrik's (2010) observation that the process of diagnosing before prescribing is rarely made explicit still seems valid today. Our study makes this process more transparent. We diagnose which behavioral obstacles are most common at scale through simple diagnostic tools – a survey of a randomly selected sample of the target population – before designing and implementing interventions. In a setting where sanctioning tools are difficult to implement, we diagnose two unexpected obstacles that can be addressed through soft, scalable, and cost-effective interventions. Our treatments target informational frictions and make use of psychological commitment techniques to narrow the gap between customers' willingness to pay and their actual payment behavior. Initially, payments increase by 29% to 55% (depending on treatment). While the impact of the *basic*

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

treatment is long-lasting, the effectiveness of the psychological commitment techniques wanes over time.

Our diagnosis led to different interventions than if we had followed the previous literature on non-payment for public utilities or advice of local experts. Merely relying on experts' advice may indeed have only limited value, as shown by DellaVigna and Pope (2018). The treatment effects of our interventions compare favorably to a much more costly intervention (a door-to-door campaign) in South Africa's water sector (Szabó and Ujhelyi, 2015), to large-scale nudges in the US (DellaVigna and Linos, 2022), and tax compliance nudges, in particular in light of their lower effectiveness in low- and middle-income countries (Antinyan and Asatryan, 2019). Against the backdrop of White's finding (2019) that 80% of large-scale interventions show weak or no positive effects, one may wonder whether the failure rate of interventions may be lowered by first surveying a random sample of the target population before designing interventions. As our study lacks a counterfactual to our diagnostic approach, it would be highly worthwhile for future research to measure its potential benefits.

Our intervention only targets customers that can be reached by phone. As mobile phone coverage is already widespread and increasing even in low-income countries, this is becoming even less of a restriction for future interventions. On the other hand, phone-based interventions may have to increasingly compete against other interventions or campaigns for the attention of phone users. A related concern is whether the treatments cause negative spillover on the payments of other utility bills. While we did not measure such spillovers in our experiment, a related study by Tonke (2023) suggests that water utility bill payments do not systematically crowd out payments of other utility bills. Given the short-lived effects of the psychological commitment treatments, future research may also address whether and how the psychological commitment treatments could be adapted to yield a longer-term impact. Possibly, repeatedly eliciting a plan to pay would have resulted in more sustainable effects.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Learning more about how to reap long-term impact is also worthwhile in light of the many domains where non-compliance is an important issue, such as tax evasion (e.g., Hallsworth et al., 2017) and non-payment of fees (e.g., Fellner et al., 2013). Little attention has hitherto been devoted to the question whether a substantial share of individuals regularly fail to act despite their earnest intention to comply. The presence of intention-to-action gaps offers policy options in contexts when it is difficult to further increase motivation to comply, e.g. because sanctioning or monitoring tools are difficult to implement

References

Abel, Martin, Rulof Burger, Eliana Carranza, and Patrizio Piraino. 2019. Bridging the Intention-Behavior Gap? The Effect of Plan-Making Prompts on Job Search and Employment. *American Economic Journal: Applied Economics* 11 (2): 284–301.

Aguilar-Benitez, Ismael, and Jean-Daniel Saphores. 2008. An Analysis of Nonpayment of Water Services in Nuevo Laredo, Mexico. *Water Resources Research* 44 (8).

Akerlof, George A., and Rachel E. Kranton. 2000. Economics and Identity*. *The Quarterly Journal of Economics* 115 (3): 715–53.

Antinyan, Armenak, and Zareh Asatryan. 2019. Nudging for Tax Compliance: A Meta-Analysis. SSRN Scholarly Paper. Rochester, NY.

Ariely, Dan, and Klaus Wertenbroch. 2002. Procrastination, Deadlines, and Performance: Self-Control by Precommitment. *Psychological Science* 13 (3): 219–24.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Atkins, Lou, Jill Francis, Rafat Islam, Denise O'Connor, Andrea Patey, Noah Ivers, Robbie Foy, et al. 2017. A Guide to Using the Theoretical Domains Framework of Behaviour Change to Investigate Implementation Problems. *Implementation Science* 12 (1): 77.

Bellemare, Marc F., and Casey J. Wichman. 2020. Elasticities and the Inverse Hyperbolic Sine Transformation. *Oxford Bulletin of Economics and Statistics* 82 (1): 50–61.

Belotti, Federico, Partha Deb, Willard G. Manning, and Edward C. Norton. 2015. Twopm: Two-Part Models. *The Stata Journal: Promoting Communications on Statistics and Stata* 15 (1): 3–20.

Bem, Daryl J. 1967. Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena. *Psychological Review* 74 (3): 183–200.

Bénabou, Roland, and Jean Tirole. 2011. Identity, Morals, and Taboos: Beliefs as Assets *. *The Quarterly Journal of Economics* 126 (2): 805–55.

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. 2010. Social Identity and Preferences. *American Economic Review* 100 (4): 1913–28.

Beshears, John, Katherine L. Milkman, and Joshua Schwartzstein. 2016. Beyond Beta-Delta: The Emerging Economics of Personal Plans. *American Economic Review* 106 (5): 430–34.

Bott, Kristina M., Alexander W. Cappelen, Erik Ø. Sørensen, and Bertil Tungodden. 2020. You've Got Mail: A Randomized Field Experiment on Tax Evasion. *Management Science* 66 (7): 2801–19.

Bridges, Sarah, and Richard Disney. 2004. Use of Credit and Arrears on Debt among Low-Income Families in the United Kingdom. *Fiscal Studies* 25 (1): 1–25.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Bruhn, Miriam, and David McKenzie. 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics* 1 (4): 200–232.

Bryan, Christopher, Gabrielle Adams, and Benoît Monin. 2013. When Cheating Would Make You a Cheater: Implicating the Self Prevents Unethical Behavior. *Journal of Experimental Psychology. General* 142 (4).

Bryan, Christopher J., Allison Master, and Gregory M. Walton. 2014. 'Helping' Versus 'Being a Helper': Invoking the Self to Increase Helping in Young Children. *Child Development* 85 (5): 1836–42.

Bryan, Christopher J., Elizabeth Tipton, and David S. Yeager. 2021. Behavioural Science Is Unlikely to Change the World without a Heterogeneity Revolution. *Nature Human Behaviour* 5 (8): 980–89.

Bryan, Christopher J., Gregory M. Walton, Todd Rogers, and Carol S. Dweck. 2011. Motivating Voter Turnout by Invoking the Self. *Proceedings of the National Academy of Sciences* 108 (31): 12653–56.

Burbidge, John B., Lonnie Magee, and A. Leslie Robb. 1988. Alternative Transformations to Handle Extreme Values of the Dependent Variable. *Journal of the American Statistical Association* 83 (401): 123–27.

Bursztyn, Leonardo, Stefano Fiorin, Daniel Gottlieb, and Martin Kanz. 2019. Moral Incentives in Credit Card Debt Repayment: Evidence from a Field Experiment. *Journal of Political Economy* 127 (4): 1641–83.

Cadena, Ximena, and Antoinette Schoar. 2011. Remembering to Pay? Reminders vs. Financial Incentives for Loan Payments. Working Paper. Working Paper Series. National Bureau of Economic Research.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Cohn, Alain, Ernst Fehr, and Michel André Maréchal. 2014. Business Culture and Dishonesty in the Banking Industry. *Nature* 516 (7529): 86–89.

Cohn, Alain, Michel André Maréchal, and Thomas Noll. 2015. Bad Boys: How Criminal Identity Salience Affects Rule Violation. *The Review of Economic Studies* 82 (4): 1289–1308.

Datta, Saugato, and Sendhil Mullainathan. 2014. Behavioral Design: A New Approach to Development Policy. *Review of Income and Wealth* 60 (1): 7–35.

DellaVigna, Stefano, and Elizabeth Linos. 2022. RCTs to Scale: Comprehensive Evidence From Two Nudge Units. *Econometrica* 90 (1): 81–116.

DellaVigna, Stefano, and Devin Pope. 2018. What Motivates Effort? Evidence and Expert Forecasts. *The Review of Economic Studies* 85 (2): 1029–69.

Dinkelman, Taryn. 2011. The Effects of Rural Electrification on Employment: New Evidence from South Africa. *American Economic Review* 101 (7): 3078–3108.

Duan, Naihua. 1983. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association* 78 (383): 605–10.

Duflo, Esther. 2017. The Economist as Plumber. *American Economic Review* 107 (5): 1–26.

Duflo, Esther, and Rohini Pande. 2007. Dams. *The Quarterly Journal of Economics* 122 (2): 601–46.

du Plessis, NP, Neels, Cathline., Anyim, Imelda, and Matros, Anna. (2005). Perceptions on affordability and willingness to pay for water services. Research Document.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler. 2013. Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information. *Journal of the European Economic Association* 11 (3): 634–60.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford University Press.

Finger, Matthias, Jeremy Allouche, and Patricia Luis-Manso. 2007. *Water and Liberalisation*. IWA Publishing.

Frey, Erin, and Todd Rogers. 2014. Persistence: How Treatment Effects Persist After Interventions Stop. *Policy Insights from the Behavioral and Brain Sciences* 1 (1): 172–79.

Gollwitzer, Peter M. 1999. Implementation Intentions: Strong Effects of Simple Plans. *American Psychologist* 54: 493–503.

Gollwitzer, Peter M., and Paschal Sheeran. 2006. Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes. In *Advances in Experimental Social Psychology* 38:69–119. Academic Press.

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

Hainmueller, Jens. 2012. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. *Political Analysis* 20 (1): 25–46.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev. 2017. The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance. *Journal of Public Economics* 148: 14–31.

Hoff, Karla, and Priyanka Pandey. 2014. Making up People—The Effect of Identity on Performance in a Modernizing Society. *Journal of Development Economics* 106: 118–31.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Iacus, Stefano M., Gary King, and Giuseppe Porro. 2012. Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis* 20 (1): 1–24.

Jack, B. Kelsey, and Grant Smith. 2015. Pay as You Go: Prepaid Metering and Electricity Expenditures in South Africa. *American Economic Review* 105 (5): 237–41.

Jensen, Olivia, and Namrata Chindarkar. 2019. Sustaining Reforms in Water Service Delivery: The Role of Service Quality, Salience, Trust and Financial Viability. *Water Resources Management* 33 (3): 975–92.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. Getting to the Top of Mind: How Reminders Increase Saving. *Management Science* 62 (12): 3393–3411.

Karlan, Dean, Melanie Morten, and Jonathan Zinman. 2016. A Personal Touch in Text Messaging Can Improve Microloan Repayment. *Behavioral Science & Policy* 1: 25–31.

Kayaga, Sam, Richard Franceys, and Kevin Sansom. 2004. Bill Payment Behaviour in Urban Water Services: Empirical Data from Uganda. *Journal of Water Supply: Research and Technology-Aqua* 53 (5): 339–49.

Kessler, Judd B., and Katherine L. Milkman. 2018. Identity in Charitable Giving. *Management Science* 64 (2): 845–59.

King, Gary, and Richard Nielsen. 2019. Why Propensity Scores Should Not Be Used for Matching. *Political Analysis* 27 (4): 435–54.

Klintenberg, Patrik, Clarence Mazambani, and Komeine Nantanga. 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

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Konow, James. 2000. Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions. *American Economic Review* 90 (4): 1072–91.

Laibson, David. 1997. Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics* 112 (2): 443–78.

List, John A. 2011. Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *Journal of Economic Perspectives* 25 (3): 3–16.

Lu, Xuefei, Lixin Wang, Ming Pan, Kudzai F. Kaseke, and Bonan Li. 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, James G., and Lonnie Magee. 1990. Transforming the Dependent Variable in Regression Models. *International Economic Review* 31 (2): 315–39.

Mazambani, Clarence, Sarah Schönbrodt-Stitt, and Patrik Klintenberg. 2006. Water supply: a gift from God or does it come with a cost? World Water Week in Stockholm. (August 20-26, 2006). Stockholm: 150-151.

McRae, Shaun. 2015. Infrastructure Quality and the Subsidy Trap. *American Economic Review* 105 (1): 35–66.

Merritt, Anna C., Daniel A. Effron, and Benoît Monin. 2010. Moral Self-Licensing: When Being Good Frees Us to Be Bad. *Social and Personality Psychology Compass* 4 (5): 344–57.

Michie, Susan, Maartje M. van Stralen, and Robert West. 2011. The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions. *Implementation Science* 6 (1): 42.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Milkman, Katherine, John Beshears, James Choi, David Laibson, and Brigitte C. Madrian. 2013. Planning Prompts as a Means of Increasing Preventive Screening Rates. *Preventive Medicine* 56 (1): 92–93.

Milkman, Katherine L., John Beshears, James J. Choi, David Laibson, and Brigitte C. Madrian. 2011. Using Implementation Intentions Prompts to Enhance Influenza Vaccination Rates. *Proceedings of the National Academy of Sciences* 108 (26): 10415–20.

Mugabi, Josses, Sam Kayaga, Ian Smout, and Cyrus Njiru. 2010. Determinants of Customer Decisions to Pay Utility Water Bills Promptly. *Water Policy* 12 (2): 220–36.

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

Nickerson, David W., and Todd Rogers. 2010. Do You Have a Voting Plan?: Implementation Intentions, Voter Turnout, and Organic Plan Making. *Psychological Science* 21 (2): 194–99.

Rodrik, Dani. 2010. Diagnostics before Prescription. *Journal of Economic Perspectives* 24 (3): 33–44.

Rogers, Todd, Katherine L. Milkman, Leslie K. John, and Michael I. Norton. 2015. Beyond Good Intentions: Prompting People to Make Plans Improves Follow-through on Important Tasks. *Behavioral Science & Policy* 1 (2): 33–41.

Savin, Eugene N. 1980. The Bonferroni and the Scheffé Multiple Comparison Procedures. *The Review of Economic Studies* 47 (1): 255–73.

Sheeran, Paschal, Thomas L. Webb, and Peter M. Gollwitzer. 2005. The Interplay Between Goal Intentions and Implementation Intentions. *Personality and Social Psychology Bulletin* 31 (1): 87–98.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Strand, Jon. 2012. Low-Level versus High-Level Equilibrium in Public Utility Services. *Journal of Public Economics* 96 (1): 163–72.

Szabó, Andrea, and Gergely Ujhelyi. 2015. Reducing Nonpayment for Public Utilities: Experimental Evidence from South Africa. *Journal of Development Economics* 117: 20–31.

Thaler, Richard H. 2018. From Cashews to Nudges: The Evolution of Behavioral Economics. *American Economic Review* 108 (6): 1265–87.

Tonke, Sebastian. 2023. Shaping Identity: Evidence from a Large-scale Field Experiment. *Working paper*.

United Nations Children's Fund and World Health Organization (2019). Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities.

Vásquez, William F. 2015. Nonpayment of Water Bills in Guatemala: Dissatisfaction or Inability to Pay? *Water Resources Research* 51 (11): 8806–16.

Vásquez, William F., and Jessica Alicea-Planas. 2017. Factors Associated with Nonpayment Behavior in the Water Sector of Nicaragua. *Utilities Policy* 47: 50–57.

Walton, Gregory M., and Mahzarin R. Banaji. 2004. Being What You Say: The Effect of Essentialist Linguistic Labels on Preferences. *Social Cognition* 22 (2): 193–213.

White, Howard. 2019. The Twenty-First Century Experimenting Society: The Four Waves of the Evidence Revolution. *Palgrave Communications* 5 (1): 1–7.

World Bank Group (2017). Towards Effective Enforcement of Uncontested Monetary Claims: Lessons from Eastern and Central Europe. Retrieved from: https://openknowledge.worldbank.org/handle/10986/28400

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Table 1. Pre-intervention Summary Statistics (ITT Sample)

	Mean	SD	P25	Median	P75	Mean diff. to Basic	p-value
Basic (N= 3,300)							
Payment in N$	100.16	237.30	0	0	100	-	-
Payment in N$ if >0	292.98	378.06	81	158	315	-	-
Consumption in N$	125.49	228.38	21.20	53.70	119.40	-	-
Number of payments	4.15	2.75	2	4	6	-	-
Debt in N$	510.36	1274.77	31.95	138.55	456.67	-	-
Account age in month	44.30	39.22	19	34	53	-	-
Self-Concept (N=3,312)							
Payment in N$	100.07	237.02	0	0	100	-0.089	0.971
Payment in N$ if >0	292.78	381.82	85	156	301	-0.202	0.979
Consumption in N$	123.15	223.92	26.70	53.70	115.70	-2.335	0.462
Number of payments	4.18	2.69	2	4	6	0.028	0.683
Debt in N$	466.68	1203.76	37.27	125.61	406.75	-43.67	0.165
Account age in month	45.37	39.70	19	34	55	1.070	0.281
Plan (N=3,264)							
Payment in N$	99.77	237.64	0	0	100	-0.392	0.875
Payment in N$ if >0	292.03	379.47	80	155.50	300	-0.953	0.903
Consumption in N$	124.31	227.86	21.66	50.75	111.65	-1.179	0.717
Number of payments	4.15	2.70	2	4	6	<0.001	0.990
Debt in N$	495.59	1334.25	31.65	127.33	417.36	-14.77	0.657
Account age in month	45.11	39.61	19	34	55	0.812	0.415

Notes: The table reports summary statistics of the pre-intervention year for the ITT sample. All continuous variables are top coded (winsorized) at the 99th percentile. The unwinsorized data can be found in Table A10. 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.

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Table 2. Summary Statistics of the Implementation

Treatment	Take-up rate	Delivery rate of text messages	Attrition rate	Most common commitment answer	2nd-most common commitment answer	Plans to pay in October	Phone call in Oshiwambo
Basic	0.990	0.883	0.014	-	-	-	0.703
Self-concept	0.972	0.885	0.013	0.627 (V. import.)	0.329 (Important)	-	0.736
Plan	0.974	0.884	0.016	0.494 (4+ steps)	0.441 (2-3 steps)	0.865	0.733

Notes: The column "Most common commitment answer" shows the answer that was given most frequently by customers when the phone callers implemented the self-concept or plan treatment (i.e., "Very important" or at least 4 concrete implementation steps). 2nd-most common commitment answer shows the second most frequently elicited answer (i.e., "important" or "2-3 concrete implementation steps").

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Table 3. ITT Effects in Comparison to Basic Treatment

	Initial month (October)				Medium term (November-June)			
	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$	(5) Payment propensity (binary)	(6) Log (Payment amount | >0)	(7) Combined Effect on payment amount	(8) IHS Payment amount in N$
Self-concept	0.028**	0.013	17.403*	0.180**	0.013***	-0.023	1.499	0.068**
	(0.011)	(0.031)	(9.631)	(0.071)	(0.005)	(0.021)	(2.017)	(0.029)
Plan	0.075***	0.035	46.573***	0.467***	0.009*	-0.033	-1.047	0.041
	(0.012)	(0.031)	(10.176)	(0.072)	(0.005)	(0.021)	(2.077)	(0.030)
Comparison mean	0.522	5.178	178.891	3.063	0.393	5.157	120.957	2.302
Observations	9,823	5,456	9,823	9,823	78,025	31,291	78,025	78,025
R-Squared	0.115	0.330	-	0.128	0.089	0.173	-	0.107
Model	Probit	OLS	Two-part model	OLS	Probit	OLS	Two-part model	OLS

Notes: The table reports ITT effects in comparison to *basic*. Column 1 shows the treatment effects of a probit regression (marginal effects at means). Column 2 shows OLS estimates on the intensive margin. Column 3 uses a two-part model (Belotti et al., 2015) to get an estimate of the combined effect. Standard errors are obtained by bootstrapping. Column 4 shows OLS estimates on the inverse hyperbolic sine (IHS) of payment amount. Regressions control for pre-treatment means of the variables 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. Columns 5-8 show treatment effects on the pooled monthly data from November to June (eight obs. per customer). Columns 5-8 also include year-month fixed effects. Robust standard errors are clustered at the customer level (if applicable). * p<0.1; ** p<0.05; *** p<0.01

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Table 4. No Heterogeneous Treatment Effects with Respect to Sanctioning Probability

	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$
Self-concept (prepayers)	0.004	0.080	19.970	0.056
	(0.030)	(0.086)	(23.880)	(0.178)
Self-concept (below cut-off threshold)	0.030*	-0.002	13.580	0.170*
	(0.017)	(0.037)	(11.496)	(0.095)
Self-concept (above cut-off threshold)	0.029	0.034	21.479	0.215*
	(0.018)	(0.064)	(16.806)	(0.126)
Plan (prepayers)	0.016	0.045	17.698	0.135
	(0.029)	(0.085)	(23.411)	(0.177)
Plan (below cut-off threshold)	0.090***	0.006	43.568***	0.484***
	(0.017)	(0.038)	(12.196)	(0.096)
Plan (above cut-off threshold)	0.081***	0.094	59.030***	0.577***
	(0.018)	(0.061)	(17.255)	(0.127)
Observations	9,823	5,456	9,823	9,823
R-squared	0.123	0.352	-	0.135

Notes: The table reports ITT effects on payment behavior for October 2015 in comparison to the *basic* treatment. We split our sample into subgroups w.r.t. their eligibility of getting sanctioned based on pre-treatment data. Column 1 shows the treatment effects of a probit regression (marginal effects at means). Column 2 shows OLS estimates on the intensive margin. Column 3 uses a two-part model (Belotti et al., 2015) to get an estimate of the combined effect. 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. Column 4 shows OLS estimates on the inverse hyperbolic sine (IHS) of payment amount. We include pre-treatment means of the variables 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. Standard errors (in parentheses) are clustered at the customer level to account for serial correlation. * p<0.1; *** p<0.01

Review of Economics and Statistics Just Accepted MS.
https://doi.org/10.1162/rest_a_01363

Table 5. Treatment Effects of Basic Treatment in Comparison to Untreated Group

	(1) Payment propensity (binary)	(2) Log (Payment amount | >0)	(3) Combined Effect on payment amount	(4) IHS Payment amount in N$
Initial Month	0.088***	0.038	43.332***	0.539***
(October)	(0.012)	(0.038)	(8.728)	(0.073)
Medium term	0.052***	-0.073***	8.929***	0.278***
(November-June)	(0.006)	(0.025)	(2.466)	(0.036)
Comp. mean (Oct.)	0.440	5.157	150.199	2.577
Comp. mean (Nov.-June)	0.347	5.244	114.154	2.060
Model	Weighted probit	Weighted OLS	Weighted Two-part model	Weighted OLS

Notes: The regressions show treatment effects of the *basic* treatment in comparison to the matching comparison group, using entropy balancing. We match on the following pre-intervention variables, which closely resemble those used for randomization: Biannual payment propensity, biannual payment amount (if payment amount is >0), debt in month before intervention, age of account, fraction of bill paid, inverse hyperbolic sine of water consumption, inactivity (no water consumption) in the month prior to intervention and total months of inactivity in the pre-intervention year. Entropy balancing uses a weighted sample of untreated customers to match the treated customers, which results in a match of 3,046 *basic* treatment customers, with 7,090 weighted untreated customers. Some customers in the basic treatment are not matched because they have missing pre-treatment values, which are necessary for the matching algorithm. All standard errors (in parentheses) are clustered at the customer level to account for serial correlation. *** p<0.01