Social Norms and Personalized Messaging to Promote Energy Conservation: evidence from a university residence hall

Erica Myers, Mateus Souza
University of Illinois at Urbana Champaign
ecmyers@illinois.edu, nogueir2@illinois.edu

Selected Paper prepared for presentation at the 2018 Agricultural & Applied Economics Association Annual Meeting, Washington, D.C., August 5-August 7

Social Norms and Personalized Messaging to Promote Energy Conservation: evidence from a university residence hall

Erica Myers, Mateus Souza*

May 9, 2018

Abstract

We aim to provide insight about behavioral responses to interventions that promote energy conservation, in a context in which monetary incentives are controlled for. We conducted 2 randomized controlled trials in a university residence hall, to test for the effects of social norms and moral suasion on energy-saving behavior. Our findings can be contrasted with those from standard residential settings, where utilities conduct similar interventions. University residents typically do not see billing or consumption information, and so energy services might not be translated into costs. Further, those residents cannot make physical capital investments to their dwellings. We focus on heating/cooling, and our main outcome of interest are thermostat settings, which we observe at the room-level. For our main intervention, weekly personalized energy reports were emailed to the subjects. Those reports included information about usage own heating/cooling energy usage, as well as usage of average and efficient neighbors. Reports were designed similarly to home energy reports typically provided by utilities. For a second intervention, the emails were simpler, including a suggestion for residents to lower thermostats before leaving the building for winter break. The main intervention produced precise null results, suggesting that energy reports, alone, are not sufficient to promote energy conservation in this context. On the other hand, the intervention prior to winter break was successful in promoting thermostat reduction of 1.1^{o}F, on average. That suggests that inattention is unlikely to have driven the null results of our main intervention.

Key words: behavioral nudges; energy conservation; moral suasion; social norms

*We are thankful for the support from the Levenick iSEE Fellowship.

1 Introduction

Social comparisons have been widely used by water and electric utilities because they have been shown to be remarkably cost effective, reducing energy consumption by 2-4% for the relatively low cost of adding an additional section to the bill (Allcott, 2011; Jessoe et al., 2017). Typically, that additional section would include information about own energy usage, how that compares with neighbors' usage, what energy-saving actions could be taken, and what would be the monetary savings in case usage is reduced. The reasons why those type of reports are successful in promoting conservation, however, are still unclear. A few hypotheses, commonly discussed in the literature (for a review, see: Andor and Fels, 2018), are that the reports could: serve as continuous reminders of the savings opportunities (addressing consumer inattention); appeal to competitiveness and (above average) consumers' desires not to feel out of the norm (social norms); serve as moral suasion (Ito, Ida, and Tanaka, 2018); simply empower subjects with information to act on previously established intention; or could combine all of those and other factors. In this study, we are able to control for the effects of potential monetary savings and physical capital investments in this context, by conducting interventions with subjects who do not directly pay for their energy bills, and who cannot invest in physical capital improvements in their dwellings.

Energy costs in a campus dormitory setting are often not well understood or salient for students. On-campus residents often do not see billing or consumption information, making it difficult to translate use of particular energy services into costs. As a result, energy consumption is often "out of sight, out of mind" as students go through their busy days. Furthermore, being tenants in university housing, these students have little choice in terms of physical capital investments in their apartments. Our research aims to estimate if energy reports including social comparisons and moral suasion effectively promote conservation in this context. There is evidence that, in university residence halls, public displays of least and most efficient consumers can reduce electricity usage (Delmas and Lessem, 2014). However, our interventions will be personalized private emails, much like the home energy reports that are being used by utilities. We focus on heating/cooling systems, which can account for over 40% of energy consumption in the residential context (EIA, 2009). To the best of our knowledge, we are the first to analyze in detail the behavioral outcome that is ultimately relevant in this context: how consumers interact with their thermostat settings.

We have conducted randomized controlled trials in a university residence hall, which houses close to 400 students, within 320 bedrooms. Our main outcome of interest are the thermostat setpoints in each bedroom, which were recorded at high frequency (every 15 minutes)[1]. For the trials, one third of the rooms were assigned to a control group, while the remaining two-thirds (2 treatment arms) were assigned to treatment. With the thermostat data, it was possible to generate personalized graphs of estimated weekly energy consumption to be provided via email to the students. The graphs also included usage of average and efficient (15th percentile) neighbors, much like the reports provided by utilities.

Results reveal no statistically significant change in behavior of the treatment groups, compared to control. Our estimates are precise, with standard errors around 0.3, which allows us to rule out reductions (or increases) greater than 0.6^{o}F (corresponding to 0.8% of the average setting). That suggests that social norms alone are not sufficient to reduce consumption of individuals in a context in which they do not directly pay for energy. Monetary incentives are lacking in this context, different from a standard residential setting in which individuals receive monthly energy bills. Any energy savings in this context would come from behavioral change alone, since the residents cannot make physical capital investments (for the heating system) in their dorms. Recent literature shows that those capital investments result in persistence of effects, and thus are crucial for energy savings in this context (Allcott and Rogers, 2014; Brandon et al., 2017).

Towards the end of the semester we also ran a second set of (randomized) trials in the same building, with the intention to promote energy conservation during the 2017/2018 winter break. For those, the focus was on moral suasion nudges, as opposed to the social norms. Throughout finals week, emails were sent to half of the residents, asking them to

[1] Considering all data losses and sample restrictions, our dataset consists of almost 3 million thermostat observations, spanning from August-December 2017.

lower their thermostats to 68^{o}F before leaving for break (when they leave the building to spend the holidays with their families).[2] These extra nudges also served as a test for subjects' attention to our message delivery method (emails).

The nudges prior to winter break were successful in promoting a 1.1^{o}F reduction on average setpoints (of the treated group), which is approximately 1.5% of the average setpoint in the building. That significant result suggests that inattention is unlikely to have driven the null results of the Fall semester intervention. Simple moral suasion, therefore, seems to have been more effective than social norms in this context. However, the timing difference between the 2 sets interventions cannot be ignored, and could have potentially driven differences in the results: for the winter break treatment, subjects did not expect to return to their rooms for a long time, which might have decreased the intrinsic value of heating in the dorms.

In the following sections we provide further details about the study. Section 2 describes the research design. Model specifications and regression results are presented in Section 3. Conclusions are outlined in Section 4.

2 Research Design

2.1 Fall Intervention (Social Norms Nudging)

We have conducted randomized controlled trials in a university residence hall, which contains 320 bedrooms.[3] For the intervention, 115 rooms were randomly assigned to a control group, 105 were assigned to treatment arm A, and 100 to treatment arm B.[4] Randomization was done at the suite level.[5] For each room in the building, we have ac-

[2]For building maintenance purposes, 68 degrees was the lowest temperature that the thermostat could be set at.

[3]The building actually has 330 rooms in total, but 7 of those were unoccupied. Furthermore, 3 rooms requested to withdraw from the study.

[4]Prior to initiating the intervention, we calculated minimum detectable effects (MDEs) through simulations of statistical power. Details about the simulations can be found in the Appendix A. With 2 treatment arms, the MDEs were calculated to be around 1%, with control setpoints being the outcome.

[5]There are 4 types of suites in the building: "single-bedroom" suites (33% of sample) consist of 4 single-bedrooms; "double-bedroom" suites (55%) have 2 double-bedrooms; "mixed-bedroom" suites (8%) include one single-bedroom and one double-bedroom; and "special units" (3%) which are a isolated single-bedrooms.

cess to high-frequency data (15-minute intervals) about the thermostat setpoints. With that data, we generated and sent out (via email) personalized graphs of estimated weekly energy consumption. The graphs included own energy usage, usage of the average neighbors, and usage of the 15th percentile most efficient neighbors. If a room had below average estimated consumption, it was assigned a "GOOD" status. If estimated usage was below the 15th percentile, the room was assigned a "GREAT" status. And if estimated usage was above average, then the rooms was rated as "BELOW AVERAGE" in terms of efficiency. Figure 1 below presents a sample of the weekly emails sent to subjects. The design is very similar to the first page of home energy reports studied in Allcott (2011) Allcott and Rogers (2014), and others.

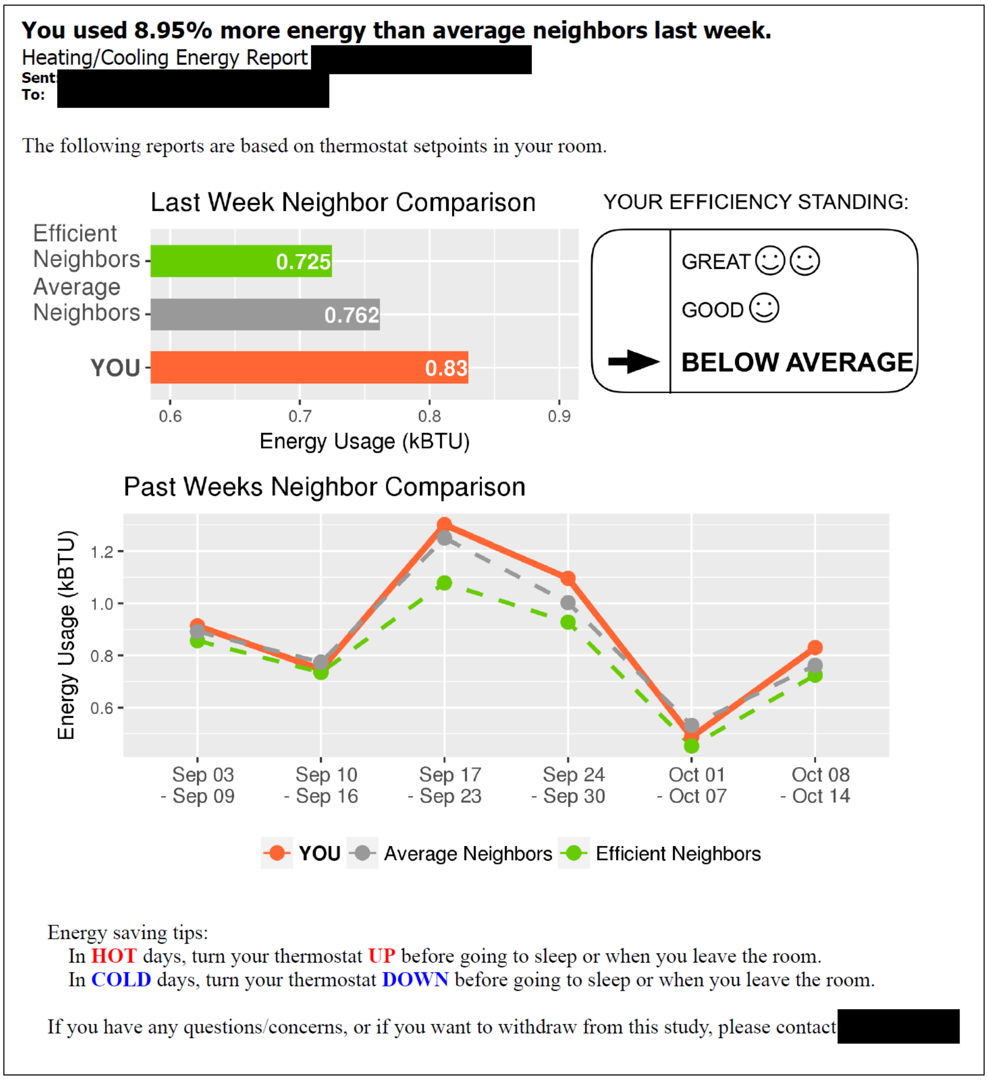

Figure 1: Sample Fall Treatment Email

For treatment arm A, we provided bedroom-level energy usage to the students, while for treatment arm B we provided information aggregated at the suite-level. Treatment B is intended to measure whether any effects dissipate if the information is provided for a group of several rooms rather than for a single room. In other words, how does the level of aggregation for the information provided impact its effectiveness for reducing energy consumption? While we expect information to have the strongest effect at the room-level, individually metering rooms is costly. Therefore, understanding the impact of aggregation can help campus planners interested in using information to reduce consumption with their decisions about how to sub-meter dorms. By comparing outcomes in treatment A to those in the control group, we can estimate the effect of the social norms on residents' behavior with respect to the thermostats. By further comparing treatment A to treatment B, we can estimate if meter aggregation is relevant in this context.

The emails were sent every Wednesdays, around 5pm. September 13th was the first day of treatment, so all data collected prior to that can be considered baseline usage. December 15th will be considered as the last day of this intervention, which is the date when subjects received emails about the winter break treatment (described in the following section).

The following Table 1 compares treated and control groups in terms of average thermostat setpoint before treatment. Some comparisons between suite physical attributes and locations are also presented. It is clear that control and treated groups are balanced in terms of the outcome of interest (setpoints). Some slight imbalance can be noted in terms of where the suites a located. For example, Treatment B suites are slightly more likely to be in the 3rd floor, but less likely to be in the 6th floor of the building. Nevertheless, those observable differences ccan be controlled for in regression specifications, either by explicitly including indicators for each characteristic, or with room fixed effects (which further control for time-invariant unobservable differences between groups).

Table 1: Balance for Fall Intervention

	Control	Treat A	P-value of diff. (Control-Treat A)	Treat B	P-value of diff. (Control-Treat B)
Pre-treat Setpoint Mean	71.1491	71.4204	0.4139	71.1147	0.9204
Double-Bed Suite %	0.4762	0.5227	0.6706	0.6842	0.0614
Mixed-Bed Suite %	0.0476	0.1818	0.0531	0.0000	0.1774
Single-Bed Suite %	0.4048	0.2727	0.1998	0.3158	0.4148
Special Suite %	0.0714	0.0227	0.2892	0.0000	0.0954
1st Floor %	0.1190	0.1364	0.8128	0.0263	0.1188
2nd Floor %	0.1905	0.1818	0.9191	0.1842	0.9437
3rd Floor %	0.2857	0.1818	0.2595	0.0789	0.0177
4th Floor %	0.1905	0.0909	0.1870	0.2632	0.4433
5th Floor %	0.1190	0.2273	0.1904	0.1842	0.4214
6th Floor %	0.0952	0.1818	0.2519	0.2632	0.0492
South Wing %	0.6905	0.5227	0.1144	0.6053	0.4313
West Wing %	0.3095	0.4773	0.1144	0.3947	0.4313
Bottom West Wing %	0.0476	0.0455	0.9626	0.0263	0.6218
Center South Wing %	0.1667	0.1818	0.8553	0.2368	0.4397
Left South Wing %	0.2619	0.2273	0.7126	0.1842	0.4124
Mid West Wing %	0.0952	0.1591	0.3814	0.2368	0.0885
Right South Wing %	0.2619	0.1136	0.0790	0.1842	0.4124
Top West Wing %	0.1667	0.2727	0.2410	0.1316	0.6656
Number of Suites	42	44		38	

2.2 Winter Break Intervention (Moral Suasion)

Prior to winter break, we sent out emails to subjects asking them to lower their thermostats down to 68 degrees. Subjects were also re-randomized and split into two groups: 159 room were assigned to control, and 161 were assigned to treatment. In this case, the randomization was done at the bedroom level (rather than suite-level). The exact wording and images used in the emails sent to the treated subjects can be found in the following Figure 2.

Figure 2: Sample Winter Break Treatment Email

Note that the energy-saving action ("lower your thermostat to 68 degrees") is clearly stated and highlighted. Also, the image and the last sentence of the emails include the word "save," to reinforce the positive/beneficial nature of the requested action. These emails were designed to appeal solely to moral suasion, and differ greatly from the Fall treatment, since subjects are not compared to each other, neither is own usage revealed. The same set of emails were sent out three times (to make the information more salient): 12/15 (Friday), 12/18 (Monday), and 12/20 (Wednesday). The final day of exams was 12/22, and most students were expected to have left the building (for break) by that weekend[6].

[6]Students are not required to leave the building during that period. Some of them may opt to stay for short classes ("winter semester"), or for any other reason.

In Table 2, we assess balance between treatment and control groups for the Winter Break intervention. It can be noted that rooms from the control group had higher pre-treatment setpoints, on average. Therefore, a difference-in-differences approach will be more robust to estimate the causal effect of the treatment in this context. Groups are well balanced in terms of room physical attributes or location. The following section presents results from analyses of our interventions.

Table 2: Balance for Winter Break Intervention

	Control	Treated	P-value of diff. (Control-Treated)
Pre-treat Setpoint Mean	72.4162	71.8175	0.0319
Double-Bedroom %	0.4906	0.4286	0.2673
Single-Bedroom %	0.5094	0.5714	0.2673
1st Floor %	0.1069	0.1242	0.6296
2nd Floor %	0.1509	0.2050	0.2078
3rd Floor %	0.2138	0.1491	0.1335
4th Floor %	0.1887	0.1491	0.3457
5th Floor %	0.1824	0.1677	0.7305
6th Floor %	0.1572	0.2050	0.2691
South Wing %	0.7044	0.6894	0.7718
West Wing %	0.2956	0.3106	0.7718
Bottom West Wing %	0.0189	0.0373	0.3211
Center South Wing %	0.1824	0.1180	0.1075
Left South Wing %	0.2453	0.2919	0.3482
Mid West Wing %	0.1006	0.1491	0.1913
Right South Wing %	0.2767	0.2795	0.9560
Top West Wing %	0.1761	0.1242	0.1950
Number of Rooms	159	161	

3 Model Specifications and Results

3.1 Average treatment effects from the main Fall intervention

We have used the following linear regression specification to test if the Fall intervention affected the students' behavior with respect to thermostat settings:

$$T_{it} = \alpha + \beta_1 treat_{it} + \boldsymbol{\beta}_2 \mathbf{X}_{it} + \varepsilon_{it} \tag{1}$$

where T_{it} is the thermostat set-point in room i and time t; α is a constant term; the indicator $treat_i$ is equal to 1 if room i was assigned to any of the treatment arms, zero otherwise; $\mathbf{X}_{it}$ are exogenous controls which can include room physical attributes and location, date and time fixed effects; and ε_{it} is an idiosyncratic error term.

Alternatively, we consider a difference-in-differences approach, as follows:

$$T_{it} = \beta_1 treat_{it} \times Post_t + \gamma_i + \delta_t + \varepsilon_{it} \tag{2}$$

where $Post_t$ indicates time periods after September 14th; γ_i are room fixed effects; and δ_t are time fixed effects.

We estimate equations (1) and (2) above through Ordinary Least Squares (OLS), with standard errors that are robust to heteroskedasticity and arbitrary autocorrelations within suites.[7] A significant β_1 would indicate that occupants from treated rooms behaved differently than the control group. Results are presented in Table 3. Specification (I) includes no independent variables, other than the treatment indicator. Specification (II) controls for room physical attributes and location (described in Table 1). Specification (III) adds day of sample fixed effects, which controls for daily trends in thermostat settings that are common across rooms. Specification (IV) further controls for finer scale (15-minute interval) common trends. Finally, column (V) presents the difference-in-differences coefficient obtained from estimating equation (2).

[7]We use standard errors clustered by suite for all analyses of the Fall intervention. For all analysis of the Winter Break intervention, we cluster at the room-level. For the sake of brevity, we omit the standard error specifications in some sections of the paper.

Table 3: Pooled Treatment Effect on Thermostat Setpoints

	(I)	(II)	(III)	(IV)	(V)
Control Average Setting	71.61	71.61	71.61	71.61	71.11
Treated=1	0.2236	0.2721	0.2695	0.2695	
	(0.2848)	(0.2773)	(0.2777)	(0.2780)	
Treated=1 × Post Sep.14=1					0.0665
					(0.1560)
Observations	3,090,537	3,090,537	3,090,537	3,090,537	3,090,537
Controls	None	Room Attributes	Day FE	Time FE	Room + Time FE

Standard errors clustered by suite.

All five specifications reveal that there was no significant change in behavior of the treated subjects, when compared to control. Adding date or time fixed effects does not significantly change points estimates (which is expected due to the randomized design of the intervention). The diff-in-diff approach further takes into account that treated rooms had relatively higher pre-intervention setpoints, and thus produces a point estimate that is closer to zero. With relatively small standard errors (close to 0.3, compared to the average setpoint of 71 degrees), it can be argued that the estimates are precise zeros. It is possible to rule out setpoint reductions (or increases) greater than 0.6 degree (0.8% of the control average setpoint).

The following graphs also summarize those results. In Figure 3, we plot average setpoints by date, for treatment and control groups. Within the intervention period, average setpoints for all groups were stable, usually ranging from 71^oF to 72^oF. It can be noted that around November 17th there is a sudden drop in setpoints for both treatment and control groups. That can be attributed to the start of Thanksgiving Break, for which many students might have lowered the settings before leaving the dorms.

The average daily outdoor temperature during the intervention period is shown in Figure 4. Significant variability in outdoor temperature can be noted, in contrast to the slight changes noted in thermostat settings. Outdoor temperature steadily falls after September, which seems to translate into a slight increase in the average setpoints.

Figure 5 presents average setpoints by hour of the day, for treatment and control groups, before and during the intervention period. Again, we note that the setpoints are stable across treatment groups, indicating that residents did not interact much with their

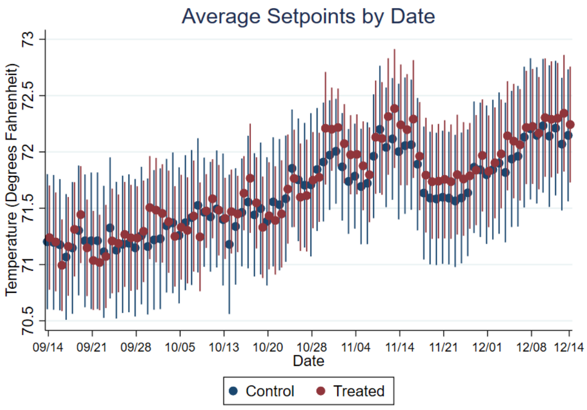

Figure 3: Average Setpoints by Date, for Treatment and Control Groups

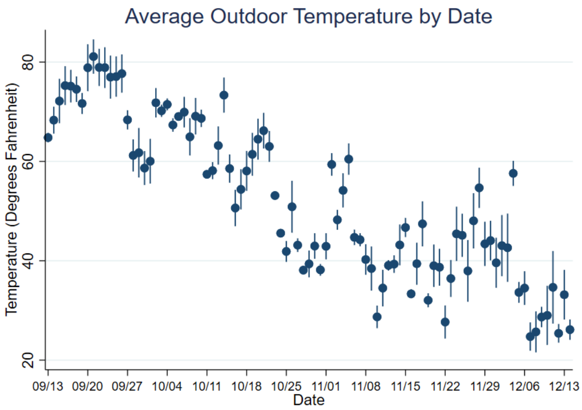

Figure 4: Average Outdoor Temperature by Date[a]

[a]Based on temperatures from the closest weather station, provided NOAA: `https://www.ncdc.noaa.gov/`

thermostats within a day, even though our interventions suggested them to do so. The average outdoor temperature by hour of the day (Figure 6) reveals a standard weather pattern: temperatures start to rise in the morning (when there is sunlight), reaching a peak towards the middle, then start to descend rapidly at night.

We have also tested if treatment arms A (room-level) and B (suite-level) have differential effects by estimating equations (1) and (2) above with separate indicators for each treatment arm, rather than a pooled indicator. Any statistically significant differences between estimated coefficients would indicate that information aggregation is relevant in this context. Results are presented in Table 4, which again reveals null treatment effects. The coefficients of treatments A and B are not statistically significant. The difference in coefficients may just be due to slight imbalance between the groups. For the diff-

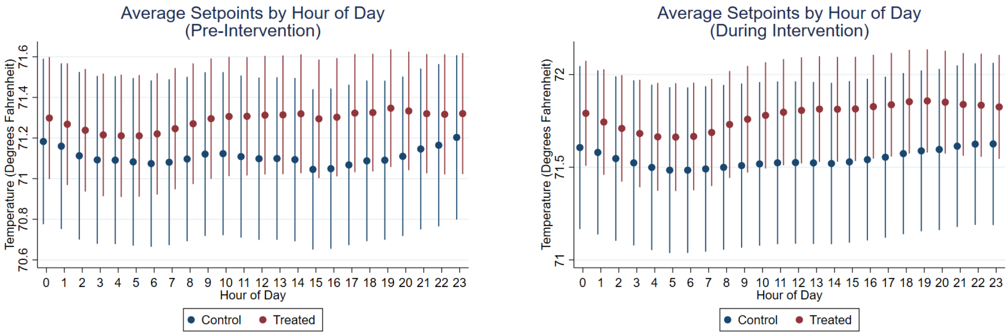

Figure 5: Average Setpoints by Hour of the Day, for Treatment and Control Groups

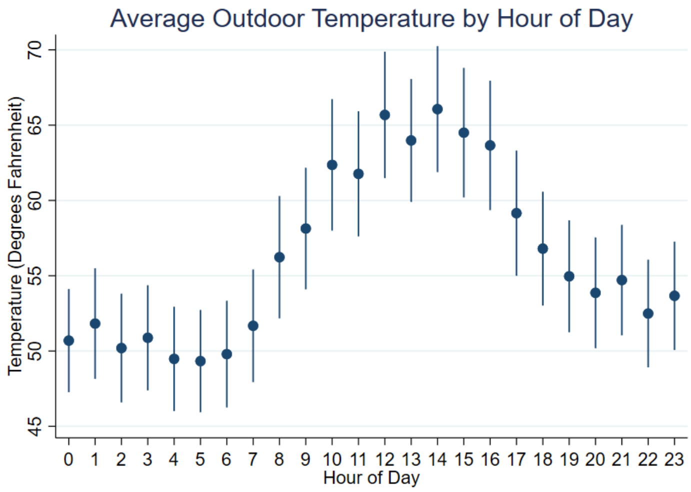

Figure 6: Average Outdoor Temperature by Hour of the Day[a]

[a]Based on temperatures from the closest weather station, provided NOAA: `https://www.ncdc.noaa.gov/`

in-diff specification takes care of those imbalances, and provides suggestive evidence of differences between the groups. In that case, the coefficient on room-level treatment has the expected sign (negative), indicating some (not statistically significant) effort towards lowering thermostats. Further analysis, and robustness checks on results from the main Fall intervention can be found in Appendix B

Table 4: Treatment Effect on Thermostat Setpoints (separated by treatment arms)

	(I)	(II)	(III)	(IV)	(V)
Control Average Setting	71.61	71.61	71.61	71.61	71.11
Room-level Treatment	0.3797	0.3592	0.3646	0.3645	
	(0.3366)	(0.3253)	(0.3254)	(0.3259)	
Suite-level Treatment	0.1072	0.2092	0.1962	0.1963	
	(0.3352)	(0.3343)	(0.3362)	(0.3367)	
Room-level Treatment × Post Sep.14					-0.0477
					(0.1867)
Suite-level Treatment × Post Sep.14					0.1041
					(0.1711)
Observations	3,090,537	3,090,537	3,090,537	3,090,537	3,090,537
Controls	None	Room Attributes	Day FE	Time FE	Room + Time FE

Standard errors clustered by suite.

3.2 Results from the Winter Break intervention

As described in section 2, rooms were re-randomized and exposed to a different treatment starting at December 15th. For this intervention, the treated subjects received moral suasion emails asking them to lower their thermostats before leaving the building for winter break. To test if the intervention had any effect in behavior, we estimate a difference-in-differences model, considering only the month of December:

$$T_{it} = \alpha_1[Post\_Dec15]_{it} + \beta_{wb}[WB\_treat]_{it} \times [Post\_Dec15]_{it} + \gamma_i + \delta_t + \varepsilon_{it} \qquad (3)$$

where T_{it} is the thermostat set-point in room i and time t; the indicator $[Post\_Dec15]_{it}$ is equal to one for all time periods after December 15th, zero otherwise; WB_treat_{it} is equal to 1 if room i was assigned to the Winter Break treatment arm, zero otherwise; γ_i are room fixed effects; and δ_t are time fixed effects. A significant β_{wb} would indicate that occupants from treated rooms behaved differently than control.

Table 5 presents results from the above regression specification. The α_1 estimate reveals that overall rooms lowered their thermostats by approximately 0.64 degree after December 15th. And the WB_treat_{it} estimate shows that treated rooms further lowered their thermostats by 1.17 degree. That corresponds to approximately 1.5% of the average thermostat setpoint in the building.

A graphical analysis showing mean comparisons by date (Figure 7) confirms that result. After the second round of emails, the average thermostats of the treatment group start to steadily decrease. That is around the time when students complete their academic activities for the semester (depending on the classes taken, they might have had exams earlier or later in the week), and thus can leave for the winter vacations. Right after the final day of exams, the average setpoints stabilize, with the treated group averages at a significantly lower level.

Table 5: Results from the Winter Break Treatment

	(1)	(2)
Control Average Setpoint	72.0039	72.0039
WB Treat=1	-0.6431*	
	(0.3700)	
Post Dec. 15=1 × WB Treat=1	-1.1691***	-1.1906***
	(0.2557)	(0.2449)
Observations	620,733	620,733
Room FE	no	yes
Time FE	yes	yes

Standard errors clustered by suite.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

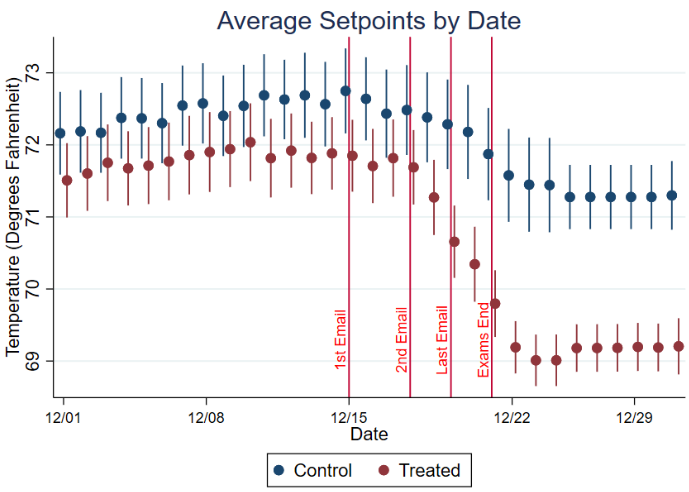

Figure 7: Average Setpoints by Date, for Winter Break Treatment and Control Groups

4 Conclusions and Future Work

Results from the Fall intervention suggest that social norms, by themselves, are not sufficient to promote energy conservation in this context. This contrasts with results from energy reports provided by utilities (in a standard residential setting), which have been shown to promote 2-4% savings. It be could be that lack of monetary incentives discourages subjects to change their behavior. Subjects also cannot make physical capital investments in this context, which limits their ability to change the energy usage patterns in their dorms. Furthermore, we focus on heating/cooling demand, which might be relatively more inelastic (to behavioral nudges), compared to other energy amenities, such as lighting, or plug load. Significant results for the Winter Break (moral suasion) treatment suggests that inattention is unlikely to have driven the null results of the Fall semester intervention. Moral suasion may be more effective in this context because it transmits a more clear message about which actions to take, and how those would be viewed as overall positive/beneficial. It is important to note that the moral suasion emails were sent right before winter break, which is a period when subjects did not expect to return to their rooms for a long time, and thus might be a period of decreased value of the heating amenity. Current ongoing research aims to estimate the effects of moral suasion during the semester (rather than before a break), when subjects are expected to normally occupy the rooms. Results from that intervention, as well as from a post-intervention survey, will be presented in a future version of this paper.

References

Allcott, Hunt (2011). "Social norms and energy conservation". In: *Journal of Public Economics* 95(9), pp. 1082–1095.

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

Andor, Mark A. and Katja M. Fels (2018). "Behavioral Economics and Energy Conservation – A Systematic Review of Non-price Interventions and Their Causal Effects". In: *Ecological Economics* 148, pp. 178–210.

Brandon, A. et al. (2017). "Do The Effects of Social Nudges Persist? Theory and Evidence From 38 Natural Field Experiments". In: *NBER Working Paper Series*(23277).

Delmas, M. and N. Lessem (2014). "Saving power to conserve your reputation? The effectiveness of private versus public information". In: *Journal of Environmental Economics and Management*(67), pp. 353–370.

Ito, Koichiro, Takanori Ida, and Makoto Tanaka (2018). "Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand". In: *American Economic Journal: Economic Policy* 10(1), pp. 240–67.

Jessoe, Katrina et al. (2017). "Spillovers from Behavioral Interventions: Experimental Evidence from Water and Energy Use". In: *Working Paper*. URL: `https://business.illinois.edu/finance/wp-content/uploads/sites/46/2015/01/Paper.pdf`.

US Energy Information Administration (2009). *Residential Energy Consumption Survey*. [Online; accessed in 2017]. URL: `https://www.eia.gov/consumption/residential/`.

Appendix

A Simulations of Statistical Power

In order to determine the number of treatment arms and sample sizes for this study, we ran simulations to assess statistical power and minimum detectable effect (MDE) sizes. For the simulations, we used (high-frequency) historical data about thermostat set-points in the intervention building. Historical data were from up to a year prior to the intervention start date, although, due to collection issues, most available data points were for the months of January through August 2017. It is also important to note that new residents move into the intervention building every academic year. Therefore the consumption data used for power calculations refers to a different set of individuals than the actual subjects of this study. Nevertheless, residents across all years should have similar distributions in terms of demographics, year of enrollment, field of study or other relevant covariates.

We tried simulations considering 2 treatment arms[8], with 1/3 of the suites randomly selected to be in treatment A, 1/3 to be in treatment B, and 1/3 to be in control. We then proceeded to add one degree Fahrenheit to the thermostat set-points of the treated rooms. With varying portions of observations that received the 1 degree change, we simulated different effect sizes. For example, if 50% of observations from treated rooms had a 1 degree change, then the simulated effect size would be of 0.5 degree. We then tried to recover that effect by estimating equation (4), as follows:

$$T_{it} = \beta_{1A} treatA_{it} \times Post_t + \beta_{1B} treatB_{it} \times Post_t + \gamma_i + \delta_t + \varepsilon_{it} \tag{4}$$

where T_{it} are thermostat settings; $treatA$ and $treatB$ are treatment indicators; $Post_t$ indicates time periods after Feb. 1st; γ_i are room fixed effects; and δ_t are time fixed effects. For each of the effect sizes, we ran 100 iterations, re-randomizing the treated

[8] We also tried simulations with 1 and 3 treatment arms. However, for the sake of brevity, we are only presenting results and more details about the specification with 2 treatment arms, which was the chosen design.

suites in each iteration. We stored the estimated coefficients $\beta_1 A$ and $\beta_1 B$, as well as their standard errors (clustered by suite). The following Figure 8 summarizes the simulation results for the effect sizes of 0.6, 0.65, 0.7, and 0.75. For the effect size of 0.75^{o}F (bottom-right panel), we were able to recover (with a 95% confidence interval) the simulated treatment effect in over 80% of simulations. The MDE, therefore corresponds to approximately 1% of the average setpoint in the building.

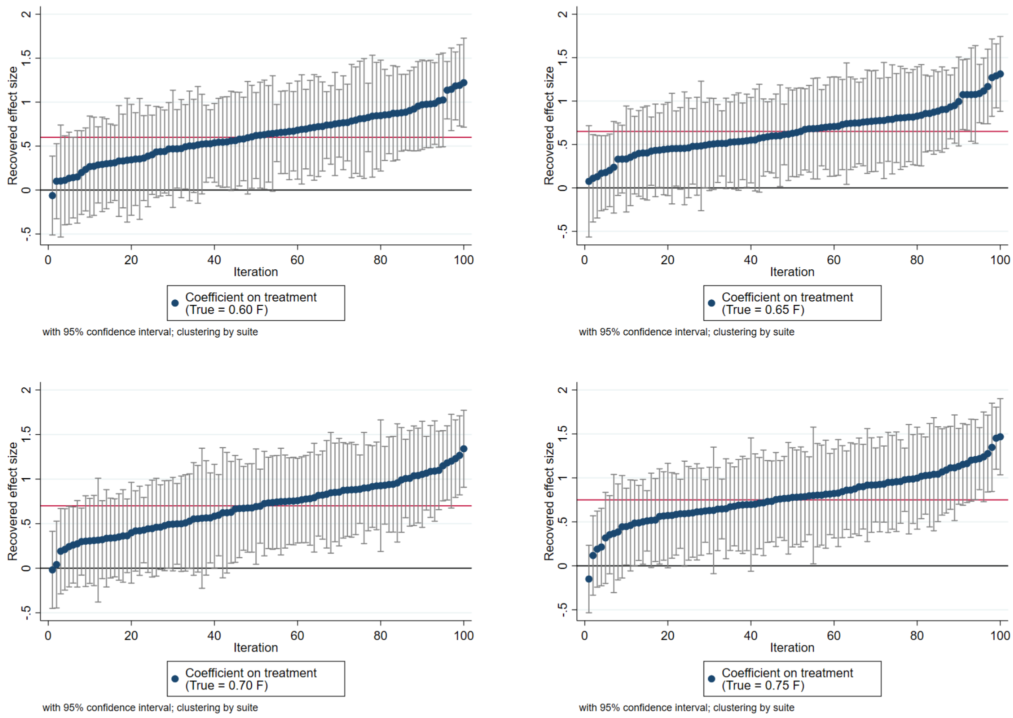

Figure 8: Results from Simulations of Statistical Power

B Robustness Checks for the Fall Intervention

B.1 Treatment effects by hour of the day

In order to test if the Fall intervention had different effects depending on hour of the day, we ran the following specification:

$$T_{it} = \boldsymbol{\beta}_1^h treat_{it} \times Hour_h + \boldsymbol{\beta}_2 \mathbf{X}_{it} + \varepsilon_{it} \tag{5}$$

where T_{it} are thermostat setpoints; $Hour_h$ indicates hour of the day; $\mathbf{X}_{it}$ are exogenous controls which include room physical attributes and location, as well as date fixed effects. The estimated coefficients $\boldsymbol{\beta}_1^h$ are plotted in Figure 9. Effects are not statistically significant across hours of the day, even though a slight reduction in thermostats can be noted during the night/early morning.

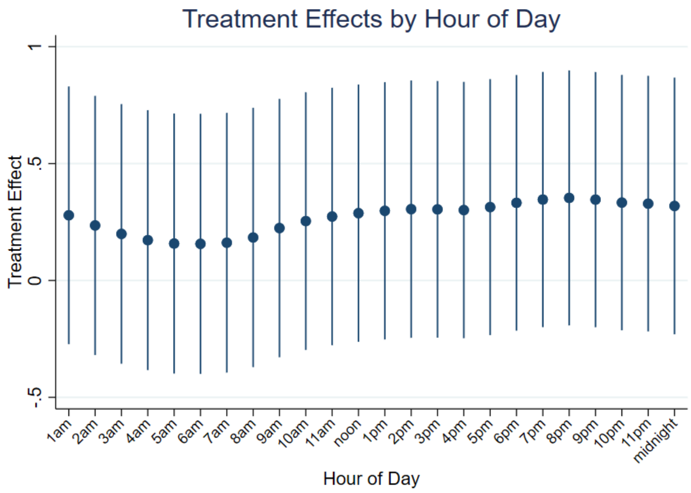

Figure 9: Treatment Effects by Hour of the Day

B.2 Treatment effects by weekday

The emails for the Fall intervention were sent weekly, on Wednesdays. The effect may be expected to be stronger on that weekday, or shortly after, assuming that the emails served as reminders about energy conservation behavior. To test for that, we consider the specification:

$$T_{it} = \boldsymbol{\beta}_1^d treat_{it} \times Weekday_d + \boldsymbol{\beta}_2 \mathbf{X}_{it} + \varepsilon_{it} \tag{6}$$

where T_{it} are thermostat setpoints; $Weekday_d$ indicates weekdays; $\mathbf{X}_{it}$ are exogenous controls which include room physical attributes and location, as well as time fixed effects. Results are presented in Table 6. No statistically significant effects emerge. The point estimate for Mondays seem slightly higher than for other days, but the differences are not significant.

Table 6: Treatment Effects by Weekday

Monday	0.3017
	(0.2800)
Tuesday	0.2693
	(0.2818)
Wednesday	0.2644
	(0.2769)
Thursday	0.2658
	(0.2791)
Friday	0.2654
	(0.2809)
Saturday	0.2652
	(0.2777)
Sunday	0.2528
	(0.2785)
Observations	3,090,537

Standard errors clustered by suite.

B.3 Treatment effects around time of treatment

The usage reports were sent to subjects on Wednesdays, at 5pm. In this section, we assess if there are heterogenous effects close to that time. In Figure 5 from section 3.1, it can be noted that treated rooms have a slightly different (although not statistically significant) pattern of behavior in terms of intraday thermostat settings. That is apparent even for the pre-treatment period. In order to avoid misattributing to treatment any pre-extisitng differences, we use the following triple-difference model:

$$T_{ih} = \boldsymbol{\beta}_1^h treat_{ih} \times Hour_h \times Post\_Treat + \boldsymbol{\beta}_2^h treat_{ih} \times Hour_h + \boldsymbol{\beta}_3^h Post\_Treat \times Hour_h + \beta_4 Post\_Treat \times treat_{ih} + \beta_5 Post\_Treat + \beta_6 treat_{ih} + \boldsymbol{\beta}_7^h Hour_h + \varepsilon_{ih} \quad (7)$$

where T_{ih} is the setpoint for room i in hour h; $treat_{ih}$ indicates if the room was in the treated or control group; $Post\_Treat$ is equal to one after Sept. 14th (first emails sent), zero otherwise; we restrict the sample to hours around treatment time (30 hour bandwidths), such that $Hour_h$ indicates hours that are close to Wednesday's, 5pm; The coefficients of interest are $\boldsymbol{\beta}_1^h$, which will reveal if there are differences in behavior (between treated and control) close to the treatment time, taking into account pre-existing differences.

Figure 10 plots the point estimates of $\boldsymbol{\beta}_1^h$ obtained from equation 7. There is slight evidence that treatment may have induced an increase in setpoints shortly after receipt of emails, followed by reductions 8 hours after that (around 1am). Nevertheless, none of the coefficients are statistically significant, and they are very small in magnitude (less than 0.15% of average setpoint).

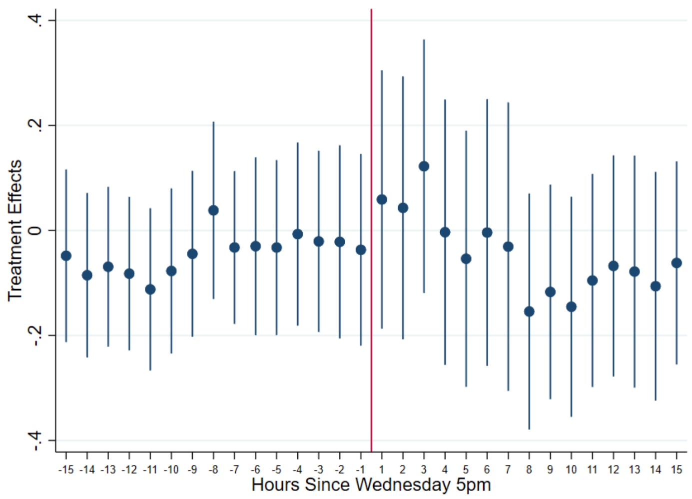

Figure 10: Treatment effects around treatment time (5pm on Wednesdays are the omitted comparison group)