E2e Working Paper 041

Social Comparison Nudges Without Monetary Incentives: Evidence from Home Energy Reports

Erica Myers and Mateus Souza
September 2018

This paper is part of the E2e Project Working Paper Series.

E2e is a joint initiative of the Energy Institute at Haas at the University of California, Berkeley, the Center for Energy and Environmental Policy Research (CEEPR) at the Massachusetts Institute of Technology, and the Energy Policy Institute at Chicago, University of Chicago. E2e is supported by a generous grant from The Alfred P. Sloan Foundation.

Social Comparison Nudges Without Monetary Incentives: Evidence from Home Energy Reports

Erica Myers[*], Mateus Souza[†]

September 25, 2018

Abstract

This paper explores the mechanisms driving the remarkable effectiveness of a widely-used behavioral intervention that reduces energy consumption by repeatedly mailing social comparison-based energy reports to households. We perform a randomized controlled trial of this home energy report (HER) intervention in a new environment, where tenants do not pay energy bills. Our results show that HERs induce almost no behavioral changes for heating demand, with precise estimates that allow us to rule out thermostat (energy) reductions greater than 0.33% (0.1%). We provide evidence that inattention is unlikely to have driven null results, since tenants reacted to simpler nudges from the same sender to conserve energy while they were away for vacation. While, in theory, social comparison nudges could help ameliorate moral hazard, our findings suggest that behavioral channels, such as competitiveness, social norms, or moral suasion, do not motivate conservation in the absence of direct monetary incentives.

Key words: behavioral nudges; energy conservation; moral suasion; social norms
JEL classification: C93, D92, L94, Q41

[*]University of Illinois at Urbana-Champaign, email: ecmyers@illinois.edu
[†]University of Illinois at Urbana-Champaign, email: nogueir2@illinois.edu

We are grateful for the support from the Levenick iSEE Fellowship. We thank Bruce Mikos, Bryan Johnson, Morgan White and all University of Illinois staff who were helpful with data provisioning and overall feedback. We are also thankful for excellent research assistance from Eli Yu, and for helpful comments from Hunt Allcott, Severin Borenstein, Peter Christensen, Tatyana Deryugina, Don Fullerton, Madhu Khanna, Julian Reif, Catherine Wolfram, and seminar participants at the University of Illinois, 2018 Midwest Energy Fest, 2018 Mannheim Energy Conference, and 2018 AAEA Meetings. This research has been pre-registered at the American Economic Association's registry for randomized controlled trials, with ID number AEARCTR-0002398.

1 Introduction

Governments and firms around the world are increasingly relying on policies motivated by behavioral insights to alter consumer and worker choices in ways that might improve welfare. Behavioral nudges or interventions such as information provision, social comparisons, commitment devices, and others have been shown to be effective in diverse contexts: smoking cessation, education, exercise and weight loss, energy and water conservation.[1] Yet, often little is understood about how and why these interventions work. Understanding the underlying mechanisms has important implications for welfare effects as well as the external validity of particular nudges for other settings.

Our study focuses on understanding the mechanisms driving an especially policy relevant behavioral intervention: Home Energy Reports (HERs). HERs provide information about a household's own energy usage, how that compares with neighbors' usage, and estimated monetary savings from several suggested conservation actions. The reports have been shown to be remarkably cost-effective: a simple additional section to consumers' monthly bills produces energy savings that range from 2 to 6% (Allcott, 2011; Ferraro and Price, 2013; Allcott, 2015; Jessoe, Lade, et al., 2017; Allcott and Kessler, 2018; Torres and Carlsson, 2018).

We introduce HERs into a new environment, where tenants do not directly pay for energy. Focusing on this setting allows us to make two contributions to our understanding of social comparison nudges: first, our interventions solely target behavioral channels for reductions, such as competitiveness, social norms, or moral suasion, which could operate independently from direct monetary incentives; second, we can estimate whether it is cost effective to use HERs in situations where tenants make choices about energy consumption, but landlords pay the bills. Approximately 21% of rented residential properties are under landlord-pay contracts for energy (EIA, 2015), and for commercial buildings that figure is close to 20% (Jessoe, Papineau, and Rapson, 2018). If tenants under these types of

[1] See, for example impacts on smoking cessation (K. G. Volpp, Troxel, et al., 2009), education (Leuven, Oosterbeek, and Klaauw, 2010; Levitt, List, and Sadoff, 2016), exercise and weight loss (K. G. Volpp, John, et al., 2008; Charness and Gneezy, 2009; Milkman, Minson, and K. Volpp, 2014; Royer, Stehr, and Sydnor, 2015; Acland and Levy, 2015), and energy and water conservation (Allcott, 2011; Ferraro and Price, 2013).

contracts respond to HERs with consumption reduction, HERs could be an important lever to help ameliorate moral hazard for consumers who otherwise have little incentive to conserve.

Social comparisons like HERs are becoming an increasingly important policy mechanism for addressing energy conservation, where price-based approaches are often difficult to design and implement. In some parts of the world, such as the United States, it has been politically infeasible to implement large-scale Pigouvan taxes or cap-and-trade programs for carbon emissions. While energy efficiency subsidies are used more widely, they are difficult to design optimally without knowledge of elasticities of demand for both energy consumption and efficiency in durables.[2] These difficulties with traditional price-based approaches, combined with the remarkable cost-effectiveness of HERs, have led to widespread use of that intervention. As of mid-2015, leading HER provider, Opower, was working with close to 100 utility companies in 9 different countries, sending regular letters to 15 million households (Allcott and Kessler, 2018).

A few hypotheses have been advanced as to why HERs have been successful in promoting conservation: they may serve as continuous reminders of the monetary savings opportunities (addressing consumer inattention); may appeal to competitiveness and (above average) consumers' desires not to feel out of the norm (social norms); may increase the moral burden of being an above average user (moral suasion, see Ito, Ida, and Tanaka, 2018); may simply empower subjects with information to act on previously established intention; or could combine all of those and other factors.[3]

In order to better understand the mechanisms underlying the effects of HERs, we ran a randomized controlled trial (RCT) in a university residence hall.[4] As in most residences where tenants do not pay for energy, participants in our study could not make

[2]Designing subsidies can be even more complicated in the presence of behavioral factors such as inattention to or biased beliefs about operating costs of energy using durables. To design these policies optimally, governments would need knowledge of the distribution of bias in the population (Allcott, Mullainathan, and Taubinsky, 2014; Farhi and Gabaix, 2017; Houde and Myers, 2018).

[3]For a review, see Andor and Fels (2018).

[4]There is evidence, from previous trials in university residence halls, that public displays of least and most efficient consumers can significantly reduce electricity usage (Delmas and Lessem, 2014). We focus, however, on the effects of HERs. Further, we use an opt-out recruitment design, to avoid selection problems that may arise from voluntary participation in the study.

significant capital investments, which allows us to estimate the effect of HERs in a setting where behavioral channels are the only mechanisms that would affect energy consumption. We had access to high-frequency thermostat setpoint data for each bedroom, which we used to generate weekly personalized heating/cooling energy reports for treatment rooms, designed to closely replicate HERs used by utilities in traditional residential contexts.

With thermostat setpoints as our main outcome of interest, our results reveal no statistically significant change in behavior for the treatment groups compared to control. Our estimates are precise enough to allow us to rule out reductions of the size found in traditional residential contexts. When residents pay for energy, an estimated 45 to 67% of the 2-6% short-run savings come through behavioral channels rather than through capital investments (Brandon et al., 2017).[5] We can rule out reductions greater than 0.33% of the average thermostat setting (close to 0.1% change in energy consumption).[6] These results suggest that HERs do not change behavior in our context.[7] An important caveat is that college students may not be representative of the general population. However, since they tend to have "greener" preferences and place a heightened importance on peers, the effects of competitiveness, social norms, and moral suasion would likely be stronger in this setting, compared to standard residential contexts.

With a separate randomized control trial, we provide evidence that students were reading and reacting to information from our sender, suggesting that our null effect is not being driven by inattention to our HER intervention. At the end of our HER study period, we re-randomized rooms into treatment and control, and sent simpler messages to treatment rooms, asking students to turn down their thermostats to 68^oF before leaving

[5]Brandon et al. (2017) show that 33 to 55% of energy reductions persist after households move and HER treatment stops, such that those savings are attributable to capital investments. Therefore, the remaining 45 to 67% short-run savings come through behavioral channels.

[6]Based on energy consumption simulations for control rooms, using the point estimates from our preferred specification. Further details on the simulations are presented in Online Appendix A.

[7]It may be argued that the effects of HERs are smaller when delivered by email rather than standard mail. However, even if emails are somewhat less effective because people are more likely to read traditional mail, our estimates are precise enough to rule out effects sizes that are even a small fraction of the size found with traditional mail. Further, as we discuss below, it is clear that this population reads and responds to emails from our sender.

for winter break.[8] Those messages came from the same sender as our previous trial, and arrived during finals week, a very busy time for students. Nevertheless, the simple nudges prior to winter break were successful in promoting a 1.1^{o}F setpoint reduction for treated rooms (approximately 1.5% of full sample average setpoint; or 43% of the average within-room, across time setpoint variance during the Fall semester).

The significant treatment effect during winter break suggests that students were still opening and reading emails from our sender at the end of our HERs study period. Further, they were willing to make a change to the thermostat setting, at least in a circumstance where they would not experience thermal discomfort (students were not expected to be in the bedrooms during break).[9] Post-study surveys of students are also suggestive that they were receiving the HERs and reading them, but they did not feel compelled to change their behavior.

These results deepen our understanding of the mechanisms driving the effectiveness of HERs and have important implications for policymakers considering using behavioral instruments for conservation in contexts where tenants do not pay their energy bills. Given their low costs and ease of implementation, simple nudges may be attractive to policymakers for changing behavior in contexts where monetary incentives are not present, such as when tenants do not pay utility bills. However, our findings suggest that social comparisons like those offered in HERs are not effective at changing consumption behavior in those environments.

The following sections provide further details about the study. In Section 2 we present a conceptual framework and derive testable hypotheses. Section 3 describes the research design. Model specifications and regression results are presented in Section 4. We assess robustness of our findings in Section 5. Conclusions are outlined in Section 6.

[8] For building maintenance purposes, 68 degrees was the lowest thermostat setting allowed. The winter break RCT was initially unrelated to our HER study and therefore is not described in our pre-analysis plan. We include a brief description of the set up and findings here because they serve as a test for subjects' attention to our message delivery method (emails).

[9] We ran a third set of trials during the Spring semester, which consisted of simple weekly emails with the same message sent prior to winter break (lower the thermostat to 68). Null results from those trials suggest that it was the special circumstance that students would be away for 4 weeks and not the difference in the content of the simple message, relative HERs, that caused the weekly nudges to be ineffective in our setting. Details from this trial can be found in Online Appendix D.

2 Conceptual Framework

Allcott and Kessler (2018) formalize a theoretical framework for a consumer's maximization problem under the HER "nudge." As a starting point, we restate their first order conditions for utility maximization under the standard residential context and then adapt the model to the case in which we are interested, where tenants do not pay for energy.

Following Allcott and Kessler (2018)'s exposition, consumers are assumed to derive utility from a numeraire good x and from energy use e. The utility obtained from e depends on a function $f(e;\alpha,\gamma)$, with $f' > 0$, $f'' < 0$. Consumer (preference) heterogeneity is captured by a taste parameter α. Behavioral biases, inattention or lack of information are incorporated through γ. Additionally, consumers face a "moral utility" $M = m - \mu e$, which is derived from social pressures or morality surrounding externalities from energy consumption. The parameter m captures any energy-independent (dis)utility generated from the nudges (e.g. subjects may not want to have their mailboxes filled with energy reports). Finally, the parameter μ represents a "moral tax" which results in higher disutility as energy consumption increases. Let $\theta = \{y, p_e, \alpha, \gamma, m, \mu\}$ be a vector of parameters that affect utility. Assuming utility is quasilinear in x, the consumers' maximization problem, for the standard residential context, can be expressed as:

$$\max_{x,e} U(\theta) = x + f(e;\alpha,\gamma) + m - \mu e$$

$$\text{subject to: } y \geq x + e p_e$$

where y is income, and p_e is the price of energy. The optimal choice of energy, denoted e^*, will thus be determined by the first order condition:

$$f'(e;\alpha,\gamma) = \mu + p_e \quad \text{(Standard residential)} \tag{1}$$

such that marginal utility should equal the moral tax plus the price of energy. In standard residential contexts, optimal energy choice e^* will depend on the taste parameter α, behavioral biases γ, the moral tax μ, energy price p_e, and how those are functionally related.

Behavioral nudges implemented through home energy reports are assumed to affect both γ and μ.[10] For example, by providing suggestions on how to save energy, or by simply making energy consumption more salient, the reports are expected to reduce behavioral biases (γ) which stem from inattention or lack of information. Further, by presenting energy usage from average and efficient neighbors, the reports can appeal to social norms and morality (increasing moral tax μ). The effects of changing γ and μ on optimal energy choice e^*, however, cannot be isolated from energy price levels p_e. It is clear that, depending on the functional form of $f(e; \alpha, \gamma)$, the solution of the first order condition (1) presented above will depend on interactions between all the parameters, including p_e.

Our research aims to shed light on the importance of behavioral parameters, rather than energy prices, by conducting randomized controlled trials with subjects that do not directly pay for energy. In our setting, residents face a modified budget constraint $y \leq x + E$, where E is a lump sum payment for energy incurred at the beginning of the housing contract. The first order condition will then be:

$$f'(e; \alpha, \gamma) = \mu \quad \text{(Tenants do not pay energy bills)} \tag{2}$$

such that, when tenants do not pay for energy, energy prices (p_e) no longer affect the optimal energy consumption level e^*. Rather, e^* depends only on the taste parameter α, behavioral biases γ, and the moral tax μ. Although our reports are assumed to directly affect only γ and μ, we also explore the effectiveness of nudges under a scenario in which the energy taste parameter α is also changed: during school breaks, our subjects have less demand for the energy services because they are less likely to be occupying their rooms.

[10] Nudges also affect m, but energy consumption is independent of that.

3 Research Design

3.1 Personalized Energy Reports

We conducted randomized controlled trials in a university residence hall, which houses over 400 undergraduate students. Some of those students (treatment group) were assigned to receive weekly reports, designed to be very similar to HERs typically provided by utilities (Allcott, 2011; Allcott and Rogers, 2014). The main difference is that, as student residents do not receive or pay monthly energy bills, our reports do not include any information about monetary savings. Thus there are no monetary incentives for behavioral changes in this context, as highlighted in Section 2.

During Fall 2017, energy report emails were sent to the treatment group, every Wednesday at 5pm. A sample email is presented in Figure 1 below. Energy usage graphs were created based on thermostat readings from each bedroom in the building.[11] Thermostat data was available at high-frequency (15-minute intervals), which implies significant variation in terms of individual-level energy usage, and allows precision in our estimates.[12] The emailed reports included graphs of a given student's own energy usage, average neighbors' (same bedroom type) usage, and the 15th percentile of neighbors' usage. The reports also included information on students' "efficiency standing," which indicated if they were "GREAT," "GOOD," or "BELOW AVERAGE" based on their energy usage percentile for a given week.[13] This information appeared both in the subject line of the email, and in an "efficiency standing box,"in the body of the email. Recommendations for adjusting thermostats for saving energy appeared in the body of the email as well.

[11] Rooms are equipped with individual fan coil units, which ventilate either hot or cold air (depending on the thermostat setpoint) into the room. Students did not have the option to completely turn off the HVAC system. More details about the building's HVAC system can be found in Online Appendix A.

[12] Estimates of energy usage within each room were based on the differences between outdoor temperature and thermostat setpoints. More details about the energy estimate specification can be found in Online Appendix A.

[13] Usage below the 15th percentile, was categorized as "GREAT," usage below the mean was categorized as "GOOD," and usage above the mean was categorized as "BELOW AVERAGE."

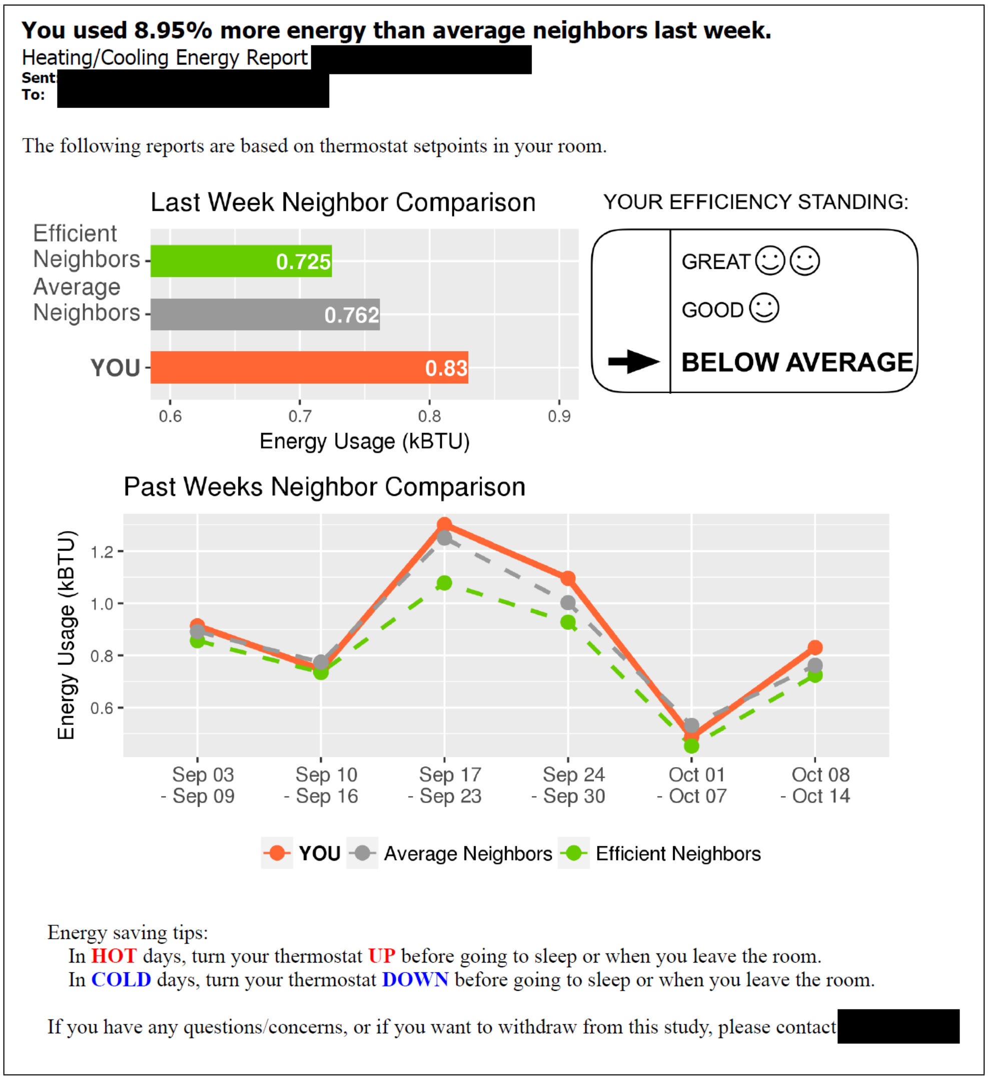

Figure 1: Sample Fall Treatment Energy Report Email

The treatment building is constituted of 330 bedrooms,[14] which are segmented into suites with shared bathrooms and living rooms.[15] Randomization was done at the suite-level (rather than bedroom-level), such that every bedroom from a same suite was assigned to one of three possible groups: control, treatment A, or treatment B.[16]

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

[14]Ten rooms were dropped from the study because they were either unoccupied (7), or asked to withdraw (3).

[15]There are 4 types of suites: "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 isolated single-bedrooms.

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

B was 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 of 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 could help planners of multi-unit buildings with their decisions about how to sub-meter (if energy conservation is a concern).

The first reports were emailed to treatment students on September 13th, so data collected prior to that can be considered the baseline (starting at midnight of August 28, the first day of the academic year, and ending at 5pm of September 13). The last day of this trial was December 15th, which is when subjects received emails about a secondary winter break treatment (which served as an attention check to our emails, described in the following section).

Figure 2 provides some insight about level of thermostat variability that exists in our data, for the 2.5 weeks prior to treatment. For building maintenance purposes, 68^{o}F is generally the lowest setpoint allowed by the system, although in some edge cases 67^{o}F was recorded. It can be noted that thermostats were set at the lowest possible level for 20% of the pre-treatment sample. For the remaining remaining 80%, there is a 10 degree range of variation in settings, with higher concentration around 70-73^{o}F. That implies significant heterogeneity in preferences for heating/cooling demand, even when all subjects are experiencing the same weather.

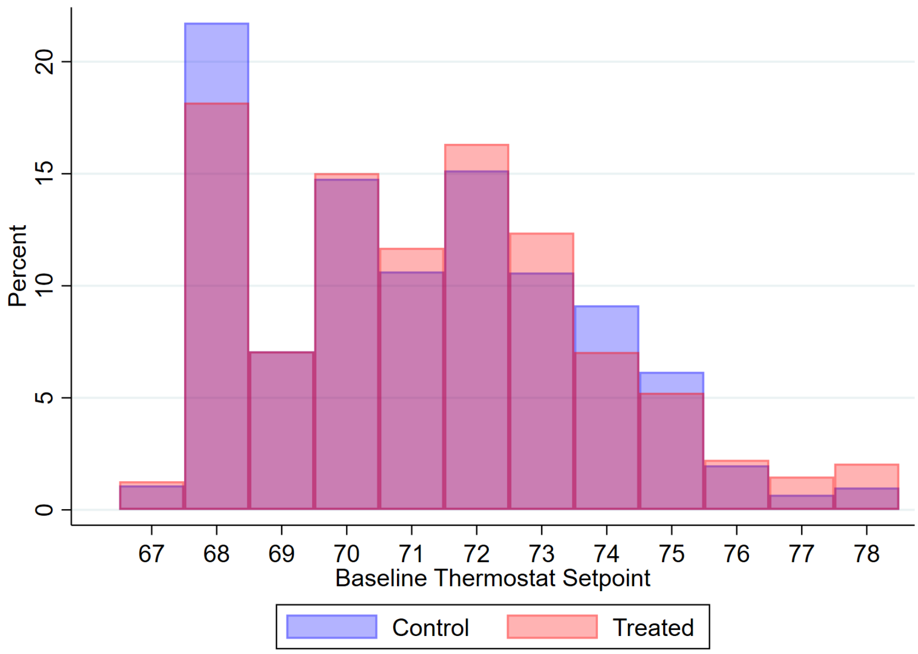

Figure 2: Histogram of Baseline Thermostat Setpoints

The balance table B.1 from Online Appendix B compares treatment and control groups in terms of average thermostat setpoints during the baseline. Some covariate mean comparisons are also presented. It is clear that control and treated groups are balanced in terms of setpoints (outcome of interest) and suite type, as well as occupants' college affiliations, sex, and residency status. Some slight imbalance was noted for rooms' location: Treatment B suites are marginally more likely to be in the 3rd floor, but less likely to be in the 6th floor of the building. Nevertheless, those observable differences can 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).

4 Model Specifications and Results

We use the following linear regression specification to test if our main trial (HERs) affected the students' behavior with respect to thermostat settings:

$$T_{it} = \alpha + \beta_1 treat_i + \boldsymbol{\beta}_2 \mathbf{X}_{it} + \varepsilon_{it} \tag{3}$$

where T_{it} is the thermostat setpoint for room i at 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, weather, date and time fixed effects; and ε_{it} is an idiosyncratic error term.

Given that we have about 2.5 weeks of baseline data, we can also consider a difference-in-differences (DID) approach,[17] as follows:

$$T_{it} = \beta_1 treat_i \times Post_t + \gamma_i + \delta_t + \varepsilon_{it} \tag{4}$$

where $Post_t$ indicates time periods after September 13th (day the first email was sent); γ_i are room fixed effects; and δ_t are time fixed effects.

We estimate equations (3) and (4) above through Ordinary Least Squares (OLS), with standard errors clustered at the suit level to account for any within suite correlations among rooms and across time. A significant β_1 would indicate that occupants from treated rooms behaved differently compared to the control group. Results are presented in Table 1. Specification (I) includes no independent variables, other than the treatment indicator. Specification (II) controls for room physical attributes and location (bedroom type, floor, building wing, and window position). Specification (III) adds hourly weather variables, such as outdoor temperature, wind speed, precipitation and relative humidity.[18] Specification (IV) further controls for fine scale (15-minute interval) time fixed effects,

[17] The DID strategy was not included in our pre-analysis plan. However, given the availability of baseline data, we chose to report DID estimates as well. Given our randomization, we do not expect the DID estimates to qualitatively differ from the post-treatment comparison, though they may improve precision.

[18] Hourly weather data from the closest station was provided by the National Oceanic and Atmospheric Administration (NOAA, 2017).

which capture common trends across rooms. For specification (V), the baseline setpoint averages were added as a control. Finally, column (VI) presents the DID coefficient obtained from estimating equation (4).

Table 1: Pooled Treatment Effect on Thermostat Setpoints

	(I)	(II)	(III)	(IV)	(V)	(VI)
Treated	0.2320	0.2800	0.2791	0.2784	0.0993	
	(0.2908)	(0.2842)	(0.2845)	(0.2849)	(0.1555)	
Treated × Post Sep.13						0.0665
						(0.1560)
Sample Average Setpoint (^{o}F)	71.69	71.69	71.69	71.69	71.69	71.62
Observations	2,591,687	2,591,687	2,564,891	2,591,687	2,591,687	3,090,708
Controls:						
Room physical characteristics	No	Yes	Yes	Yes	Yes	No
Weather	No	No	Yes	No	No	No
Date/Time fixed effects	No	No	No	Yes	Yes	Yes
Avg. pre-treatment setpoint	No	No	No	No	Yes	No
Room fixed effects	No	No	No	No	No	Yes

Note: This table presents estimates of behavior change induced by the weekly energy reports, sent to treated subjects during Fall 2017. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by suite.

All six specifications reveal that there was no significant change in behavior of the treated subjects, when compared to control. Adding time fixed effects does not significantly change points estimates (which is expected due to the randomized design of the trial). Specifications (V) and (VI) are more precise because they further take into account baseline setpoints. With small standard errors (ranging from $0.155 - 0.290$, compared to the average setpoint of 71 degrees), it can be argued that the estimates are precise zeros. With our preferred specification (VI), we can rule out thermostat reductions greater than 0.24^{o}F (lower bound of a 95% confidence interval), which is only 0.33% of the sample average setpoint. That translates into a less than 0.1% change in energy consumption.[19]

[19] Based on simulations of counterfactual energy reductions using control rooms. Further details are presented in Online Appendix A.

The following figures summarize the results graphically. In Figure 3, we plot average setpoints by date, for treatment and control groups. Within the treatment period, there is no statistically distinguishable difference between treatment and control groups. It can further be noted that average setpoints for both groups were slightly increasing, ranging from 71^{o}F to 72^{o}F. That contrasts with the steady decreases in outdoor temperatures observed during that period (presented in Online Appendix E). Around November 17th there is a sudden drop in setpoints for both groups, which can be attributed to the start of Thanksgiving break: many students lowered the thermostats before leaving the dorms for break, even though we did not nudge them to do so.

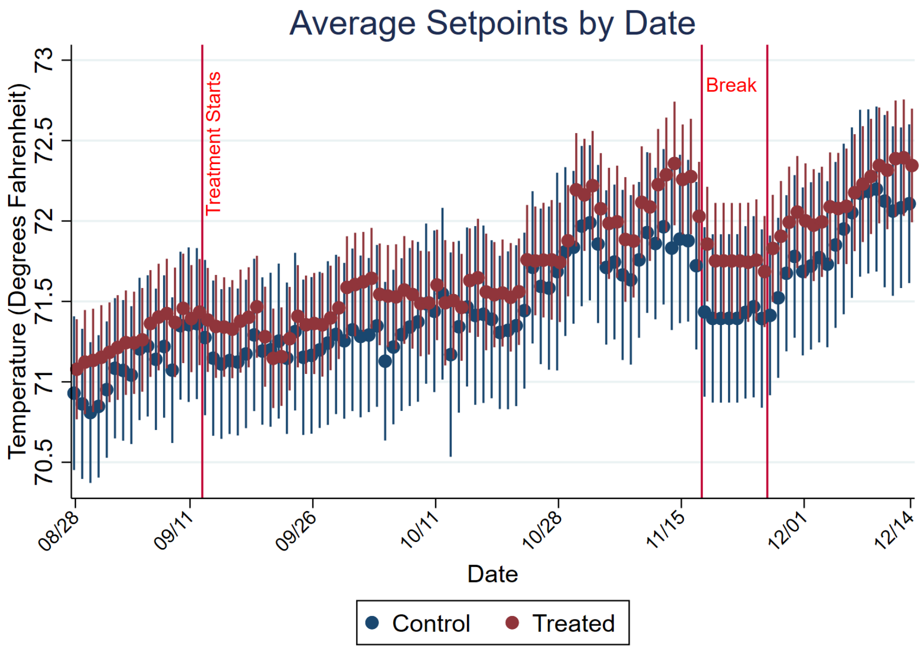

Figure 3: Average Setpoints by Date, for Treatment and Control Groups

Figure 4 presents average setpoints by hour of the day, during the treatment period. Again, we note that the setpoints are stable across both groups groups. There is little evidence of thermostat reductions during late night/early morning, even though our nudges promoted that as an energy-saving behavior. Compared to control, the treatment group seems to be setting the thermostats higher during the middle of the day. That difference is not statistically significant, and does not appear to be attributable to our treatment, since average hourly setpoints during the baseline, pre-intervention period

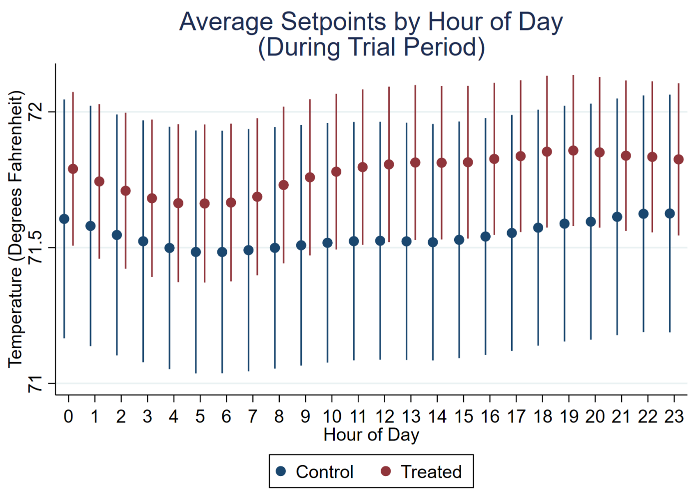

Figure 4: Average Setpoints by Hour of the Day, during treatment period

already revealed that same pattern (Figure 5).

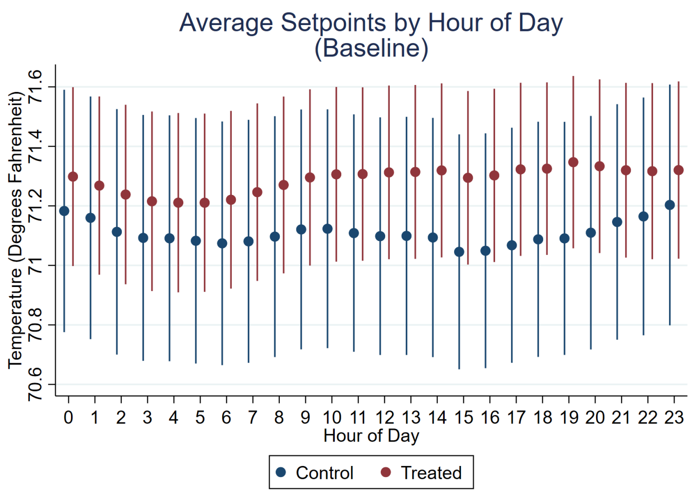

Figure 5: Average Setpoints by Hour of the Day, during baseline

We also tested if effects were different for treatment arm A (room-level), compared to treatment arm B (suite-level), by estimating equations (3) and (4) above with separated (rather than pooled) treatment indicators. Any statistically significant differences between estimated coefficients for each group would indicate that information aggregation is relevant in this context. Results are presented in Table 2, which again reveals precise null treatment effects. The lower bounds of 95% confidence intervals, according to specification (VI), allow us to rule out thermostat reductions greater than 0.413^{o}F (0.58%) and 0.231^{o}F (0.32%), for room-level treatment and suite-level treatment, respectively.

Table 2: Treatment Effect on Thermostat Setpoints (separated by treatment arms)

	(I)	(II)	(III)	(IV)	(V)	(VI)
Control Average Pre-Treatment Setpoint	71.1	71.1	71.1	71.1	71.1	71.1
Room-level Treatment	0.3686	0.3460	0.3504	0.3501	-0.0142	
	(0.3466)	(0.3364)	(0.3366)	(0.3369)	(0.1941)	
Suite-level Treatment	0.1236	0.2249	0.2171	0.2160	0.1449	
	(0.3415)	(0.3429)	(0.3447)	(0.3454)	(0.1800)	
Room-level Treatment $\times$ Post Sep. 13						-0.0477
						(0.1867)
Suite-level Treatment $\times$ Post Sep. 13						0.1041
						(0.1711)
Observations	2,591,687	2,591,687	2,564,891	2,591,687	2,591,687	3,090,708
Controls:						
Room physical characteristics	No	Yes	Yes	Yes	Yes	No
Weather	No	No	Yes	No	No	No
Date/Time fixed effects	No	No	No	Yes	Yes	Yes
Avg. pre-treatment setpoint	No	No	No	No	Yes	No
Room fixed effects	No	No	No	No	No	Yes

Note: This table presents estimates of behavior change induced by the weekly energy reports, sent to treated subjects during Fall 2017. The outcome variable is thermostat setpoints. Indicators for both treatment arms were included in the regressions. Standard errors (in parentheses) are clustered by suite.

5 Robustness Checks

Null results in this type of research may arise if data or contextual limitations prevent us from observing marginal changes in behavior. In this section, we explore potential drivers of our null results. For example, it could be that residents do not pay attention whatsoever to energy-saving nudges provided by emails (our method of information delivery), or do not understand how to operate their thermostats. We rule out those hypotheses with results from simpler conservation nudges emailed prior to winter break, which provide evidence that subjects were willing to lower thermostats before leaving for vacations. Further, a post-treatment survey suggests that students were not blocking/ignoring our emails, and that they were possibly not willing to sacrifice their thermal comfort during the regular semester.

It could also be the case that students predominantly leave temperatures at a default level, set before they moved in. Alternatively, they may choose to leave the thermostats at the lowest (or highest) temperature allowed by the HVAC system. We find no evidence to support those hypotheses by looking at the frequency of students' weekly interactions with thermostats, which suggests that students: believed the thermostats functioned; understood how to use them; and regularly changed temperature settings.

Finally, there might be a concern about spillovers, in a sense that treated and control subjects may interact and discuss the energy reports, or if there are significant heat transfers across rooms/suites. To address that concern, we look at average setpoints for control rooms from a year prior to our treatment. The trends in setpoints for Fall 2016 and Fall 2017 look remarkably similar and are statistically indistinguishable, suggesting that treatment spillovers are unlikely to have occurred.

5.1 Attention check: simple conservation nudges (winter break trial)

Prior to winter break, we sent simple emails to subjects, asking them to lower their thermostats down to 68 degrees. These messages were conceptualized after the

Fall trial had already started, thus they were not included in our pre-analysis plan. Nevertheless, these simple messages serve as a robustness check, providing insight about possible mechanisms driving the results from our main trial. They constitute a change in the timing of treatment (right before an absence from the room during winter break, as opposed to during the semester), as well as an attention check (to test if subjects simply ignore energy conservation messages).

For this secondary trial, subjects were re-randomized and split into two groups: 159 rooms were assigned to control, and 161 were assigned to treatment, with randomization done at the bedroom level (rather than suite level). The exact wording and image included in the emails can be found in the following Figure 6.

Dear [redacted]

Please, remember to **lower your thermostat to 68 degrees** before leaving for Winter Break. This is a very simple task that can **help save a lot of energy!**

Thank You!

Figure 6: 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 act primarily as moral suasion, and differ greatly from the Fall reports, since subjects are not compared to each other, neither is own usage revealed. The same set of emails was 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[20].

We assume that the baseline period for this winter break trial spans from December 1st to December 15th. In Table B.2 from Online Appendix B, we assess baseline balance. We find that rooms from the control group had slightly higher initial setpoints, on average. Therefore, we present results from a difference-in-differences approach as well as a simple comparison of means to estimate the causal effect of the treatment. Groups are well balanced in terms of room physical attributes, location, as well as occupants' sex, college affiliation, and residency status.

To test if the simple nudges had any effect on residents' behavior, we can estimate equations (3) and (4), with the adequate treatment indicator. For this analysis, we restrict the sample to the month of December. For the DID specification, December 15th is the cutoff that determines start of treatment, since that is the date when the first moral suasion email was sent. We cluster standard errors at the level of randomization (room level), to account for any within room correlations over time.

Table 3 presents results from the winter break trial. Specifications (I) through (IV) suggest that the nudges promoted average thermostat reductions ranging from 1.61^{o}F to 1.72^{o}F. However, recall that our analysis of balance (Table B.2) revealed that control rooms already had slightly higher average setpoints prior to the start of this treatment. The more robust specifications (V) and (VI), which control for pre-treatment imbalance, suggest that the average treatment effect is closer to 1.1^{o}F. That corresponds to: approximately 1.5% of the average thermostat setpoint in the building; or 43% of the average within-room, across time setpoint variance during the Fall semester.

[20]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.

Table 3: Average treatment effects from simple conservation messages sent prior to winter break 2017/2018

	(I)	(II)	(III)	(IV)	(V)	(VI)
Winter Break Treatment	-1.7209***	-1.6127***	-1.6240***	-1.6216***	-1.2520***	
	(0.2572)	(0.2712)	(0.2721)	(0.2718)	(0.1872)	
Winter Break Treatment × Post Dec. 15						-1.0786***
						(0.1967)
Sample Average Setpoint (^{o}F)	70.98	70.98	70.98	70.98	70.98	71.50
Observations	508,506	508,506	499,665	508,506	508,506	933,850
Controls:						
Room physical characteristics	No	Yes	Yes	Yes	Yes	No
Weather	No	No	Yes	No	No	No
Date/Time fixed effects	No	No	No	Yes	Yes	Yes
Avg. pre-treatment setpoint	No	No	No	No	Yes	No
Room fixed effects	No	No	No	No	No	Yes

Note: This table presents estimates of behavior change induced by simple nudges asking treated students to lower their thermostats to 68^{o}F prior to leaving for winter break. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by room. Significance at 1% is represented by ***.

A graphical analysis of mean comparisons by date (Figure 7) confirms that result. After the second round of nudging emails, the average thermostats decreased steadily for the treatment group. Around that date, students probably completed their academic activities for the semester,[21] and thus could leave for winter vacation. Shortly after the final day of exams (12/21), the setpoints stabilize, with the treated group averages remaining significantly lower than control. Once the Spring semester started (January 16th), the setpoints for both groups quickly converge back to their pre-treatment levels, indicating that our nudges persisted only through the break period, while students were away.

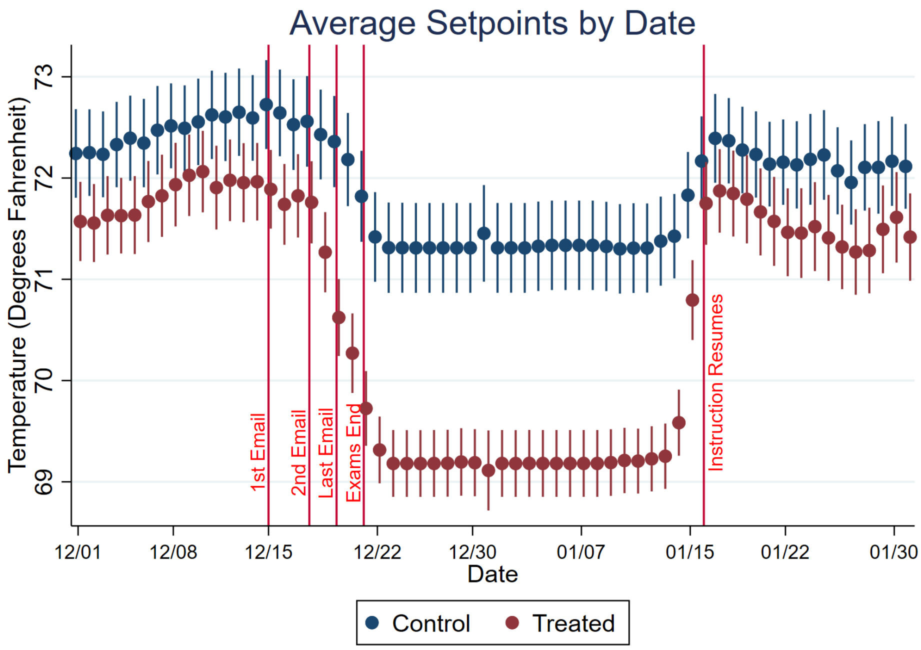

Figure 7: Average Setpoints by Date, for Winter Break Treatment and Control Groups

Given the success of these simple nudges, it can be argued that our method of information delivery (emails) did not drive null results from the Fall trial (i.e. students do pay attention to some forms of emails). It also follows that, compared to the Fall trial, significant results during the winter break could be due to differences in: 1) design of emails; or 2) timing of treatment (right before break as opposed to during the semester). To provide insight about which of those 2 was more important, we ran another randomized

[21] Depending on class different schedules, students might have had exams earlier or later in the week.

trial, during Spring 2018. For the Spring semester, treated students received weekly moral suasion nudges, similar to those sent prior to winter break. Results for the Spring trial were null, suggesting that the timing of treatment is crucial in this context. Subjects were probably more willing to lower thermostats prior to leaving for break, since that would not cause them any thermal discomfort. During a regular semester, however, the students may wish to use their heating more intensively (which constitutes a higher energy taste parameter α from the conceptual framework, Section 2).

5.2 Post-treatment survey

Towards the end of the academic year (May 2018), we sent an online post-treatment survey to any subject who received the weekly energy reports. We obtained valid responses from 82 out of 336 subjects (24.4% response rate).[22] Questions were meant to assess subjects' attention to the reports and how those affected their daily lives.

Students' attentiveness to the energy reports was measured with the questions presented in Figure 8. The vertical red lines in the Figures indicate the mean response. Participants reported that they rarely deleted or blocked our emails without reading their contents. Furthermore, many of them stated that they often opened the report emails, or at least read the subject lines. Thus, most post-treatment survey respondents were appropriately exposed to the relevant energy information.

[22]Responses for the post-survey were incentivized through a lottery. Participants had a 7.5% chance of winning a $25 electronic gift card.

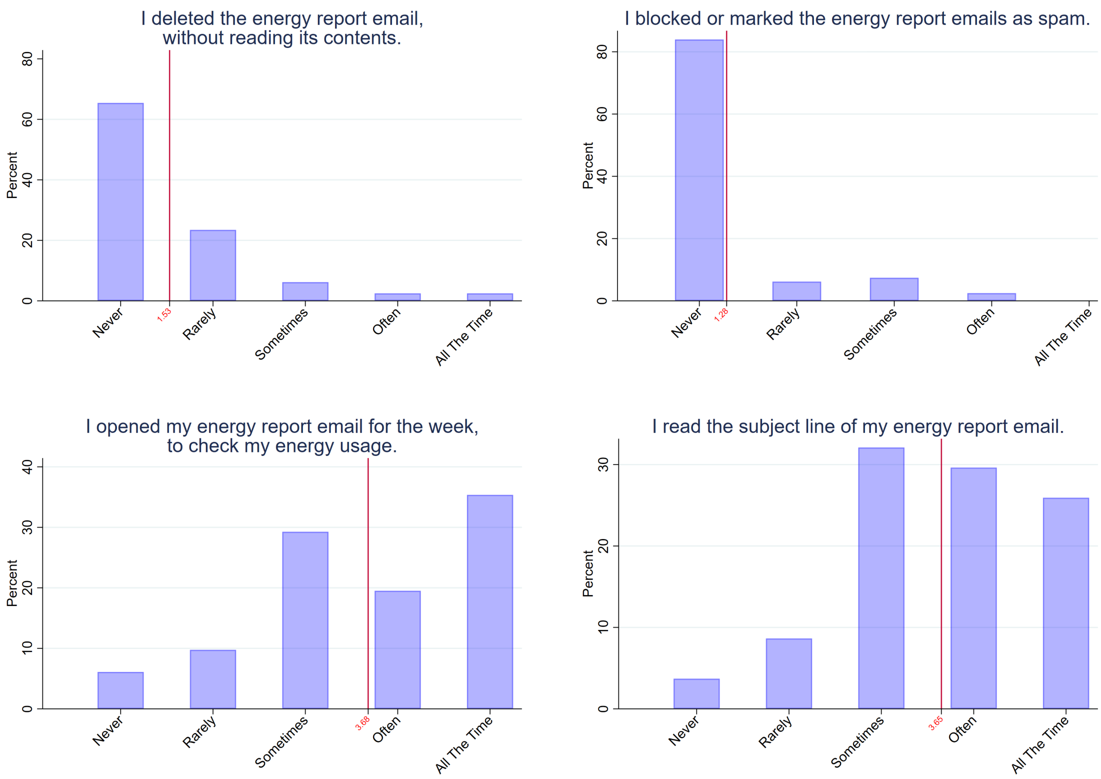

Figure 8: Did students ignore the energy report emails?

Finally, Figure 9 provides suggestive evidence that students were not willing to lower their thermostats (sacrificing thermal comfort) to promote conservation. The post-treatment survey respondents are likely different than non-respondents in important ways, so while these results are supportive of the findings from the winter break treatment, they are not necessarily representative of all participants in the study.

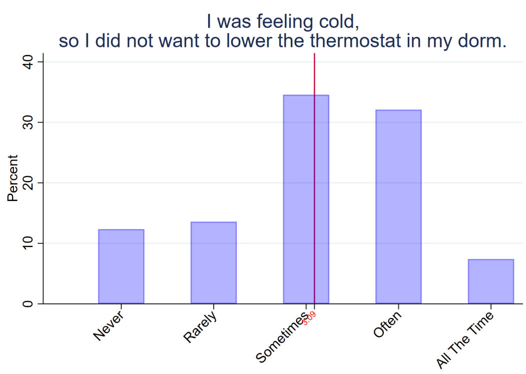

Figure 9: Why didn't subjects lower their thermostats?

5.3 Students' interactions with thermostats

We reiterate that if students do not 1) believe thermostats work, 2) know how to use them, or 3) routinely interact with them, it could be difficult to observe any behavioral changes caused by our nudges, due to small variability in our outcome variable. To further address that concern, Figure 11 below plots the average number of times that thermostat setpoints were changed within a room, for each week of the Fall trial. For both treated and control groups, the averages range from 6 to 11 setpoint changes within a week (except during Thanksgiving break). Most residents, therefore, knew how to use their thermostats and routinely used them to improve comfort, giving them many opportunities to engage in energy-saving behavior. However, they chose not to do so during the Fall semester.

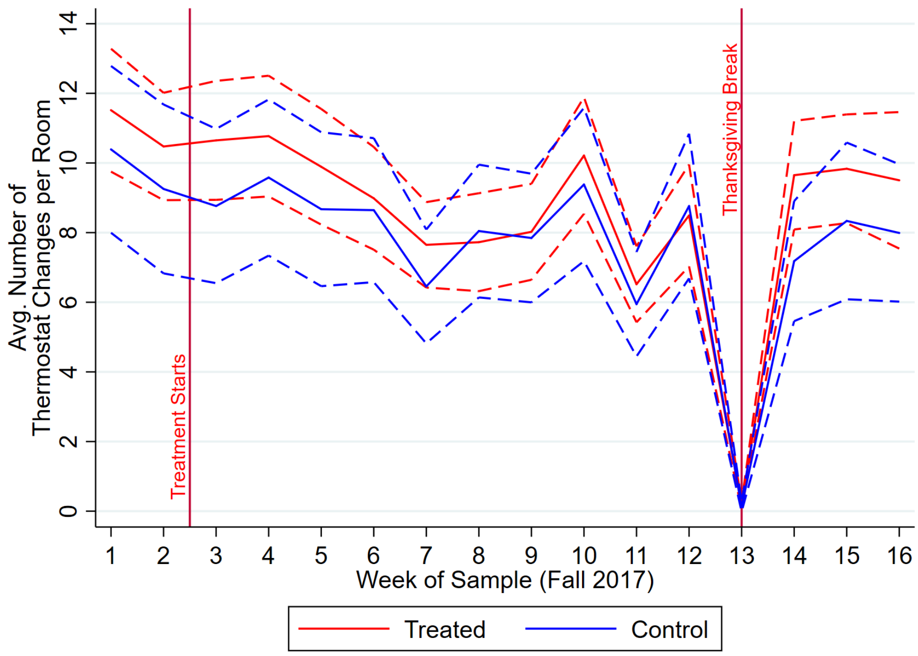

Figure 10: Average Number of Thermostat Changes per Room

We further check if prior year (Fall 2016) residents from treated and control rooms similarly interacted with their thermostats.[23] Figure 10 reveals that in Fall 2016 the average number of weekly thermostat changes ranged between 6 to 12, which is very similar to what we observe during our trial period (Fall 2017). The nudges, therefore, do

[23]For this analysis, we restrict the sample to 128 rooms with available data in Fall 2016. During that period, data was not being recorded for the full building.

not seem to have significantly affected students' engagement with temperature settings in their rooms.

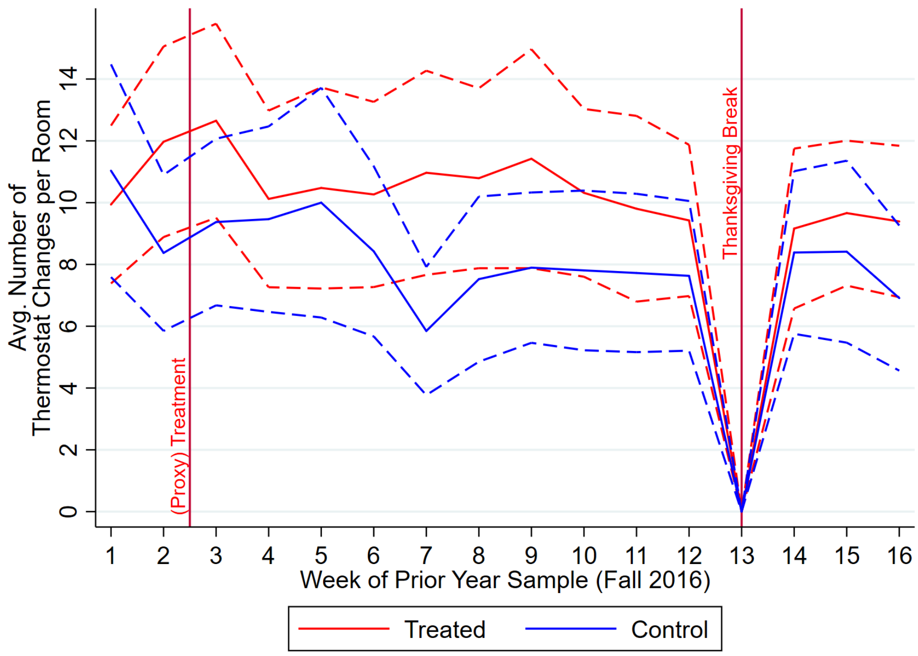

Figure 11: Average Number of Thermostat Changes per Room - 1 Year Prior to Treatment (Fall 2016)

5.4 Spillovers

It may be argued that treated and control subjects had opportunities to interact about the HERs, such that the control subjects were also made aware of the ongoing conservation efforts. If that were the case, control rooms might also feel nudged to lower thermostats, leading to attenuation bias in our estimates. Alternatively, heat exchanges between treated and control rooms may have also provoked changes in behavior of the control subjects. For example, if treated rooms were setting the thermostats lower than normal, then adjacent control rooms may have become colder, leading the control residents to set the thermostat higher. Note that suite-level randomization reduces the chances of both types of spillovers: roommates (more likely to interact) were always assigned to the same group; and there is greater physical separation across suites, compared to across rooms.

The post-intervention survey also included questions to assess if spillovers are likely in this context. Figure 12 presents the distributions of responses, with the red vertical line indicating the mean response. Results reveal that close to 50% of survey respondents rarely or never compared/talked about their energy reports with roommates or other hall residents.

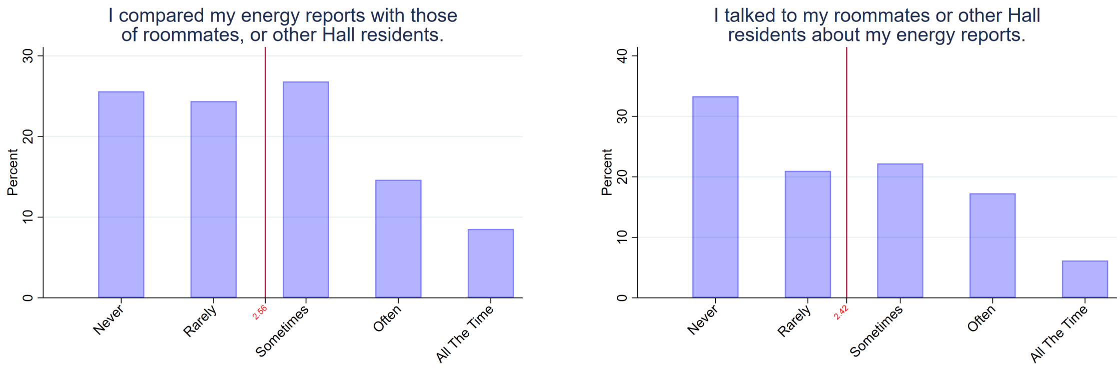

Figure 12: Did students talk to each other about the energy reports?

To test if the control rooms were significantly impacted by the trial, and if their consumption patterns may have been affected, we use data from Fall 2016 (one year prior to our trials). That constitutes an extension of the pre-trial period, with a caveat that data for fewer rooms were available going further back, and residents in 2016 were different from residents in 2017. We restrict our sample to the 72 control rooms with data available for both academic semesters, and graphically inspect the setpoint patterns.

From Figure 13, it is clear that setpoints were lower for control rooms in 2017, compared to 2016. That could be due to differences in weather across years, or simply because the rooms were occupied by different individuals. It is important to note, however, that the setpoint gap between 2016 and 2017 remains the same throughout most dates, including during the pre-trial period.

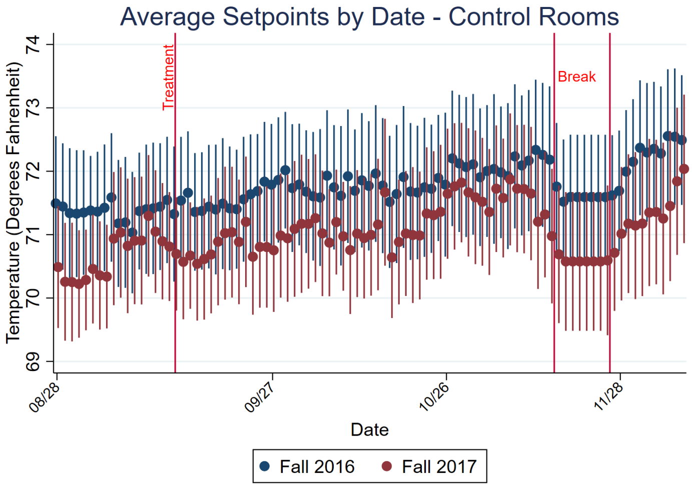

Figure 13: Average Setpoints by Date, for Control Rooms

We also consider the following DID specification, to estimate if control room trends were different in Fall 2017, compared to 2016:

$$T_{it} = \beta_1 Fall17_t + \beta_2 Post_t + \beta_3 Fall17_t \times Post_t + \gamma_i + \varepsilon_{it} \tag{5}$$

where T_{it} are thermostat settings; $Fall17_t$ is equal to one for observations during Fall 2017 semester, and equal to zero during Fall 2016; $Post_t$ indicates time periods after September 13th, which is the (proxy) first day of treatment for Fall (2016) 2017; and γ_i are room fixed effects. For this model, the sample was restricted to control rooms. A significant β_3 would therefore indicate if control room residents behaved differently after the treatment date in Fall 2017, compared to the same period during Fall 2016.

Results from model specification 5 are presented in Table 4. It can be noted that none of the coefficients are statistically significant. Therefore, it is unlikely that control rooms in our sample were affected by our nudges during Fall 2017.

Table 4: Difference-in-differences comparing control room setpoints in Fall 2016 and 2017, pre and post-trial start date

Fall 2017	-0.4504
	(0.5194)
Post Sep. 13	0.2433
	(0.1691)
Fall 2017 × Post Sep. 13	0.1548
	(0.2139)
Sample Average Setpoint (^oF)	71.58
Observations	1,559,540
Room FE	Yes

Note: This table presents results from a DID specification, to assess if control rooms may have been affected by our nudges. The outcome variable is thermostat setpoints. We check if setpoint trends are different in Fall 2017, compared to 2016, before and after the treatment start date (September 13th). Standard errors (in parentheses) are clustered by suite.

Further analyses and some additional concerns not addressed in this Section 5 are presented in Online Appendix F. First, we consider an alternative outcome variable (distance between thermostat setting and outdoor temperature), and confirm that results from our main specification hold. We then analyze heterogeneity based on: hour of the day, weekday, time of treatment, "efficency standing" categorization, and subjects' degree of environmental concern (assessed through a pre-intervention survey). Across all of those dimensions, we find no significant heterogeneity in treatment effects. Finally, we test if long-term exposure to treatment produces changes in behavior, by looking at a subsample of subjects who continued to receive HERs through the Spring semester (rather than Fall only). Again, we find no treatment effects across the sample period.

6 Conclusions

This paper explores the mechanisms driving the effectiveness of HERs in reducing energy consumption. Results from our randomized control trials suggest that HERs have no effect on behavior when tenants do not pay for energy and cannot make capital investments. That contrasts with results from traditional residential settings, for which 45-67% of the 2-6% energy savings from HERs come through behavioral channels (Brandon et al., 2017). In our setting, we can rule out thermostat reductions larger than 0.33% (which implies 0.1% reduction in energy usage), such that the nudge induced almost no behavioral changes. This suggests that the behavioral channels (commonly discussed in the literature as to why HERs have been successful), such as competitiveness, social norms, or moral suasion, do not lead to conservation in absence of monetary incentives.

We report results from a second randomized controlled trial, which reveals that inattention is unlikely to have driven the null results of the HER nudges. The second treatment consisted of simply asking students to turn their thermostats down to 68°F before leaving for winter break. That nudge was successful and resulted in setpoint reductions of 1.5%. It appears students were still opening and reading emails from our sender at the end of the HERs study period, and they were willing to change the thermostat setting. Robustness checks further suggest that issues of statistical power, low variability of our outcome variable, or spillovers are unlikely to have driven the null results from the HER intervention.

It is important to note that the timing of the winter break treatment is relevant in this context, since subjects did not expect to return to their rooms for 4 weeks, and thus the break might be a period of decreased value of the heating amenity. That hypothesis is strengthened by the fact that the same simple conservation messages produced null results when they were sent during Spring 2018 (when subjects were occupying their rooms).

Collectively, these results have important implications for policymakers considering behavioral instruments for conservation. Behavioral nudges may be attractive in contexts where monetary incentives do not exist, such as when tenants do not pay their energy bills. However, we find that the social comparisons in HERs may not be effective at reducing energy consumption in those environments.

References

Acland, Dan and Matthew R. Levy (2015). "Naiveté, Projection Bias, and Habit Formation in Gym Attendance". *Management Science* 61(1), pp. 146–160.

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

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

Allcott, Hunt and Judd Kessler (2018). "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons". *American Economic Journal: Applied Economics* (Forthcoming).

Allcott, Hunt, Sendhil Mullainathan, and Dmitry Taubinsky (2014). "Energy policy with externalities and internalities". *Journal of Public Economics* 112, pp. 72–88.

Allcott, Hunt and Todd Rogers (2014). "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation". *American Economic Review* 104(10), pp. 3003–3037.

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

Brandon, A., P. Ferraro, J. List, R. Metcalfe, M. Price, and F. Rundhammer (2017). "Do The Effects of Social Nudges Persist? Theory and Evidence From 38 Natural Field Experiments". *NBER Working Paper* 23277.

Charness, Gary and Uri Gneezy (2009). "Incentives to Exercise". *Econometrica* 77(3), pp. 909–931.

Clee, Mona A. and Robert A. Wicklund (1980). "Consumer Behavior and Psychological Reactance". *Journal of Consumer Research* 6(4), pp. 389–405.

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

Dunlap, Riley, Kent Van Liere, Angela Mertig, and Robert Jones (2000). "Measuring Endorsement of the New Ecological Paradigm: A Revised NEP Scale". *Journal of Social Issues* 56(3), pp. 425–442.

Farhi, Emmanuel and Xavier Gabaix (2017). "Optimal Taxation with Behavioral Agents". *Working Paper*. Revise and resubmit, American Economic Review.

Ferraro, Paul J. and Michael K. Price (2013). "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-Scale Field Experiment". *The Review of Economics and Statistics* 95(1), pp. 64–73.

Houde, Sebastien and Erica Myers (2018). "Heterogeneous Misperceptions of Energy Costs: Implications for Measurement and Policy Design". *Working Paper*.

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

Jessoe, Katrina, Gabriel E Lade, Frank Loge, and Edward Spang (2017). "Spillovers from Behavioral Interventions: Experimental Evidence from Water and Energy Use". *Working Paper*. URL: `https://business.illinois.edu/finance/wp-content/uploads/sites/46/2015/01/Paper.pdf`.

Jessoe, Katrina, Maya Papineau, and David Rapson (2018). "Utilities Included: Split Incentives in Commercial Electricity Contracts". *E2e Working Paper* 029.

Leuven, Edwin, Hessel Oosterbeek, and Bas van der Klaauw (2010). "The Effect of Financial Rewards on Students' Achievement: Evidence from a randomized experiment". *Journal of the European Economic Association* 8(6), pp. 1243–1265.

Levitt, Steven D, John A List, and Sally Sadoff (2016). "The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment". *NBER Working Paper* 22107.

Milkman, Katherine, Julia Minson, and Kevin Volpp (2014). "Holding the Hunger Games Hostage at the Gym: An Evaluation of Temptation Bundling". *Management Science* 60(2), pp. 283–299.

Pienaar, Elizabeth, Daniel Lew, and Kristy Wallmo (2015). “The importance of survey content: Testing for the context dependency of the New Ecological Paradigm Scale”. *Social Science Research* 51, pp. 338–349.

Royer, Heather, Mark Stehr, and Justin Sydnor (2015). “Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company”. *American Economic Journal: Applied Economics* 7(3), pp. 51–84.

Torres, Mónica M. Jaime and Fredrik Carlsson (2018). “Direct and Spillover Effects of a Social Information Campaign on Residential Water-Savings”. *Journal of Environmental Economics and Management*.

US Energy Information Administration (2015). *Residential Energy Consumption Survey 2015*. [Online; accessed in 2018]. URL: `https://www.eia.gov/consumption/residential/data/2015/`.

US National Oceanic and Atmospheric Administration (2017). *Local Climatological Data (LCD)*. [Online; accessed in 2017]. URL: `https://www.ncdc.noaa.gov/cdo-web/datatools/lcd`.

Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein (2008). “Financial Incentive-Based Approaches for Weight Loss: A Randomized Trial”. *Journal of the American Medical Association* 300(22), pp. 2631–2637.

Volpp, Kevin G., Andrea B. Troxel, Mark V. Pauly, Henry A. Glick, Andrea Puig, David A. Asch, Robert Galvin, Jingsan Zhu, Fei Wan, Jill DeGuzman, Elizabeth Corbett, Janet Weiner, and Janet Audrain-McGovern (2009). “A Randomized, Controlled Trial of Financial Incentives for Smoking Cessation”. *New England Journal of Medicine* 360(7), pp. 699–709.

Online Appendix

A Estimates of Bedroom-Level Energy Usage

Although we have access to highly detailed thermostat setpoint data, the bedrooms selected for our randomized controlled trials were not individually metered in terms of energy usage. Rather, energy data was available at the building-level. The basic set-up of the HVAC system of the building can be described as folows: each bedroom is equipped with a fan coil unit, which uses chilled water for cooling, and steam for heating. The fan (which blows either hot or cold air into the room) turns on or off depending on the difference between a room's thermostat setpoint and actual indoor temperature. For example, if the thermostat is set at 68^oF, and indoor temperature is initially at 71^oF, then the fan will remain turned on (blowing cold air into the room) until indoor temperature reaches the desired level of 68^oF. Conversely, if the setpoint is at 71^oF, and indoor temperature is at 68^oF, then the fan will remain on (blowing hot air into the room) until the desired setpoint of 71^oF is reached. Students do not have the option of completely shutting down this system, or the fan.

To determine the contribution of a single room to the overall building's usage of chilled water or steam, first we established a correlation between energy consumption and thermostat changes with the following model:

$$y_{it} = \alpha + \beta_1[T_{set} - T_{outdoor}]_{it} + \beta_2[T_{set} - T_{outdoor}]_{it} \times [\text{Double Bed}]_i + \boldsymbol{\beta_3}\mathbf{X}_{it} + \varepsilon_{it} \quad \text{(A.1)}$$

where y_t is either the building's chilled water consumption, or steam consumption at time t; α is a constant; T_{set} is the thermostat setpoint for room i, at time t; $T_{outdoor}$ is outdoor temperature; $[\text{Double Bed}]_i$ is equal to one for double-bedrooms, zero otherwise; $\mathbf{X}_{it}$ are controls including percent values (and lagged percent values) of steam and chilled water circulation intensity, the fan voltage, outdoor dew point temperature, relative humidity, and precipitation.

Equation A.1 above was estimated with data from the academic year prior to when our trials were conducted. Table A.1 below presents the results. The coefficient β_1 provides an estimate of how much chilled water or steam is used depending on the difference between a single-bedroom's thermostat setpoint and outdoor temperature. The coefficient β_2 should be added to the estimates for double-bedrooms. Those coefficients should be interpreted for hours by degree. So, for example, if the setpoint of a single-bedroom remains 1 degree lower than outdoor temperature for 1 hour, then the chilled water usage within that hour would be of 0.794 BTU. Conversely, if the setpoint for a single-bedroom remains 1 degree higher than outdoor temperature for 1 hour, then the steam usage within that hour would be of 0.454 BTU. Note that, probably because of generation of more bodily heat, less energy is required to heat double-bedrooms (compared to single-bedrooms) in our sample.

Table A.1: Estimates of Chilled Water and Steam Usage, Based on Differences Between Thermostat Setpoint and Outdoor Temperature

	Chilled Water Usage (BTU)	Steam Usage (BTU)
(β_1, single-bedrooms)	-0.794	0.454
	(0.009)	(0.017)
(β_2, additional for double-bedrooms)	-0.174	-0.090
	(0.002)	(0.004)

Note: This table presents estimates of how much energy for chilled water or steam is used if a bedroom's setpoint temperature remains above the outdoor temperature by 1 degree, for 1 hour. Standard errors are in parentheses. All coefficients are significant at 1%.

A simplifying assumption for our energy estimates is that chilled water is only used for cooling, and steam is only used for heating (which generally holds for our sample). Further, estimated parameters are heterogenous only by bedroom types. Therefore, if 2 same-type bedrooms maintain thermostats at the same temperatures throughout the entire sample, then their estimated energy usage will be the same. That implies that the energy usage reported to residents, as well as the intensity of the nudges, depend solely on the differences between thermostat setpoints and outdoor temperatures. We are not able to capture differences in room occupancy patterns (and how those affect energy usage), unless they lead to changes in the thermostats within the bedrooms.

Recall that the point estimate from Table 1, specification (VI), suggests a lower bound of 0.24^oF in thermostat reductions from treatment. We simulate how much that represents in energy space by applying the $0.24^o F$ reduction to control rooms, and then estimating weekly energy consumption with parameters from Table A.1 above. Comparing energy estimates from actual thermostats versus reduced thermostats, the energy impact is minimal: close to 0.1% of average weekly usage of control rooms during the intervention period.

B Balance Tables

Table B.1: Balance for the Fall Trial

	Control	Treat A	P-value of diff. (Control-Treat A)	Treat B	P-value of diff. (Control-Treat B)
Baseline Average Setpoint	71.1072	71.4619	0.2772	71.0984	0.9795
Suite Type:					
Double-Bed Suite %	0.3545	0.4144	0.5716	0.5200	0.1515
Mixed-Bed Suite %	0.0364	0.1441	0.0550	0.0000	0.1601
Single-Bed Suite %	0.5818	0.4324	0.1946	0.4800	0.3878
Special Suite %	0.0273	0.0090	0.3252	0.0000	0.0935
Suite Location:					
1st Floor %	0.1546	0.1622	0.9331	0.0200	0.0506
2nd Floor %	0.1546	0.1802	0.7588	0.2000	0.6177
3rd Floor %	0.2909	0.1802	0.2607	0.0600	0.0070
4th Floor %	0.1636	0.0901	0.3210	0.2600	0.3142
5th Floor %	0.1455	0.1892	0.6110	0.2000	0.5668
6th Floor %	0.0909	0.1981	0.1881	0.2599	0.0603
South Wing %	0.7727	0.6216	0.1163	0.7000	0.4405
West Wing %	0.2273	0.3784	0.1163	0.3000	0.4405
Bottom West Wing %	0.0273	0.0360	0.7880	0.0200	0.8002
Center South Wing %	0.1273	0.1441	0.8044	0.1800	0.4793
Left South Wing %	0.2636	0.2973	0.7588	0.2400	0.8296
Mid West Wing %	0.0727	0.1261	0.3638	0.1800	0.1175
Right South Wing %	0.3818	0.1802	0.0725	0.2800	0.4019
Top West Wing %	0.1273	0.2162	0.2400	0.1000	0.6718
Occupants' College Affiliations:					
Ag., Consumer & Env. Sciences %	0.0265	0.0661	0.2577	0.0400	0.5197
Applied Health Sciences %	0.0452	0.0360	0.6892	0.0350	0.6669
College of Business %	0.0701	0.0796	0.7660	0.0617	0.7682
College of Media %	0.0327	0.0225	0.6833	0.0150	0.4541
Division of General Studies %	0.2430	0.2260	0.7902	0.3083	0.2571
Education %	0.0234	0.0135	0.6436	0.0233	0.9990
Engineering %	0.1106	0.1284	0.6658	0.1183	0.8632
Fine & Applied Arts %	0.0467	0.0195	0.1884	0.0550	0.7801
Liberal Arts & Sciences %	0.3972	0.3994	0.9742	0.3433	0.4201
School of Social Work %	0.0047	0.0090	0.6692	0.0000	0.3118
Female %	0.5327	0.3559	0.0840	0.4100	0.2509
Residency Status:					
In State %	0.7998	0.8251	0.6339	0.8350	0.5109
Out of State %	0.0903	0.1021	0.7605	0.0500	0.2345
International %	0.1098	0.0728	0.3505	0.1050	0.8980
Number of Suites	42	44		38	

Note: This table compares control and treatment groups from the Fall randomized control trial in terms of baseline thermostat settings and covariate means. The p-values are based on standard errors clustered at the suite level, and indicate if differences in means are significant.

Table B.2: Balance for the Winter Break Trial

	Control	Winter Break Treatment	P-value of diff. (Control-Treated)
Baseline Average Setpoint	72.4504	71.8173	0.0231
Suite Type:			
Double-Bed Suite %	0.4591	0.3975	0.2662
Mixed-Bed Suite %	0.0566	0.0683	0.6655
Single-Bed Suite %	0.4717	0.5217	0.3715
Special Suite %	0.0126	0.0124	0.9900
Suite Location:			
1st Floor %	0.1069	0.1242	0.6290
2nd Floor %	0.1509	0.2050	0.2068
3rd Floor %	0.2138	0.1491	0.1332
4th Floor %	0.1887	0.1491	0.3453
5th Floor %	0.1824	0.1677	0.7302
6th Floor %	0.1572	0.2050	0.2681
South Wing %	0.7044	0.6894	0.7714
West Wing %	0.2956	0.3106	0.7714
Bottom West Wing %	0.0189	0.0373	0.3194
Center South Wing %	0.1824	0.1180	0.1073
Left South Wing %	0.2453	0.2919	0.3473
Mid West Wing %	0.1006	0.1491	0.1902
Right South Wing %	0.2767	0.2795	0.9560
Top West Wing %	0.1761	0.1242	0.1946
Occupants' College Affiliations:			
Ag., Consumer & Env. Sciences %	0.0520	0.0373	0.4880
Applied Health Sciences %	0.0446	0.0331	0.5476
College of Business %	0.0743	0.0673	0.7806
College of Media %	0.0223	0.0248	0.8745
Division of General Studies %	0.2665	0.2490	0.6890
Education %	0.0149	0.0248	0.4761
Engineering %	0.1136	0.1247	0.7248
Fine & Applied Arts %	0.0435	0.0362	0.6918
Liberal Arts & Sciences %	0.3684	0.3934	0.6105
School of Social Work %	0.0000	0.0093	0.1789
Female %	0.4554	0.4099	0.4008
Residency Status:			
In State %	0.8408	0.7992	0.2900
Out of State %	0.0616	0.1014	0.1517
International %	0.0945	0.0963	0.9524
Number of Rooms	159	161	

Note: This table compares control and treatment groups from the winter break trial in terms of baseline thermostat settings and covariate means. The p-values are based on standard errors clustered at the bedroom level, and indicate if differences in means are significant.

C Simulations of Statistical Power

Prior to initiating the trials, we ran simulations to assess statistical power and minimum detectable effect (MDE) sizes. With that, it was possible to determine the appropriate number of treatment arms and sample sizes for this study. For the simulations, we used (high-frequency) historical data about thermostat set-points in the treated building. Historical data were from up to a year prior to the treatment 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 treated 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 are likely to have similar distributions in terms of demographics, year of enrollment, field of study or other relevant covariates.

We tried simulations considering 2 treatment arms[24], 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 (C.1), as follows:

$$T_{it} = \beta_{1A} treatA_i \times Post_t + \beta_{1B} treatB_i \times Post_t + \gamma_i + \delta_t + \varepsilon_{it} \tag{C.1}$$

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 suites in each iteration. We stored the estimated coefficients β_{1A} and β_{1B}, as well as their standard errors (clustered by suite). The following Figure C.1 summarizes the simulation

[24]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.

results for the effect sizes of 0.6, 0.65, 0.7, and 0.75. The vertical axes show the values of recovered simulated treatment effects, while the horizontal axes indicate the iteration number. The red horizontal lines represent the true simulated effect (which we are trying to recover).

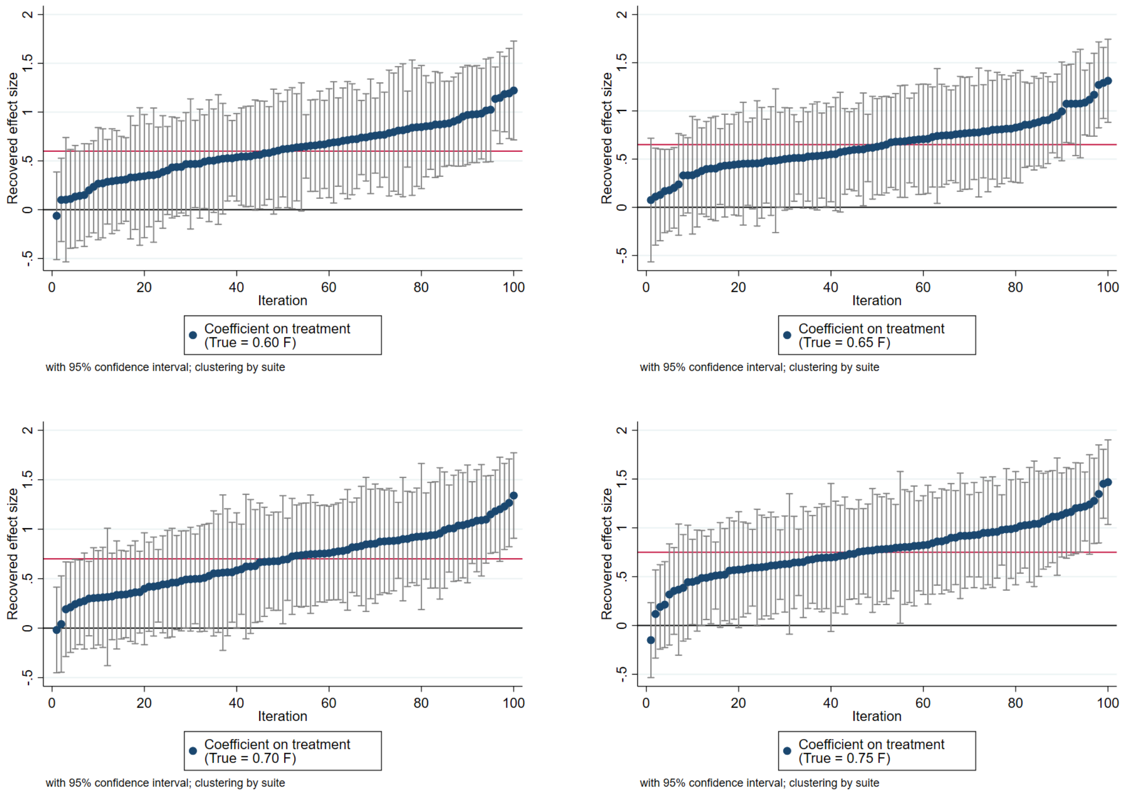

Figure C.1: Results from Simulations of Statistical Power

For the effect size of 0.75°F (bottom-right panel), we were able to recover (with a 95% confidence interval) a treatment effect in over 80% of simulations. The simulated MDE using pre-treatment data, therefore corresponds to approximately 1% of the average setpoint in the building.

D Spring 2018 Nudges

The significant effects shown in Table 3 could be attributed to the timing of treatment. Since students were not expecting to occupy their rooms during winter break, they might have been more willing to comply with the requests of lowering thermostats.

We therefore conducted another randomized controlled trial during Spring 2018, to test if moral suasion emails would work during a regular semester. Rooms were re-randomized, with 1/3 assigned to treatment, and 1/3 assigned to a control group.[25] Balance for the pre-treatment setpoints, and for relevant covariates are presented in Table D.1.

Four different email designs (shown in Figure D.1) were used during Spring 2018, to ask students to lower their thermostats. Emails were sent weekly, either in the morning (8-10 am) or evening (5-10 pm). We varied designed as an effort to promote more attention to the emails, and so that students would not feel that the information was too repetitive.

Figure D.1: Simple Emails Sent During Spring 2018

[25]Another 1/3 of rooms were assigned to a treatment arm that continued to receive energy report emails. Results were null, thus were omitted from this paper.

Table D.1: Balance for the Spring Trial

	Control	Spring Treatment	P-value of diff. (Control-Spring Treatment)
Baseline Average Setpoint	72.2824	72.3359	0.8831
Suite Type:			
Double-Bed Suite %	0.4815	0.3953	0.2301
Mixed-Bed Suite %	0.0370	0.0930	0.1247
Single-Bed Suite %	0.4722	0.5000	0.7016
Special Suite %	0.0093	0.0116	0.8732
Suite Location:			
1st Floor %	0.1376	0.0698	0.1166
2nd Floor %	0.1651	0.1744	0.8646
3rd Floor %	0.1927	0.2093	0.7746
4th Floor %	0.1835	0.1977	0.8033
5th Floor %	0.1376	0.1860	0.3666
6th Floor %	0.1835	0.1628	0.7048
South Wing %	0.7156	0.6512	0.3397
West Wing %	0.2844	0.3488	0.3397
Bottom West Wing %	0.0367	0.0233	0.5811
Center South Wing %	0.2018	0.1047	0.0572
Left South Wing %	0.2569	0.2791	0.7297
Mid West Wing %	0.1284	0.1279	0.9912
Right South Wing %	0.2569	0.2674	0.8684
Top West Wing %	0.1193	0.1977	0.1416
Occupants' College Affiliations:			
Ag., Consumer & Env. Sciences %	0.0642	0.0446	0.4974
Applied Health Sciences %	0.0168	0.0640	0.0697
College of Business %	0.0719	0.0620	0.7457
College of Media %	0.0138	0.0523	0.1319
Division of General Studies %	0.2638	0.2364	0.6202
Education %	0.0260	0.0174	0.6173
Engineering %	0.1491	0.0640	0.0247
Fine & Applied Arts %	0.0398	0.0252	0.4830
Liberal Arts & Sciences %	0.3502	0.4341	0.1849
School of Social Work %	0.0046	0.0000	0.3176
Female %	0.4817	0.3953	0.2117
Residency Status:			
In State %	0.8379	0.8682	0.5140
Out of State %	0.0765	0.0504	0.4028
International %	0.0810	0.0814	0.9921
Number of Rooms	108	86	

Note: This table presents covariate mean comparisons between control and treatment groups from the Spring randomized control trial. The p-values are based on standard errors clustered at the bedroom level, and indicate if differences in means are significant.

Effects of the Spring 2018 simple nudges were estimated with equations analogous to (3) and (4) from the main text. Results are presented in Table D.2 below. Although the coefficients are negative (suggesting effort to lower thermostats), they are not statistically significant. Null results for Spring, along with significant results for winter break (Table 3), therefore indicate that these types of nudges may only promote conservation for individuals who expect to vacate their rooms for longer periods of time (e.g. during winter break). The students might not be willing to sacrifice thermal comfort (in favor of conservation) during the regular semester routine.

Table D.2: Spring 2018 Treatment Effects

	(I)	(II)	(III)	(IV)	(V)	(VI)
Spring Treatment	-0.1750	-0.0553	-0.0638	-0.0616	-0.1963	
	(0.3793)	(0.3774)	(0.3798)	(0.3812)	(0.1511)	
Spring Treatment × Post Jan. 31						-0.2403
						(0.1556)
Sample Average Setpoint (^{o}F)	72.14	72.14	72.14	72.14	72.14	72.17
Observations	1,386,111	1,386,111	1,361,355	1,386,111	1,386,111	1,677,513
Controls:						
Room physical characteristics	No	Yes	Yes	Yes	Yes	No
Weather	No	No	Yes	No	No	No
Date/Time fixed effects	No	No	No	Yes	Yes	Yes
Avg. pre-treatment setpoint	No	No	No	No	Yes	No
Room fixed effects	No	No	No	No	No	Yes

Note: This table presents estimates of behavior change induced by simple nudges asking treated students to lower their thermostats to 68^oF during Spring 2018. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by room.

E Outdoor Temperature During Treatment Period

This section presents the temperature patterns observed during Fall 2017. Weather data from the station closest to the treated building was obtained from the National Oceanic and Atmospheric Administration (NOAA, 2017). The average daily outdoor temperature during the treatment period is shown in Figure E.1. Significant variability can be noted, in contrast to the stability in average thermostats observed in Figure 3. Outdoor temperature steadily falls after September, which seems to translate into a slight increase in the setpoints.

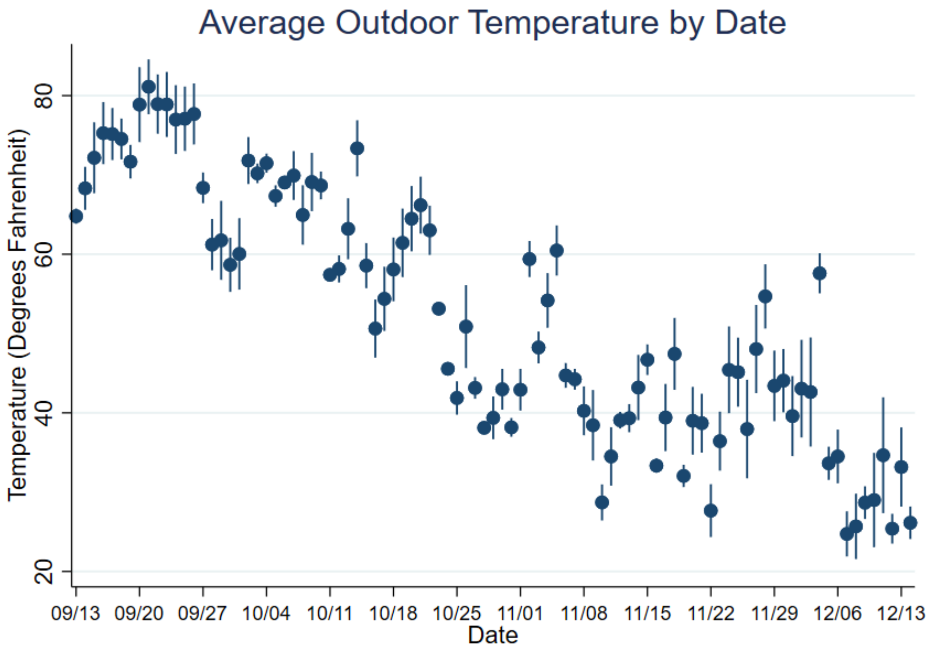

Figure E.1: Average Outdoor Temperature by Date

Average outdoor temperatures by hour of the day are presented in Figure E.2, which 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. Ideally, residents should take that pattern into consideration to conserve energy. Students were advised to lower setpoints at night (for cold days), lowering demand for heating, but there is no evidence of compliance.

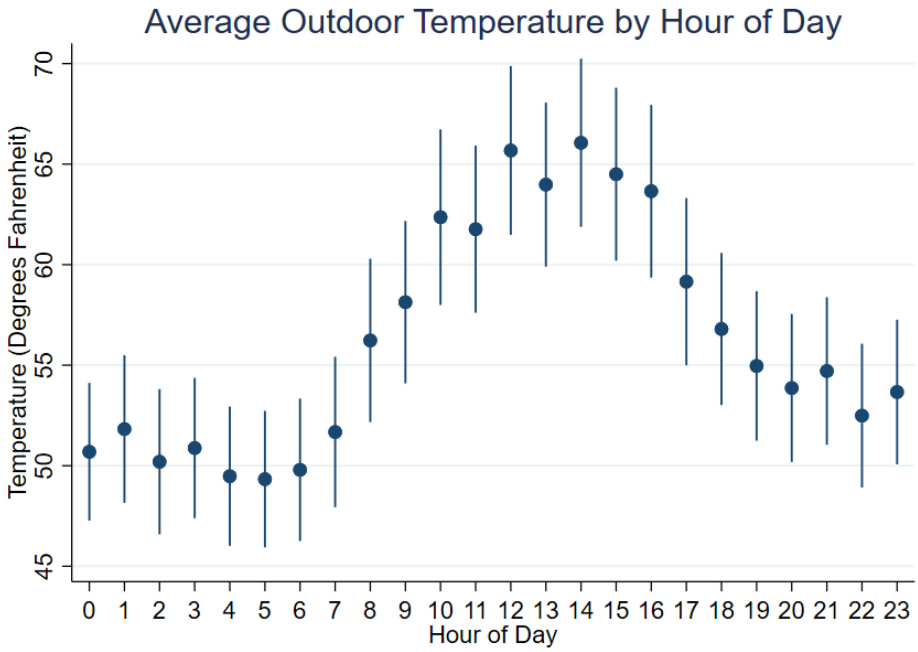

Figure E.2: Average Outdoor Temperature by Hour of the Day, During Treatment Period

F Further Robustness Checks and Treatment Heterogeneity

F.1 Alternative outcome variable

The expected sign of our main treatment effect coefficient varies depending on weather. For hot days, we would have evidence to support that the nudges were effective if the coefficient is positive and significant. Conversely, for cold days the coefficient is expected to be negative. That is because less energy is used when the thermostats are set closer to outdoor temperatures. To take that into account, we consider an alternative outcome variable for our specifications. We substitute the outcome in equations (3) and (4) with the absolute distance (D_{it}) between thermostat set-points (T_{it}) and outdoor temperature (O_{it}), such that: $D_{it} = |T_{it} - O_{it}|$. Larger values of D_{it} indicates increased demand for the heating/cooling system, and thus more energy use. Therefore, when using D_{it} as an outcome, we expect the treatment coefficient to be negative, indicating an effort towards energy conservation.

Results with D_{it} as an outcome variable are presented in Table F.1. Again, we obtain precisely estimated null treatment effects.

Table F.1: Treatment Effect on Absolute Difference Between Setpoints and Outdoor Temperature

	(I)	(II)	(III)	(IV)	(V)	(VI)
Treated	0.1227	0.1728	0.1778	0.1774	0.0739	
	(0.2135)	(0.2032)	(0.2064)	(0.2052)	(0.1388)	
Treated × Post Sep. 13						0.0999
						(0.1309)
Observations	2,602,592	2,602,592	2,591,389	2,602,592	2,602,592	3,102,082
Controls:						
Room physical characteristics	No	Yes	Yes	Yes	Yes	No
Weather	No	No	Yes	No	No	No
Date/Time fixed effects	No	No	No	Yes	Yes	Yes
Avg. pre-treatment setpoint	No	No	No	No	Yes	No
Room fixed effects	No	No	No	No	No	Yes

Note: This table presents estimates of behavior change induced by the weekly energy reports, sent to treated subjects during Fall 2017. The outcome variable is the absolute difference between thermostat setpoints and outdoor temperature. Standard errors (in parentheses) are clustered by suite.

F.2 Treatment effects by hour of the day

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

$$T_{it} = \boldsymbol{\beta}_1^h treat_i \times Hour_h + \boldsymbol{\beta}_2 \mathbf{X}_{it} + \varepsilon_{it} \tag{F.1}$$

where T_{it} are thermostat setpoints; $treat_i$ is the treatment indicator; $Hour_h$ 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 F.1. 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.

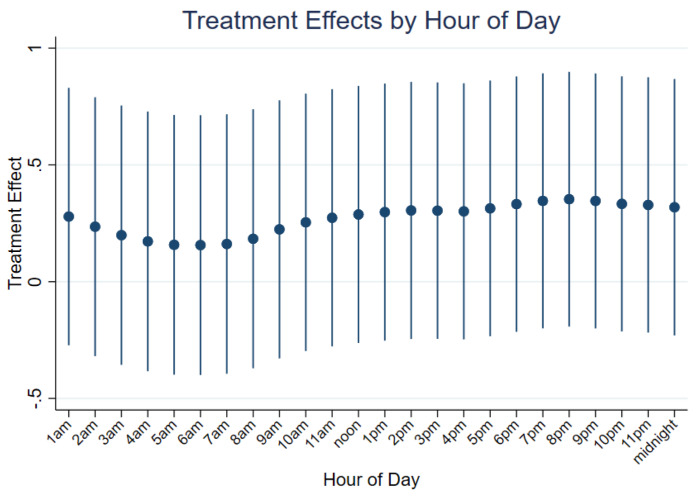

Figure F.1: Treatment Effects by Hour of the Day

F.3 Treatment effects by weekday

The energy reports 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:

$$T_{it} = \boldsymbol{\beta}_1^d treat_i \times Weekday_d + \boldsymbol{\beta}_2 \mathbf{X}_{it} + \varepsilon_{it} \tag{F.2}$$

where T_{it} are thermostat setpoints; $treat_i$ is the treatment indicator; $Weekday_d$ 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 F.2. No statistically significant effects emerge. The point estimate for Mondays seem slightly higher than for other days, but the differences are negligible.

Table F.2: 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

Note: This table presents estimates of treatment effect heterogeneity by weekday, for the HERs sent during Fall 2017. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by suite.

F.4 Treatment effects around time of treatment

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 4 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:

$$T_{ih} = \boldsymbol{\beta}_1^h treat_i \times Hour_h \times Post\_Treat + \boldsymbol{\beta}_2^h treat_i \times Hour_h + \boldsymbol{\beta}_3^h Post\_Treat \times Hour_h + \beta_4 Post\_Treat \times treat_i + \beta_5 Post\_Treat + \beta_6 treat_i + \boldsymbol{\beta}_7^h Hour_h + \varepsilon_{ih} \quad \text{(F.3)}$$

where T_{ih} is the setpoint for room i in hour h; $treat_i$ indicates if the room was in the treated or control group; $Post\_Treat$ is equal to one after Sept. 13th (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 F.2 plots the point estimates of $\boldsymbol{\beta}_1^h$ obtained from equation F.3. 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).

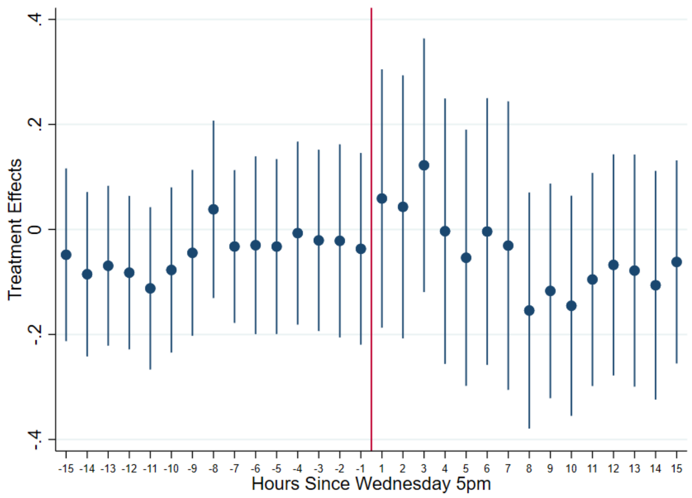

Figure F.2: Treatment effects around treatment time (5pm on Wednesdays are the omitted comparison group)

F.5 Treatment heterogeneity based on "efficiency standing"

Recall that for the Fall 2017 trial subjects were categorized, each week, according to their "efficiency standing" (Great, Good, or Below Average). In order to test if those categorizations differentially affected students' behavior, we run the following regression model:

$$T_{it} = \alpha_1 treat_i + \alpha_{2g}[\text{Good}]_{iw-1} + \alpha_{2b}[\text{Below Average}]_{iw-1} + \alpha_{3g} treat_i \times [\text{Good}]_{iw-1} + \alpha_{3b} treat_i \times [\text{Below Average}]_{iw-1} + \boldsymbol{\alpha}_4 \mathbf{X}_{it} + \varepsilon_{it} \quad \text{(F.4)}$$

where T_{it} are thermostat setpoints for room i, in time t; $treat_i$ is equal to 1 if the room was assigned to treatment, zero otherwise; $[\text{Good}]_{iw-1}$ indicates if the room received a "Good" efficiency rating (consumption below average, but above the 15th percentile) in week $w-1$ (most recent previous week); and $[\text{Below Average}]_{iw-1}$ indicates if the room received a "Below Average" efficiency rating (consumption above average) in week $w-1$; $\mathbf{X}_{it}$ are exogenous controls. Negative and significant α_{3g} and α_{3b} would suggest that treated subjects with the corresponding rating in week $w-1$ (most recent previous week) exerted more effort to change their behavior in (current) week w. All coefficients should be interpreted relative to the omitted comparison group: non-treated most efficient users from week $w-1$.

Results are presented in Table F.3 below. It can be noted that the coefficient on treatment is positive, and even marginally significant for specification (III), which includes date/time fixed effects and controls for room physical attributes. That suggests that treated efficient residents may have increased their thermostats (and usage) due to the nudges. That could therefore constitute a "boomerang effect" (Clee and Wicklund, 1980), for which consumers that are doing well end up moving in the opposite direction than what is expected form the nudges.

The coefficients on "Good Rating" and "Below Average Rating" are positive and highly significant, as expected. Those just reveal the mechanical relationship between the rating received and actual thermostat setpoints. It can be noted, for example, that above average consumers set their thermostats, on average, 2.4^oF higher than efficient consumers. The interactions between the treatment indicator and the efficiency indicators are negative, but not significant. The direction of those effects suggest that subjects who were not in the "Great" category may have exerted more effort to conserve (by reducing thermostats). However, we cannot rule out that the effects are null.

However, Allcott (2011) argues that specification (F.4) above is naive, and does not truly recover the causal effect of the categorizations on treated subjects' behavior. There could be unobservable differences between residents which lead them to be above or below average consumers, and which simultaneously affect how they respond to the nudges. For example, high users might have higher preference for heating, thus are less willing to change their behavior. To address this issue, we exploit the fact that the efficiency categorizations are based on sharp cutoffs of estimated usage. We can then implement a Regression Discontinuity (RD) design, with the assumption that residents around the cutoff are similar in both observable and unobservable characteristics. With the RD, it will then be possible to recover a Local Average Treatment Effect (LATE) for being above the usage cutoffs (i.e. receiving a "Below Average" rating, rather than "Good"; or receiving a "Good" rating, rather than "Great").

Table F.3: Treatment Heterogeneity by "Efficiency Standing" Rating

	(I)	(II)	(III)
Treated	0.5841	0.5945	0.7398*
	(0.3674)	(0.3993)	(0.3949)
Good Rating in Previous Week	0.7398***	0.7532***	0.7429***
	(0.2558)	(0.2818)	(0.2818)
Below Average Rating in Previous Week	2.4090***	2.4551***	2.4017***
	(0.2586)	(0.2784)	(0.2830)
Treated × Good Rating in Previous Week	-0.1638	-0.1460	-0.1950
	(0.3148)	(0.3444)	(0.3357)
Treated × Below Average Rating in Previous Week	-0.4426	-0.4818	-0.5740
	(0.3368)	(0.3575)	(0.3577)
Observations	2,570,694	2,570,694	2,570,694
Controls:			
Room physical characteristics	No	No	Yes
Day of Sample fixed effects	No	Yes	No
Date/Time fixed effects	No	No	Yes

Note: This table presents estimates of treatment heterogeneity for the Fall 2017 trial, based on the "efficiency standing" ratings received by subjects. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by suite. Significance at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

For the RD, we start with a graphical analysis to assess if there are any observable discontinuities (in thermostat settings) around the usage cutoffs that determine a subject's category. Recall that there are 2 cutoffs to consider: for the Average user, and for the Efficient (15th percentile) user. We can therefore consider 2 sets of RDs. In the following Figures F.3 and F.4, we normalize the (previous week's) Average usage and the Efficient usage to 0 for all weeks of treatment, and look at the (current week's) thermostat setpoints of treated subjects who were near the 0 cutoff. The red lines represent simple linear fits, between thermostat setpoints and distance from the normalized cutoffs, estimated separately to the left and to the right. There is no evidence of a significant discontinuity in setpoints. Further, the linear fits even suggest an effect that is opposite of what we expect: subjects who received bad efficiency ratings in the previous week may even do worse than similar subjects who received a better rating.

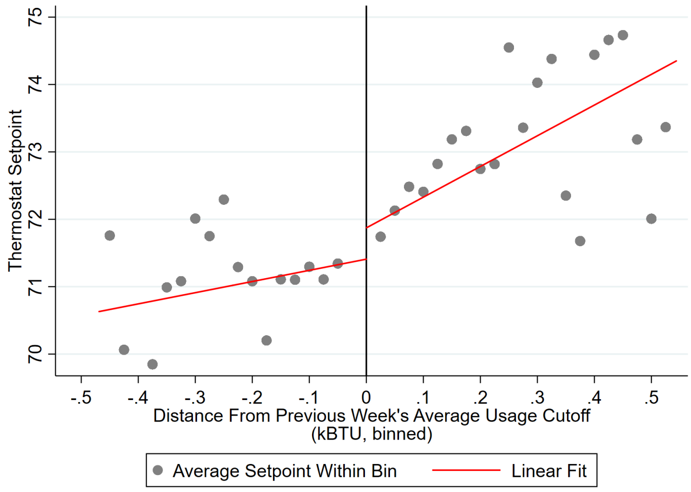

Figure F.3: Thermostat Setpoints Around the Average Usage Cutoff, for Treated Rooms

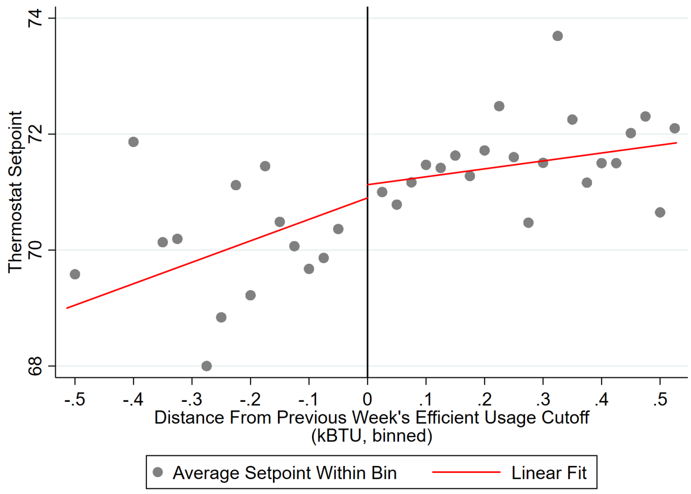

Figure F.4: Thermostat Setpoints Around the Efficient Usage Cutoff, for Treated Rooms

To provide further robustness for the RD analysis, we consider a triple-difference framework to estimate if treated subjects above the cutoffs behave significantly different than those below. We run, separately for each of the 2 cutoffs, the following RD specification:

$$\begin{aligned} T_{it} = \alpha\beta_1 treat_i + \beta_2[Usage - c]_{iw-1} + \beta_3 C_{iw-1} + \\ \beta_4 treat_i \times C_{iw-1} + \beta_5[Usage - c]_{iw-1} \times C_{iw-1} + \beta_6[Usage - c]_{iw-1} \times treat_i + \\ \beta_7 treat_i \times [Usage - c]_{iw-1} \times C_{iw-1} + \varepsilon_{it} \end{aligned} \quad \text{(F.5)}$$

where T_{it} are thermostat setpoints for room i, in time t; α is a constant; $treat_i$ is equal to 1 if the room was assigned to treatment, zero otherwise; $[Usage - c]_{iw-1}$ represents subjects i usage distance from the cutoff (either Average or Efficient), in the most recent previous week $w - 1$; C_{iw-1} is an indicator equal to 1 if the subject was above the cutoff, zero if below. We restrict the regression sample to a bandwidth of 0.2 kBTU around the cutoffs. The RD coefficient $beta_7$ reveals if treated subjects just above the cutoff behaved differently form treated subjects just below. We further employ triangular kernel weighting, such that observations closer to the cutoffs receive higher importance.

Results are presented in Tables F.4 and F.5. The coefficients on the triple interactions are not statistically significant. Thus we cannot infer that treated subjects just above the cutoffs (who received worse ratings) were differentially affected by our nudges. The signs of the coefficients (positive) are in line with the graphical analysis, suggesting that subjects above the cutoffs may have performed worse due to the categorization.

Table F.4: Regression Discontinuity Estimates Around the Average Usage Cutoff

Treated	0.3520
	(0.2399)
Distance from "Average Usage" Cutoff	4.1256**
	(1.5959)
Above "Average Usage" Cutoff	0.1277
	(0.1957)
Treated × Above "Average Usage" Cutoff	-0.0618
	(0.2498)
Distance from "Average Usage" Cutoff × Above "Average Usage" Cutoff	6.2074**
	(2.9451)
Distance from "Average Usage" Cutoff × Treated	-0.9139
	(2.1251)
Distance from "Average Usage" Cutoff × Above "Average Usage" Cutoff × Treated	0.4272
	(3.5270)
Constant	71.1014***
	(0.1928)
Observations	2,136,954

Note: This table presents regression discontinuity estimates to assess if treated subjects just above the "Average Usage" cutoff behaved differently from those just below. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by suite. Significance at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Table F.5: Regression Discontinuity Estimates Around the Efficient Usage Cutoff

Treated	0.3943
	(0.2904)
Distance from "Efficient Usage" Cutoff	11.6175***
	(1.4424)
Above "Efficient Usage" Cutoff	0.0696
	(0.1783)
Treatment × Above "Efficient Usage" Cutoff	0.0642
	(0.2311)
Distance from "Efficient Usage" Cutoff × Above "Efficient Usage" Cutoff	-10.2316***
	(2.0746)
Distance from "Efficient Usage" Cutoff × Treated	-3.5467*
	(1.9083)
Distance from "Efficient Usage" Cutoff × Above "Efficient Usage" Cutoff × Treated	2.1356
	(2.8610)
Constant	70.6244***
	(0.2309)
Observations	1894338

Note: This table presents regression discontinuity estimates to assess if treated subjects just above the "Efficient Usage" cutoff behaved differently from those just below. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by suite. Significance at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

F.6 Pre-treatment survey

Prior to starting the Fall trial, we sent an online survey to all residents of the trial building. We were able to collect 98 responses, out of the 445 students contacted (22% response rate). Analyses in this subsection are therefore only valid for a (self-)selected subsample of residents.

The pre-intervention survey was meant to assess individuals' degree of concern about conservation or environmental issues in general. For that, we used the revised version of the New Ecological Paradigm (NEP) scale (Dunlap et al., 2000). The NEP consists of 15 statements, for which subjects are asked to indicate their level of agreement, with a Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree). Agreement with the 8 odd-numbered statements indicates pro-environmental beliefs, while carelessness for environmental degradation is revealed by agreement with the 7 even-numbered statements. When compiling a single score per person, values from all responses are added to form a scale from 15-75.[26] Individuals who score above 45 are considered to lean more towards environmental conservation. The following Figure F.5 presents histograms of NEP scores from the individuals who completed the survey, separated by treatment and control groups.

Most respondents obtained scores above 45, with the averages ranging form 53.4 (treatment) to 57.1 (control). That is above an average of 51.3 obtained from a representative household survey of the US population (Pienaar, Lew, and Wallmo, 2015). Our subjects may therefore be considered more environmentally concerned than typical US residents.

To test if higher concerned individuals responded more strongly to the Fall reports, we consider the following regression model:

$$T_{it} = \alpha_1 treat_{it} + \alpha_2 treat_{it} \times [\text{NEP scale}]_i + \boldsymbol{\alpha}_3 \mathbf{X}_{it} + \varepsilon_{it} \tag{F.6}$$

where $[\text{NEP scale}]_i$ is a measure of the New Ecological Paradigm scale for room i, obtained

[26] The highest score for (even)odd-numbered questions is 5, when the subject 'strongly (dis)agrees.'

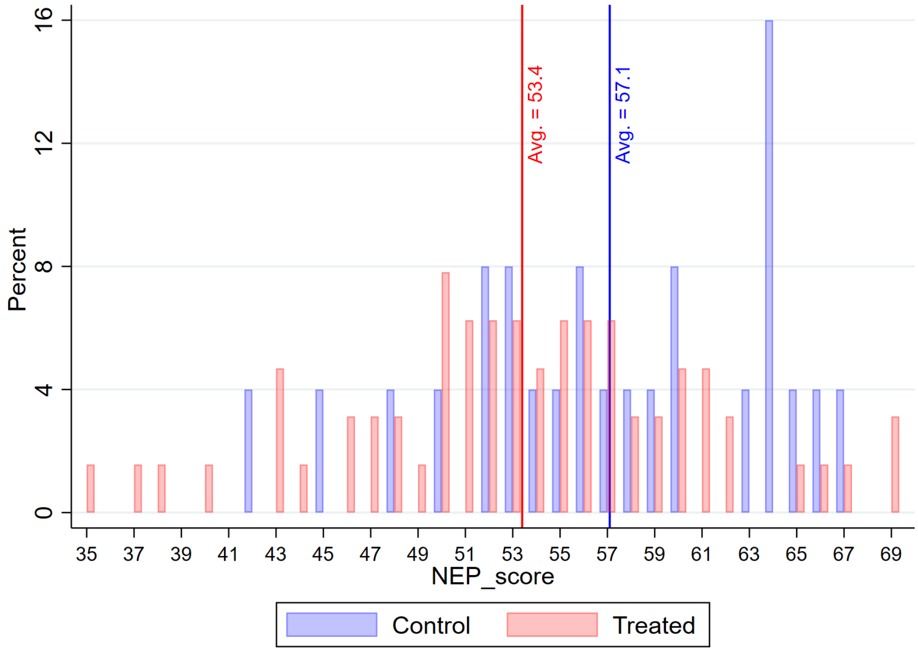

Figure F.5: Histograms of NEP Scores, for treated and control groups

from the pre-intervention survey.[27] If α_2 is found to be negative, then we would have evidence that individuals with higher NEP scores are more likely to lower thermostats in response to the nudges. We also consider a DID framework, for which equation F.6 above is modified to include interactions with the indicator for the date of treatment start (Post Sep. 14th). Results are presented in Table F.6 below. Note that the sign on α_2 is, as expected, negative in all specifications. However, those are not statistically significant, thus the level of environmental concern is negligible in this context.

[27]Note that for double-bedrooms we use the average NEP of the two residents.

Table F.6: Treatment Effect Heterogeneity by NEP Scores

	(I)	(II)	(III)
Treated	4.3648	4.3439	-
	(3.0214)	(3.0349)	
NEP Score	0.0619	0.0618	-
	(0.0402)	(0.0404)	
Treated × NEP Score	-0.0763	-0.0759	-
	(0.0524)	(0.0526)	
Treated × Post Sep. 13			0.3815
			(2.0985)
Post Sep. 13 × NEP Score			-0.0108
			(0.0268)
Treated × Post Sep. 13 × NEP Score			-0.0091
			(0.0370)
Observations	632,880	632,880	752,415
Controls:			
Date/Time fixed effects	No	Yes	Yes
Room fixed effects	No	No	Yes

Note: This table presents treatment effect heterogeneity estimates, based on New Ecological Paradigm (NEP) scores from subjects, acording to a pre-intervention survey. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by suite.

F.7 Long-term effects

We test if subjects may have had a delayed response to treatment, by looking at a subsample of 65 rooms that received HERs during both Fall and subsequent Spring semesters (referred to as "Always Treated"). Those were compared to 35 rooms that were never assigned to any of the treatments arms (referred to as "Always Control"), other than the winter break nudges. Again, we run regression specifications 3 and 4, with the adequate treatment indicators and sample restrictions.

Results are presented in Table F.7. None of the estimated coefficients are statistically significant, and point estimates are close to zero. A complementary graphical analysis of average setpoints by date (Figure F.6), produces the same results: there is no evidence of significant change in behavior of subjects who were continuously exposed to treatment during the whole academic year.

Table F.7: Treatment Effects for Long-Term Exposure to HERs

	(I)	(II)	(III)	(IV)	(V)	(VI)
Always Treated	-0.1764	-0.1815	-0.2153	-0.2196	0.0619	
	(0.4616)	(0.4286)	(0.4235)	(0.4260)	(0.2424)	
Always Treated × Post Sep. 13						-0.1317
						(0.2603)
Observations	1,654,930	1,654,930	1,633,233	1,654,910	1,654,633	1,808,961
Controls:						
Room physical characteristics	No	Yes	Yes	Yes	Yes	No
Weather	No	No	Yes	No	No	No
Date/Time fixed effects	No	No	No	Yes	Yes	Yes
Avg. pre-treatment setpoint	No	No	No	No	Yes	No
Room fixed effects	No	No	No	No	No	Yes

Note: This table presents estimates of treatment effects for subjects who received HERs during both Fall and Spring semesters, compared to subjects who were never assigned to treatment. The outcome variable is thermostat setpoints. Standard errors (in parentheses) are clustered by suite.

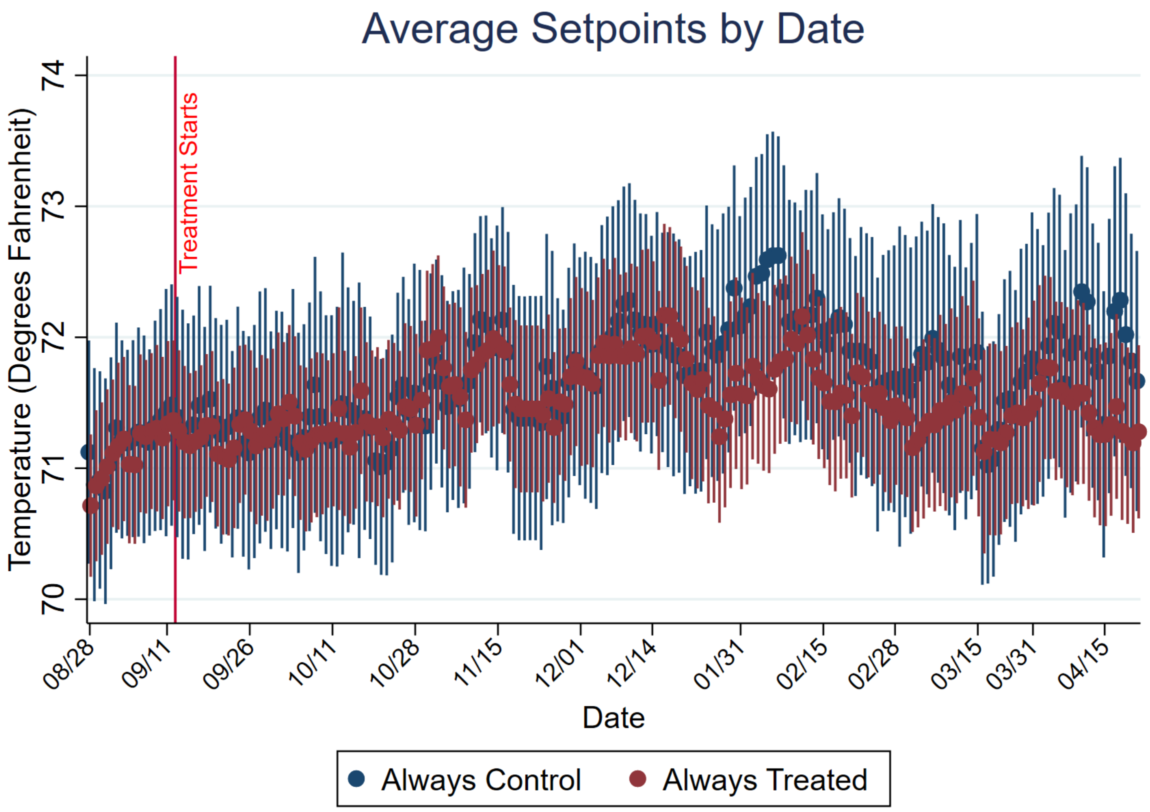

Figure F.6: Average setpoints by date, comparing subjects that were "Always Treated" with those that were "Always Control"