CEWP 19-10

Visualizing Energy Efficiency: A Picture is Worth More Than 1,022 Words

Maya Papineau
Carleton University

Nicholas Rivers
University of Ottawa

November 2019; revised 26 May 2020

CARLETON ECONOMICS WORKING PAPERS

Carleton UNIVERSITY

Department of Economics

1125 Colonel By Drive
Ottawa, Ontario, Canada
K1S 5B6

Visualizing energy efficiency: A picture is worth more than 1,022 words*

Maya Papineau†
Carleton University

Nicholas Rivers‡
University of Ottawa

May 26, 2020

Abstract

We deploy a randomized controlled trial involving approximately 12,500 households showing that providing consumers with a visual depiction of heat loss on utility bills leads to considerably larger energy savings compared to a popular social comparison "nudge". Images showing roof heat loss were provided to approximately 4,000 randomly selected households in on-bill messaging. Heat loss is visualized using infrared images taken from an aircraft-mounted infrared sensor during the winter heating season. A similarly-sized randomly selected group received bill messaging with a 'traditional' social norm comparing their consumption to similar homes. We also find that the heat loss treatment results in a higher rate of realized energy efficiency durables investment and leads households to conserve in a manner consistent with private and social efficiency: the most inefficient households exhibit much larger energy reductions relative to the traditional social comparison.

Keywords: Energy efficiency; Social norms; Nudges; Randomized field experiments

*We acknowledge the substantial contributions of Grant Guenther and Mike Dykstra at the City of Medicine Hat; Daygan Fowler and Darren Jones at MyHEAT; and funding from Natural Resources Canada and the Social Sciences and Humanities Research Council. Massimo Filippini, Ken Gillingham, Arik Levinson, Leslie Martin, Erica Myers, Matthew Webb, and seminar and conference participants at Carleton University, McGill University and the Canadian Resource and Environmental Economics Association provided valuable comments.

†Department of Economics, Carleton University (maya.papineau@carleton.ca)

‡Corresponding author. Graduate School of Public and International Affairs and Institute of the Environment, University of Ottawa (nicholas.rivers@uottawa.ca)

1 Introduction

Information programs aimed at affecting consumer decisions using insights from the behavioral sciences are used across many different domains, including personal finance, health, and education (Carroll et al., 2009; Johnson and Goldstein, 2003; Damgaard and Nielsen, 2018). The intent of these programs is to shape individual behavior in a manner considered to be welfare-improving from the perspective of both society and the individual. These "nudges" often take the form of selectively revealing information to individuals or conveying normative messages about peer behavior, and have been shown in some cases to exert a strong influence on individual behavior. However, while it is clear that in some cases nudges do motivate behavior, their effect sizes are often small and much remains to be learned about how best to use nudges to promote welfare-improving behavioral change.

In this paper we show that behavioral insights can be used to significantly magnify the energy savings of a widely deployed peer-comparison nudge through simple changes in the framing and visualization of consumer consumption decisions. Our study is motivated by findings in the psychological literature that images may be better recalled and promote a stronger response than the same information presented in figures or text. We test this motivation in a randomized controlled trial focused on household energy efficiency, and find that visual presentation of heat loss information together with a social comparison elicits a considerably stronger response, as measured by energy consumption and energy-using durables investments, than figures including peer comparisons about energy consumption.

Home Energy Reports (HERs) that motivate consumers to reduce energy consumption through normative feedback comparing a household's energy usage to that of comparable households have become increasingly popular over the last decade (Allcott, 2011; Nolan et al., 2008; Schultz et al., 2007). The most prominent example is Opower's HERs, which are currently sent regularly to over 60 million customers at more than 100 utilities (Greentech Media, 2018; Oracle, 2019). These programs are seen by utilities and regulators as a cost-effective way to reduce energy consumption with few barriers to implementation, and a number of peer-reviewed studies have found that HERs are effective at modestly reducing energy consumption.

Early HERs rolled out by Opower between 2008 and 2010 lowered electricity consumption by about 2 percent for an average household (Allcott (2011)), and these reductions persisted for several years (Allcott and Rogers (2014)). On the other hand, Allcott (2015) found that utilities who were among the early adopters of Opower's home energy reports had larger shares of high income and environmentalist consumers, such that evaluations performed using data from early-adopter utilities, despite their high internally validity due to randomization, overstate future program efficacy in other customer populations. As a result, despite their popularity, the magnitude of savings from existing HER programs means they are not likely to make meaningful

contributions towards meeting the ambitious greenhouse-gas reduction commitments by several U.S. states, Canada, and the European Union.[1]

Psychological theory suggests visual processing can improve both the salience of information and the motivation to take action (Slovic et al., 2002; Epstein, 1994). Our treatment combines this insight with a novel aerial remote-sensing application that identifies house-level thermal envelope characteristics (Hay et al., 2011; Rahman et al., 2015). More specifically, we provide consumers with high-resolution infrared images of their house on their monthly utility bills. The infrared images are taken at night from a small aircraft in the heating season in a Canadian city with a cold climate, and capture heat loss from customers' roofs and walls. The thermal images clearly illustrate sources of heat loss for each home, and are accompanied with a personalized heat loss score and messaging communicating each customer's score in terms of heat loss. Together with the images, households are provided with information comparing their own score to their neighbourhood score.

If, in addition to being motivated by normative peer comparisons, consumers respond both to framing energy consumption as leading to heat loss and the visualization of this heat loss, we expect our treatment to magnify the conservation impact of a traditional HER. Images of home energy loss may provide unique actionable information (for example, on the location of an air leak) and induce some customers to respond in a way that numerical information on home energy consumption cannot replicate. A number of studies provide a basis for the hypothesis that thermal images may provide behavioural cues that motivate consumers to take action to improve energy efficiency. In particular, thermal images may provide "vivid" representations of energy loss that can draw viewers in and hold attention in a way that tabular information fails to do (Slovic et al., 2002; Nisbett and Ross, 1980). Images are also considered to affect behaviour by being more available for recall during decision making, and to convert abstract ideas (heat loss) into more concrete terms that can be acted upon (Sheppard et al. (2011)). Two prior studies using small samples of voluntarily recruited participants in the UK offer suggestive evidence that improving the visibility of energy use may be more effective at reducing consumption compared to energy audits or textual information (Goodhew et al., 2015; Boomsma et al., 2016).

For the study, we partnered with a municipally-owned natural gas and electricity distribution company in Medicine Hat, Alberta and MyHEAT, an Alberta-based firm that provided us with thermal imaging measurements for each household. Roughly 12,500 single-detached households from across the city were randomly selected to participate in the study, and randomly divided into two treatment groups and one control group of roughly equal sizes. One treated group, which we call the 'traditional' social comparison group, was shown a figure comparing their

[1] The Canadian federal government and the European Union have both endorsed carbon neutrality goals by 2050, and several U.S. states have also committed to similar goals. These goals are summarized in Environment and Climate Change Canada (2019), European Commission (2018), and Center for Climate and Energy Solutions (2019).

electricity and natural gas consumption to both average and energy efficient similar homes.[2] The second treatment group, which we call the MyHEAT social comparison group, received a high-resolution infrared image of their roof indicating areas of heat loss, their thermal image-based heat loss score ranging from 1 to 10, and a comparison of their score to their average neighbourhood score. Finally, like many conventional HERs, recipients from both treatments were provided with information about their potential bill savings from energy efficiency improvements as well as tips for improving energy performance.

We evaluate these interventions using daily data on natural gas and electricity consumption for households in the control and both treatment groups. Our data covers the period from about one year prior to treatment to one year following initial treatment. Our analysis produces several findings. First, we find that the MyHEAT social comparison caused consumers to reduce natural gas consumption by about 0.8 percent on average, but had small and statistically insignificant impacts on electricity consumption, whereas the traditional social comparison had small negative but statistically insignificant impacts on both natural gas and electricity consumption. However, we find substantial heterogeneity within each of the treatment groups. While there are either small or insignificant aggregate impacts of each treatment, across both groups low-efficiency/high consumption households respond by reducing electricity and natural gas consumption by more than 5 percent on average, while high efficiency/low consumption households respond by increasing usage by roughly 3 percent, a "boomerang" effect that has been documented in other studies (Byrne et al., 2018; Delmas et al., 2013).[3]

Second, a key finding emerges when we account for household heterogeneity driven by pre-existing consumption levels. As noted above, for both treatments, households were informed how much they could save annually on their energy bills by improving their energy efficiency. These annual savings were personalized for each customer such that more inefficient or high consuming households were shown higher annual savings from undertaking energy saving measures. When we take into account the heterogeneity in potential savings, we find that while both treatments induced a statistically significant reduction of natural gas consumption per dollar of estimated savings, the effect of the MyHEAT social comparison was considerably larger. For a household with $150 of estimated potential annual savings from improving energy efficiency, the traditional social comparison reduced natural gas consumption by 2 percent, whereas the MyHEAT treatment reduced consumption by 4.4 percent – more than double the traditional HER.

The savings from the MyHEAT treatment are further magnified when we include separate coefficients for the most efficient households who were shown potential savings of zero, and

[2] The reference to "1,022 words" in the title refers to the word count in the bill shown to the traditional social comparison group.

[3] The high/low-efficiency groups were defined on the basis of annual data and controls for pre-treatment consumption levels do not change our results so these results are not driven by mean-reversion.

for those that saw non-zero savings. Aside from the most efficient households, the MyHEAT treatment reduced gas consumption by 8.1 percent per $100 of savings, whereas the traditional social comparison reduced natural gas consumption by 1.3 percent per $100 of savings, about six times less than the MyHEAT treatment. Similar results are observed for electricity, though the savings magnitudes are smaller. Given that the annual cost of thermal imaging is about one dollar per home whereas traditional HERs cost one dollar per customer report, typically sent 6-12 times per year, we conclude that the thermal imaging treatment holds promise for cost-effectively magnifying the savings achieved from home energy reports.

Finally, by linking the household addresses to a database maintained by the provincial energy efficiency agency, we show that households that receive the MyHEAT treatment are more likely to participate in energy efficiency programs following treatment than either the control group or the households in the traditional home energy report group. These energy efficiency programs are targeted at improving the thermal integrity of the building shell. As a result, these results are suggestive that the intervention produced gains in energy efficiency and not just transient changes in behaviour.

In addition to implementing a novel nudge program that magnifies HER savings, our paper also contributes to a growing literature that documents divergences between the predicted versus realized performance of low-carbon investments. In some instances predictions about the returns to residential energy efficiency or renewable energy investments have been found to overestimate returns, underestimate the pre-existing efficiency of homes, or, as in the case of solar financing, use metrics that imperfectly predict repayment performance, sometimes by a large margin (Levinson, 2016; Papineau, 2017; Fowlie et al., 2018; Davuluri et al., 2019). These findings point to the importance of developing strategies that accurately predict home energy efficiency to help target programs to the most inefficient homes, and reduce the regressivity of energy efficiency and renewable energy subsidies (Allcott and Kessler, 2019; Drehobl and Ross, 2016; Allcott and Greenstone, 2012).

Our study utilizes what we term a hybrid engineering and realized consumption approach to make predictions about energy usage and bill reductions from improved building envelope energy efficiency. Our use of infrared technology to detect long wave thermal radiation mapped to individual homes allows us to identify residence-level heat loss and generate a simple metric to identify the most energy inefficient homes: a heat loss score ranging from 1 to 10. We link this residence-level heat loss score to customer consumption data to predict annual bill savings from a reduction in heat loss (or equivalently, an improvement in building envelope energy efficiency). As discussed later in the paper, we find a strong relationship between measured heat loss scores and annual pre-treatment household energy consumption. This suggests that a targeting strategy aimed at improving the efficiency of high-score homes is likely to improve the realized returns from the myriad weatherization and demand-side management programs

deployed by utilities and governments worldwide.

The rest of the paper is organized as follows. Section 2 begins with a brief literature review on the impact of HERs on natural consumption and the psychological basis for using imagery to motivate action, then describes our experimental treatments that are the focus of this paper. Section 3 provides an assessment of the relationship between our measured heat loss ratings and realized energy consumption. Section 4 describes our data sources for the variables used in the analysis, and Section 5 presents our results. Section 6 briefly concludes.

2 Background and experiment overview

The treatment we employ is targeted at magnifying the natural gas savings from home energy reports, one of the most prominent non-price energy conservation programs employed among utilities worldwide. While most of the peer-reviewed HER literature has focused on electricity savings, studies that have evaluated their effects on fossil fuel usage for space heating, primarily natural gas, have found that natural gas HERs tend to induce smaller conservation effects (Allcott and Kessler, 2019, Smith and Morris, 2014, Kerr and Tondro, 2012). This suggests that in northern countries with high heating demands met mostly by fossil fuels, there are large potential gains from strategies that improve the effectiveness of HERs. Space heating is the largest single contributor to residential energy consumption across the thirty International Energy Agency member countries (International Energy Agency, 2019).[4] In Canada, where our study takes place, close to 62 percent of residential sector end-use energy and 63 percent of residential greenhouse gases (GHGs) are from space heating, whereas 1.5 percent of GHGs arise from space cooling (Natural Resources Canada, 2019).[5]

Our use of infrared imagery and heat loss messaging is influenced by academic work from psychology and the behavioral sciences. Several studies have documented the role of imagery at invoking affective responses from our emotionally driven "experiential" cognitive processes rather than more rational systems of cognition, and that these responses are both effective at motivating action and predictive for the type of action undertaken (Epstein, 1994). Affective reactions are emotional responses that occur automatically in response to a stimulus that captures a decision-maker's attention, and subsequently guides information processing and judgment (Zajonc, 1994). This cognitive system of decision-making has been referred to as the "affect heuristic", and images that trigger affective responses have been found in several settings to both predict and guide judgment, from stock purchases to adolescent behavior to raising awareness of environmental change (Slovic et al., 2002; MacGregor et al., 2000; Sheppard et al.,

[4] There are 30 IEA member countries, primarily countries located in Europe and North America, as well as Japan and Korea.

[5] About fifteen percent of GHGs derive from lighting and appliance use and another twenty percent from water heating.

2011; Benthin et al., 1995). Framing effects may also interact with affective responses generated from infrared images of customers' roofs (DellaVigna, 2009). Relative to traditional social comparisons, which focus messaging on energy consumption relative to peers, the MyHEAT treatment instead communicates household consumption and the social comparison exclusively in terms of heat loss, which is an approach to framing energy use that is likely unfamiliar to most customers.

Our experiment was deployed in Medicine Hat, Alberta. Medicine Hat is a city of about 60,000 located in southeast Alberta, with relatively hot summers and cold winters. The municipally-owned utility provides gas, electricity, and water to all City residents and businesses, and was responsible for coordinating the experiment whose results we report here. Until this experiment, the City had not implemented other behavioural feedback on energy bills, such as Home Energy Reports.

The experiment consists of providing on-bill feedback to randomly selected households. Households in Medicine Hat receive monthly utility bills (including natural gas, electricity, and water), and the intervention began by including the treatments on the February 2018 billing cycle.[6] The treatments were repeated on the March, April and November billing cycles of 2018. Figure 1 illustrates the timeline of the experimental intervention. The intervention was run by the City of Medicine Hat as a *natural field experiment*. As described in Czibor et al. (2019), in a natural field experiment subjects include the relevant population (rather than a sub-sample of voluntary participants) and subjects are not aware of being part of an experiment, which eliminates Hawthorne effects that have been shown to be important in similar studies (Schwartz et al., 2013).[7]

The experiment population includes about 12,500 single family residential buildings randomly selected from the population of municipally-served single-family residential buildings in Medicine Hat. This population was randomly assigned to two experimental groups and a control group using pre-intervention gas and electricity consumption, year of construction, assessed value, and building size as balancing variables. Table 1 shows balance statistics for the three groups, and confirms that the randomization delivers groups that are balanced on observable covariates. For each experimental group, the normalized difference comparing the treatment group with the control group is well below the 0.25 threshold recommended by (Imbens and Wooldridge, 2009).[8] Before the experiment deployment we also completed

[6] Like most utilities, Medicine Hat stratifies its customers into groups, who are on different billing cycles, so not all treated households receive the treatment on the same day.

[7] To address the possibility that control group members may have called the utility to inquire why they did not receive the treatment, the utility's call center was instructed to state that the treatment bills were a pilot project that would eventually be rolled out to the rest of the population if it was successful.

[8] A technical glitch in the utility's billing system resulted in one of the treatment waves sending out erroneous treatments to a small share of customers, primarily in the heat loss and control groups. As a result these households are not included in our analysis. However, as shown in Table 1, this did not demonstrably affect our covariate

calculations of statistical power. The power calculations are available upon request and show that our experimental design has a high power to recover the effects of treatment if they are of similar magnitude to estimates from prior interventions.

The experiment involves two treatment groups and a control group. Treatment 1 is similar to the home energy reports that have been used extensively in recent years to inform households about their energy consumption relative to that of their neighbours (Allcott, 2011; Ayres et al., 2013). Households in this 'traditional' HER treatment group received on-bill messaging that includes a month-to-month consumption comparison, over the previous 12 month period, between a given household and the 50 most similar households, as well as the top quintile of most efficient households among the group of similar households.[9]

A sample treatment 1 is included in Appendix A.1. Large text in a yellow box on the first page of the bill provides a statement comparing the household's energy consumption in the prior month relative to energy consumption in the group of similar homes in the same month, along with an indication of potential annual bill savings from reducing energy consumption to the mean consumption level of similar households. For households with consumption above the mean consumption of similar households, this number is presented as excess annual billing expenditures due to higher energy use, whereas for households with consumption below the mean consumption of similar households, this value is presented as annual expenditures saved as a result of relatively lower usage. This page also prompts customers to see more detail on their relative consumption on page 4 of their bill, with the statement 'See page 4 for your personalized comparison and options to save energy'. Page 4 of the bill presents graphical information on natural gas and electricity consumption over the past year for the household compared to similar households and energy efficient neighbors.[10] A list of suggestions for reducing energy consumption are also included on page 4, including two provincial rebate programs for energy efficient window and insulation investments.[11]

Households in treatment group 2, the 'MyHEAT' HER, were shown infrared images of their roof. The infrared images were taken at night in the heating season three months before the experiment, and measure heat loss from the home's roof. Thermal images were acquired using the MyHEAT technology platform, which is a combination of image acquisition equipment

balance.

[9] 'Similar' homes were the 50 homes with the smallest differences with the comparison home in terms of year of construction and size. 'Similar' homes were also restricted to homes that were on the same billing cycle as the treated household. More precisely, define an index of similarity (IS_{hj}) between a target house j and a possible comparison house h, which compares the attributes of house h to attributes of house j: $IS_{hj} = (Nsize_h - Nsize_j)^2 + (Nyear_h - Nyear_j)^2$, where $Nsize_h = (size_h - mean(size))/sd(size)$ and $Nyear_h = (year_h - mean(year))/sd(year)$. The group of similar homes consists of the 50 homes with the smallest index of similarity.

[10] Annual consumption prior to the first treatment roll-out is used to identify the comparison groups.

[11] We omit pages 2 and 3 of the bill as they are the same for both treatments and do not differ from a typical utility bill.

and processing software designed for the purpose of measuring heat loss from buildings.[12] Thermal images are gathered using an aircraft-mounted infrared sensor, which detects emitted long-wave radiation. These images are used in conjunction with other measurements (e.g., temperature, elevation) as well as building shapefiles to create a thermal profile for all buildings in a municipality. The combined process is able to produce extremely high-resolution thermal images of building heat loss, accurate to within 0.05°C at a sub-one metre resolution. Using the thermal images, each dwelling is assigned a heat loss score in discrete units ranging from 1 to 10, which indicates the amount of heat loss: 1 indicates very low heat loss; 10 indicates very high heat loss.[13] Example thermal images corresponding to heat loss scores of 1 and 10 are provided in Figure 2. The following section of the paper uses pre-program energy consumption data to verify that the thermal images and associated image-based heat loss ratings convey meaningful information about the relative energy performance of dwellings in the experimental region.

A sample bill for treatment 2 is included in Appendix A.2. These households were also provided with on-bill messaging, including text on page 1 of their bill informing them of their heat loss performance and potential annual bill savings from improving their MyHEAT score. Potential bill savings were calculated using the regression coefficients estimated in the following section, and were determined based on the energy savings from an improvement to a heat loss score of 1 (the best possible score). This text is accompanied by a prompt to find further information on page 4, as in treatment one. However, the page four information differs from the traditional social comparison in treatment group 1. It includes the thermal image of their house, along with brief instructions for interpreting the image. Households were also provided with their heat loss score, along with the average heat loss score for houses in their neighbourhood and the average heat loss score for houses in the City of Medicine Hat. Finally, households were provided with the same list of suggestions for reducing home energy consumption as in treatment group 1.

As noted above, households in both treatment groups received on-bill messaging for the first time starting in February 2018 and were provided with messaging for three consecutive months in February through April 2018. Another bill treatment was included in the November 2018 bill.[14] These months were chosen as they cover the heating season, when building heat loss is most important for determining natural gas consumption. The treated unit in the analysis

[12] See https://MyHEAT.ca/technology.

[13] The heat loss scores are assigned using a proprietary algorithm developed by MyHEAT. The heat loss measurements are not affected by customers not being home on the night of the fly-over. Because the measurements are taken during the heating season when temperatures are near or below freezing, most home thermostats would still be on and set to a temperature sufficient to prevent pipes from freezing (typically a setpoint above 10-12 degrees celsius), which is sufficient to record an accurate heat loss reading. When the fly-over was completed the average overnight low at that time of year was minus 4 degrees celsius.

[14] Due to budget constraints the February paper bill mailout was provided in color, while subsequent mailouts were grayscale. Email billing customers, approximately 20% of total customers, received color bills for all four treatments.

is the physical location, and so we do not have concerns about attrition of treated units from the experiment.[15] In addition, we construct a balanced sample by only including locations for which consumption data is available over the entire analysis period to ensure that entry has no effect on our results.

In the following section, we compare pre-intervention natural gas and electricity consumption across homes with different thermal image-based heat loss ratings. We show that buildings with higher (worse) heat loss ratings consume substantially more natural gas, as well as somewhat more electricity, compared to similar homes with lower (better) heat loss ratings.

3 Relationship between MyHEAT rating and energy consumption

This section documents the relationship between infrared image-based MyHEAT ratings and building energy consumption. To do this, MyHEAT scores, which were collected overnight on October 31 2017, are compared to building energy consumption data for 2015 and 2016. For each dwelling, we merge monthly billed energy consumption in 2015 and 2016, MyHEAT rating, and tax assessment information. Tax assessment information includes data on building size, building type,[16] year built, assessed value, as well as neighborhood and street name. The full merged data set consists of 12,304 observations. Each observation represents a single dwelling, and contains the MyHEAT score, average annual natural gas and electricity consumption over 2015 and 2016, and building characteristics.[17]

We assess the relationship between MyHEAT rating and building energy consumption using a regression framework. We consider a cross-sectional regression of the form:

$$\log(Y_i) = \beta_0 + \beta_1 \text{HEATSCORE}_i + \beta_2 \log(\text{Size}_i) + \phi \mathbf{X_i} + \epsilon_\mathbf{i}, \tag{1}$$

where i indexes dwellings, HEATSCORE_i is the MyHEAT score ranging from one to ten, Y_i is average annual energy (electricity or gas) consumption, Size_i is building size in m^2, and $\mathbf{X_i}$ includes other observable variables, such as building age, building type, etc.

Table 2 summarizes the results of the analysis in which the logarithm of natural gas con-

[15] To the extent that the intervention causes changes in *customer* behavior as opposed to *physical location*, our analysis will result in an underestimate of the treatment effect on customers. Billing data from the utility indicates that 12% of customers moved during the one-year treatment window.

[16] There are 16 building types in the assessment data with at least 100 observations. Building types are descriptors such as: 1 1/2 storey with basement; 1 storey multi side x side basementless; 2 storey duplex with a basement; etc.

[17] We remove 21 dwellings for which the building size is listed as less than 10m^2, as well as roughly 1,500 dwellings for which we are missing a MyHEAT score or for which annual gas, electricity, or water consumption is zero over either 2015 or 2016. Because we dropped homes with zero consumption in 2015 or 2016, the total number of observations reported in this section is slightly smaller than our experimental sample in the reported results from Section 5.

sumption is the dependent variable. Column (1) includes a control for building size only. The estimate suggests that a one unit improvement (i.e. decrease) in the MyHEAT score is associated with a 4.3 percent reduction in natural gas consumption. The second column also includes building type as an explanatory variable, such that only buildings of the same type are compared to each other. This column suggests that a one unit reduction in the MyHEAT score is associated with a 4.8 percent reduction in natural gas consumption. The third column adds controls for year built. This column suggests that each one-unit MyHEAT score improvement reduces gas consumption by 3.2 percent. The fourth column adds a control for neighbourhood name and street name. The fifth column adds a control for the (log of the) assessed value of the house. In each of these last two columns, the coefficient remains unchanged and indicates that a one unit reduction in the MyHEAT score is associated with a 3.4 reduction in natural gas consumption. Based on these estimates, we estimate that each one unit improvement in the MyHEAT score is associated with 3.2-4.8 percent less natural gas consumption. In each case, the standard errors indicate we are able to estimate the effect with a substantial amount of precision and reject the null hypothesis that there is no relationship between MyHEAT rating and energy consumption.

Table 3 estimates the same regression using electricity consumption, rather than gas, as a dependent variable. Aside from the first column, which does not include controls for anything except building size, the regression coefficients suggest each unit of MyHEAT improvement is associated with savings of electricity of 1.4-1.9 percent. Once again, the effect is estimated precisely.

Tables 2 and 3 treat the MyHEAT rating as a continuous variable, and find that reductions in the MyHEAT score of a dwelling (i.e., decreases in measured heat loss) are associated with reductions in energy consumption. In Figure 3, we re-estimate the models above, but treat the MyHEAT rating as a discrete variable:

$$\log(Y_i) = \beta_0 + \sum_{n=1}^{10} \beta_n \mathbb{1}(HEATSCORE_i == n) + \beta_2 \log(SIZE_i) + \theta X_i + \epsilon_i \tag{2}$$

where $n = \{1, ..., 10\}$ indicates the set of possible MyHEAT ratings. We treat dwellings with a MyHEAT rating of 5 as the reference category, and measure energy consumption relative to that category. We adopt the formulation in column (2) of the tables above, which conditions energy consumption on both building size as well as building type. Results are qualitatively unchanged when we include additional controls.

Consistent with the prior analysis, Figure 3 shows a strong relationship between natural gas consumption and MyHEAT rating. Buildings with a MyHEAT rating of 10 consume on average about 50% more natural gas than similarly sized buildings of the same type with a MyHEAT rating of 1. For electricity, the results show a distinct relationship between MyHEAT rating and

electricity consumption as well, although the standard errors are larger, particularly on houses with extreme MyHEAT ratings, such that the relationship is not as clear as for gas. This is not surprising, since the primary space heating fuel is natural gas, rather than electricity, in the city under study (the utility refers to Medicine Hat as 'The Gas City' on its bills).

Overall, the findings in this section indicate that the MyHEAT rating is a significant predictor of residential energy consumption.

4 Experiment Data

We combine data from a number of sources to conduct the analysis that follows. Altogether, after cleaning data, we observe bills and consumption from 12,407 households. The control group includes 3,963 households, treatment group one includes 4,441 households, and treatment group two includes 4,003 households.[18]

Our dataset is constructed on the basis of a number of different data sources. The first data source is monthly consumption and expenditure data for each household and billing period. Utility bills provide monthly information on natural gas, electricity, and water consumption, and are available starting in 2015. Second, in addition to monthly billing data, we also obtain a separate source of consumption data from household meters. Medicine Hat uses digital (smart) electricity meters that record electricity and natural gas consumption at both daily and hourly intervals. Our main analysis is based on daily natural gas and electricity consumption data. The consumption data starts in January 2017 and extends until March 2019, roughly one year prior to and after treatment (see Figure 1). We retain only households with a complete set of daily consumption data, such that there is no entry to the sample nor attrition from the sample.

Third, we obtain tax assessment data, which provides information on building size, assessed value, building type (e.g., split level, bungalow, etc.), neighbourhood, and year of construction. Fourth, we obtain thermal imaging data for all residential dwellings from MyHEAT. As described above, each dwelling is given a heat loss score ranging in discrete units from 1 (low heat loss; high efficiency) to 10 (high heat loss; low efficiency). We observe this score for all buildings in the population.

The unit of observation in our analysis is the residential building-day, rather than the customer. This is useful since it makes it straightforward to ensure that there is no attrition from or selection into our sample. However, it does not take into account possible moving of customers into and out of houses during the period covered by our analysis. To the extent that households respond to the intervention with behavioral or habit changes, rather that permanent

[18] As noted previously, the original randomization was designed to include about 4,500 homes in each group but a technical problem with the City's billing system in the first treatment month results in about 10 percent of the observations being generated with errors. As a result we have excluded these observations.

physical changes to the housing equipment or envelope, moves of customers between households after the treatment is initiated will attenuate the treatment effect we seek to estimate.

5 Analysis

5.1 Treatment impacts on natural gas and electricity consumption

We begin the analysis by using the data described above to estimate a model that captures the average impact of treatment on energy consumption:

$$Y_{it} = \beta_0 + \sum_{k=1}^{2} \beta_k T_{ik} \times P_{it} + \mu_i + \lambda_t + \epsilon_{it}, \tag{3}$$

where Y_{it} is consumption (either gas or electricity), normalized by average post-treatment consumption in the control group.[19] T_{ik} is a treatment dummy, which indicates whether household i is in treatment group $k = \{1, 2\}$, and P_{it} is a post-treatment dummy that indicates whether the observation is in the post-treatment period (note that since households are mailed bills on different dates, the post-treatment period differs by household). We define the post-treatment period as any time after the mailout of the first treated bill to the household. We also include a location (house) fixed effect μ_i and day-of-sample fixed effect λ_t. The location fixed effect absorbs any constant differences between households (number of occupants, thermal properties of the dwelling, etc.). The day-of-sample fixed effect absorbs common factors that shift over time that impact households (weather, holidays, etc.). In specification (3) β_k is then the average effect of treatment k – the effect of treatment on electricity or natural gas consumption in the post-treatment period. Given our specification in equation (3), β_k is identified from within-household and within-day differences between the treatment and control groups. For all reported specifications, standard errors are two-way clustered by household and day-of-sample.

We report the results of estimating (3) in Table 4. Columns (1) and (2) pertain to the impact of the interventions on daily natural gas consumption, and columns (3) and (4) describe the impact on daily electricity demand. Columns (1) and (3) aggregate the impacts of the traditional HER and and the thermal image interventions together, and columns (2) and (4) separately identify the effects of each of these interventions, relative to the control group. The results in columns (1) and (3) indicate that the two treatments combined reduced natural gas consumption by 0.6 percent (significant at the 10 percent level), but had no significant effect on electricity

[19] This normalization is the same as that used by Allcott (2011). We use this normalization instead of taking logarithms to avoid dropping zero daily consumption observations or using an ad-hoc procedure to include zero observations when taking logs. Coefficients can be interpreted identically to a model with a logged left-hand side variable. In practice, we find little difference between the results of a model estimated with a normalized left hand side and a model estimated with a logged left hand side variable.

consumption. The coefficients in column (2) show that on average the traditional HER treatment did not significantly reduce daily natural gas consumption. In contrast, the coefficient on the MyHEAT social comparison is almost three times larger and significant at the 5 percent level. In the case of electricity, column (4) indicates that both treatment effects are small and neither are statistically different from zero. In total, the average effects of the traditional social comparison are small and on the low side of other estimates in the HER literature reviewed in the introduction, while the MyHEAT intervention has somewhat larger impacts than the traditional social comparison.

Our findings relating to the average impact of traditional HER interventions are fairly small compared to the extant literature. There are three potential explanations. First, Medicine Hat, where the experiment took place, is amongst the most conservative regions in Canada. In the 2019 federal election, right-leaning (Conservative and People's Party) parties garnered 82% of the vote in this electoral district, whereas the winning Liberal party only received a 6% vote share. Costa and Kahn (2013) show that in the US conservative (republican) voters respond three times less to HER interventions compared to progressive voters. Second, Medicine Hat is a key hub of Canada's natural gas industry (it is referred to as "The Gas City"), and has one of the lowest shares of green voters in the country (2.3% in the 2019 federal election). Again, Costa and Kahn (2013) show that environmentalists respond much more to HER interventions compared to non-environmentalists. Third, Allcott (2015) shows that regions that adopt HERs early tend to have favourable conditions for their success, and have larger impacts than later-adopting regions. Medicine Hat is representative of a "later-adopting region," and may be expected to have lower impacts from such an intervention than typical early-adopting regions.

Another important point to consider is that the small average treatment effects reported above mask important within-group differences in the treatments that these households receive. Specification (3) groups all households in treatment group 1 together, and all households in treatment group 2 together. In particular, a key feature of both treatments is an estimate of the potential monetary savings from improving household energy efficiency and social cues relating to how much energy they consume relative to their neighbours. In the traditional social comparison treatment, households are told how much they would save (or are already saving) on an annual basis if their energy consumption was the same as the average similar household. In the MyHEAT treatment group, households are told how much they could save on an annual basis if they were able to improve their MyHEAT rating to 1 (the best possible rating). In both treatments, households are also provided with normative cues that rank their energy consumption relative to that of their neighbours. Clearly, households that are informed that there are large potential savings from improvements in energy efficiency and that they are large energy consumers relative to their neighbours may respond differently to treatment than households who are told there are small (or negative) savings and that they consume less than

their neighbours.

We evaluate the hypothesis that treatment effects are heterogeneous depending on messaging received by estimating the following equation,

$$Y_{it} = \sum_{g=1}^{3} T_{i1} P_t \theta_i^g + \sum_{h=1}^{3} T_{i2} P_t \theta_i^h + \mu_i + \lambda_t + \epsilon_{it}. \tag{4}$$

where θ_i^g is a set of dummy variables that allocate each household in treatment 1 into one of three groups g corresponding to whether households saw messaging that they were spending more, less or the same amount on their annual utility bills relative to the average of similar households, and θ_i^h is a set of dummy variables that groups households in treatment 2 into low, medium, and high, based on their MyHEAT scores.[20] We report the results from estimating this equation in Figures 4 and 5.

Figure 4 illustrates heterogeneity in the treatment effect for households in the traditional social comparison treatment. 'Negative' denotes customers who were told they were consuming less than average and therefore saving on billing expenditures relative to similar households; 'Zero' denotes customers who were told they were saving zero dollars relative to similar households; and 'Positive' denotes customers who were consuming more than similar households and therefore told they were paying more money than average. For both gas and electricity consumption, we observe that customers who were told they were saving money relative to the average household increased their consumption, whereas customers who were told they were spending more money relative to the average household decreased their consumption.

Figure 5 illustrates heterogeneity in the treatment effect for households in the MyHEAT treatment. Households with the highest MyHEAT scores (worst energy efficiency) respond to treatment by reducing natural gas consumption, whereas households with the lowest MyHEAT scores respond to treatment by increasing gas consumption. A similar pattern is observed for electricity consumption, though the relative changes are small and not statistically significant.

Both of these figures confirm our hypothesis that the treatment effect is heterogeneous within each program, depending on specific messaging received. This is a finding that has been reported in the literature previously, and is sometimes referred to as the "boomerang effect," in which more efficient households actually increase their consumption in response to comparisons with their neighbours that reveal their relative efficiency (Schultz et al., 2007).

Figures 4 and 5 are useful for illustrating the heterogeneity in responses to different messaging in each intervention. However, they don't allow for the two treatments to be directly compared against one another. To do this, we estimate a model in which we interact each

[20] The low group includes MyHEAT ratings 1-3; the medium group includes MyHEAT ratings 4-7; and the high group includes MyHEAT ratings 8-10.

treatment with the dollar value of savings each household was told they could save on energy if they made improvements in their dwelling. In particular, specification (5) controls for the heterogeneity in each household's estimated annual savings by including variable D_{ikm}, the dollar savings estimate household i in treatment group k was shown on both pages one and four of their utlility bill in billing month m, in units of hundreds of dollars. The interpretation of coefficient α_k is the percent reduction in consumption in treatment k per hundred dollars of estimated savings.

$$Y_{it} = \alpha_0 + \sum_{k=1}^{2} \alpha_k D_{ikm} \times T_{ik} \times P_{it} + \mu_i + \lambda_t + \epsilon_{it}. \tag{5}$$

Specification (6) allows for a heterogeneous response to treatment by estimated dollar savings as above, but also allows for a different response for households who were shown zero dollar savings, by incorporating the term $T_{ik} \times P_{it}$.[21] The interpretation of coefficient ψ_k is the percent reduction in consumption in treatment k when dollar savings are zero.

$$Y_{it} = \alpha_0 + \sum_{k=1}^{2} \alpha_k D_{ikm} \times T_{ik} \times P_{it} + \sum_{k=1}^{2} \psi_k \times T_{ik} \times P_{it} + \mu_i + \lambda_t + \epsilon_{it}. \tag{6}$$

We report the results from estimating equations (5) and (6) in Table 5. Columns (1) and (2) estimate the impact on natural gas consumption, and columns (3) and (4) on electricity. Columns (1) and (3) include only the interaction term between dollars and treatment (equation (5)), and columns (2) and (4) include both an interaction term and a main effect of treatment, as in equation (6).

In column (1) of Table 5, we report that the traditional social comparison treatment reduced daily natural gas consumption by 1.3 percent per hundred dollars of estimated savings, and the MyHEAT treatment reduced daily natural gas consumption by about 2.9 percent per hundred dollars of estimated savings. Both of these estimates are significant at the 99 percent level of confidence and the MyHEAT effect is statistically significantly larger than the traditional HER. To put this in context, at the mean estimated MyHEAT annual potential saving of \$150, the traditional HER reduced gas consumption by 2 percent, whereas the MyHEAT treatment reduced gas consumption by 4.4 percent, more than double the traditional social comparison. As shown in the electricity results from column (3) of Table 5, the MyHEAT social comparison also brought about larger, statistically significant reductions in daily electricity consumption per hundred dollars of savings, whereas the traditional HER treatment led to small statistically

[21] The households in treatment 1 who were shown zero potential savings were those with consumption equal (or very close to) the average similar household, whereas in treatment 2 it was households with a heat score of 1.

insignificant reductions.[22]

A similar general pattern is observed in columns (2) and (4) of Table 5, where we estimate specification (6). In column (2), the traditional HER treatment reduced daily natural gas consumption by about 1.3 percent per hundred dollars of estimated savings, and the MyHEAT treatment reduced daily natural gas consumption by 8.1 percent per hundred dollars of estimated savings. The traditional social comparison for customers who were shown savings of zero did not change their gas consumption, however MyHEAT customers who were shown savings of zero (i.e., the most efficient households) exhibited a substantial boomerang effect by increasing their consumption by 11.3 percent. In column (4) of Table 5, the MyHEAT treatment brought about daily electricity reductions of 2.9 percent per hundred dollars of estimated savings, whereas households who were shown savings of zero rebounded by increasing consumption by 4 percent per hundred dollars of savings. The traditional social comparison did not significantly affect electricity consumption.[23]

The column (2) and (4) result discussed above suggest that HER treatments with heat loss messaging and imagery targeting relatively inefficient households hold promise in increasing gas savings relative to traditional HERs. The same is true for electricity, though to a lesser extent. The heterogeneity we observe among households who are informed that they are relatively energy efficient, implying they have less potential to gain from reducing consumption, respond differently from households who are less efficient, is not surprising. As stated above, this type of heterogeneity in the impact of information on energy consumption has been documented in the prior literature (Byrne et al., 2018; Allcott, 2011; Costa and Kahn, 2013).

5.2 Treatment impacts on energy efficiency program participation

In addition to examining impacts of our treatments on gas and electricity consumption, we extend our analysis by estimating whether treatment caused households to participate in other energy efficiency programs. The on-bill treatments (both the traditional social comparison as well as the MyHEAT treatment) were accompanied with information about provincial energy efficiency programs that were available to the household. These on-bill energy efficiency program suggestions are aimed at improving the energy efficiency of the household by providing investment subsidies for high-efficiency windows and insulation. As a result, participation in these programs should be expected to induce an on-going reduction in energy consumption.

We evaluate the impact on energy efficiency program participation using administrative data on program participation rates from Energy Efficiency Alberta. The provincial agency provided

[22] For both gas and electricity, an F-test reveals that the difference between the HER and MyHEAT treatment is statistically significant at the 99 percent confidence level.

[23] In appendix B we show that our results from Tables 4 and 5 are not affected by controlling for differential pre-treatment consumption trends. We also show that there is no significant heterogeneity based on messaging indicating positive or negative expenditures relative to the average in the traditional social comparison treatment

us with participation information for all households in Medicine Hat, linked to the treatment groups by matching on addresses.[24] Program participation information is from spring 2018 through early 2019, and therefore captures applications up to one year after the treatment is initiated. We analyze the results using a regression-based approach, in which we regress a program participation dummy on a treatment indicator. We focus on Home Improvement grants, since these are the types promoted by the on-bill consumer information. Results are provided in Table 6.

Following the approach above, we conduct two analyses: one in which we regress program participation on a treatment dummy variable, and one in which we regress program participation on a treatment dummy variable interacted with the on-bill estimate of savings that was presented to the customer. The second specification is aimed at testing whether consumers that we shown higher estimated savings from pursuing an energy efficiency project were more likely to pursue such a project.

The first column evaluates the impact of treatment on participation in the Home (energy) Improvement programs offered by Energy Efficiency Alberta. As indicated in the table, participation in the control group is 2.1%. Participation in Treatment group 1 is 0.03% lower than in the control group, a non-significant difference. Participation in the MyHEAT group is 0.6% percentage points higher, or almost 30% higher (the difference is significant at the 10% level). Thus, there is evidence that the MyHEAT program caused increases in uptake of energy efficiency programs.

Using the linked data, we are also able to determine whether households that received large on-bill estimates of potential savings respond more to treatment than households with smaller estimates of potential savings. As above, we do this by interacting the estimated savings from adopting energy efficiency measures – which consumers were provided with on their bills – with a treatment indicator. Results are provided in column (2) of Table 6. The results suggest that households in treatment 1 are not affected by the amount of savings shown on their bills. In contrast, households in the MyHeat group (treatment group 2) experience higher uptake of energy efficiency measures when they are provided with a larger estimate of energy efficiency potential. Our estimate suggests that each $100 of savings shown to consumers increases the control group program participation rate by 0.3%, a difference that is significant at the 5% level. Again, this provides evidence that the MyHeat treatment provides consumers with more salient or actionable information than the standard home energy report.

[24] An 85% success rate in matching was achieved, such that 15% of energy efficiency program participants could not be matched to a treatment group. Energy Efficiency Alberta informed us customer application data was entered by hand into their database which is likely why the address match rate with municipal addresses in Medicine Hat was below 100%.

6 Conclusion

This paper reports on a randomized controlled trial that compares the effects of two different social comparisons among customers of a natural gas and electric utility in Canada: traditional home energy reports versus infrared imagery and messaging on home heat loss. Both treatments also included personalized messaging on the estimated annual billing expenditures they could save (or were saving) relative to energy efficient homes, from reducing their consumption.

We find that the heat loss treatment led to significantly larger consumption reductions among relatively inefficient households, relative to the traditional home energy report social comparison. The thermal imaging treatment is also more effective at motivating customers to participate in energy efficiency retrofit programs. Compared to the cost of a traditional HER, of one dollar per customer report, typically mailed on a bi-monthly or monthly basis, the annual cost of thermal imaging is only about one dollar per home. This is strongly suggestive evidence that home heat loss imaging, together with framing consumption in terms of heat loss, hold promise for increasing the savings achieved from the home energy reports that have become ubiquitous among utility customers in North America and Europe. While our results have direct implications for the large literature focused on behavioral interventions to improve energy efficiency, our results likely also have relevance for nudges in other domains as well.

References

Allcott, Hunt, "Social Norms and Energy Conservation," *Journal of Public Economics*, 2011, *95* (9-10), 1082–1095.

_ , "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 2015, *130* (3), 1117–1165.

_ **and Judd B. Kessler**, "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *American Economic Journal: Applied Economics*, 2019, *11* (1), 236–276.

_ **and Michael Greenstone**, "Is There An Energy Efficiency Gap?," *Journal of Economic Perspectives*, 2012, *26* (1), 3–28.

_ **and Todd Rogers**, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," *American Economic Review*, 2014, *104* (10), 3003–37.

Ayres, Ian, Sophie Raseman, and Alice Shih, "Evidence From Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage," *The Journal of Law, Economics, and Organization*, 2013, *29* (5), 992–1022.

Benthin, Alida, Paul Slovic, Patricia Moran, Herbert Severson, C.K. Mertz, and Meg Gerrard, "Adolescent Health-Threatening and Health- Enhancing Behaviors: A Study of Word Association and Imagery," *Journal of Adolescent Health*, 1995, *17*, 143–152.

Boomsma, Christine, Julie Goodhew, Steve Goodhew, and Sabine Pahl, "Improving the Visibility of Energy Use in Home Heating in England: Thermal Images and the Role of Visual Tailoring," *Energy Research & Social Science*, 2016, *14*, 111–121.

Brandon, Alec, Paul J. Ferraro, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer, "Do The Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments," 2017. NBER Working Paper 23277.

Byrne, David P., Andrea La Nauze, and Leslie A. Martin, "Tell Me Something I Don't Already Know: Informedness and the Impact of Information Programs," *Review of Economics and Statistics*, 2018, *100* (3), 510–517.

Carroll, Gabriel D, James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick, "Optimal defaults and active decisions," *The Quarterly Journal of Economics*, 2009, *124* (4), 1639–1674.

Center for Climate and Energy Solutions, "U.S. State Greenhouse Gas Emissions Targets," 2019. https://www.c2es.org/document/greenhouse-gas-emissions-targets.

Costa, Dora L and Matthew E Kahn, "Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment," *Journal of the European Economic Association*, 2013, *11* (3), 680–702.

Czibor, Eszter, David Jimenez-Gomez, and John A List, "The Dozen Things Experimental Economists Should Do (More of)," Technical Report, National Bureau of Economic Research 2019.

Damgaard, Mette Trier and Helena Skyt Nielsen, "Nudging in education," *Economics of Education Review*, 2018, *64*, 313–342.

Davuluri, Sruthi, Rene Garcia Franceschini, Christopher Knittel, Chikara Onda, and Kelly Roache, "Machine Learning for Solar Accessibility: Implications for Low-Income Solar Expansion and Profitability," 2019. NBER Working Paper 26178.

DellaVigna, Stefano, "Psychology and Economics: Evidence from the Field," *Journal of Economic Literature*, 2009, *47* (2), 315–372.

Delmas, Magali A., Miriam Fischlein, and Omar I. Asensio, "Information Strategies and Energy Conservation Behavior: A Meta-Analysis of Experimental Studies From 1975 to 2012," *Energy Policy*, 2013, *61*, 729–739.

Drehobl, Ariel and Lauren Ross, "Lifting the High Energy Burden in America's Largest Cities: How Energy Efficiency Can Improve Low Income and Underserved Communities," 2016. American Council for an Energy Efficient Economy.

Environment and Climate Change Canada, "Government of Canada releases emissions projections, showing progress towards climate target," 2019. News Release.

Epstein, Seymour, "Integration of the Cognitive and Psychodynamic Unconscious," *American Psychologist*, 1994, *49* (8), 709–724.

European Commission, "A Clean Planet for all: A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy," 2018. COM(2018) 773 final.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram, "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program," *The Quarterly Journal of Economics*, 2018, *133* (3), 1597–1644.

Goodhew, Julie, Sabine Pahl, Tim Auburn, and Steve Goodhew, "Making Heat Visible: Promoting Energy Conservation Behaviors Through Thermal Imaging," *Environment and Behavior*, 2015, *47* (10), 1059–1088.

Greentech Media, "After Two Years as an Oracle Company, What's Next for Opower?," April 2018. Accessed February 6, 2020.

Hay, Geoffrey J., Christopher Kyle, Bharanidharan Hemachandran, Gang Chen, Mir Mustafizur Rahman, Tak S. Fung, , and Joseph L. Arvai, "Geospatial Technologies to Improve Urban Energy Efficiency," *Remote Sensing*, 2011, *3*, 1380–1405.

Imbens, Guido W and Jeffrey M Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," *Journal of economic literature*, 2009, *47* (1), 5–86.

International Energy Agency, "Energy Efficiency Indicators, available at https://bit.ly/2UEqCdm," 2019.

Johnson, Eric J and Daniel Goldstein, "Do defaults save lives?," *Science*, 2003, *302*, 1338–1339.

Kerr, R. and M. Tondro, "Residential feedback devices and programs: Opportunities for natural gas," 2012. Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy.

Levinson, Arik, "How Much Energy Do Building Energy Codes Save? Evidence from California Houses," *American Economic Review*, 2016, *106* (10), 2867–94.

MacGregor, Donald G., Paul Slovic, David Dreman, and Michael Berry, "Imagery, Affect, and Financial Judgment," *The Journal of Psychology and Financial Markets*, 2000, *1* (2), 104–110.

Natural Resources Canada, "National Energy Use Database, Natural Resources Canada," 2019.

Nisbett, Richard E and Lee Ross, *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, NJ: Prentice Hall, 1980.

Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius, "Normative Social Influence is Underdetected," *Personality and Social Psychology Bulletin*, 2008, *34* (7), 913–923. PMID: 18550863.

Oracle, "Consumers reach nearly 23 TWh of energy savings with Oracle Utilities Opower," February 2019. Oracle Inc. Accessed February 6, 2020.

Papineau, Maya, "Setting the Standard? Evaluating the Cost-Effectiveness of Building Energy Standards," *Energy Economics*, 2017, *64*, 63–76.

Rahman, Mir Mustafizur, Geoffrey J. Hay, I. Couloigner, Bharanidharan Hemachandran, and J. Bailin, "A Comparison of Four Relative Radiometric Normalization (RRN) Techniques for Mosaicing H-Res Multi-temporal Thermal Infrared (TIR) Flight-Lines of a Complex Urban Scene," *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, *106*, 82–94.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius, "The Constructive, Destructive, and Reconstructive Power of Social Norms," *Psychological Science*, 2007, *18* (5), 429–434. PMID: 17576283.

Schwartz, Daniel, Baruch Fischhoff, Tamar Krishnamurti, and Fallaw Sowell, "The Hawthorne Effect and Energy Awareness," *Proceedings of the National Academy of Sciences*, 2013, *110* (38), 15242–15246.

Sheppard, Stephen RJ, Alison Shaw, David Flanders, Sarah Burch, Arnim Wiek, Jeff Carmichael, John Robinson, and Stewart Cohen, "Future Visioning of Local Climate Change: A Framework For Community Engagement and Planning with Scenarios and Visualisation," *Futures*, 2011, *43* (4), 400–412.

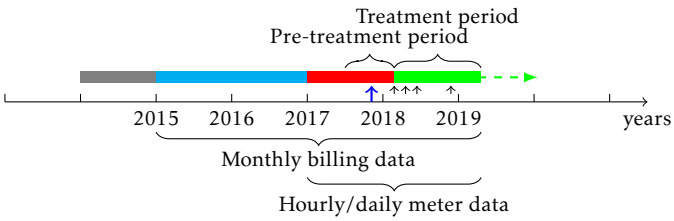
Slovic, Paul, Melissa L. Finucane, Ellen Peters, and Donald G. MacGregor, "The Affect Heuristic," in T. Grilovitch, D. Griffin, and D. Kahneman, eds., *Heuristics and Biases: The Psychology of Intuitive Judgement*, New York, NY: Cambridge University Press, 2002, chapter 23, pp. 397–420.

Smith, Brian Arthur and Lucy Morris, "Neighbor Comparison Reports Produce Savings, But HOW?," 2014. ACEEE Summer Study on Energy Efficiency in Buildings.

Zajonc, R.B., "On the Primacy of Affect.," *American Psychologist*, 1994, *39* (2), 117–123.

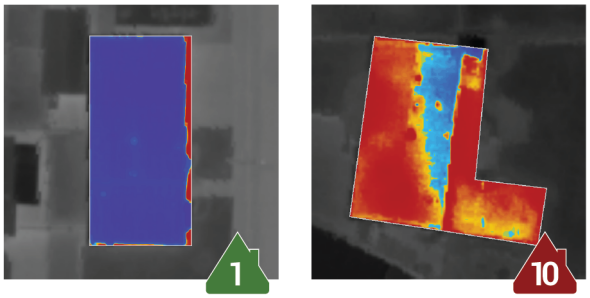
Tables and Figures

Figure 1: Timeline of experiment.



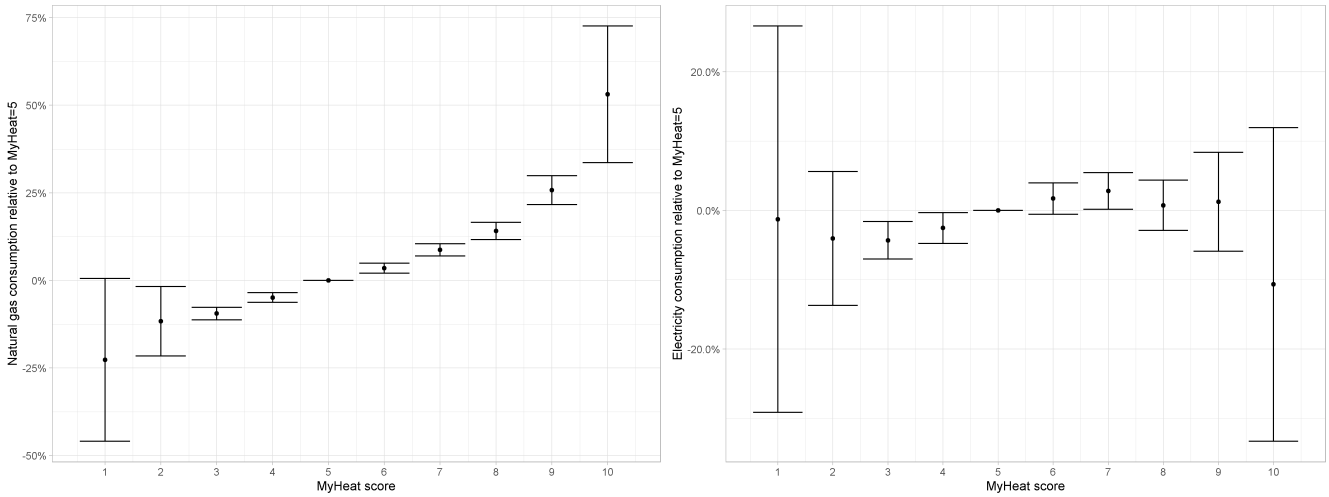
Notes: The small black arrows indicate when bill inserts were included in monthly household energy bills. The larger blue arrow indicates when the thermal images were gathered.

Figure 2: Examples of thermal images for buildings with heat scores of 1 and 10



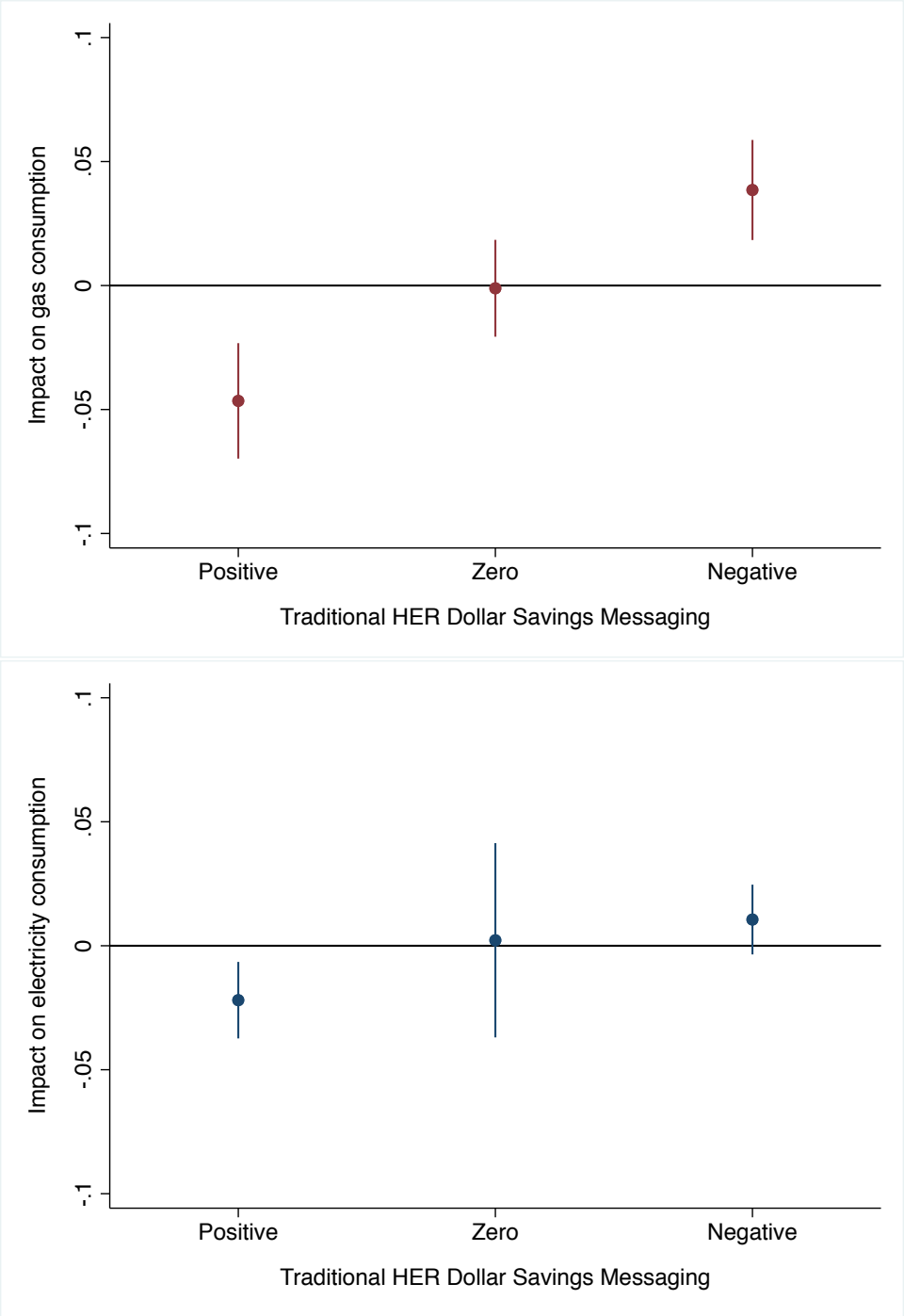
Notes: The image on the left shows a heat score of 1 (lowest heat loss); the image on the right shows a heat score of 10 (highest heat loss). As shown in Appendix Section A.2, the heat loss treatment included a legend explaining how to interpret the images.

Figure 3: Comparison of MyHEAT rating and building energy consumption



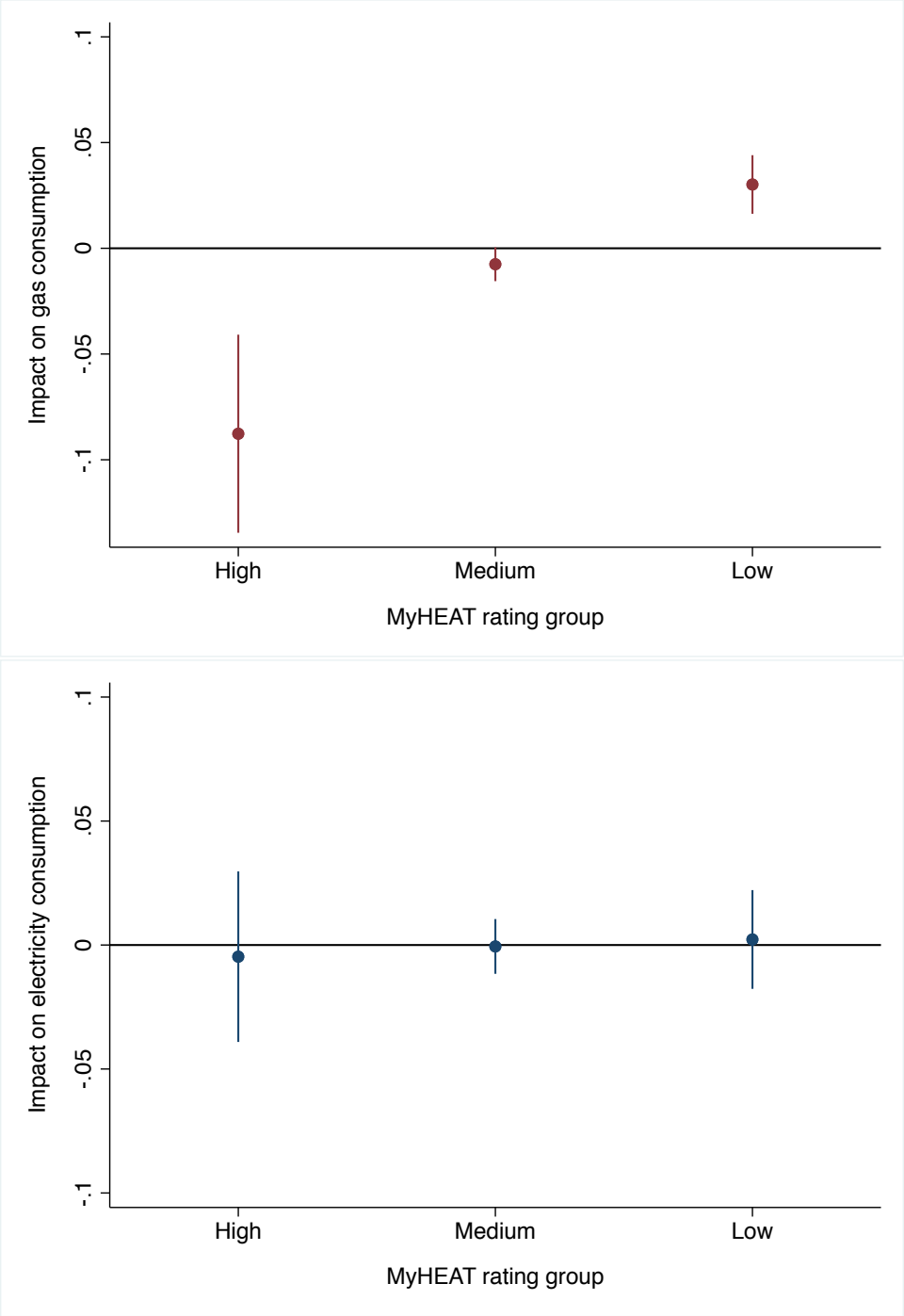
Notes: This figure shows the results of a regression of annual natural gas consumption (left panel) and electricity consumption (right panel) on MyHEAT score, building size, and building type. In each case, the reference category is houses with a MyHEAT rating of 5.

Figure 4: Heterogeneous effects of traditional HER by dollar savings messaging.



Notes: Point estimates of the effect on consumption and confidence intervals for the treatment effect across traditional HER dollar savings messaging. 'Negative' denotes customers who were told they were consuming less and therefore saving on billing expenditures relative to an average similar household; 'Zero' denotes customers who were told they were saving zero dollars relative to similar households; and 'Positive' denotes customers who were consuming more than similar households and therefore told they were paying more money than average.

Figure 5: Heterogeneous effects of MyHEAT treatment according to MyHEAT score.



Notes: Point estimates of the effect on consumption and confidence intervals for the treatment effect across MyHEAT rating groups. 'High' denotes ratings of 8, 9 and 10; 'Medium' denotes ratings of 4,5,6 and 7; and 'low' group denotes ratings of 1,2 and 3.

Table 1: Balance statistics

Experimental Groups and Pre-Treament Balance	Control	Treatment	
	0	1	2
Electricity (kWh/day)			
Mean	23.7	23.9	24.2
s.d.	14.9	15.1	15.0
Norm. Diff.		0.01	0.02
Natural gas (mcf/day)			
Mean	0.5	0.5	0.5
s.d.	0.3	0.3	0.3
Norm. Diff.		0.00	0.05
Size (m^2)			
Mean	123	121	123
s.d.	42	42	43
Norm. Diff.		-0.03	0.00
Assessed value			
Mean	304,732	298,691	303,981
s.d.	109,559	109,533	108,837
Norm. Diff.		-0.04	0.00
Year built			
Mean	1982	1980	1979
s.d.	23	23	24
Norm. Diff.		-0.07	-0.11
Heastcore			
Mean	5.2	5.2	5.2
s.d.	1.5	1.5	1.5
Norm. Diff.		0.01	0.01
Number of Households	3,963	4,441	4,003

Notes: The mean and standard deviation for the consumption and hedonic characteristic variables are presented, along with normalized difference balance statistics for each treatment group relative to the control. The normalized difference is a measure of overlap among the covariates in the treated and control samples. A normalized difference less than 0.25 is typically considered good overlap (Imbens and Wooldridge (2009)). Electricity and gas consumption are measured before the treatment was initiated.

Table 2: Estimated relationship between MyHEAT rating and natural gas consumption in the pre-treatment period

Dependent variable:	log(Gas)				
	(1)	(2)	(3)	(4)	(5)
Heat Score	0.043***	0.048***	0.032***	0.034***	0.034***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Building Size	X	X	X	X	X
Building Type		X	X	X	X
Year Built			X	X	X
Neighborhood				X	X
Street Name				X	X
Assessed Value					X
Observations	12,304	12,304	12,304	12,304	12,304
R-squared	0.22	0.33	0.38	0.43	0.43

Notes: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3: Estimated relationship between MyHEAT rating and electricity consumption in the pre-treatment period

Dependent variable:	log(Electricity)				
	(1)	(2)	(3)	(4)	(5)
Heat Score	0.000	0.014***	0.019***	0.018***	0.017***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Building Size	X	X	X	X	X
Building Type		X	X	X	X
Year Built			X	X	X
Neighborhood				X	X
Street Name				X	X
Assessed Value					X
Observations	12,304	12,304	12,304	12,304	12,304
R-squared	0.13	0.17	0.19	0.22	0.22

Notes: Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Regression results by treatment type

Dependent variable:	Daily Gas Use		Daily Electricity Use	
	(1)	(2)	(3)	(4)
Any treatment	-0.006*		-0.003	
T_i x P_{it}	(0.003)		(0.004)	
Traditional Social Comparison		-0.003		-0.005
T_{i1} x P_{it}		(0.003)		(0.005)
MyHeat Social Comparison		-0.008**		-0.000
T_{i2} x P_{it}		(0.004)		(0.005)
Observations	3,476,492	3,476,492	3,476,492	3,476,492
R-squared	0.86	0.86	0.68	0.68

Notes: As explained in the text the dependent variables are normalized by average post-treatment consumption in the control group. Standard errors are two-way clustered by household and day-of-sample, shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5: Regression results by treatment type with heterogeneity by dollar savings

Dependent variable:	Daily Gas Use		Daily Electricity Use	
	(1)	(2)	(3)	(4)
Traditional Social Comparison		-0.000		-0.005
T_{i1} x P_{it}		(0.003)		(0.005)
MyHeat Social Comparison		0.113***		0.042***
T_{i2} x P_{it}		(0.022)		(0.013)
Traditional Social Comparison x Dollar Savings	-0.013***	-0.013***	-0.002	-0.002
D_{i1m} x T_{i1} x P_{it}	(0.004)	(0.004)	(0.003)	(0.003)
MyHeat Social Comparison x Dollar Savings	-0.029***	-0.081***	-0.008**	-0.029***
D_{i2m} x T_{i2} x P_{it}	(0.007)	(0.017)	(0.004)	(0.009)
Observations	3,476,492	3,476,492	3,476,492	3,476,492
R-squared	0.86	0.86	0.69	0.69

Notes: As explained in the text the dependent variables are normalized by average post-treatment consumption in the control group. Standard errors are two-way clustered by household and day-of-sample, shown in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Columns (1) and (3) show estimates of equation (5) and columns (2) and (4) show estimates of equation (6).

Table 6: Impact of treatments on energy efficiency program participation

Dependent variable:	Program participation	
	(1)	(2)
Traditional Social Comparison	-0.0003	
	(0.003)	
MyHEAT Social Comparison	0.006*	
	(0.003)	
Traditional Social Comparison x Dollar Savings		-0.0001
		(0.001)
MyHEAT Social Comparison x Dollar Savings		0.003**
		(0.002)
Constant	0.021***	0.022***
	(0.002)	(0.002)
Observations	12,407	12,407
R-squared	0.0003	0.0003

Notes: To obtain these results we regress a program participation dummy on a dummy treatment indicator. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

A Sample utility bills with experimental treatments

A.1 Treatment 1

Medicine Hat
The Gas City

Utility Statement February 14 2018
580 1 St SE, Medicine Hat, AB T1A 8E6
customer_accounts@medicinehat.ca
403 529 8111

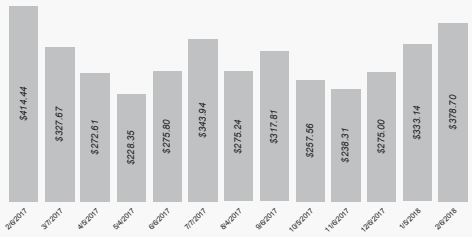
C-10

[redacted]

[redacted]

Bill Period **Jan 06 to Feb 06**

Your billed amounts history:



Knowledge Saves Power

Last month you used **30% more energy** than similar homes. Over the course of a year, your **high** consumption could **cost** you **$390**.

See page 4 for your personalized comparison and options to save energy.

You currently owe 378.70
Automatic withdrawal date Mar 8 2018

Your account activity

Amount on your last bill	333.14
Payment (Feb 8, 2018)	-333.14
Your balance forward	0.00

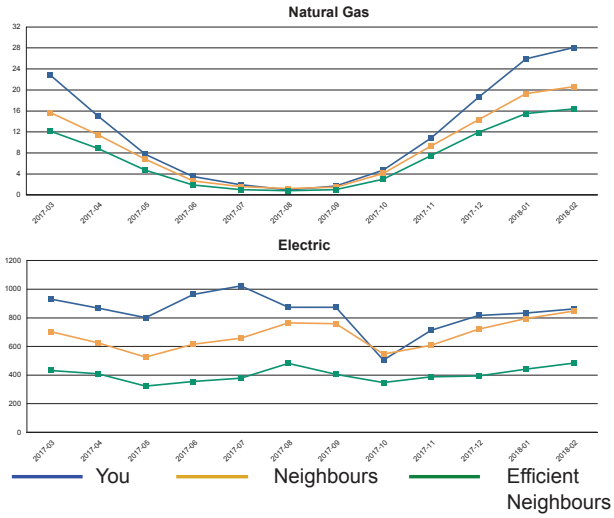
Current Charges

*Electric (862 kwh)	99.56
*Gas (28.05 GJ)	169.79
Water (3.00 CM)	30.79
Sewer	42.18
Solid Waste	22.91
*GST(Registration 121408967 RT0001)	13.47

Total new charges	**378.70**
Total you now owe	**378.70**
After March 8 pay	386.27

Automatic withdrawal date Mar 8 2018

You used 30% more energy than similar homes in Medicine Hat last month.



The graphs above show the comparison of your home's energy usage to similar homes in Medicine Hat. Your Neighbours are others in the city with homes that are similar in size and age. Your energy efficient neighbours are those that fall within the lowest 20% of energy users for similar homes in the city.

Based on your energy usage last year, you could end up spending **$390** more per year on your utility bills when compared to your neighbours.

See a breakdown of your home's energy consumption by signing up for the City's eUtility service.

KNOWLEDGE SAVES POWER

What can you do to save?

Seal Air Leaks	You may be eligible for a rebate of up to $700 from HAT Smart for reducing air leakage in your home.
Turn Down the Heat	Avoid heat loss by simply turning down the heat to 16°C when you leave home.

Learn more at **www.hatsmart.ca**

Upgrade Your Insulation	You may be eligible for a rebate of up to $3,500 from Energy Efficiency Alberta for upgrading insulation in your home.
Install New Windows	You may be eligible for a rebate of up to $1,500 from Energy Efficiency Alberta for switching to efficient windows.

Learn more at **www.efficiencyalberta.ca**

For more information on the Knowledge Saves Power project, visit **www.hatsmart.ca** or call **403.502.8799**.

A.2 Treatment 2

Medicine Hat
The Gas City

Utility Statement February 14 2018
580 1 St SE, Medicine Hat, AB T1A 8E6
customer_accounts@medicinehat.ca
403 529 8111

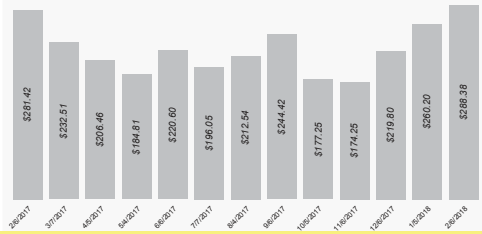
C-10

Utility bill for:

[redacted]

Bill Period **Jan 06 to Feb 06**

Your billed amounts history:



Knowledge Saves Power

Your home's heat loss rate is **average.** You could **save $125** per year on your bills by improving this score.

See page 4 for your personalized comparison and options to save energy.

You currently owe 288.38
Please pay by March 13 2018

Your account activity

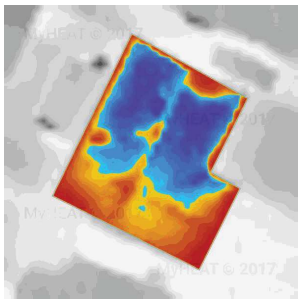
Amount on your last bill	260.20
Payment (Feb 1, 2018)	-260.20
Your balance forward	0.00

Current Charges

*Electric (518 kwh)	64.46
*Gas (18.40 GJ)	112.85
Water (9.00 CM)	37.12
Sewer	42.18
Solid Waste	22.91
*GST(Registration 121408967 RT0001)	8.86
Total new charges	**288.38**

Total you now owe	**288.38**
After March 13 pay	294.15

Your home has a medium heat loss rate with a score of 5/10



Low High Low High

The lower the rating, the less heat is leaving your home. You could save **$125** per year on your bills by lowering this score.

The thermal image was taken of your home's roof using an infrared camera in fall 2017. This image can help you identify air leaks that may be wasting energy in your home and resulting in higher bills.

Red areas on your heat map show potential heat loss and can be improved with simple weatherization techniques.

For more information on your home's MyHeat score, visit **www.myheat.ca/thehat/EJMDXA.**

KNOWLEDGE SAVES POWER

What can you do to save?

Seal Air Leaks	You may be eligible for a rebate of up to $700 from HAT Smart for reducing air leakage in your home.
Turn Down the Heat	Avoid heat loss by simply turning down the heat to 16°C when you leave home.

Learn more at **www.hatsmart.ca**

Upgrade Your Insulation	You may be eligible for a rebate of up to $3,500 from Energy Efficiency Alberta for upgrading insulation in your home.
Install New Windows	You may be eligible for a rebate of up to $1,500 from Energy Efficiency Alberta for switching to efficient windows.

Learn more at **www.efficiencyalberta.ca**

For more information on the Knowledge Saves Power project, visit **www.hatsmart.ca** or call **403.502.8799**.

B Robustness

B.1 Heterogeneous pre-treatment consumption trends

In this section we report a robustness check that consists of evaluating whether pretreatment household energy use levels are a confounding variable in our analysis. We do this by estimating equations (7) and (8), which incorporate day-of-sample fixed effects that vary by low, medium, and high heat score categories:

$$Y_{it} = \beta_0 + \sum_{k=1}^{2} \beta_k T_{ik} \times P_{it} + \mu_i + \sum_{g=1}^{3} \lambda_{gt} + \epsilon_{it}, \tag{7}$$

$$Y_{it} = \alpha_0 + \sum_{k=1}^{2} \alpha_k D_{ikm} \times T_{ik} \times P_{it} + \sum_{k=1}^{2} \psi_k \times T_{ik} \times P_{it} + \mu_i + \sum_{g=1}^{3} \lambda_{gt} + \epsilon_{it}. \tag{8}$$

The low group includes MyHEAT ratings 1-3; the medium group includes MyHEAT ratings 4-7; and the high group includes MyHEAT ratings 8-10. As discussed in Section 3 of the paper, these scores were calculated before the experiment was deployed and are a statistically significant predictor of pre-treatment annual household energy consumption.

The results are reported in Tables A1 and columns (1)-(2) and (4)-(5) of Table A2. The estimated coefficients are very similar to the corresponding main results from Tables 4 and 5 in the paper.

B.2 Heterogeneous response to savings messaging in traditional social comparison

We also evaluate whether we observe heterogeneity based on messaging indicating positive or negative expenditures relative to the average in the traditional social comparison treatment. As described in Section 2, low consumption households in treatment 1 were shown messaging indicating they were saving on billing expenditures relative to an average similar household, whereas high consumption households were told their high consumption resulted in excess billing expenditures.

To test whether customers in this treatment had heterogeneous responses from being told they were saving or paying extra due to their consumption levels, we estimate equation (**??**),

$$Y_{it} = \alpha_0 + \sum_{s=1}^{2} \alpha_s D_{i1sm} \times T_{i1s} \times P_{it} + \alpha_3 D_{i2m} \times T_{i2} \times P_{it} + \sum_{k=1}^{2} \psi_k \times T_{ik} \times P_{it} + \mu_i + \lambda_t + \epsilon_{it}. \quad (9)$$

In this specification, s indexes whether customers in treatment 1 saw excess expenditures or expenditure savings, so that the coefficients α_1 and α_2, respectively, measure the per-dollar effect of seeing each of these two types of messaging. The results are reported in columns (3) and (6) of Table A2. As shown, there is little evidence of heterogeneous response to being told you are spending more than or less than similar households: messaging that you are spending more leads to a 1.4 percent reduction in gas consumption per \$100 of estimated savings, whereas messaging that you are spending less leads to a 1.1 percent increase in gas consumption per \$100 of savings.[25] The results for electricity are insignificant.

[25] These two coefficients are not statistically different.

Table A1: Results with heterogeneity by pre-treatment MyHEAT score

Dependent variable:	Daily Gas Use		Daily Electricity Use	
	(1)	(2)	(3)	(4)
Any treatment	-0.007**		-0.003	
T_i x P_{it}	(0.003)		(0.004)	
Traditional Social Comparison		-0.004		-0.005
T_{i1} x P_{it}		(0.003)		(0.005)
MyHeat Social Comparison		-0.009**		-0.001
T_{i2} x P_{it}		(0.004)		(0.005)
Observations	3,476,492	3,476,492	3,476,492	3,476,492
R-squared	0.86	0.86	0.68	0.68

Notes: To obtain these results we incorporate day-of-sample fixed effects that vary by low, medium, and high heat score categories, as described in Appendix Section B.1. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A2: Results with heterogeneity by pre-treatment MyHEAT score and treatment 1 sign of dollar savings

Dependent variable:	Daily Gas Use			Daily Electricity Use		
	(1)	(2)	(3)	(4)	(5)	(6)
Traditional Social Comparison		-0.001	0.004		-0.005	0.009
		(0.003)	(0.005)		(0.005)	(0.012)
MyHeat Social Comparison		0.101***	0.113***		0.042***	0.042***
		(0.022)	(0.022)		(0.014)	(0.013)
Traditional Social Comparison x Dollar Savings	-0.013***	-0.013***		-0.002	-0.002	
	(0.004)	(0.004)		(0.003)	(0.003)	
MyHeat Social Comparison x Dollar Savings	-0.026***	-0.074***	-0.081***	-0.008*	-0.029***	-0.029***
	(0.006)	(0.016)	(0.017)	(0.004)	(0.010)	(0.009)
Traditional x Extra Cost x Dollar Savings			-0.014***			-0.006
			(0.004)			(0.004)
Traditional x Saving x Dollar Savings			-0.011***			0.006
			(0.003)			(0.007)
Observations	3,476,492	3,476,492	3,476,492	3,476,492	3,476,492	3,476,492
R-squared	0.86	0.86	0.86	0.69	0.69	0.69

Notes: To obtain the results in columns (1)-(2) and (4)-(5) we incorporate day-of-sample fixed effects that vary by low, medium, and high heat score categories. Columns (3) and (6) report estimates of specification (6) from the text, augmented with separate coefficients for treatment 1 participants who were shown messaging reporting billing expenditures they saved relative to the average household versus those who saw messaging reporting extra expenditures relative to the average. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1