NBER WORKING PAPER SERIES

THE POWER TO CONSERVE:
A FIELD EXPERIMENT ON ELECTRICITY USE IN QATAR

Omar Al-Ubaydli
Alecia W. Cassidy
Anomitro Chatterjee
Ahmed Khalifa
Michael K. Price

Working Paper 31931
http://www.nber.org/papers/w31931

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2023

We thank the following scholars for helpful feedback: Dylan Brewer, Traviss Cassidy, Kenneth Gillingham, Elisabeth Gsottbauer, Ganga Shreedhar, and Jeffrey Smith. We thank KAHRAAMA for providing us with the electricity consumption data. This research was made possible through grant numbers NPRP9-232-5-026 and NPRP12C-0821-190017 from the Qatar National Research Fund, Qatar Foundation. The research was approved under Qatar University's Institution Review Board QU-IRB-999-EA/18. The statements and opinions expressed hereis are those of the authors and do not necessarily reflect those of Qatar Foundation, or KAHRAAMA. No party had any right of first refusal over our results. This study is registered in the AEA RCT Registry and the unique idnetifying number is AEARCTR-0012431. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Power to Conserve: A Field Experiment on Electricity Use in Qatar
Omar Al-Ubaydli, Alecia W. Cassidy, Anomitro Chatterjee, Ahmed Khalifa, and Michael K. Price
NBER Working Paper No. 31931
December 2023
JEL No. C93,D90,Q4

ABSTRACT

High resource users often have the strongest response to behavioral interventions promoting conservation. Yet, litlle is known about how to motivate them. We implement a field experiment in Qatar, where residential customers have some of the highest energy use per capita in the world. Our dataset consists of 207,325 monthly electricity meter readings from a panel of 6,096 customers. We employ two normative treatments priming identity - a religious message quoting the Qur'an, and a national message reminding households that Qatar prioritizes energy conservation. The treatments reduce electricity use by 3.8%and both messages are equally effective. Using machine learning methods on supplemental survey data, we elucidate how agency, motivation, and responsibility activate conservation responses to our identity primes.

Omar Al-Ubaydli
Department of Economics and Mercatus Center
George Mason University
omar@omar.ec

Alecia W. Cassidy
Department of Economics, Finance
and Legal Studies
University of Alabama
200 Alston Hall, Box 870224, 361 Stadium
Tuscaloosa, AL 35487
awcassidy1@cba.ua.edu

Anomitro Chatterjee
London School of Economics and
Political Science
a.chatterjee6@lse.ac.uk

Ahmed Khalifa
Department of Finance and Economics
College of Business and Economics
Qatar University
Doha 2713
aliabdelkh91@gmail.com

Michael K. Price
Department of Economics, Finance,
and Legal Studies
The University of Alabama
250 Alston Hall
Box 870224
Tuscaloosa, AL 35487
and Australian National University
and also NBER
mkprice2@ua.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/12431

1 Introduction

While there have been great advances in understanding the beliefs, behaviour, and habits that impact residential energy use (Attari et al., 2010; Allcott, 2011; Byrne et al., 2018; Delmas et al., 2013; Ferraro and Price, 2013; Ito et al., 2018; Murakami et al., 2022), the determinants of energy use among the top 1% of global users remains an unanswered question. These super-users have a greater potential to reduce energy use and carbon emissions than the average consumer. Strong evidence suggests the highest energy users respond differently to both pecuniary and non-pecuniary incentives (Ferraro and Price, 2013), underscoring the need for further investigation of the determinants of their energy use behavior. Moreover, the literature on these interventions has established high users to be most responsive to treatment, suggesting that utility companies would get the most bang for their buck by targeting these users (Allcott, 2011; Byrne et al., 2018; Knittel and Stolper, 2021; Gerarden and Yang, 2023).

Yet, little is known on how to most effectively motivate high users. In particular, correcting errors in beliefs about one's own use among high users does not lead to increased conservation (Byrne et al., 2018; Murakami et al., 2022). If anything, high users are the least sensitive to information (Byrne et al., 2018). Despite this, they respond the most to information treatments.[1] This is puzzling, and leaves open the question of how and why high users reduce the most in response to information treatments and other behavioral interventions.[2,3]

Two related questions are most interesting to us. First, if information is ineffective for high users, but treatment effects are nevertheless large, would injunctive norms or encouragement work without providing information? Second, what motivates the high-users that are most responsive? The answer to the second question could help policymakers better activate the most effective channels promoting conservation among high users.

We answer these questions by studying a context where the entire population consists almost exclusively of high users- Qatar. Qatar has one of the highest levels of per capita energy consumption and per capita CO_2 emissions in the world (International Energy Agency, 2018). Electricity is provided at subsidized rates to non-nationals and free of cost to Qatari nationals,[4] which poses a unique challenge to reduction of energy use where prices do not reflect marginal cost of production.

This is a high-stakes context for which research on the efficacy of non-pecuniary interventions is critical. Current estimates by the IEA rank Qatar as the country with the highest per-capita fossil fuel subsidies in the world, at US$2,326 per person. The electricity sector in

[1] Byrne et al. (2018) find that while low users who tend to overestimate use increase their use when information about their relative use is provided, high-users who tend to underestimate use do not exhibit a symmetric effect. High baseline use itself exerts a large and independent effect on conservation.

[2] Murakami et al. (2022) did not find that high users responded the most, instead finding no effect among high users. A back-of-the envelope calculation suggests that confidence intervals for their high and low income users overlap, but we cannot say definitively whether they are able to rule out a significant response among high users as compared to low users.

[3] To add to the puzzle, Brewer (2023) and Elinder et al. (2017) find similar effects in response to price incentives.

[4] Qatari nationals' primary residence is free of cost; they pay a subsidized rate for any secondary residences.

Qatar accounts for 48% of per capita subsidies. The total fiscal cost of subsidies was estimated to be US$6.82 billion, which is 3.6% of GDP.[5]

We implement a natural field experiment and detailed customer-level surveys on electricity use in Doha. Our dataset consists of 207,325 monthly electricity meter readings from a panel of 6,096 customers.

Our intervention consists of two randomized 'nudge'-style interventions to motivate reductions in electricity use by evoking both identity and agency, which we designed in partnership with Kahraama, Qatar's national utility company. The first is a message quoting a passage from the Qur'an stressing the importance of conservation (Religious message). The second is a message reminding households that the government of Qatar prioritizes energy conservation (National message). The treatments that we use leverage injunctive norms (what people "ought" to do) rather than descriptive norms (what people actually do), and are intended to leverage individuals' identity.

In our main analysis, we estimate two parameters- an intent-to-treat, which describes the effect for the customers we sent the message to, and an IV, which captures the effect for those receiving the message. The estimated effects are 3.2% and 3.8% of baseline use, respectively. We do not find that the religious and national message produce differential effects. This could be because religious and national identity are intertwined since Qatar is an Islamic state- laws and customs are rooted in Islam, and Islam is the official national religion. In the context of high per-capita energy use in Qatar, the impacts we find translate to a sizeable reduction in electricity use (about 100 kWh per month).[6]

To delve deeper into what drives our treatment effects, we use machine learning methods proposed by Chernozhukov et al. (2022) to investigate potential heterogeneity in our treatment effects using a pre-intervention survey of a randomly selected sub-sample of customers. Two findings stand out. First, we show that customers who respond most to the treatment are more likely to believe that conserving energy is both easy and effective. Second, the most responsive customers are also more likely to acknowledge anthropogenic climate change and are motivated to change their own behavior to mitigate it. This suggests that agency and personal responsibility are important mechanisms through which interventions designed to evoke identity could impact conservation.

There is a large literature on behavioural interventions to reduce energy use. A meta-analysis reviewing four popular categories of these interventions found that these have the potential to reduce energy use, although there are large differences in effect sizes (Andor and Fels, 2018). However, motivations that drive these effects remain under-explored, yet crucial

[5]Fossil Fuel Subsidy Database, International Energy Agency, 2022. Available at https://www.iea.org/product/download/012730-000298-012432, last accessed on August 24, 2023.

[6]To put this in perspective, the average monthly usage for control units in our study is approximately 3.2 times as much electricity as the typical US household. As of 2021, the average US household uses 866 kWh per month according to the EIA (source: https://www.eia.gov/tools/faqs/faq.php?id=97&t=3). See Figure 2 for the distribution of electricity use in our study compared to that of the US.

for better design and targeting of nudges.

In that vein, moral incentive interventions have emerged as a promising way to promote pro-social action. However, within the literature on moral incentives, there is a dearth of research on the role of identity. Seminal work by Akerlof and Kranton (2000) elucidated the role of identity in determining social behavior, and a growing body of work has explored how to leverage identity in a variety of contexts. For instance, messages priming religious identity reduce credit card defaults among Muslims in Indonesia (Bursztyn et al., 2019), and increase voluntary contribution to public goods among Protestant Christians (Benjamin et al., 2016). However, in the domain of energy use, moral incentives aimed at identity have received little attention, with injunctive norms often being tested as complementary to information or descriptive social norm interventions.[7]

Our novel intervention is designed to fill this gap in the literature by leveraging religious and national identity in particular. Given that Muslims accounted for 24% of the world's population in 2015, and are projected to comprise about 26% of the world's population by 2030[8], and given the political economy constraints that prevent removal of electricity subsidies in many predominantly Muslim countries, it is vital to understand whether Islamic religious identity can be leveraged to conserve natural resources.

Four papers that evaluate injunctive appeals to conserve energy, Ito et al. (2018), Ferraro and Price (2013), Bonan et al. (2021) and Murakami et al. (2022), are most closely related to ours. Ito et al. (2018) finds that moral appeals do not significantly change energy use behaviour compared to information treatments or financial incentives. Ferraro and Price (2013) find that a weak social norm that mentions consumers' responsibility to conserve water does reduce water use compared to the control or an information-only treatment, but less than the reduction brought about by a strong social norm treatment that includes social comparisons. Bonan et al. (2021) find that making environmental identity more salient does not strengthen the effectiveness of a social norm treatment. Murakami et al. (2022) compare an injunctive norm treatment with a price incentive, and find that the nudge generates significant heterogeneity in treatment effects.[9] We build on this work by focusing specifically on a novel intervention that primes injunctive norms of customers' identity as Muslims or as residents of Qatar. Importantly, the messages we use do not include any information on customers' own or relative energy use, nor any specific 'tips' on ways to reduce electricity use. This helps us to cleanly detect effects coming through moral incentives without conflating these effects with information on use or

[7]For example, Allcott (2011) does not find any additional effect of the injunctive norm component of Home Energy Reports. Schultz et al. (2007) found that combining a descriptive norm (average energy use in the neighborhood) with an injunctive norm (a positive-valence emoticon for below-average users and negative-valence emoticon for above-average users) could eliminate the boomerang effect of increased energy use in below-average energy use households.

[8]See, e.g., https://www.pewresearch.org/religion/2011/01/27/future-of-the-global-muslim-population-muslim-majority/

[9]Murakami et al. (2022) do not provide explicitly negative injunctive norm feedback on own use to high users, since it might backfire. Thus our paper is a good complement to their findings.

social comparisons.

We also contribute to the literature on how agency might motivate prosocial behavior. Providing rich potential donors with a sense of agency on how their donations could be utilised has been shown to increase charitable giving (Kessler et al., 2019). Personal agency, i.e. emphasising the importance of individual action as opposed to shared contributions has also been shown to be an effective motivator among wealthy donors (Whillans and Dunn, 2018). In the context of climate change, perceived behaviour (how many others attempt to fight climate change) and norms (how many others should fight climate change) are found to be important predictors of prosocial behavior (Falk et al., 2021). Adding to this literature, we find agency to be a significant motivator for high-users to conserve energy.

Moreover, our work is the first to examine energy use behavior change among primarily high-income residential consumers who face a low or zero marginal price for electricity. In Qatar and other Gulf nations, electricity is provided at highly subsidized rates, which poses a unique challenge to the reduction of energy use where prices do not reflect the marginal cost of production.[10] Existing studies on behavioral interventions in low (or zero) marginal price settings have focused on students in dormitories and hotel guests, with interventions like competitions to reduce energy use (Petersen et al., 2015), daily and real-time feedback (Bekker et al., 2010; Tiefenbeck et al., 2019; Fang et al., 2022), and social comparisons (Bator et al., 2019). These have had moderate effects in terms of reducing electricity use. While this literature advances our understanding of non-pecuniary incentives to save energy, they are unlikely to apply to other settings in which customers face a low or no marginal cost of electricity, because students are limited in their ability to install appliances and affect energy efficiency of their environment, and are typically not high-income.

Notably, our paper contributes to the growing literature on scaling experimental interventions to the population level, and on the risks of "voltage drops", i.e., the tendency for treatment effects to be attenuated (Al-Ubaydli et al., 2017; List, 2022). This is reflected in the distinctions we draw between the intent-to-treat effects estimated for the larger sample versus the IV estimates that account for problems in message receipt and focus on our preferred sample. Understanding the causes of differences between these two sets of findings is of import to government officials looking to leverage our results in the policy domain.

Additionally, we contribute to a growing body of work that uses machine learning to investigate heterogeneity in treatment effects. Employing machine learning methods allows us to relax parametric assumptions and leverage predictive algorithms to uncover heterogeneity in treatment effects, thus enhancing our ability to effectively target interventions. Machine learning has been applied to estimate treatment effect heterogeneity in a broad array of contexts, including energy efficiency upgrades in schools (Burlig et al., 2020), youth employment (Davis and Heller, 2020), and loans to small businesses (Bryan et al., 2023). Notable recent papers applying these methods to consumer energy use include Knittel and Stolper (2021) and

[10] See Section 2 for more details.

Murakami et al. (2022). We employ the technique proposed in Chernozhukov et al. (2022), which addresses common issues economists face when using machine learning, such as overfitting.

The paper proceeds as follows. In section 2, we summarize context on electricity consumption in Qatar. In section 3, we describe our data and experiment design. In section 4, we present our main specifications. In section 5, we present main results from our field experiment, and also investigate whether response differed by customer group and number of messages. In section 6, we use machine learning to understand heterogeneity in treatment effects. Section 7 concludes.

2 Background: electricity use and subsidies in Qatar

Two features of the Qatari energy context are most important to our study. The first is the presence of subsidies, which reduce pecuniary incentives. The second is the fact that Qatar represents a high-use population compared to the world, or even developed countries.

The most notable feature of Qatar's energy environment is the presence of energy subsidies. Many countries deploy similar subsidies, exceeding 6% of global GDP (Coady et al., 2015). In some countries, in conjunction with their being means-tested, these subsidies represent efforts at improving living standards for the poor, given the large weight that energy consumption typically has in the consumption basket of poor households.

Qatar's energy subsidies are unique in two regards. First, for nationals, they are absolute, meaning that Qatari citizens' electricity bills are always zero, irrespective of their energy consumption levels or their material means. Moreover, even for non-nationals, electricity is highly subsidized.[11] Second, living standards for middle- and upper-income people in Qatar are exceptionally high on average, and even more so were one to restrict the sample to citizens. In fact, its GNI per capita is so high that Qatar consistently scores the maximum in the income-related sub-index of human development (Al Muftah, 2018).

Therefore, poverty relief is not a rationale for Qatar's idiosyncratic energy subsidies. Instead, they are better explained by the rentier economic model: an implicit social contract exists whereby the state is expected to provide citizens with a comfortable life, and in return those citizens provide political acquiescence (Tsai and Mezher, 2020). Similar tacit social contracts operate in the remaining five Gulf countries (Reiche, 2010), though none are as generous as the Qatari government (or can afford to be so), as they involve electricity tariffs that are considerably below the cost price while still being substantively above zero.

Qatar is able to fund its expansive energy subsidies due to its abundant income from natural resources, as it has the world's third largest reserves of natural gas and significant oil reserves, while having a population of fewer than three million, including fewer than half a million citizens. Nevertheless, the sharp decline in oil and natural gas prices that occurred in

[11] Electricity tariffs for non-nationals follows a tiered system that increases from 0.11 QR (US$0.03) per kWh for 1 - 2,000 kWh per month, to 0.13 QR (US$0.036) per kWh for 2,001 - 4,000 kWh per month, to 0.18 QR (US$0.05) per kWh for 4,001 - 15,000 kWh per month, to a maximum marginal price of 0.26 QR (US$0.07) per kWh for use in excess of 15,000 kWh per month.

2014 created significant fiscal pressure in Qatar, exacerbated by the need to step up capital expenditure in preparation for the 2022 FIFA World Cup. This led to subsidy reforms that included charging expatriates for their electricity consumption, albeit at a subsidized rate (Al-Saidi, 2020). The other Gulf countries increased the tariffs paid by nationals, but the budgetary pressure was not enough for Qatari citizens to undergo similar reforms. Through our discussions with key stakeholders in Qatar, it is evident that the government is reluctant to introduce electricity tariffs due to the potential socio-economic impact.

Unsurprisingly, given the absence of electricity bills, Qatar has one of the highest levels of per capita energy consumption and CO2 emissions in the world (International Energy Agency, 2018), with the demand for air conditioning induced by the arid climate and the need to desalinate seawater contributing to these high levels. The Qatari electricity sector accounts for 48% of per capita subsidies, with a total outlay of approximately $6.8 billion, equaling 3.6% of GDP.[12] The significant fiscal cost of electricity consumption is exacerbated by a considerable diplomatic cost, too: the country will struggle to fulfill its commitments to the Sustainability Agenda 2030 unless Qatari households and businesses become more energy efficient, and it will continue to draw negative media attention for this presumed profligacy (De Oliveira and Smith, 2022).

Putting these attributes together, Qatar finds itself in a situation where it has a strong fiscal and diplomatic incentive to decrease energy consumption, while at the same time facing political forces that constrain its ability to use the most straightforward tool of raising electricity tariffs. Though the underlying circumstances are highly unusual, this final outcome is actually consistent with the current experience of many other countries. For example, many advanced economies in the European Union have economic and political forces that push them toward continuing to improve energy efficiency, for example due to the increasing popularity of green parties (Muller-Rommel, 2019), while also facing an electorate that is keen on exploring alternatives to increased indirect taxes on energy consumption (Douenne and Fabre, 2020).

To sum up, the Qatari context severely limits the feasibility of reducing price distortions and incorporating the external damages of carbon emissions into electricity tariffs, which would be the first-best solution to aligning energy use to socially optimal levels. In the absence of adequate financial incentives to conserve, behavioural interventions that leverage moral suasion provide the potential to drive reduction in energy use.

3 Experimental Design and Data

Since Akerlof and Kranton's seminal work, economists have theorized that identity can affect economic choices through taking actions that preserve individuals' self-image (Akerlof and Kranton, 2000). Normative prescriptions that are inherent in certain aspects of identity can increase personal utility when individuals take actions conforming to such prescriptions. We

[12] Data from the International Energy Agency's Fossil Fuel Subsidies database, available at https://www.iea.org/data-and-statistics/data-product/fossil-fuel-subsidies-database.

use this idea to design our treatments in partnership with Qatar General Electricity & Water Corporation - "KAHRAAMA". The treatments use two explicit primes leveraging injunctive norms of conservation inherent in two aspects of identity in Qatar.

First, we leverage the fact that religion is an important part of Qatari life, and its residents are predominantly Muslim. Therefore, priming religious values promoting conservation could lead to behavior change if people value their self-image as Muslims. In our treatment message, we use a specific verse from the Qur'an that asks its followers to 'waste not by excess'. This treatment is similar in spirit to the message used to encourage credit card debt repayment in Bursztyn et al. (2019).

Second, residents of Qatar may see themselves as playing a role in the country's stated desire to develop its economy in a more sustainable manner.[13] Qatar instituted the National Program for Conservation and Energy Efficiency in 2012 - a campaign to encourage conservation of electricity and water.[14] We reference this program and its patronage by the Amir of Qatar in our second treatment message to prime individuals' identity as Qatari residents and remind them that conserving energy is congruent with this identity.

Our treatment messages include text highlighting that customers have the ability to conserve energy. This language is similar to prior work using nudges in the energy domain (Ferraro and Price, 2013), and congruent with the utility's general messaging to its customers to 'consume wisely'.[15] Figure 1 shows the content of the two treatment messages.

Figure 1: Text Message Content

"O Children of Adam... Eat and drink: But waste not by excess, for God loveth not the wasters." **Al-a'araf, , Verse#31)** Please, use electricity wisely. You have the power to conserve. **Regards,** **TARSHEED- KAHRAAMA & QU**	۞ يَا بَنِي آدَمَ خُذُوا زِينَتَكُمْ عِندَ كُلِّ مَسْجِدٍ وَكُلُوا وَاشْرَبُوا وَلَا تُسْرِفُوا ۚ إِنَّهُ لَا يُحِبُّ الْمُسْرِفِينَ **سورة الأعراف , اية 31** **نرجو استخدام الكهرباء والماء بحكمة. لأننا نثق في قدرتكم علي الترشيد**
The Energy Conservation Program is under the patronage of HE (The Amir of Qatar), Shaikh Tamim Bin Hamed Al-Thani). Please, use electricity wisely. You have the power to conserve. **Regards,** **TARSHEED- KAHRAAMA & QU**	**ان برنامج الترشيد – كهرماء. تحت رعاية الأمير، الشيخ تميم بن حمد ال ثاني- أمير البلاد المفدى:** **نرجو استخدام الكهرباء والماء بحكمة. لأننا نثق في قدرتكم علي الترشيد** **مع تحيات برنامج الترشيد – كهرماء وجامعة قطر**

Notes: This figure shows our message content, in both English and Arabic. The top panel is the religious message and the bottom panel is the national message.

Prior to the experiment, Kahraama shared monthly electricity use data for customers in three areas of the capital city of Doha- specifically Al Saad, Al Dafna and Al Qassar. These areas are located in the central part of the city and comprise several residencies, including many

[13]Individuals' identity as Qatari residents may have been made more salient due to the geopolitical blockade of Qatar by its neighboring countries from 2017 to 2021.

[14]For more information on the program, see https://www.km.qa/Tarsheed/Pages/TarsheedIntro.aspx.

[15]See https://www.km.qa/Tarsheed/pages/default.aspx#front

newly built homes. The areas were chosen by the utility as being appropriate for our experiment since they comprise high electricity and water use, and meter readings are taken relatively frequently. The utility's customer database includes their registered cellphone numbers, which we use to deliver our interventions as text messages.

To select customers into the sample for randomized assignment, we proceeded in the following way. First, we keep only those meters in the database for which there exist at least one bill-month observation for the period between April 2018 to March 2019. Second, we keep only those meters that are billed as either a flat or a villa within this time period. Third, we include only those meters that are registered as belonging to "Regular Customers" or "Qatari Owners", excluding properties that are registered as being "Rented out by Qataris". Fourth, we consider only those customers who have a cellphone number registered with the utility to randomize into treated and control groups while balancing on observed average monthly electricity use over April 2018 to March 2019. Finally, we note that electricity use varies not just across months of the year, but also by type of residence (flats or villas) and ownership category (national or non-national). We anticipated heterogeneous effects among these groups and, hence, stratify the experiment on type of residence and ownership category.[16] This results in three strata – (i) flats (n=4,803), (ii) villas owned by non-Qatari individuals (n=647), and (iii) villas owned by Qatari individuals (n=665).[17] Overall, and within each strata, we divided customers into the two treatment groups and a control group in a 2:2:1 ratio within the sample of customers who had registered phone numbers in the utility's database.

Our intended experimental sample comprises 2,438 customers in each of the two treatment groups, and 1,220 customers in the control group. Given a 2:2:1 assignment ratio, and that our treatments are clustered at the customer level with an intra-cluster coefficient of 0.77 in the logarithm of electricity use at the customer level, we are powered to detect a minimum effect of 11.1% (7.9% of the standard deviation) change at 80% power and 10% level of significance, when using a parametric t-test for differences between treated and control groups. The interventions started in May 2019, with two messages to be sent each month until October 2019, for a total of twelve planned messages.

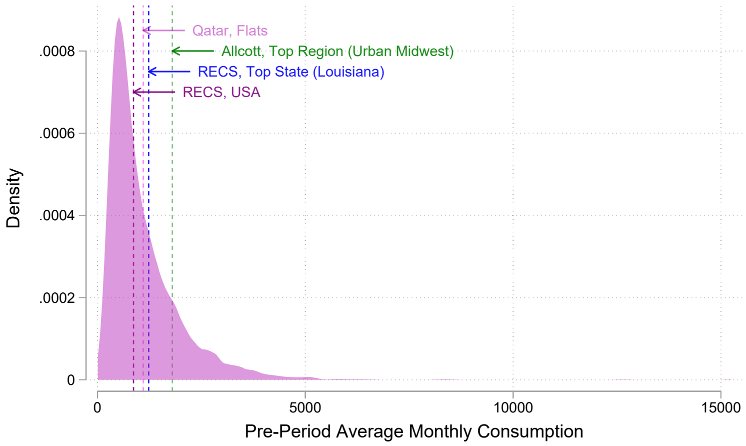
However, we encountered several issues with the electricity data in our sample after the experiment was administered, which led us to increase our sample observations to more customer-month observations (January 2016 - February 2020) and focus on a preferred sample throughout most of our analysis. Below, we discuss two important checks on our preferred sample and difference-in-differences specification - customer composition and parallel pre-trends; a detailed discussion of the construction of the preferred sample can be found in Section

[16]See Section 5.2 for a description of differences in electricity use by strata.

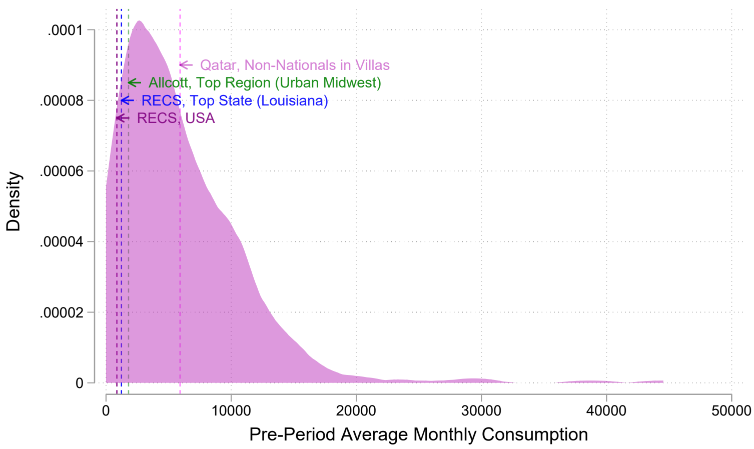
[17]We do not include Qatari owners of flats as a separate stratum due to the low numbers of such premises relative to other categories in the customer database. Stratifying these residences would have negatively affected the statistical power of our experiment.

Figure 2: Distribution of electricity use by strata

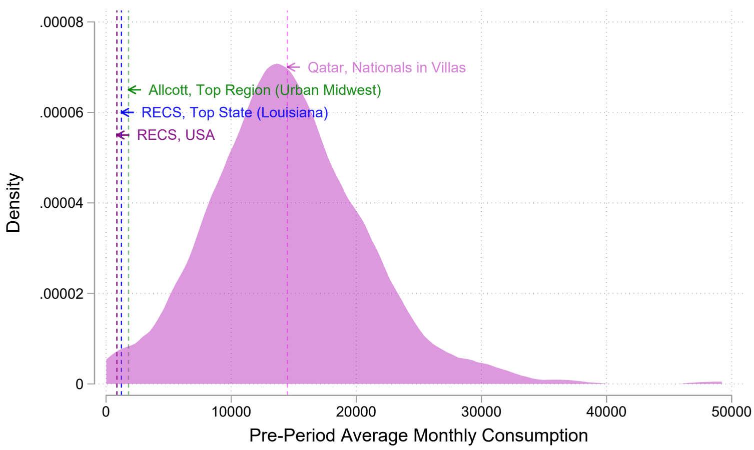
(a) Flats



(b) Non-nationals in Villas



(c) Nationals in Villas



Notes: The figure shows kernel density plots of the distribution of monthly pre-intervention electricity use at the customer level, separately for each of the three strata (Flats, Non-nationals in Villas, and Nationals in Villas) we utilise in the experiment. Reference lines show the highest monthly electricity consumption among the regions included in the OPower experiments (Allcott, Top Region (Urban Midwest)), as well as monthly per capita residential electricity consumption estimated by the Residential Energy Consumption Survey (RECS) for Louisiana as the highest per-capita use state in the US (RECS Top State (Louisiana)) and average per-capita use for the entire US (RECS, USA).

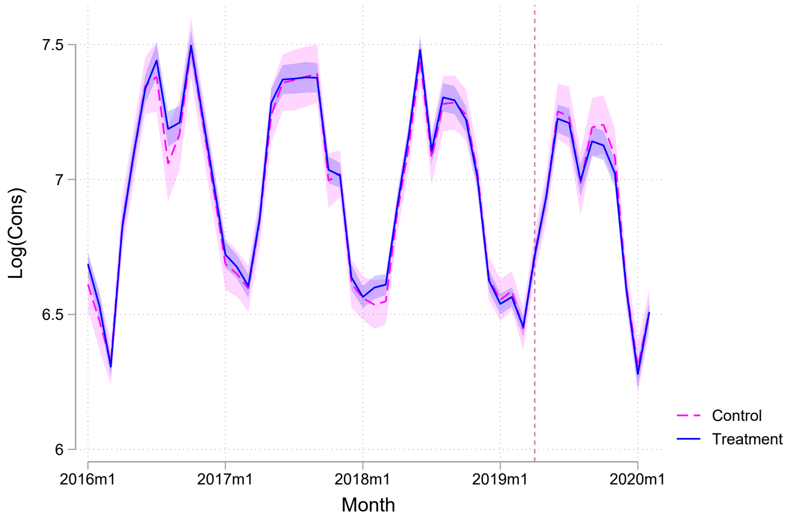
A of the appendix.[18, 19] Our preferred sample retains 4,836 customers, each with at least two non-missing electricity use observations.[20]

Figure 2 shows that electricity use in our study sample is considerably high, especially for residents of villas. In fact, Qatari nationals residing in villas use an order of magnitude more electricity compared to the average US consumer.

Table 1 shows the number of customers by customer group and assigned treatment group in both our experimental sample and our preferred sample. We confirm that the intended ratios for the respective assigned treatment groups and strata are maintained in our preferred sample. Further, both samples have similar pre-treatment electricity use - in both levels and logarithmic terms - among treated and control groups.

Since our preferred sample differs from the initial randomized assignment, we further verify that the treatment and control groups have similar electricity use patterns over time prior to the treatment. In Figure 3, we show monthly electricity use over time for both the control and treatment groups in our preferred sample. The two groups appear to have parallel pre-intervention trends in electricity use.

Figure 3: Monthly Electricity Use Over Time



Notes: This figure depicts the treated and control groups over time, with the start of our treatment indicated by the vertical dashed line.

Some of the messages we sent out were not received by the participants.[21] We have unique data that allows us to detect receipt of the messages. We discuss message receipt in depth in Appendix section A.2. We will apply an instrumental variables approach to adjust for the fact that message receipt is imperfect.

[18] For completeness, we show the estimate of our treatment effect for the unrestricted sample in our table of main results (Table 2, col 1), but we are most confident of the results that use our preferred sample.

[19] We also discuss implications of using the preferred sample in Appendix Section A.

[20] As noted in Table 1, 4,832 customers have electricity use information before the intervention, while 4 customers have electricity use observations only after the intervention.

[21] Receipt of a message means the message was successfully delivered by the phone company.

Table 1: Balance of Baseline Electricity Use by Treatment and Strata

Panel A: Experimental Sample (April 2018 - March 2019)

Variable	(1) Control N/[Customers]	(1) Control Mean/SE	(2) Religious message N/[Customers]	(2) Religious message Mean/SE	(3) National message N/[Customers]	(3) National message Mean/SE	T-test P-value (1)-(2)	T-test P-value (1)-(3)	T-test P-value (2)-(3)
Cons	14640	3103.717	29256	3091.988	29256	3146.896	0.950	0.821	0.727
	[1220]	(153.833)	[2438]	(108.802)	[2438]	(113.562)			
Flats	11532	1150.556	23052	1129.029	23052	1125.540	0.622	0.570	0.919
	[961]	(36.615)	[1921]	(23.846)	[1921]	(24.423)			
Non-nat Villas	1537	6875.917	3029	6540.895	3073	6639.351	0.670	0.767	0.871
	[130]	(667.881)	[256]	(418.699)	[261]	(441.839)			
Nat Villas	1571	13750.424	3175	14053.685	3131	14601.383	0.691	0.275	0.415
	[133]	(607.339)	[266]	(461.402)	[266]	(488.981)			
log(Cons)	12653	7.124	25063	7.132	25021	7.125	0.847	0.975	0.844
	[1204]	(0.037)	[2400]	(0.025)	[2406]	(0.026)			
Flats	10085	6.682	19975	6.690	19923	6.674	0.809	0.797	0.547
	[956]	(0.026)	[1907]	(0.018)	[1907]	(0.019)			
Non-nat Villas	1248	8.226	2397	8.300	2433	8.261	0.606	0.815	0.707
	[123]	(0.125)	[239]	(0.072)	[247]	(0.078)			
Nat Villas	1320	9.452	2691	9.374	2665	9.459	0.338	0.929	0.231
	[129]	(0.061)	[258]	(0.054)	[261]	(0.045)			

Panel B: Preferred Sample (January 2016 - April 2019)

Variable	(4) Control N/[Customers]	(4) Control Mean/(SE)	(5) Religious message N/[Customers]	(5) Religious message Mean/(SE)	(6) National message N/[Customers]	(6) National message Mean/(SE)	T-test P-value (4)-(5)	T-test P-value (4)-(6)	T-test P-value (4)-(5)
Cons	25642	2776.258	52581	2772.809	52145	2885.042	0.986	0.602	0.495
	[944]	(169.953)	[1951]	(111.056)	[1937]	(121.446)			
Flats	21359	1134.289	43444	1135.691	42906	1115.263	0.973	0.644	0.544
	[790]	(33.641)	[1609]	(23.978)	[1610]	(23.669)			
Non-national Villas	1995	6659.159	4230	5897.752	3824	5991.145	0.418	0.492	0.877
	[77]	(857.228)	[172]	(389.926)	[147]	(460.862)			
National Villas	2288	14718.749	4907	14573.187	5415	14714.481	0.858	0.996	0.832
	[77]	(686.029)	[170]	(445.617)	[180]	(498.152)			
log(Cons)	25642	6.949	52581	6.964	52145	6.969	0.750	0.666	0.884
	[944]	(0.039)	[1951]	(0.026)	[1937]	(0.028)			
Flats	21359	6.592	43444	6.594	42906	6.580	0.945	0.721	0.597
	[790]	(0.026)	[1609]	(0.018)	[1610]	(0.019)			
Non-national Villas	1995	8.083	4230	8.087	3824	8.061	0.984	0.892	0.829
	[77]	(0.142)	[172]	(0.081)	[147]	(0.088)			
National Villas	2288	9.289	4907	9.267	5415	9.279	0.829	0.924	0.877
	[77]	(0.089)	[170]	(0.055)	[180]	(0.058)			

Notes: This table shows the number of customer-month observations, number of customers, means, and standard errors clustered at the customer level for monthly electricity use (Cons) in kWh and logarithm of monthly electricity use (log(Cons)). Panel A represents information for the experimental sample: the sample of customers who were randomly assigned into Control, Religious message or National message treatment group. Treatment assignment and electricity use of customers within each strata (Flats, Non-nationals in Villas and Nationals in Villas) and p-values corresponding to pairwise t-tests for differences in baseline outcomes between groups are provided. Note that 19 customer IDs registered to villas switch nationality class - 4 in the Control group, 5 in the Religious message group and 10 in the National message group. These customer IDs are accounted for in both nationality strata. Panel B represents the same information for the preferred sample of customers. Note that the total number of customers in the preferred sample is 4,832 and not 4,836 as indicated in Table 2. 4 customer IDs (3 in the Religious treatment group and 1 in the National treatment group do not have actual meter reads in the pre-intervention period. Therefore, they do not contribute towards the estimates in Table 2. Detailed criteria for exclusion of customers from the experimental sample is discussed in Appendix A.

We additionally match our electricity use data with a survey that occurred prior to treatment to understand heterogeneity. We describe the survey data and show descriptive statistics for the supplemental survey in Appendix Section A.3.

4 Main Specifications

4.1 ITT Specification

Our first specification is a simple OLS regression that estimates the intent-to-treat effect.

$$log(Cons_{it}) = \alpha Treat_i \cdot Post_t + \delta_i + \tau_t + \kappa_{sm} + \epsilon_{it} \tag{1}$$

In the above, $log(Cons_{it})$ is the natural logarithm of customer i's energy consumption in month t, $Treat_i$ is an indicator for the customer being assigned to either the religious or national message group,[22] and $Post_t$ is an indicator for the post-April 2019, which is when all treated customers were supposed to receive their first text message. We include fixed effects for customer (δ_i), month of sample (τ_t), and strata by month of year (κ_{sm}). We cluster standard errors at the customer level.

In the above, the parameter of interest is α, which can be interpreted as the intent-to-treat effect on energy consumption of having received at least one religious or national message. We would expect these effects to be attenuated as compared to the true average treatment effect of the messages because of the way that message receipt was imperfect. This specification would be equal to the ATE if all households assigned to treatment received and read all twelve messages.

4.2 IV Specification

Because of the way the ITT is expected to be attenuated compared to the ATE, we also employ an IV specification, which is our preferred specification.

The first stage consists of:

$$\mathbb{1}\left\{\text{Received Msg}\right\}_{it} = \gamma Treat_i \cdot Post_t \quad + \xi_i + \theta_t + \eta_{sm} + \nu_{it} \tag{2}$$

In the above, $\mathbb{1}\left\{\text{Received Msg}\right\}_{it}$ is a dummy for having received at least one of the two messages.

The second stage is specified as:

$$log(Cons_{it}) = \alpha \mathbb{1}\left\{\text{Received Msg}\right\}_{it} + \delta_i + \tau_t + \kappa_{sm} + \epsilon_{it} \tag{3}$$

We also show results from the IV where we break out the treatment into the religious and national message types. In that case, we have two first-stage equations- one corresponding to

[22] We also show results breaking out the treatment assignment into the religious and national treatment groups.

each of the two text messages.[23,24]

When interpreting the parameters from the IV setup specified in (2) and (3), we allow for heterogeneous gains to text messages that induce endogenous selection into message receipt. In the standard framework, α would be interpreted as the local average treatment effects (LATE), or the causal effect of receiving at least one message in prior months on the marginal complier. The marginal complier is the customer that is just indifferent between opening and not opening the text message.

However, our IV actually identifies the causal effect of receiving the messages for a broader population- the entire set of compliers. We do not have always-takers in this experiment, as customers were not allowed to sign up for our text messages if they were not in the treatment group. Therefore, the local average treatment effect also equals the average treatment effect on the treated (ATT), which is the causal effect of treatment for the entire population of compliers (Bloom, 1984). That is, IV estimation of (3) recovers the causal effect of receiving a message on energy consumption *for the entire set of customers who receive the text messages.*

5 Results

5.1 Main Results

Results from our intent-to-treat specification are found in the top panel of Table 2. The first column shows results for the experimental sample, while column 2 shows results for "AC" reads only within the experimental sample. Columns 3-6 show the results using our preferred sample of customers. In column 1, the intent-to-treat estimate is -0.026, indicating around a two-percent reduction in energy use due to our treatment. The estimate is statistically significant at the 10% level. In column 2, we see that the estimate does not change when considering actual reads only. The estimated effects are more statistically precise when we limit to our preferred sample, starting in column 3. In column 4, we add strata by month-of-year fixed effects. This is our preferred specification, since it accounts for strata-level behavior that occurs on a seasonal basis (e.g. nationals in villas vacationing during the summer months). Our preferred estimate of the ITT indicates a reduction of 3.2% on average for those assigned to treatment, which is significant at the 5% level.

In column 5, we separate out the treatment effect for our religious and national message to see if there is heterogeneity in the treatment effect by message type. We find no evidence of heterogeneity. While the effect of the national message is slightly larger, the two messages produce similar reductions, both statistically and economically. This is perhaps unsurprising since Qatar is an Islamic state- laws and customs are rooted in Islam, and Islam is the state religion. Therefore, religious and national identities are likely to be inextricably linked.

[23]Note that, due to our design, the first stage will contain zeros with this two-instrument and two-message setup since no customers can receive the other treatment.

[24]We also conduct a heterogeneity analysis where we estimate effects separately by strata. In that analysis, the first stage instruments and second stage dummies are also interacted with dummies for each of the three strata, with three first-stage equations.

Table 2: Main Results

	(1)	(2)	(3)	(4)	(5)
ITT:					
Pooled	-0.026^{*}	-0.026^{*}	-0.032^{**}	-0.032^{**}	
	(0.014)	(0.014)	(0.016)	(0.016)	
Religious					-0.029^{*}
					(0.017)
National					-0.035^{**}
					(0.017)
IV:					
Pooled	-0.033^{*}	-0.034^{*}	-0.038^{**}	-0.038^{**}	
	(0.018)	(0.019)	(0.019)	(0.019)	
Religious					-0.035^{*}
					(0.020)
National					-0.042^{**}
					(0.021)
First Stage:					
Pooled	0.774^{***}	0.777^{***}	0.834^{***}	0.834^{***}	
	(0.006)	(0.007)	(0.007)	(0.007)	
Religious					0.839^{***}
					(0.009)
National					0.830^{***}
					(0.010)
FE:					
Cust	*Y*	*Y*	*Y*	*Y*	*Y*
Month	*Y*	*Y*	*Y*	*Y*	*Y*
Reading Type	*Y*				
Strata × MOY				*Y*	*Y*
Sample	Experimental	Exp: AC reads only	Preferred	Preferred	Preferred
Cragg-Donald F	131,497.69	131,497.69	158,692.08	158,786.96	79,353.58
Avg Cons (Ctrl)	3,564.91	3,126.18	2,780.67	2,780.67	2,780.67
Customers	5,797	5,785	4,836	4,836	4,836
Observations	207,325	191,933	161,254	161,254	161,254

Notes: The table shows our estimates of the intent-to-treat (ITT) and local average treatment effect (IV) on log(Consumption). Column (1) shows the effects over the entire experimental sample, column (2) restricts the experimental sample to actual reads only, column (3) presents estimates for the preferred sample, while column (4) shows estimates from our preferred specification with customer, month and strata by month-of-year fixed effects. Column (5) breaks down the treatment into the two types of treatment messages - Religious and National. Standard errors are clustered at the customer level. Note that the first stage in Column (5) technically contains two equations and four coefficients. But, due to the fact that the treatments are mutually exclusive and there is no possibility of assignment to the other treatment, they are only nonzero for own treatments, and thus are presented here as two coefficients. Singletons do not contribute to main estimates and are thus dropped from the cluster count.

The middle panel of Table 2 shows the instrumental variable results, which are larger than the ITT results but produce substantively similar conclusions. As discussed above, column 4 presents results from our preferred specification. Our preferred estimate is significant at the 5% level and indicates a 3.8% reduction in electricity use for those who received the message. We again find that the two messages are equally effective.

The bottom panel of Table 2 shows the first stage for each regression of interest.[25] The first stage is extremely strong for each specification, with Cragg-Donald F-statistics at the bottom of the table all in excess of 79,000. We summarize our results in Figure 4.

A reduction of 3.2% for those we sent messages to and 3.8% for those receiving messages is sizeable when considering typical electricity use in Qatar. Given that the average monthly consumption for control units is 2,780.67 kWh, our treatment effects translate to an average reduction of 88.98 kWh for those assigned to treatment and 105.67 kWh for those receiving our text messages.

It is worth dwelling on the epistemological differences between the experimental and preferred samples. The construction of the preferred sample makes it more likely to detect a treatment effect – if it exists. The sample restrictions we employ should at least partially address the issues that lead to downward bias in the ITT compared to the LATE. From the perspective of scaling our intervention to the entire population, the larger experimental sample in Columns 1 and 2 are nominally more relevant (Al-Ubaydli et al., 2017). This is because many of the factors that undermine the intervention, such as people having multiple telephones, or not receiving the text message, are ones that organically emerge whenever one scales the intervention. In fact, they represent the reason why researchers typically initiate their scientific investigations in the highly-controlled confines of a laboratory. This may give the impression that from a policy perspective, it is ultimately only the larger sample that matters. However, that is based on the assumption that a government adopting the intervention makes no effort at combating the factors that lead to an organic attenuation in the treatment effect when scaling. In practice, there exists a growing literature that explains exactly how to prevent such "voltage drops" (Al-Ubaydli et al., 2020; List, 2022), affirming the intellectual and policy importance of the results emerging from the preferred sample.

5.2 Heterogeneity by Customer Group

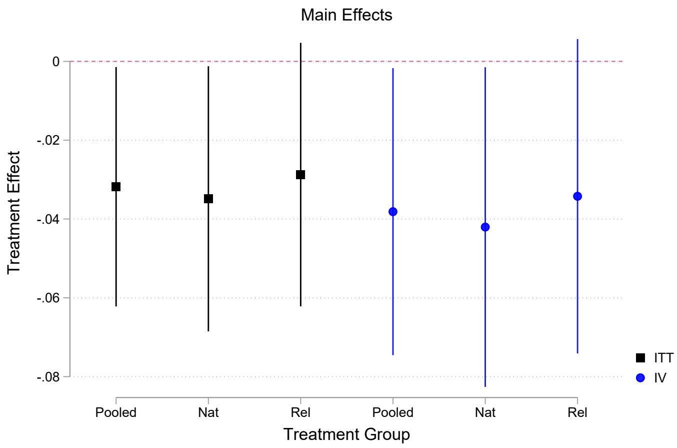
We now examine whether there is treatment effect heterogeneity by customer group, for several reasons.

First, the three customer groups have very different average baseline consumption (see Figure 2). In our preferred sample, baseline consumption for flats is 1,127.28 kWh per month, whereas non-nationals and nationals in villas use 6,084.45 and 14,660.27 kWh per month, respectively.[26] Higher baseline use could translate to more "low-hanging fruit" when it comes

[25] We have displayed only non-zero first stage coefficients in column 4; see table notes.

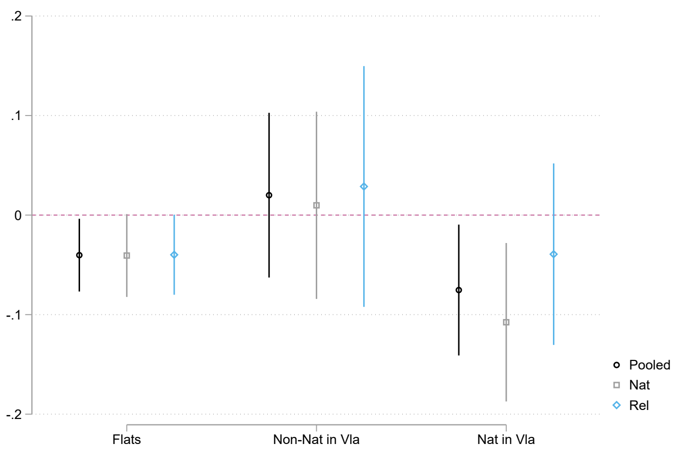
[26] So, an average national in villa uses nearly 17 times the electricity of the average US household.

Figure 4: Effect of Messages on Electricity Use



Notes: This figure shows our estimated treatment effects. Estimated coefficients and 95% confidence intervals corresponding to our intent-to-treat (ITT) effects are indicated in black markers and lines, while the estimates from the instrumental variable (IV) specifications are indicated with blue circles and lines. We show estimates both for pooled treatment (Column 4 in Table 2) and separated for each message type (Column 5 in Table 2) The magnitude of effects represent percentage changes relative to the control group.

Figure 5: Treatment Effects by Customer Group



Notes: In this figure, we plot coefficients from interacting both message receipt and instruments with customer group dummies in our IV specifications. The black circles are estimated coefficients from a single IV specification breaking the pooled treatment into 3 strata-based categories, analogous to column 3 in Table 2. The gray square and blue diamond come from a single IV model breaking out the two treatments into 6 categories, analogous to column 4 in Table 2.

to actions to conserve.[27]

Second, nationals in villas do not typically pay for electricity,[28] whereas the other two customer categories typically do. Therefore, nationals in villas are not normally incentivized to conserve, and thus may have paid less attention to actions they could take in the past, increasing the extent of potential "low hanging fruit."

Third, our national treatment is expected to appeal most to Qatari nationals (the majority of whom live in villas), whereas our religious treatment might appeal to all three customer groups since the majority of individuals living in Qatar are Muslim.

As a parametric test of heterogeneity by customer group, we interact customer group dummies with both message receipt and the instruments in our IV specification. The results are plotted in Figure 5. The magnitudes of our estimates suggest that nationals in villas conserve more in response to the messages than flats or non-nationals in villas. Further, they indicate that nationals in villas respond more to the national message compared to the religious message. However, the standard errors are too large to statistically reject the null of no differences in response between the two messages at 95% level of significance.

It is worth noting that even without heterogeneity in treatment effects in percentage terms, the implied reductions in levels are very different between the three customer groups because of their different levels of baseline consumption. For example, a 3.8% monthly reduction for a national in a villa is 557.09 kWh - more than half the monthly consumption of the typical US household.

5.3 Heterogeneity by Number of Messages

Next, we adapt our IV strategy in (3) to detect the effect of the *number* of messages received in place of whether the individual received at least one message. We might expect heterogeneity for two reasons. First, it could be that the messages become more salient and more likely to trigger action after multiple messages are received. Second, customers might experiment with actions to conserve energy upon receipt of the first few messages, with the gains from that experimentation appearing later.

We present results assuming three different polynomial functional forms in number of messages received- quadratic, cubic, and quartic. To parallel our main specification, we use the number of messages the individual was supposed to have received given treatment status as the instrument, expressed using the same polynomial transformation that we apply to the number of messages they received.

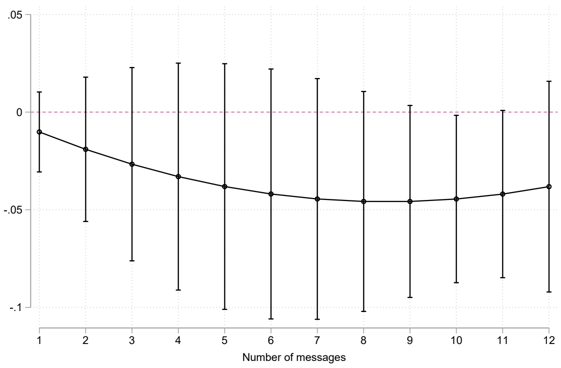
Results are shown in Figure 6. Our conclusion is that we cannot reject the null hypothesis of no heterogeneity by number of messages- this conclusion holds across all three specifications. While the figures generated by the quadratic and cubic specifications suggest that more messages results in a larger point estimate of the reduction in electricity use, the quartic graph

[27] Additionally, other work has found differential electricity use behavior by strata in Qatar, see, e.g. Bernstein et al. (2023).

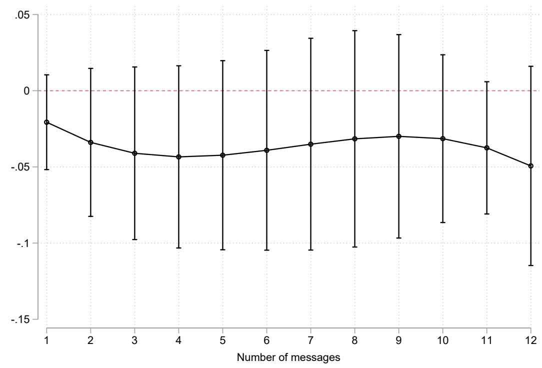
[28] They do not pay for utilities in their primary residence, but do in secondary residences.

Figure 6: Effect by Number of Messages

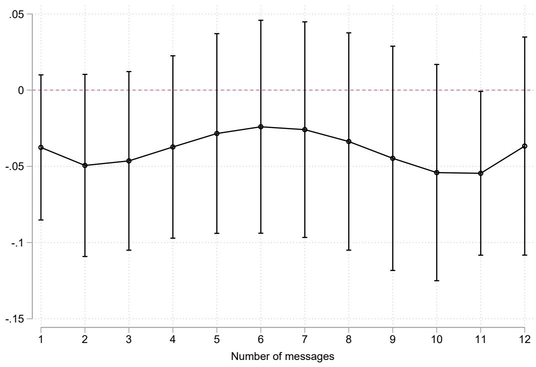
(a) Quadratic Polynomial



(b) Cubic Polynomial



(c) Quartic Polynomial



Notes: This figure shows the impact of receiving messages from several polynomial IV specifications, where the number of messages received is presented on the horizontal axis and the treatment effect is presented on the vertical axis.

does not corroborate that pattern, and confidence intervals overlap in all three cases.[29] While the shape of the response to messages may depend on the functional form assumption, the conclusion of no significant heterogeneity by number of messages does not.

The finding of no heterogeneity in the number of messages received is surprising given that Allcott and Rogers (2014) find significant heterogeneity in effects with respect to the passage of time. One difference is that our setup instead investigates heterogeneity by number of messages received. A drawback of studying effects in the number of messages in our setting is that we cannot differentiate treatment effects differing by the number of messages from the treatment effects simply being different in different months (e.g. seasonal variation in treatment effects). Additionally, we do not have enough purely post-period observations to disentangle temporal persistence from the intensity effect of additional messages.

6 Using Machine Learning to Investigate Heterogeneity in Treatment Effects

Prior to our field experiment interventions, we conducted a survey of a randomly selected group of customers using the phone numbers registered in the utility's database. We collect information on customers' attitudes toward climate change, beliefs about energy use of appliances, energy savings of popular actions, along with detailed demographic information. In this section, we use the survey data matched with our field experiment sample to examine whether there is heterogeneity in our treatment effects, as well as what characteristics of customers might explain that heterogeneity.

6.1 Overview of Machine Learning Procedure

We implement the Generic Machine Learning procedure due to Chernozhukov et al. (2022). The technique uses random splits of the data to avoid overfitting and increase the validity of the results. In each split, the training set is used to determine the relationship between consumption and a set of covariates for both the control and treatment groups, and then a test dataset is used to estimate the treatment effect and heterogeneity using those relationships.

We employ the technique on 247 customers that appear in both our preferred sample and a supplemental survey occurring prior to the treatment. Missing survey responses are imputed within strata as the strata-level average for that variable. See supplemental appendix section A.3 for more on our survey and the matched sample.

We limit data on electricity use to the post-period of our experiment for the machine learning investigation following Chernozhukov et al. (2022). This produces a total of 1,704 customer-month level observations. We present all the details on the implementation of the machine learning procedure in Section 6.2 of the Appendix for interested readers; below, we summarize our parameters of interest.

[29] We also explored a fully flexible specification, but found that it suffered from weak instruments (Cragg-Donald F-statistic = 0.12).

There are three important sets of target parameters. The first is the Best Linear Predictor (BLP) of the Conditional Average Treatment Effect (CATE) on the machine learning proxy predictor. We will shorten this to "BLP" following Chernozhukov et al. (2022). This encompasses two parameters, one representing the (conditional) average treatment effect, and one representing the heterogeneity in treatment effects. We will denote these as "ATE" and "HTE" respectively. These objects are computed in each split of the dataset. The estimates of ATE and HTE as well as the confidence intervals and p-values that we report are the medians of those objects over all the splits.

The second is the Sorted Group Average Treatment Effect, which will be referred to as 'GATES'. This is the average treatment effect by heterogeneity groups (as classified by the machine learning proxy predictor). In each split, observations are divided into four quartiles. The treatment effect, confidence intervals, and p-values are then defined as the medians of those objects for a particular group over all the splits.

The third is the Classification Analysis, which we will call 'CLAN' following Chernozhukov et al. (2022). This is a description of the average characteristics of the most and least affected units defined in terms of the machine learning proxy predictor. For each split, we calculate the average value of each covariate for each GATES group, as well as the confidence interval and p-value. The values we report are the medians of the average value of a particular covariate for a particular group over all the splits.

We obtain all of the above estimates using five machine learners (Support Vector Machines, Random Forest, Neural Networks, Elastic Net, and Gradient Boosting), and store the results. Chernozhukov et al. (2022) recommend using the "best" machine learner, as defined by the values of Λ and $\bar{\Lambda}$, which are statistics quantifying the correlation between the ML proxy predictor and the best predictor. Since the best predictor may not be the same for the BLP and GATES, it is recommended that one use the best learner in each case, and use the best predictor for GATES for the CLAN.

6.2 Machine Learning Implementation Details

In this section, we describe the machine learning technique we use in more detail for interested readers.

In concrete terms, our target model is:

$$log(Cons_{it}) = \beta_1(T_i - P_i) + \beta_2(T_i - P_i)S(Z_i) + B(Z_i) + \theta_t + \epsilon_{it} \tag{4}$$

The dependent variable in our machine learning exercise is the log of consumption in each post-period month. To parallel our main specification, we net out the estimated consumer-level and strata-by-month fixed effect from the log of consumption.[30]

[30]The machine learning exercise compares consumption between the treatment and control group in the post-period. Two sets of fixed effects from our main regression net out characteristics that do not vary between the pre-and-post period- the individual and strata by month of year effects. These cannot be accounted for (at

In (4), T_i is a dummy for assignment to either of the two treatments in our field experiment. $S(Z_i)$ is a function of customer-level variables Z_i. In an ideal world, it would be known, and would quantify how baseline features Z_i would affect treatment effects if a customer were to be assigned to the treatment group. $B(Z_i)$ is a function of variables Z_i that represents the baseline relationship between variables Z_i and consumption. As with $S(Z_i)$, in an ideal world, $B(Z_i)$ would be known, and would quantify how baseline covariates would impact consumption if a customer were to be assigned to the control group.

θ_t is a bill month fixed effect. P_i is the probability of assignment to treatment, so $P_i = 0.8$, by design, for all customers in the sample.

We proceed with estimation as follows. We use 100 splits of the dataset. In each split of the data, we designate a training and test dataset. Our training dataset consists of 50% of observations, and the test dataset consists of the other 50%.[31]

In each split, using only the training dataset, we train an ML method to predict $B(Z_i)$ and $S(Z_i)$. $\hat{B}(Z_i)$ is the predicted expected baseline consumption for customers with characteristics Z_i if they were assigned to the control group. $\hat{S}(Z_i)$ is the predicted treatment effect if they were assigned to the treatment group.

Then, using only the test dataset, we plug in the predicted $\hat{S}(Z_i)$ and $\hat{B}(Z_i)$, and we estimate the empirical analogue of (4) using these predictions:

$$log(Cons_{it}) = \beta_1(T_i - P_i) + \beta_2(T_i - P_i)\hat{S}(Z_i) + \nu' X_i + \theta_t + \epsilon_{it} \quad (5)$$

In the above, X_i is a vector that includes $\hat{S}(Z_i)$ and $\hat{B}(Z_i)$, as well as an additional control for the number of values imputed for that household.[32] $\hat{\beta}_1$ and $\hat{\beta}_2$ are our BLP parameters (ATE and HTE). Standard errors are clustered at the customer level for this analysis to parallel the ITT and IV specifications used earlier.

We then break the predicted values of $\hat{S}(Z_i)$ into 4 quartiles, and with these quartiles, estimate:

$$log(Cons_{it}) = \sum_{j=1}^{4} \gamma_j \mathbb{1}\left\{\hat{S}(Z_i) \in I_j\right\}(T_i - P_i) + \nu' X_i + \theta_t + \epsilon_{it} \quad (6)$$

In the above, $\mathbb{1}\left\{\hat{S}(Z_i) \in I_j\right\}$ is an indicator for the predicted $s(Z_i)$ falling in the jth quartile.

The medians over the γ_js from each split will be the GATES parameters.

Additionally, we calculate the average of covariates Z_i from our test dataset for each quartile

least in the same spirit they were originally intended) if we apply them using only the post period. Therefore, to account for individual and strata by month of year effects in the machine learning, we estimate them in the ITT regression and then net them out of the consumption variable prior to the machine learning exercise. This ensures that we have adequately controlled for baseline consumption and differences in behavior of different strata. We still include bill month fixed effects (for each post-period observation) in the machine learning specification.

[31] In each split, we stratify splits by the strata in our experiment (flats, non-nationals in villas, or nationals in villas). The stratified splitting chooses 50% of each strata to use for the training in each split, and 50% to use for the testing (outcome model) stage.

[32] The empirical analogue of (4) contains a couple of modifications following equation 3.3 in Chernozhukov et al. (2022): namely, we control for $\hat{S}$ as well as $\hat{B}$, and also add in the number of imputed values as a control.

to characterize the heterogeneity groups according to covariates of interest. These averages constitute our CLAN.

6.3 Comparison of Learners

Before providing results quantifying the best linear predictor, we identify the best learners. Table 3 compares statistics quantifying the correlation between the ML proxy and the best predictor, separately for the BLP and GATES estimates. In our case, Random Forest performs best for the BLP but Support Vector Machines performs best for the GATES.

Table 3: Comparison of Learners

	BLP (Λ)	GATES ($\bar{\Lambda}$)
Support Vector Machines	0.067	0.055
Random Forest (10 trees)	0.077	0.044
Neural Net	0.022	0.038
Elastic Net	0.018	0.029
Gradient Boosting	0.001	0.017

Notes: This table shows the comparison of learners in terms of Λ and $\bar{\Lambda}$ statistics found in Chernozhukov et al. (2022). These quantify the correlation between the ML proxy predictor and the best predictor.

6.4 Best Linear Predictor Results

Table 4: Best Linear Predictor of CATE across machine learning models

	SVM	Forest	N Net	E Net	Boost
ATE	-0.131^{**}	-0.131^{**}	-0.110	-0.108	-0.102
	$[-0.240, -0.020]$	$[-0.232, -0.025]$	$[-0.249, 0.025]$	$[-0.242, 0.023]$	$[-0.238, 0.035]$
	$\{0.021\}$	$\{0.016\}$	$\{0.108\}$	$\{0.111\}$	$\{0.150\}$
HTE	0.889^{***}	0.623^{***}	0.795	1.159^{***}	0.006
	$[0.363, 1.427]$	$[0.286, 0.970]$	$[-0.155, 1.877]$	$[0.415, 1.822]$	$[-0.655, 0.682]$
	$\{0.001\}$	$\{0.000\}$	$\{0.116\}$	$\{0.003\}$	$\{0.973\}$
Clusters	247	247	247	247	247
Obs	1,706	1,706	1,706	1,706	1,706

Notes: This table shows the estimates for the Best Linear Predictor (BLP) for each of the five machine learners we compared in our heterogeneity analysis. ATE shows the estimated average treatment effect. HTE is the estimate of heterogeneity. SVM=Support Vector Machines, Forest=Random Forest with 10 trees, N Net= Neural Network, E Net = Elastic Net, Boost=Gradient Boosting, 90% CIs directly below estimates are computed as the medians over 100 splits; p-values in brackets below confidence intervals are also computed as the medians over 100 splits. Results are clustered at the customer level.

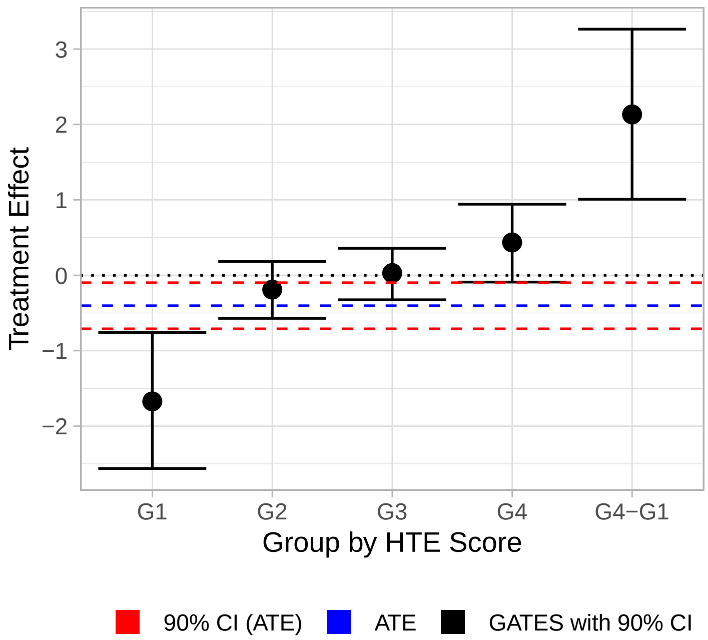
In this section, we first present the results from the BLP (Best Linear Predictor) of the Conditional Average Treatment Effect (CATE) for all five learners in Table 4. We find that the treatment effect is negative for all learners investigated, and similar across learners. Random Forest was the best learner for the BLP, and thus should be our preferred estimate. The point estimate is -0.131, with a 90% confidence interval of $[-0.232, -0.025]$ ($p = 0.021$). It is worth noting that that best learner for the GATES, SVM, exhibits a nearly-identical point estimate of -0.131, with a similar 90% confidence interval of $[-0.240, -0.020]$. The estimates are larger than the treatment effect we found in our main results, but 90% confidence intervals include the magnitudes we found in the main results.

We find significant heterogeneity in the treatment effect across all columns. The coefficient on the heterogeneity is 0.623 and we reject the null of no heterogeneity at the level $p = 0.001$.

6.5 Group Average Treatment Effect Results

Next, we explore the group average treatment effects. Figure 7 depicts the effects graphically. We reject the null that the first quartile of treatment effects is equal to the fourth quartile, in favor of the finding that the groups are very different. Treatment effects are mostly concentrated in G1, which is the group that experienced the highest magnitude of treatment effects. The confidence intervals for the three other groups (G2, G3, and G4) all include 0. The estimated treatment effect for G4 is positive, indicating that defiance is possible. However, it is not statistically significantly different from 0, so we cannot make any definitive conclusions. The overall takeaway is that treatment is highly concentrated in the most treated group (G1).

Figure 7: Group average treatment effects of best learning model



Notes: This figure depicts treatment effect heterogeneity. The vertical axis shows the estimated coefficient corresponding to the overall treatment effect for each of the four treatment groups (as defined by the HTE score), as well as the difference between the most and least treated groups. Point estimates and 90% confidence intervals are constructed as medians over 100 splits.

6.6 Classification Analysis Results

We next characterize the groups according to the features used to construct our heterogeneity proxies. We compare the most and least treated groups (groups 1 and 4) in this section of the paper. The results are presented in Tables 5 through 11. We break the results into subsections based on categories of the variables we used to predict potential heterogeneity: motivation, responsibility, and climate change opinions; easiness and effectiveness of actions; energy consumption and beliefs about relative consumption; and demographics and house characteristics.

Recall that group 1 is the most treated (has the most negative treatment effect) and group 4 is the least treated (has the most positive treatment effect) according to Figure 7. In light of this, we will focus only on the extremes, presenting the mean of the feature for groups 1 (δ_1) and 4 (δ_4), as well as their difference ($\delta_4 - \delta_1$).[33]

6.6.1 Motivation, Responsibility, and Climate Change Opinions

First, we evaluate to what extent motivation, responsibility, and climate change opinions explain heterogeneity in treatment effects. We present the level of agreement with four statements about motivation to change, climate change, and responsibility. Respondents were asked if they agreed with each of the following statements: "Humans are responsible for climate change," "Humans don't need to change," "I am responsible for climate change," and "I need to change."

Table 5: CLAN, Opinions About Climate Change and Responsibility

	δ_1 (Most Treated)	δ_4 (Least Treated)	$\delta_4 - \delta_1$ (Difference)
Humans responsible	4.754	4.318	-0.421^{***}
for climate change	[4.65, 4.86]	[4.13, 4.51]	[−0.64, −0.22]
			{0.000}
Humans don't	1.549	2.050	0.514^{***}
need to change	[1.38, 1.71]	[1.83, 2.27]	[0.25, 0.77]
			{0.000}
I am responsible	4.278	3.791	-0.497^{***}
for climate change	[4.15, 4.40]	[3.58, 3.99]	[−0.74, −0.24]
			{0.000}
I need to change	4.321	3.874	-0.452^{***}
	[4.18, 4.47]	[3.67, 4.07]	[−0.69, −0.21]
			{0.000}

Notes: This table depicts opinions about climate change and personal responsibility for the most and least treated customers. 90% CIs directly below estimates are computed as the medians over 100 splits. P-values in brackets below confidence intervals are also computed as the medians over 100 splits.

The results are found in Table 5. The most treated households are significantly more likely than the least treated to think that they personally are responsible for climate change and to

[33] It is worth noting that, since we take the median of these means over 100 splits, the estimates of the difference $\delta_4 - \delta_1$ need not equal the difference of δ_4 and δ_1 exactly.

agree that they need to change. They are more likely to believe that humans are responsible for climate change and that humans need to change. This suggests a significant role for motivation, personal responsibility, and views on societal responsibility in nudges that work through moral suasion.

6.6.2 Easiness and Effectiveness of Actions

Next, we delve into whether agency plays a role in explaining heterogeneity in our treatment effects. Survey respondents were asked how easy a variety of actions are, and also asked how effective the same actions were. The actions analyzed were: turning off lights, changing to energy-efficient lightbulbs, changing the AC temperature, and consuming less.

We find convincing evidence that our treatment works better for consumers who already believe that actions are easy and effective. Table 6 shows that our most treated group is more likely to think taking actions to conserve is easy to do. Table 7 shows that respondents in the most treated group also are more confident in the effectiveness of the actions on the whole. This suggests that heterogeneity in our treatment effect is also driven by agency.

Recall that the phrase "You have the power to conserve!" appears in both messages. The finding that the participants who view taking action as both easy and effective respond the most suggests that this empowering part of the messages may have be a moderating channel via which these identity-based primes work.

Table 6: CLAN, How Easy are Actions?

	δ_1 (Most Treated)	δ_4 (Least Treated)	$\delta_4 - \delta_1$ (Difference)
How easy to turn off lights	7.786	7.636	-0.160^{***}
	[7.73, 7.84]	[7.55, 7.71]	[−0.26, −0.06]
			{0.002}
How easy to change to EE bulbs	7.827	7.475	-0.328^{***}
	[7.75, 7.91]	[7.33, 7.62]	[−0.50, −0.17]
			{0.000}
How easy to change AC temp	7.516	7.225	-0.278^{**}
	[7.41, 7.63]	[7.05, 7.40]	[−0.50, −0.06]
			{0.011}
How easy to consume less	7.699	7.578	−0.114
	[7.59, 7.79]	[7.45, 7.70]	[−0.28, 0.05]
			{0.185}

Notes: This table shows customer beliefs about how easy conservation actions are for the most and least treated customers. 90% CIs directly below estimates are computed as the medians over 100 splits. P-values in brackets below confidence intervals are also computed as the medians over 100 splits.

6.6.3 Baseline Energy Consumption and Beliefs about Relative Consumption

Next, we examine whether the most and least treated groups differ in their baseline consumption or their relative beliefs about how their baseline consumption compares to the consumption of others. Effective conservation has been shown to depend on knowledge of own energy use Jessoe and Rapson (2014). Byrne et al. (2018) found that a social comparison

Table 7: CLAN, Effectiveness of Actions

	δ_1 (Most Treated)	δ_4 (Least Treated)	$\delta_4 - \delta_1$ (Difference)
How effective is turning off lights?	6.758	6.809	0.041
	[6.68, 6.84]	[6.74, 6.88]	[−0.06, 0.15]
			{0.445}
How effective is using EE bulbs?	6.767	6.476	-0.298^{***}
	[6.69, 6.84]	[6.35, 6.60]	[−0.44, −0.14]
			{0.000}
How effective is changing AC temp?	6.677	6.252	-0.422^{***}
	[6.58, 6.77]	[6.11, 6.40]	[−0.60, −0.26]
			{0.000}
How effective is consuming less?	6.889	6.821	-0.072^{*}
	[6.84, 6.94]	[6.76, 6.88]	[−0.16, 0.01]
			{0.084}

Notes: This table shows customer beliefs about effectiveness of actions for the most and least treated customers. 90% CIs directly below estimates are computed as the medians over 100 splits. P-values in brackets below confidence intervals are also computed as the medians over 100 splits.

treatment led to higher energy use by low users and by those who overestimate their position in the energy use distribution, and that social comparisons led to lower energy use by high users.

We test whether beliefs about relative consumption explain heterogeneity in treatment in our context by comparing the accuracy of relative beliefs (one's belief about how much they use minus their true quintile of use) between the most and least treated groups. We do this both for one's national quintile of use in the pre-period and one's quintile among their customer group in the pre-period, since survey respondents might conceptualize their electricity consumption relative to others they consider similar to themselves. The results are presented in Table 8. We find no evidence of differential responses according to over or under-estimation of consumption, in contrast with Byrne et al. (2018). This is not surprising because our messages do not include any information about one's true use or comparison with use by others.

6.6.4 Knowledge About Electricity Use of Appliances and Savings Associated with Actions

We also examine participants' knowledge about energy-using appliances and energy savings from conservation activities. Our survey asked participants how much electricity several common appliances use, as well as how much one could save from conservation actions. We produced two statistics from the survey results- a measure of bias, and a measure of overall accuracy. We define bias to be 0 if the participant's belief overlaps with the true range of electricity use of an appliance, and equal to the distance between their belief and the true range[34] when they are not overlapping (so, beliefs are less biased when they are closer to zero, and the sign of our measure indicates the sign of the bias). We define accuracy as the fraction of

[34] We provide sources for true ranges in Appendix Section A.3.

Table 8: CLAN, Consumption and Beliefs about Consumption

	δ_1 (Most Treated)	δ_4 (Least Treated)	$\delta_4 - \delta_1$ (Difference)
Cons, Pre	2.634	2.932	0.247
	[2.04, 3.24]	[2.25, 3.61]	[−0.66, 1.14]
			{0.585}
Var(Cons), Pre	6731.545	8725.034	1961.556
	[4465.50, 9216.99]	[5645.32, 11811.12]	[−2092.81, 5555.35]
			{0.310}
Belief-True Quint	0.159	0.312	0.136
	[−0.06, 0.38]	[0.09, 0.54]	[−0.18, 0.46]
			{0.351}
Rel Cons Belief	2.563	2.724	0.166*
	[2.42, 2.70]	[2.59, 2.86]	[−0.03, 0.37]
			{0.093}
Belief-True Strata Quint	−0.380	−0.163	0.233
	[−0.61, −0.16]	[−0.41, 0.06]	[−0.09, 0.56]
			{0.137}

Notes: This table shows the values of various consumption-related variables for the most and least treated customers. 90% CIs directly below estimates are computed as the medians over 100 splits. P-values in brackets below confidence intervals are also computed as the medians over 100 splits. Pre-period consumption has been scaled by 1000 kWh to improve readability.

an individual's responses that overlapped with the true range. We break the accuracy fraction out by use and savings of electricity, since knowledge about use may differ from knowledge about savings.

The setup of our survey questions on beliefs about energy is similar to that of Attari et al. (2010), a study that found evidence that people mis-perceive energy use and savings- in particular, they found that survey participants underestimated the use and savings associated with the highest-using appliances and the actions that saved the most, respectively.

We present results on perceptions about electricity use and electricity savings in Tables 9 and 10, respectively. In the tables, we ordered appliances by their true use (savings), from lowest to highest.

For appliance use, we find that the most treated group does not necessarily have less biased beliefs. Their beliefs are closer to the truth than the least treated group for energy use of CFLs, window ACs, and Wall ACs. Their beliefs are farther from the truth for other appliances, and they are most biased for the highest-using appliances. The overall accuracy of their beliefs is higher, though, which indicates they are more correct on average.

When it comes to energy savings beliefs, the degree of bias is not different (in a statistically significant sense) for any of the actions analyzed, and the level of bias does not appear to follow a pattern with regard to high-or-low savings actions, except for the most energy-intensive action (reducing dryer use).

In sum, we find mixed evidence on whether the treatment is working through a knowledge channel in general. What we can rule out, however, is that bias is relatively lower among the most treated group for the appliances and actions that use and save the most. Attari et al.

Table 9: CLAN, Bias and Accuracy in Beliefs about Energy Use of Appliances

	δ_1 (Most Treated)	δ_4 (Least Treated)	$\delta_4 - \delta_1$ (Difference)
CFL	9.378	10.826	0.777
	[6.76, 12.51]	[8.12, 13.67]	[−3.13, 4.87]
			{0.666}
Laptop	24.118	27.365	2.278
	[17.78, 30.06]	[21.58, 32.40]	[−5.91, 10.38]
			{0.575}
Stereo	36.820	23.986	-14.209^{***}
	[30.87, 42.58]	[19.68, 27.83]	[−21.14, −6.44]
			{0.000}
Desktop	2.924	1.663	−1.722
	[−0.88, 6.86]	[−1.28, 4.46]	[−6.39, 3.24]
			{0.481}
Window AC	387.105	493.913	106.712^{**}
	[306.80, 463.27]	[423.88, 565.77]	[1.14, 218.31]
			{0.045}
Wall AC	418.306	468.439	41.830
	[330.40, 504.34]	[389.22, 550.71]	[−77.31, 158.93]
			{0.471}
Dishwasher	−202.499	−123.267	85.552^{**}
	[−256.16, −149.81]	[−175.28, −70.85]	[3.83, 158.55]
			{0.037}
Dryer	−430.241	−169.938	260.235^{***}
	[−509.27, −351.14]	[−279.01, −53.46]	[126.18, 388.95]
			{0.000}
Central AC	−757.015	−493.208	311.014^{**}
	[−930.99, −569.49]	[−670.87, −328.80]	[53.28, 560.88]
			{0.016}
Accrcy, Use	0.454	0.380	-0.072^{***}
	[0.42, 0.49]	[0.34, 0.42]	[−0.12, −0.02]
			{0.007}

Notes: This table shows the bias in energy use perceptions for the most treated and least treated customers, followed by the overall accuracy of these perceptions. 90% CIs directly below estimates are computed as the medians over 100 splits. P-values in brackets below confidence intervals are also computed as the medians over 100 splits.

Table 10: CLAN, Bias and Accuracy in Beliefs about Energy Savings

	δ_1 (Most Treated)	δ_4 (Least Treated)	$\delta_4 - \delta_1$ (Difference)
100 → 75W Bulb	2.193	3.533	1.140
	[0.95, 3.40]	[1.76, 5.33]	[−1.00, 3.62]
			{0.291}
Inc → CFL	−1.010	−1.049	−0.053
	[−2.82, 0.55]	[−4.74, 1.56]	[−3.96, 3.63]
			{0.966}
↓ Washer Use	1563.699	1553.407	−32.805
	[1372.02, 1767.26]	[1391.64, 1749.24]	[−291.50, 240.01]
			{0.812}
↑ AC in Summer	1336.725	1482.766	177.171
	[1119.02, 1555.87]	[1240.37, 1732.07]	[−159.66, 514.00]
			{0.288}
↓ Dryer Use	−323.386	−192.452	120.741*
	[−415.68, −223.92]	[−324.92, −74.22]	[−27.26, 268.63]
			{0.089}
Accrcy, Savings	0.550	0.440	−0.108***
	[0.51, 0.59]	[0.41, 0.47]	[−0.16, −0.06]
			{0.000}

Notes: This table shows the bias in energy savings perceptions for the most treated and least treated customers, followed by the overall accuracy of these perceptions. 90% CIs directly below estimates are computed as the medians over 100 splits. P-values in brackets below confidence intervals are also computed as the medians over 100 splits.

(2010)'s findings suggest that correcting mis-perceptions for the highest-using appliances and highest-savings actions could move the needle on promoting conservation. But, our findings show that the most treated group is actually *more* biased in their relative perceptions about highest-using appliances and highest-savings actions. This casts doubt on whether differential mis-perceptions between appliances and actions are a significant barrier to conservation.[35]

6.6.5 Demographics and House Characteristics

Finally, we assess the remaining variables used for machine learning prediction: demographics and residence characteristics (Table 11). We find suggestive evidence that the most treated consumers are likely to have more bedrooms and bathrooms in their house, which indicates they have larger houses on average. We do not see differences in the fraction of Qatari nationals or the fraction of villas.[36] This accords with our finding of no heterogeneity in percentage reductions by customer group in Section 5.2.

One surprising finding from Table 11 is that the most treated customers are less likely to be Muslim. The fraction of Muslims is 0.768 in the most treated group and 0.908 in the least treated group. On average, Muslims comprise 79% of the customers in the sub-sample used for ML. This means that, relative to the overall sample, the most treated group is not

[35] After all, our most treated group achieved significant conservation despite exhibiting more bias for the most energy-intensive appliances and actions.

[36] Recall that strata have different baseline consumption, so similar treatment effects in percentages translate to very different treatment effects in terms of overall consumption saved.

significantly less Muslim; instead, the least treated group is significantly *more* Muslim. This suggests possible defiance among Muslims.[37]

Part of the lack of effectiveness among Muslims could be explained by the fact that these messages were sent by a state institution. Following the disruption of the Arab spring, many leaders in the Gulf region moved to co-opt religious groups, and some used corrupted interpretations of Islam to legitimize their rule. We conjecture that there could be negative mental associations being activated by a state-sent Islamic message.

Table 11: Demographic and House-Related Variables

	δ_1 (Most Treated)	δ_4 (Least Treated)	$\delta_4 - \delta_1$ (Difference)
Qatari	0.089	0.093	0.012
	[0.05, 0.13]	[0.05, 0.13]	[−0.04, 0.06]
			{0.602}
Muslim	0.768	0.908	0.144^{***}
	[0.71, 0.82]	[0.87, 0.95]	[0.08, 0.21]
			{0.000}
Villa	0.189	0.205	0.003
	[0.14, 0.24]	[0.15, 0.26]	[−0.07, 0.08]
			{0.861}
No. of people in house	5.988	5.545	−0.385
	[5.31, 6.63]	[5.07, 6.03]	[−1.17, 0.45]
			{0.379}
Bedrooms	4.292	3.560	-0.740^{***}
	[3.85, 4.69]	[3.25, 3.85]	[−1.28, −0.20]
			{0.007}
Full Baths	5.206	3.715	-1.498^{**}
	[3.95, 6.43]	[3.30, 4.13]	[−2.75, −0.21]
			{0.021}
Central AC	0.500	0.426	-0.081^{*}
	[0.43, 0.57]	[0.36, 0.49]	[−0.17, 0.01]
			{0.094}

Notes: This table shows the values of various demographic variables for the most and least treated customers. 90% CIs directly below estimates are computed as the medians over 100 splits. P-values in brackets below confidence intervals are also computed as the medians over 100 splits.

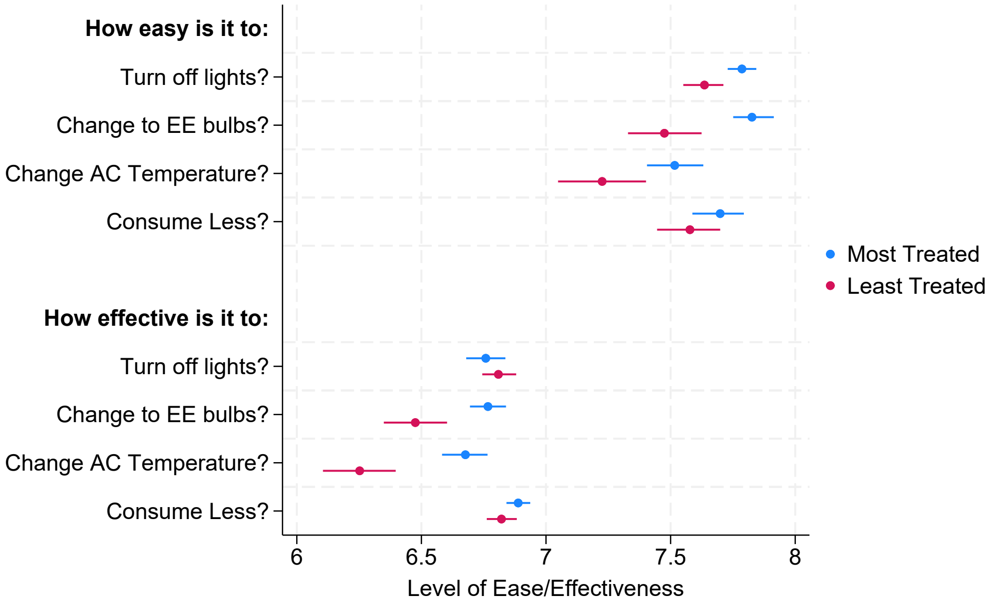
6.6.6 Overall takeaway from CLAN

Our overall takeaway is that the most treated households are those who feel responsible for climate change, are motivated to change, feel humans are responsible for climate change and should change, and find actions easy and effective.[38] This resonates with recent literature showing that perceived social norms are strongly correlated with individual willingness to

[37]To delve deeper, we estimated $\delta_4 - \delta_2$ and $\delta_4 - \delta_3$ for the Muslim variable. Our finding was that the null hypothesis was rejected in both cases ($p = 0.016$ and $p = 0.000$ respectively), confirming that the least treated group is different from the rest.

[38]One concern is that these results are not separate findings, but driven by strong correlations between these responses. To investigate this possibility, we show correlations between the three groups of variables in Figure A.3. Correlations are very weak, with the strongest correlations occurring between beliefs about how easy and how effective actions are. We take this as suggestive evidence that these findings are not merely driven by associations between these three groups of variables.

Figure 8: Differences in median ease and effectiveness of energy saving actions



Notes: This figure depicts the differences in reported ease and effectiveness of energy saving actions between the most treated (in blue) and least treated (in red) groups of customers. Higher values on the horizontal axis correspond to greater ease and effectiveness reported in the survey. Dots represent the median estimates among 100 splits and lines represent corresponding 90% confidence intervals. The figure shows that customers in the most treated group are likely to believe that energy saving actions are easier and more effective than the corresponding beliefs among the least treated customers.

pay to combat climate change, measured through donations to an environmental charity (Falk et al., 2021).

We summarize our findings in Figure 8 below, showing differences on perceived ease and effectiveness of actions between the most and least treated groups of customers.

7 Conclusion

We find that injunctive norms priming religious values and national identity decrease energy consumption by around 3.8% on average for customers in Qatar. This translates to sizeable average reductions of over 100 kWh per month per customer.[39] We find that even Qatari nationals living in villas conserve in response to our messages. This result is particularly striking given that nationals generally do not pay for electricity, and suggests that low cost non-pecuniary interventions can have sizeable effects on emission reductions for super-users of electricity. We find no statistically significant differences in effects by number of messages; future work is needed to understand whether the persistence effects documented by the literature (e.g. Allcott and Rogers (2014)) also reflect reinforcement effects due to the receipt of multiple messages.

[39]For comparison, the effect of Home Energy Reports was estimated at around 0.62 kWh per day (Allcott, 2011) or 18.6 kWh per month.

We can further contextualize the magnitude of our main effects in terms of value of emission reductions in Qatar. Electricity and heat production for the residential sector resulted in 11.9 million tons of CO_2 emissions in 2021. A 3.8% reduction would imply 452,200 tons of CO_2 avoided emissions – equivalent to the energy consumed by 56,992 homes over one year in the United States.[40] At a social cost of carbon of US $15 per ton of CO_2 in Qatar (Ricke et al., 2018), this amounts to benefits of approximately US$ 6.78 million.

Combining data on electricity use with a supplementary customer survey, we employ the Machine Learning technique from Chernozhukov et al. (2022) to investigate heterogeneity. We find significant evidence for heterogeneous effects, and examine a host of potential predictors of heterogeneity. Customers who respond the most (a) believe that both they themselves and humans in general are responsible for climate change, (b) believe that they themselves need to change and humans need to change, and (c) believe actions to conserve electricity are both easy and effective. These suggest roles for responsibility, motivation to change, and agency in the response to injunctive norm treatments.

We do not find evidence that the most responsive customers are more knowledgeable– either about their own electricity use relative to others, or regarding the energy consequences of popular household conservation 'tips' like turning off lights. Further, we also test for heterogeneous effects based on pre-intervention consumption level and, unlike some recent work (Knittel and Stolper, 2021; Gerarden and Yang, 2023), find no evidence for heterogeneous responses on this dimension in percentage terms. It is worth noting that the entire population we study could be considered high-use. This suggests that perhaps after a certain baseline use threshold, there is enough low-hanging fruit for households to conserve relative to their baseline use.

A few important caveats are in order. First, the effects of our religious and national treatment messages are not statistically different. Hence, we cannot speak to which type of message is more effective. Indeed, while the two messages delivered different content, they both contain the text "You have the power to conserve!" Our findings suggest that the messages may have appealed most to those who believed that they had this power. Second, we focus on high-income customers whose levels of electricity use are well above the world average. So, our findings may not generalize to low-income (and consequently lower electricity use) customers who may be more price elastic or find information on their own use to be more salient. Third, our dataset contains less than a year of data following the start of our intervention in May 2019, and four months of data following the end of the messages,[41] and electricity use is highly seasonal as shown in Figure 3. With such a short timespan of data, we are unable to credibly measure whether effects persist after treatment is withdrawn.[42] We anticipate that they do,

[40] EPA equivalencies calculator.

[41] The messages were delivered over the period May 2019 - October 2019, leaving only November 2020 - February 2020 as post-period untreated months.

[42] An additional empirical challenge to measuring persistence is that the seasonal nature of electricity use is differential by customer group, so we could not disentangle customer group impacts from impacts of seasonality

given the extraordinary persistence documented in the literature (Allcott and Rogers, 2014).[43]

Future research is needed on how to leverage agency to induce more conservation. Perhaps information treatments could be more effective if they focused on information about easy actions to take and highlighted how effective simple actions are at conserving energy. Our study shows that information is not needed for motivated individuals to act. Further research could evaluate whether information about effective and easy actions works better than an empowering message on its own. Further research is also needed on whether these attitudes are mutable. Finally, future work should explore the extent to which one can leverage the identified relationship between environmental attitudes and effectiveness of treatment to see if programs targeting both attitudinal change and behavioral change are more effective than one of the two approaches.

on treatment effects. Note that this motivates our use of strata by month fixed effects.

[43]Even if we were to somehow find persistence in our short panel, it would not be particularly informative for policy given the very long time horizon of persistence documented in the literature. Allcott and Rogers (2014) find a decay rate of only 10-20% after the messages are ceased.

References

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, 115(3):715–753.

Al Muftah, H. (2018). Human development in qatar. *Global Encylopedia of Public Administration, Public Policy, and Governance. Cham: Springer.*

Al-Saidi, M. (2020). Instruments of energy subsidy reforms in arab countries—the case of the gulf cooperation council (gcc) countries. *Energy Reports*, 6:68–73.

Al-Ubaydli, O., List, J. A., and Suskind, D. (2020). 2017 klein lecture: The science of using science: Toward an understanding of the threats to scalability. *International Economic Review*, 61(4):1387–1409.

Al-Ubaydli, O., List, J. A., and Suskind, D. L. (2017). What can we learn from experiments? understanding the threats to the scalability of experimental results. *American Economic Review*, 107(5):282–286.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9-10):1082–1095.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–37.

Andor, M. A. and Fels, K. M. (2018). Behavioral economics and energy conservation–a systematic review of non-price interventions and their causal effects. *Ecological economics*, 148:178–210.

Attari, S. Z., DeKay, M. L., Davidson, C. I., and De Bruin, W. B. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences*, 107(37):16054–16059.

Bator, R. J., Phelps, K., Tabanico, J., Schultz, P. W., and Walton, M. L. (2019). When it is not about the money: Social comparison and energy conservation among residents who do not pay for electricity. *Energy Research & Social Science*, 56:101198.

Bekker, M. J., Cumming, T. D., Osborne, N. K., Bruining, A. M., McClean, J. I., and Leland Jr, L. S. (2010). Encouraging electricity savings in a university residential hall through a combination of feedback, visual prompts, and incentives. *Journal of applied behavior analysis*, 43(2):327–331.

Benjamin, D. J., Choi, J. J., and Fisher, G. (2016). Religious identity and economic behavior. *Review of Economics and Statistics*, 98(4):617–637.

Bernstein, D. H., Cassidy, A., and Khalifa, A. (2023). Work-from-home, electricity, and water: Evidence from covid-19 in qatar. *Energy Strategy Reviews*, 49:101119.

Bloom, H. S. (1984). Accounting for no-shows in experimental evaluation designs. *Evaluation Review*, 8(2):225–246.

Bonan, J., Cattaneo, C., d'Adda, G., and Tavoni, M. (2021). Can social information programs be more effective? the role of environmental identity for energy conservation. *Journal of Environmental Economics and Management*, 108:102467.

Brewer, D. (2023). Household responses to winter heating costs: Implications for energy pricing policies and demand-side alternatives. *Energy Policy*, 177:113550.

Bryan, G., Karlan, D. S., and Osman, A. (2023). Big loans to small businesses: Predicting winners and losers in an entrepreneurial lending experiment. *Working Paper*.

Burlig, F., Knittel, C., Rapson, D., Reguant, M., and Wolfram, C. (2020). Machine learning from schools about energy efficiency. *Journal of the Association of Environmental and Resource Economists*, 7(6):1181–1217.

Bursztyn, L., Fiorin, S., Gottlieb, D., and Kanz, M. (2019). Moral incentives in credit card debt repayment: Evidence from a field experiment. *Journal of Political Economy*, 127(4):1641–1683.

Byrne, D. P., Nauze, A. L., and Martin, L. A. (2018). Tell me something i don't already know: Informedness and the impact of information programs. *Review of Economics and Statistics*, 100(3):510–527.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernández-Val, I. (2022). Generic machine learning inference on heterogenous treatment effects in randomized experiments.

Coady, D., Parry, I. W., Sears, L., and Shang, B. (2015). How large are global energy subsidies? *Working Paper*, 15-105.

Davis, J. M. and Heller, S. B. (2020). Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs. *The Review of Economics and Statistics*, 102(4):664–677.

De Oliveira, A. and Smith, P. (2022). Qatar world cup lays bare the huge environmental cost of tournament. *NBC News*.

Delmas, M. A., Fischlein, M., and Asensio, O. I. (2013). Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012. *Energy Policy*, 61:729–739.

Douenne, T. and Fabre, A. (2020). French attitudes on climate change, carbon taxation and other climate policies. *Ecological Economics*, 169:106496.

Elinder, M., Escobar, S., and Petre, I. (2017). Consequences of a price incentive on free riding and electric energy consumption. *Proceedings of the National Academies of Science*, 114:3091–3096.

Falk, A., Andre, P., Boneva, T., and Chopra, F. (2021). Fighting climate change: The role of norms, preferences, and moral values. *CESifo Working Paper No. 9175.*

Fang, X., Goette, L., Rockenbach, B., Sutter, M., Tiefenbeck, V., Schob, S., and Staake, T. (2022). Complementarities in behavioral interventions evidence from a field experiment on energy conservation. *Available at SSRN 4147196.*

Ferraro, P. and Price, M. K. (2013). Using non-pecuniary strategies to influence behavior: Evidence from a large scale field experiment. *Review of Economics and Statistics*, 95(1):64–73.

Gerarden, T. D. and Yang, M. (2023). Using targeting to optimize program design: evidence from an energy conservation experiment. *Journal of the Association of Environmental and Resource Economists*, 10(3):687–716.

International Energy Agency (2018). *World Energy Outlook 2018, International Energy Agency.* "OECD/IEA".

Ito, K., Ida, T., and Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, 10(1):240–267.

Jessoe, K. and Rapson, D. (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review*, 104(4):1417–38.

Kessler, J. B., Milkman, K. L., and Zhang, C. Y. (2019). Getting the rich and powerful to give. *Management Science*, 65(9):4049–4062.

Knittel, C. R. and Stolper, S. (2021). Machine learning about treatment effect heterogeneity: The case of household energy use. In *AEA Papers and Proceedings*, volume 111, pages 440–444. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale.* Currency, Penguin Random House.

Marghetis, T., Attari, S. Z., and Landy, D. (2019). Simple interventions can correct misperceptions of home energy use. *Nature Energy*, 4(10):874–881.

Muller-Rommel, F. (2019). *New politics in Western Europe: The rise and success of green parties and alternative lists.* Routledge.

Murakami, K., Shimada, H., Ushifusa, Y., and Ida, T. (2022). Heterogeneous treatment effects of nudge and rebate: causal machine learning in a field experiment on electricity conservation. *International Economic Review*, 63(4):1779–1803.

Petersen, J. E., Frantz, C. M., Shammin, M. R., Yanisch, T. M., Tincknell, E., and Myers, N. (2015). Electricity and water conservation on college and university campuses in response to national competitions among dormitories: Quantifying relationships between behavior, conservation strategies and psychological metrics. *PloS one*, 10(12):e0144070.

Reiche, D. (2010). Energy policies of gulf cooperation council (gcc) countries—possibilities and limitations of ecological modernization in rentier states. *Energy Policy*, 38(5):2395–2403.

Ricke, K., Drouet, L., Caldeira, K., and Tavoni, M. (2018). Country-level social cost of carbon. *Nature Climate Change*, 8(10):895–900.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, 18(5):429–434.

Tiefenbeck, V., Wörner, A., Schöb, S., Fleisch, E., and Staake, T. (2019). Real-time feedback promotes energy conservation in the absence of volunteer selection bias and monetary incentives. *Nature Energy*, 4(1):35–41.

Tsai, I.-T. and Mezher, T. (2020). Rationalizing energy policy reforms in the gulf cooperation council: Implications from an institutional analysis. *Energy Policy*, 142:111545.

Whillans, A. V. and Dunn, E. W. (2018). Agentic appeals increase charitable giving in an affluent sample of donors. *PloS one*, 13(12):e0208392.

A Appendix

A.1 Construction of Preferred Sample and Implications for Estimation

In this appendix, we discuss the construction of our preferred sample. Our experimental sample of customers included several problems with their electricity data which we found after the experiment was implemented. In Table 1, we show that our experimental sample and preferred sample are both balanced in terms of pre-intervention outcomes. Below, we discuss both how we excluded customers to arrive at our preferred sample and the implications of doing so.

Table A.1 provides a summary of our exclusion criteria, including how many customers are dropped in each step when constructing our preferred sample. It should be kept in mind, that these issues with our data overlap, so for example, excluding observations where pre-period consumption is 0 also means excluding many customers with no “AC” readings.

Table A.1: Construction of Preferred Sample

Reason	Customers Remaining	Customers Dropped
Entire experimental sample	6,096	
Multiple meters for same customer ID	5,859	237
Pre-period cons = 0	5,781	78
Mult nationalities for same customer ID	5,679	102
Multiple residences for same customer ID	5,570	109
No AC readings	5,559	11
Account shares phone number with someone else in the sample	5,197	362
Customer ID associated with multiple phone numbers	4,841	356

Notes: This table shows the number of observations remaining and dropped in each step of construction of the preferred sample. The sequence of dropped observations for each reason depends on the order, and there is overlap between these issues. Also note that observations for which pre-period cons = 0 would be naturally dropped as our specifications are in logarithms.

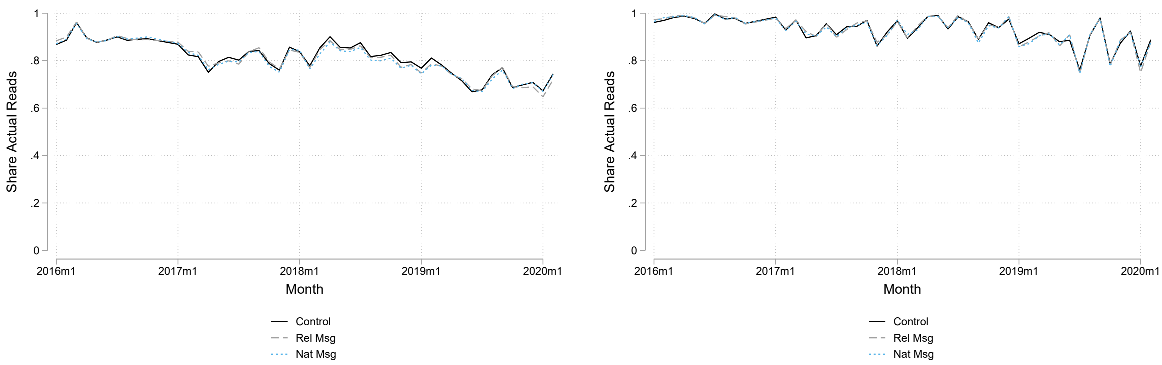
First, our preferred sample eliminates customers for whom it is difficult to accurately assign electricity use for a specific month to a registered meter and a specific strata and those customers for whom pre-period consumption in all months equals 0. The former problems are caused by customer IDs that do not have a unique electricity meter, customers IDs that do not have a unique nationality category, or customer IDs that do not have a unique residence category (flat or villa) within the period of our experiment. The latter should not affect our results as we use the logarithm of electricity use as our outcome variable.

We then move on to removing issues that may attenuate our estimates. To that end, we drop any electricity reading that is not classified as an actual (“AC”) electricity read. An AC reading means that the electricity consumption reported for a specific customer for a specific month only occurred in the corresponding billing month. Two other reading types exist in the data: reconciled and missing. Reconciled reading types could represent consumption in months other than the current month, and thus might make our estimates less precise. In particular, reconciled readings in the months after treatment could also pick up use in months before treatment. This will tend to bias treatment effects downward. Missing readings are coded as 0 but typically are included in future reconciled readings. Figure A.1 shows that the proportion of actual electricity read types among the three types of electricity reads. We see that the proportion of actual reads is around 80% over the period of our analysis and this proportion does not differ significantly between treated and control groups. However, if we only include non-zero consumption observations, the share hovers around 90%. This measure is also uncorrelated to treatment.

Figure A.1: Actual Meter Reads by Month and Treatment Assignment

(a) All Obs

(b) Non-zero Obs



Notes: This figure depicts the share of electricity meter reads that are recorded as actual reads in each month by treatment status. In panel A.1a, we include all observations, whereas in panel A.1b, we only use observations with non-zero consumption to calculate the share of actual reads. This is in line with our use of consumption in logarithms in the regressions as our outcome variable.

Next, we only include customers with just one phone number associated with the electricity company for their account. If a customer has just one phone number, and we record the phone received the message, then we can be reasonably sure the customer received the message. However, for customers with two phone numbers recorded, if we record that one of the phones received the message, we do not know if it is the phone the customer is using. For example, in the Qatari context, it could be the phone number of a servant responsible for dealing with utility bills. In such cases, our treatment effect would be attenuated if these customer IDs were included in the estimation.

Our preferred sample retains 4,841 customers. Out of this, 5 customers have singleton observations, i.e. only one customer ID-by-month electricity use observation in our panel. These 5 observations do not contribute to our estimates in Table 2, where the total number of customers is indicated as 4,836.

It is worth noting that we conduct this exercise to have more accurate electricity use observations in our preferred sample. Exclusion criteria like using AC reads only and eliminating the possibility of sharing phone numbers or having multiple phone numbers corresponding to the same customer ID reduce potential attenuation of our treatment effects. In total, we exclude 1,255 customers from the experimental sample - 276 customers from the Control group, 483 from the Religious treatment, and 495 from the National treatment group.

A.2 Message Receipt

In this appendix, we discuss message receipt in detail. Not receiving text messages could be due to the phone numbers registered on the utility's database being invalid, or 'Blacklisted' which means that they have been blocked by the respective cellphone service providers or because the messages are 'Undeliverable'. The message delivery company defines 'Undeliverable' as numbers that are 'out of coverage area, [have] bill issues, they [are not] available to receive calls or messages, or unused'.

Table A.2 tabulates the percentage of customer-month observations in each of the two treatment groups that received each number of messages. Table A.3 shows that the percentage of households within each treatment group that receive at least one message over the treatment period to be around 84% for households in the regression sample, and this proportion is very similar between the two treatment groups.

In the sample we use for estimation, 12.5% of observations assigned to the religious message group did not receive a single message, and 13.5% of observations assigned to the national group never received a message. Perfect treatment occurred in about half of the sample, with 52.72% of the religious message group receiving all 12 messages over the course of the sample and 52.38% of the national message group receiving all 12 messages. No customers in the sample were sent the wrong message, and no customers assigned to the control group received messages.

The probability of a given customer in our sample having received at least one message in prior months should in general be slightly higher in later months of our sample. In Figure A.2, we show the percent of customer-month observations in our sample that received at least one message in prior months. Because not every customer is observed in every month, this relationship is not always monotonically increasing over time. We also note that nationals in villas are less likely to receive either treatment.

A.3 Description of Survey and Survey Data

We used a phone survey implemented over three waves prior to our experiment to collect data on demographic characteristics and beliefs from Qatari residents. The overall number of survey participants was 328, but we only use the 247 who match to our field experimental sample. We will refer to these 247 as the "survey sample." A link to all survey questions is available here.

As a cursory check that the survey and field experiment population are similar, we tabulate the number and percentage of participants in each strata in both the overall estimation sample and the sample matched to the supplemental survey. The results appear in Table A.4. The proportions are very similar.

Next, we present supplementary material on how we constructed our variables. We imputed all survey variables within strata where missing. Table A.5 describes how we define correct

Table A.2: Number of Messages Received as Percentage of Observations in Each of the Two Treatment Groups

	Assigned Treatment Group:		
Number of Messages Received:	Religious Message	National Message	Total
0	12.54	13.53	13.03
1	0.56	0.72	0.64
2	0.76	0.57	0.67
3	0.71	0.87	0.79
4	0.56	0.67	0.61
5	0.87	1.03	0.95
6	1.38	1.59	1.48
7	1.83	1.70	1.77
8	2.75	2.26	2.51
9	4.18	5.04	4.61
10	7.95	7.97	7.96
11	13.20	11.73	12.47
12	52.70	52.31	52.51
Total	100.00	100.00	100.00

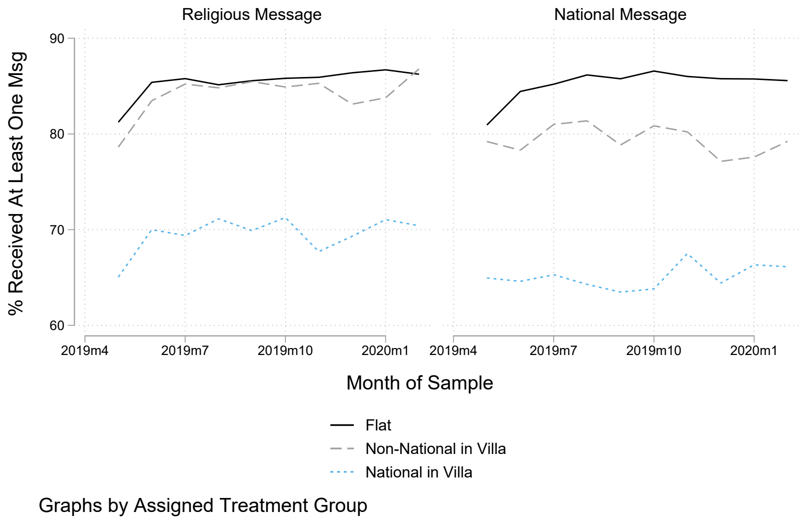
Notes: Each cell is the percentage of customer-month observations in the assigned treatment group that received the number of messages at left.

Table A.3: Customers receiving at least one message by treatment status

	(1) *Experimental sample*		(2) *Preferred sample*	
	Obs	Percentage of customers assigned	Obs	Percentage of customers assigned
Religious Treatment	1,837	75.34	1,654	84.65
National Treatment	1,835	75.26	1,618	83.49
Total	3,672	75.31	3,272	84.07

Notes: The first and third columns show the counts of houses receiving at least one message, for the experimental and preferred sample respectively. The second and fourth columns show the percent that received at least one message, out of all customers assigned to a given treatment group.

Figure A.2: Message Receipt by Assigned Treatment Group and Strata



Notes: This figure depicts the percent of customers in each of the two treatment groups observed in each month who have received at least one message in prior months. It only uses observations from our regression sample (N=161,104). Note that not every customer is observed in every month.

Table A.4: Overlap between Regression and Survey Samples

(a) Main Estimation Sample

	Treatment Group:	
	Control	Treatment
Flats	901(20%)	3,625(80%)
Non-Nat in Vla	109(19%)	450(81%)
Nat in Vla	108(19%)	451(81%)

(b) Matched to Supplemental Survey

	Treatment Group:	
	Control	Treatment
Flats	41(21%)	151(79%)
Non-nat in Villa	9(30%)	21(70%)
Nat in Villa	8(32%)	17(68%)

Notes: This table depicts how the main estimation sample and supplemental survey compare in terms of the number and percentage of customers by strata and treatment assignment.

ranges for beliefs about energy savings and energy use.

Table A.5: Actual Energy Usage and Savings of Appliances and Technologies

Appliance	**Attari 2010**	**Qatar Range**	**Source**
CFL	27	20-33	Link.[44]
Desktop PC	140	60-300	Link.[45]
Laptop	48	60	Link. [46]
Stereo	128	50-74[47]	
Dryer	3400	1800-5000	Link.[48]
Central AC	3500	3750-5000[49]	
Window AC	1000	500-1500 [50]	Link [51]
Wall AC	1000	500-1500[52]	Link [53]
Dishwasher	1800	1200-2400	Link.[54]
CFL	73	67-80	Link.[55]
Incandescent 75W	25	25[56]	
Dryer	3400	1800-5000	Link. [57]
AC Summer	115	1000-1300[58]	
Washer Cycle	4000	330-1000[59]	

Notes: This table shows the assumptions we have made about the "correct" usage ranges, and how they compare to assumptions made in Attari et al. (2010). These assumptions are used to construct the bias and accuracy measures. We define bias to be 0 if the participant's belief overlaps with the true range of electricity use of an appliance, and equal to the distance between their belief and the true range. We define accuracy to be the fraction of a customer's responses where the belief overlapped with the "correct" range.

[44] CFLs use 1/3rd to 1/5th the electrical power of incandescent lighting and can last 8 to 15 times longer

[45] An average desktop computer uses between 60 and 300 watts.

[46] Estimate that 60 watts is average power consumption for a 14-15 inch laptop when plugged in.

[47] Attari et al. (2010) has average use of stereo as 128 and Marghetis et al. (2019) has average use of stereo as 33.

[48] The energy use of a dryer varies between 1800 watts and 5000 watts, a typical dryer will use around 3000 watts.

[49] See Figure 6 in Alrawi et al (2016) for AC power consumption in a villa that uses central air conditioning for a broad indication. We have adjudged the range of actual use to lie between 90-120 kWh a day.

[50] Refer to Qatar Standards document and example labels for ACs in Qatar/Saudia Arabia.

[51] Single room air conditioners come in different sizes and use from 500 to 1500 watts.

[52] Refer to Qatar Standards document and example labels for ACs in Qatar/Saudia Arabia.

[53] Single room air conditioners come in different sizes and use from 500 to 1500 watts.

[54] Dishwashers use between 1200 and 2400 watts of power, with an average dishwasher using 1800 watts.

[55] CFLs use 1/3rd to 1/5th the electrical power of incandescent lighting and can last 8 to 15 times longer"

[56] Replacing a 100W incandescent bulb with a 75W incandescent would save exactly 25 units.

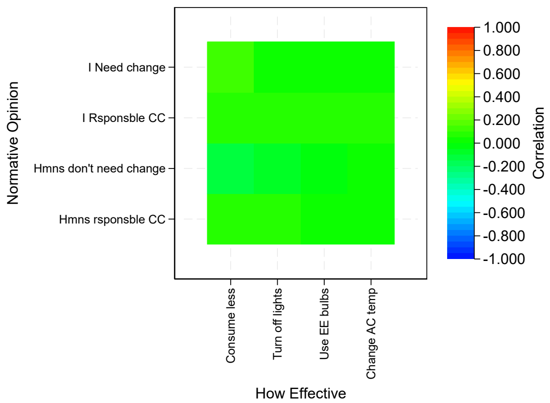
[57] The energy use of a dryer varies between 1800 watts and 5000 watts, a typical dryer will use around 3000 watts.

[58] AC use in summer in Qatar is very high compared to the US average. Changing temperature setting from 22C to 24C for one hour should conserve around 1000-1300 Wh using degree days data and average use of an AC according to information in degreedays.net

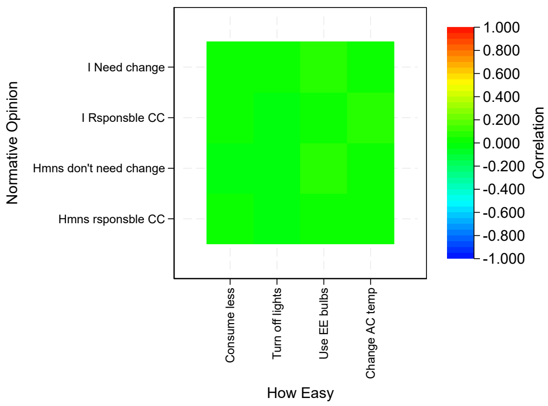
[59] Attari et al. (2010) uses estimates from Rocky Mountain Institute. Perusal of two dishwasher manuals for models sold in Qatar suggests total energy consumption of 1000 Watt-hours per load. The default "recommended" connection to water supply is to the cold water faucet. Direct Energy, a retail electricity provider, estimates that heating water increases this use level – with the amount of increase varying between 33 - 100% depending on energy efficiency rating of the unit.

Figure A.3: Correlations Between Opinions on Easiness of Actions, Effectiveness of Actions, and Normative Questions

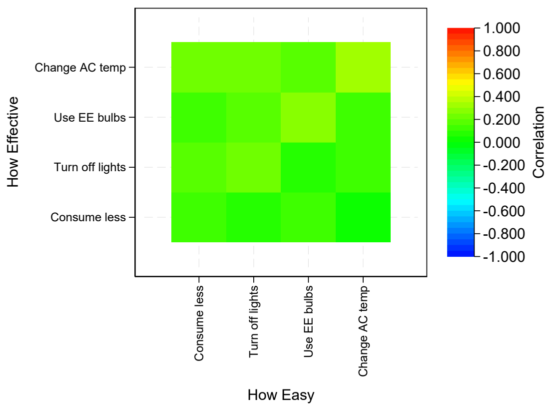
(a) Normative Opinions vs. How Effective



(b) Normative Opinions vs. How Easy



(c) How Easy vs. How Effective



Notes: This figure depicts correlation between the variables we find matter most in explaining treatment effect heterogeneity. If correlations were high, then we would worry that the heterogeneity might all be driven by one of the three sets of variables. We find most correlations are close to 0.

In Table A.6, we present summary statistics from the sample that matches between our estimation sample and our supplemental survey, which is the sample we use for our machine learning analysis.

Figure A.3 presents correlations between the groups of variables that we find the strongest evidence of heterogeneity in, to address the concern that strong correlations between the groups of variables drive our results.

Table A.6: Summary Statistics from Supplemental Survey

	Control			Treatment			
	Mean	Std Dev	Customers	Mean	Std Dev	Customers	Diff
Cons, Pre	4.16	6.04	58	2.95	4.83	189	-1.204
Belief-True Strata Quint	-0.30	1.87	58	-0.46	1.47	189	-0.161
Turn off lights	7.61	0.55	58	7.77	0.52	189	0.161*
Use EE bulbs	7.30	1.47	58	7.68	0.92	189	0.383
Change AC temp	7.10	1.45	58	7.43	1.11	189	0.322
Consume less	7.47	1.05	58	7.72	0.73	189	0.245
Belief-True Quint	0.07	2.00	58	0.13	1.50	189	0.059
No. of people in house	6.40	4.98	58	5.22	3.36	189	-1.184
Rel Cons Belief	2.72	1.05	58	2.64	1.05	189	-0.080
Var(Cons), Pre	12985.55	26822.77	58	8984.33	23910.60	189	-4,001.220
Central AC	0.36	0.47	58	0.53	0.50	189	0.168**
Bias UseCFL	11.18	23.94	58	9.35	17.47	189	-1.830
Bias UseLaptopPC	16.91	44.00	58	29.25	45.68	189	12.332
Bias UseStereo	28.63	34.22	58	30.21	48.00	189	1.582
Bias UseDesktopPC	-0.68	20.98	58	1.26	22.67	189	1.938
Bias UseWindowAC	502.62	600.84	58	440.40	568.63	189	-62.219
Bias UseWallAC	498.42	655.87	58	453.44	603.06	189	-44.974
Bias UseDishwasher	-135.52	398.72	58	-145.39	437.32	189	-9.866
Bias UseDryer	-269.14	801.42	58	-308.93	629.44	189	-39.787
Bias UseCentralAC	-524.11	1279.11	58	-670.75	1350.48	189	-146.633
Bias SavingsIncandescent75W	5.55	15.69	58	2.62	11.34	189	-2.930
Bias SavingsCFL	4.65	49.95	58	-4.12	15.88	189	-8.773
Bias SavingsWasherCycle	1562.81	1456.06	58	1629.20	1426.41	189	66.382
Bias SavingsACSummer	1619.73	1934.94	58	1423.67	1636.00	189	-196.056
Bias SavingsDryer	-290.11	845.84	58	-269.16	836.14	189	20.943
Accrcy, Use	0.42	0.31	58	0.40	0.24	189	-0.018
Accrcy, Savings	0.46	0.25	58	0.49	0.25	189	0.029
Bedrooms	3.90	2.19	58	3.69	2.67	189	-0.208
Full Baths	5.16	9.28	58	3.89	3.73	189	-1.268
Muslim	0.83	0.38	58	0.82	0.38	189	-0.010
Qatari	0.15	0.36	58	0.10	0.31	189	-0.050
Villa	0.30	0.46	58	0.19	0.40	189	-0.109
Turn off lights	6.89	0.32	58	6.73	0.78	189	-0.160**
Use EE bulbs	6.60	0.79	58	6.64	0.85	189	0.036
Change AC temp	6.23	1.13	58	6.49	0.92	189	0.267
Consume less	6.82	0.43	58	6.85	0.65	189	0.031
Hmns rsponsble CC	4.50	1.20	58	4.69	0.88	189	0.191
Hmns don't need change	1.96	1.56	58	1.73	1.38	189	-0.234
I Rsponsble CC	3.99	1.41	58	4.13	1.23	189	0.138
I Need change	4.05	1.45	58	4.18	1.20	189	0.127

Notes: These are descriptive statistics on all covariates used for the machine learning analysis.