Do The Effects of Nudges Persist? Theory and Evidence from 38 Natural Field Experiments

Alec Brandon* Johns Hopkins University Paul J. Ferraro Johns Hopkins University

John A. List University of Chicago, ANU, & NBER Robert D. Metcalfe Columbia University & NBER

Michael K. Price University of Alabama, ANU, RWI-Essen, & NBER Florian Rundhammer Georgia State University

February 14, 2025

Abstract

We formalize a research design to uncover the mechanisms underlying long-term reductions in energy consumption caused by a widely implemented nudge. We consider two channels: technology adoption and habit formation. Using data from 38 natural field experiments, we isolate the role of technology adoption by comparing treatment and control homes after the initial resident moves, which discontinues the treatment for a home. We find that the majority of energy reductions persist in the home after treatment ends and show this persistence is consonant with a technology adoption channel. The role of technology in creating persistent behavior change has important implications for designing behavioral interventions and evaluating their long-term social impacts.

^{*}We thank Arhan Gunel and Nancy Hersch for sharing the data. Hunt Allcott, Jonathan Davis, Michael Greenstone, Justin Holz, Jim Kapsis, Marc Laitin, Jeff Livingston, David Novgorodsky, Steve Puller, David Rapson, Sally Sadoff, Jeroen van de Ven, and Meng Zhu offered helpful comments. Michael Cuna and Ariel Listo provided excellent research assistance. We would also like to thank participants of the American Economic Association annual conference, the Muenster Energy Conference, the Association of Environmental and Resource Economists annual conference, Camp Resources XXII, POWER Conference 2018, the Grantham Institute at the London School of Economics, the University of Chicago, Marquette University, ZEW, RWI-Essen, and Georgia State University.

1. Introduction

A growing literature has established that nudges (Thaler and Sunstein, 2008) are a highly cost-effective approach to changing an array of behaviors in the short-term (Allcott and Mullainathan, 2010; Benartzi et al., 2017; Hummel and Maedche, 2019; DellaVigna and Linos, 2022).¹ Less, however, is known about the long-term effectiveness of nudges. In many of the contexts in which nudges are applied, such as education, health or the environment, success requires persistent behavior change.

We study the mechanisms underlying persistent energy reductions produced by one of the most widely studied nudges: the Home Energy Report (HER). The HER provides a social comparison that contrasts the recipient's energy consumption to the energy consumption of their neighbors. The HER has been evaluated in dozens of randomized trials conducted by residential energy providers across the United States (U.S.).

Studies of randomized trials find the HER is highly cost-effective. Although energy consumption is notoriously price inelastic, Allcott (2011); Ayres et al. (2013); Costa and Kahn (2013); Allcott (2015) report that average energy consumption declined by one to two percent among households who received HERs over a period of a year. The evidence for HER effectiveness has led energy providers in the U.S. to widely adopt the HER and policymakers to herald the HER as an important tool to fight against climate change (IEA, 2021). As a further testament to the success the HER, the company that developed it, Opower, was acquired by Oracle for more than \$500 million.

Follow-up studies report that that the HER effect on energy consumption persists beyond a single year. After five years of exposure to HERs, a

¹Thaler and Sunstein (2008, pp. 6) defines a nudge as an intervention designed to alter, "... behavior in a predictable way without forbidding any options or significantly changing their economic incentives."

difference in energy use between households in the treatment group (HER recipients) and the control group (untouched) can still be detected (Bell et al., 2020, and the citations therein). Furthermore, the majority of the short-run effect persists two years after HERs are discontinued (Allcott and Rogers, 2014).

The persistence of the HER effect stands in marked contrast to the persistence of the effects of analogous social comparison nudges in other contexts (Figure A1). In the short term, these nudges increase charitable giving, financial savings, tax and other types of compliance, water conservation, and voter turnout. However, only the effects on water conservation persist after the nudges are discontinued (Shang and Croson, 2009; Apesteguia et al., 2013; Ferraro and Miranda, 2013; Bernedo et al., 2014; Hallsworth et al., 2017; Coppock and Green, 2016; Rogers et al., 2017; Kast et al., 2018).

The challenge of designing nudges that produce persistent effects can be seen in a recent meta-analysis. DellaVigna and Linos (2022) find that the estimated effect of the nudge and the time horizon over which a nudge is evaluated are negatively correlated. After controlling for a variety of observable features, they find that each additional day over which a nudge is evaluated correlates with a 0.7 percent reduction in the average effect of the nudge. While this estimated effect is statistically imprecise (standard error = 0.4), it suggests that the average short-term effect of nudges would disappear after an additional year or two.²

Academics and policymakers who wish to induce persistent behavioral change would thus benefit from understanding the mechanisms that underlie the persistent effects of HERs on energy consumption. Yet the evidence about the channels through which HERs affect long-run patterns of energy consumption is limited. In two HER experiments, Allcott and Rogers (2014)

²See also Choukhmane (2021) for evidence on long-term effects of savings defaults that dwarf short-term effects.

found that no more than 2 percent of the HER's long-term effectiveness can be explained by participation in utility run energy efficiency programs. Under the assumption that adopters of energy efficient technologies would use these programs to facilitate adoption, the finding suggests that technology adoption is unlikely to be the channel driving the persistence of the HER effect.

Likewise, evaluations of an HER-like intervention for water conservation also fail to provide any evidence of technology adoption driving the persistence of the effect. Ferraro and Miranda (2013) and Bernedo et al. (2014) report that that the estimated effect is no longer statistically significant in the subgroup of homes in which the initial treated resident had moved. They conclude that a change in habits is the most plausible channel for the persistence of the intervention's effect.

These analyses suggest that the long-term effectiveness of the HER reflects changes to something in the people residing in a home, such as their habits or skills, as opposed to something in the home, such as more efficient technologies. However, these results are only suggestive. The research designs are informal, and the identifying assumptions are not clearly defined or tested. Moreover, in the analyses of movers in Ferraro and Miranda (2013) and Bernedo et al. (2014), the samples are small and thus potentially underpowered.

We formalize a research design that decomposes the long-run effect of the HER into components attributable to technology adoption and habit formation. This decomposition is accomplished by exploiting a feature of how HERs were administered in the experiments. If the initial resident in an HER experiment moved to a new home, then the HER was immediately discontinued. Moving, however, did not discontinue observations of energy consumption in the home. We show that, under certain conditions, the postmove HER effect identifies the fraction of the treatment effect attributable to technology adoption. The fraction attributable to habit formation is then the fraction of the HER's long-term impact that is not explained by technology adoption.

Our decomposition of the HER's long-term effectiveness depends on the validity of three assumptions. First, treatment assignment did not influence residents' decisions to move from a home in the experimental sample. Second, treatment assignment did not influence the types of residents that moved *into* a home in the experimental sample. Third, the technology adopted in response to the HER remained in a home after the initial resident moved. While we argue that these assumptions are plausible for a "light touch" informational intervention like the HER, we also consider their plausibility with data from our experiments and find evidence that is broadly consistent with the assumptions.

Using data on nearly 140,000 movers observed across 38 HER experiments, we apply our research design and decompose the long-term effectiveness of the HER. We find that, over the long-term, movers respond to receipt of the HER by reducing their energy consumption by 2.1 percent. Moreover, we find that fifty-one percent of this reduction remains in the home after a move, and we show this result is robust to a battery of alternative specifications. Under our decomposition assumptions, these results imply that technology adoption, as opposed to habit formation, was the primary channel responsible for the long-term energy reductions produced by the HER.

Our study makes three contributions. First, it provides a simple explanation for the variation in the persistence of social comparison effects in the literature: variation in the availability of technologies across contexts. In the contexts of energy and water conservation, households can respond to the nudge by adopting long-lived technologies that have long-term impacts by reducing the marginal cost of conservation. Such technologies, however, are scarce for households that wish to donate more to charitable organizations, evade their taxes, contribute to their financial savings, and vote in an election. The contrast between the rapid fade-out of effects produced by nudges that target these behaviors and the persistence of effects produced by nudges that target energy and water conservation can thus be explained by the variation in availability of enabling technologies.

Second, the identification of technology adoption as a critical channel for persistent behavioral change provides policymakers with an insight that can be leveraged to induce more persistent effects from nudges (or avoid such persistence when the goal is only temporary behavior change). Policymakers can target nudges towards behaviors that can be changed with the adoption of technologies. For example, we conjecture that the effect of voter turnout efforts will persist longer when a municipality allows households to default into easier modes of voting in future elections, such as mail-in or on-line ballots. When such technologies do not already exist, policymakers can encourage the development of new technologies that can be paired with nudges. For example, a social nudge promoting charitable giving or financial savings could be combined with an option to set a default donation or contribution rate in the future (Madrian and Shea, 2001; Thaler and Benartzi, 2004; Goswami and Urminsky, 2016; Altmann et al., 2019).

Third, our study illustrates an application of a new approach to decompose the mechanisms of policy effectiveness. Prior studies advocate for experimental designs that directly test for a hypothesized mechanism (Ludwig et al., 2011) or econometric analyses that rely on the collection of data that proxy for hypothesized mechanisms (Heckman and Pinto, 2015). Our approach complements these recommended designs and analyses, particularly when there is uncertainty about whether changes in human or physical forms of capital are driving an intervention's effect and when the intervention is a relatively light touch, such as a nudge, and thus will satisfy the three identifying assumptions of our design.

Our study also contributes to several other strands of research. First, it contributes to the nascent literature on the determinants of persistent responses to policy interventions (Frey and Rogers, 2014; Rogers and Frey, 2016). Second, by presenting a cost-benefit analysis of the HER that incorporates the indirect cost of the technology adopted, our study contributes to the literature on identifying the full effect of policy interventions (Heckman and Smith, 1997). Third, our study also contributes to the literature on energy efficient technology adoption by highlighting that nudges like the HER can stimulate the take up of such technologies (Jaffe and Stavins, 1994; Allcott and Greenstone, 2012; Gerarden et al., 2017; Gillingham et al., 2018). Finally, our study contributes to the theoretical and empirical literature on habit formation (Pollak, 1970; Becker and Murphy, 1988; Becker, 1992; Charness and Gneezy, 2009; John et al., 2011; Acland and Levy, 2015; Royer et al., 2105; Fujiwara et al., 2016; Levitt et al., 2016; Beshears et al., 2021; Vollaard and van Soest, 2024; Allcott et al., 2020; Bursztyn et al., 2021; Allcott et al., 2022). We contribute to the literature on habit formation by developing an approach to decompose the relative importance of changes in human factors, such as habits, and changes in non-human factors, such as technologies, for the long-run effectiveness of a policy intervention.

The remainder of this study proceeds as follows. In Section 2 we formalize our identification strategy. Section 3 describes the HER experiments and mover sample. We present our empirical findings and discuss their implications in Section 4. Section 6 concludes by considering other contexts where our identification strategy can be applied.

2. Identification Strategy

In this section, we formalize our strategy for decomposing the long-term effectiveness of the HER into components attributable to habits and technol-

2.1 Setting and Notation

Consider a subsample of homes in an HER experiment from which the initial resident will eventually move. During a baseline period, the electricity consumption of each home is observed. After this period, homes are randomly assigned to remain in the controlled state of the baseline period or enter a treated state, wherein the home receives an HER in the mail. Receipt of the HER continues for homes in the treated state until the initial resident moves, at which point the HER is discontinued.

More formally, let $i \in \{1, 2, ..., I\}$ index each home. Let $t \in \{-12, -11, ..., T\}$ index each unit of time over which a home's outcome of interest is observed and suppose this index is measured relative to the end of the baseline period (i.e., t = 0 is the time at which the treatment begins to be administered). The outcome of interest in the HER experiments is electricity consumption, which we denote with $Y_{it} \in \mathbb{R}$. Let $D_{it} \in \{0, 1\}$ be a treatment indicator that denotes whether home *i* has entered the treated state at or before time *t*. That is, during the baseline period, this treatment indicator equals o for every home. It then switches to 1 for the homes that receive the HER and stays at 1, regardless of whether the initial resident eventually moves. Let $M_{it} \in \{0, 1\}$ indicate whether the initial resident has moved out of home *i* at time *t*. It will also prove convenient to define $\tilde{M}_i \in \{1, 2, ..., T\}$ as the value of the time index at which the initial resident of home *i* moves. This variable is related to the move indicator according to $M_{it} = 1(t > \tilde{M}_i)$, where $1(\cdot)$ is the indicator function.

The relationship between the outcome of interest, Y_{it} , the treatment indicator, D_{it} , and the move indicator, M_{it} , can be described with potential outcomes notation. Let $Y_{it}(d, m)$ denote the potential outcome of electric-

ogy.

ity consumption in home *i* at time *t* if the treatment indicator is fixed at $d \in \{0,1\}$ and the move indicator is fixed at $m \in \{0,1\}$. The observed outcome is thus related to the observed treatment and move indicators according to the following expression,

$$Y_{it} = (1 - M_{it})(D_{it}Y_{it}(1,0) + (1 - D_{it})Y_{it}(0,0)) + M_{it}(D_{it}Y_{it}(1,1) + (1 - D_{it})Y_{it}(0,1)).$$
(1)

2.2 Mechanisms

Our analysis considers two broad classes of mechanisms that could give rise to the long-term effectiveness of the HER. The first mechanism is a change in the stock of habits or skills in the resident of a home. Let $H_{it}(d, m)$ denote a measure of this stock in the resident of home *i* at time *t* when the treatment and move indicators are fixed at $d \in \{0,1\}$ and $m \in \{0,1\}$. The second mechanism is a change in the stock of energy efficient technology in the home. Let $K_{it}(d,m)$ denote a measure of this stock in the home *i* at time *t* when the treatment and move indicators are fixed at $d \in \{0,1\}$ and $m \in \{0,1\}$ and $m \in \{0,1\}$. For simplicity of notation, and without loss of generality, we assume both of these stock variables are measured in units of electricity consumption.

We assume a linear relationship between habits and technology in the production of the potential outcomes,

$$Y_{it}(d,m) = H_{it}(d,m) + K_{it}(d,m) + V_{it},$$
(2)

where the variable V_{it} captures features that are relevant to electricity consumption but invariant to receipt of the HER and the decision to move, such as the weather. Some of these features may be observable, in which case we can express $V_{it} = \gamma X_{it} + U_{it}$, where X_{it} is a vector of observables and U_{it} is unobserved. The linear formulation in equation 2 is a plausible approximation of the true relationship given that the HER targets small changes in behavior that would be locally linear under a more general all-causes model of the potential outcomes.

2.3 Parameters of Interest

The objective of our analysis is to decompose the long-term effectiveness of the HER into components that can be attributed to changes in habits and technology. Accordingly, we have three parameters of interest: The long-term average treatment effect, the long-term average treatment effect attributable to changes in habits, and the long-term average treatment effect attributable to changes in technology.

The first parameter describes the effectiveness of the HER after a home and its initial resident have been exposed to the HER for a long period of time. We refer to this parameter as the long-term average treatment effect, or *ATE* for short, and define it as,

$$ATE \equiv E[Y_{it}(1,0) - Y_{it}(0,0)|t > l^*],$$
(3)

where $E[\cdot]$ is the expectations operator and l^* is a threshold that denotes long-term exposure to the HER. We delay characterizing this threshold until Section 3.2, as the theory underlying our identification strategy only requires the existence of such a threshold.

The second and third parameters of interest respectively capture the extent to which the effectiveness of the HER was caused by a change in the stock of habits and skills in the residents (H_{it}) or a change in the stock of technologies in the home (K_{it}). The relationship between these parameters and the *ATE* is obtained by plugging equation 2 into the definition of the ATE,

$$ATE = E[H_{it}(1,0) - H_{it}(0,0)|t > l^*] + E[K_{it}(1,0) - K_{it}(0,0)|t > l^*]$$

= ATH + ATK, (4)

where the parameters $ATH \equiv E[H_{it}(1,0) - H_{it}(0,0)|t > l^*]$ and $ATK \equiv E[K_{it}(1,0) - K_{it}(0,0)|t > l^*]$ capture the effect of the HER on electricity consumption that is mediated by habits and technology, respectively.

2.4 Assumptions and Identification

The primary challenge in identifying our parameters of interest is that habits and technology are unobserved. This challenge can be overcome by using the post-move effect of the HER to point identify the effect of the HER on technology (ATK). Netting the ATK out of the pre-move effect allows for the point identification of the effect attributable to habits (ATH).

The validity of this approach depends on three assumptions, which we present below. The first assumption requires that treatment assignment did not influence residents' decisions to move from a home in the experimental sample. More formally, it requires that the potential outcomes are mean independent of the treatment indicator and the time at which the initial resident moves,

$$E[Y_{it}(d,m)|D_{it},\tilde{M}_{i},X_{it}] = E[Y_{it}(d,m)|X_{it}] \text{ for } d,m \in \{0,1\},$$
(5)

where, henceforth, conditioning on the long-term is left implicit. The assumption in equation 5 has been implicitly invoked in every analysis of HER experiments. It holds if, conditional on the observables in the vector X_{it} , receipt of the HER was randomized and the decision to move homes was not made with reference to the HER.

The second assumption requires that treatment assignment did not influence the habits of the residents who moved into a home in the experimental sample. In other words, after the initial resident moves, the habits of the subsequent resident are balanced across treated and controlled homes. Because the HER was immediately discontinued after the initial resident moved, this assumption restricts sorting behavior. Formally, it imposes the following restriction,

$$E[H_{it}(1,1)|X_{it}] = E[H_{it}(0,1)|X_{it}].$$
(6)

If $E[H_{it}(1,1)|X_{it}] > E[H_{it}(0,1)|X_{it}]$, then high-energy users would be more likely to sort into treated homes and our research design would over-estimate the effect of the HER attributable to habits. If, instead, $E[H_{it}(1,1)|X_{it}] < E[H_{it}(0,1)|X_{it}]$, then low-energy users would be more likely to sort into treated homes and our research design would over-estimate the effect of the HER attributable to technology.

Before describing an empirical test of this second assumption, or how it may be relaxed, we consider its plausibility. If the HER did not alter premove technology adoption, then it is unclear how the subsequent resident of a treated home would have sorted on the basis of the initial resident's treatment status. In the United States, housing transactions are conducted at arm's length (i.e., buyers and sellers do not directly interact) and information on electricity consumption is costly to obtain (for more on this, see Section 3.2). Even if such information were freely available, buyers would have to mistakenly infer that it signaled something about the home rather than the departing resident.

If the HER did alter technology adoption, then sorting on the basis of habits would be difficult unless the HER led to significant upgrades that were salient to potential buyers of the home. While the average effect of the HER is small, this could belie significant upgrades that are undertaken by a small proportion of recipients. However, Allcott and Rogers (2014) examine data on upgrades that are likely to capture significant changes. They find the HER increased the likelihood of an upgrade by 0.4 percent and this increase can only explain 0.3 to 1.7 percent of the HER's long-term effectiveness. That is, Allcott and Rogers (2014) examine whether the HER altered participation in energy efficient subsidy programs run by utilities and find effects that explain a negligible proportion of the HER's effectiveness. Allcott and Rogers (2014) also argue that, while this participation data likely misses small upgrades because the subsidies are small for these investments and the process to receive the subsidies onerous, it is more reliable for more significant investments. More significant investments are more likely to reliably appear in the participation data because the subsidies are large and the contractors who make the investments handle the process of receiving the subsidy. Therefore, it is unlikely that the HER caused a sufficient number of significant upgrades to facilitate sorting on the basis of habits.

Despite the plausibility of the second assumption, we cannot rule out the possibility of sorting on habits. Thus, we take two approaches. First, we argue in Section 2.6.2 that a less restrictive version of the balanced habits assumption in equation 6 allows for partial identification of our parameters of interest. Second, in Section 4.4.1 we examine whether heterogeneity in the pre- and post-move effect of the HER are consistent with the balanced habits assumption.

The third assumption requires that the effect of the HER on technology adoption remains, or is stable, after the initial resident moves. Formally, this assumption implies that,

$$E[K_{it}(1,0) - K_{it}(0,0)|X_{it}] = E[K_{it}(1,1) - K_{it}(0,1)|X_{it}].$$
(7)

Intuitively, this assumption requires that a move does not cause the technol-

ogy adopted in response to the HER to exit the home, depreciate, or spread to control homes. The implications of this assumption are consistent with the HER overcoming persistent frictions in the adoption of long-lived energy efficient technology. Yet, we can also formulate a less restrictive version of the assumption that allows for partial identification of the parameters of interest. We consider this less restrictive assumption in greater detail in Section 2.6.2.

Under these three assumptions, we can use the effect of the HER before and after the initial resident moves to point identify the *ATE*, *ATH*, and *ATK*. The *ATE* is identified with the pre-move effect of the HER,

Pre-Move Effect
$$\equiv E[Y_{it}|D_{it} = 1, M_{it} = 0] - E[Y_{it}|D_{it} = 0, M_{it} = 0]$$

= ATE.

The *ATK* is identified with the post-move effect of the HER,

Post-Move Effect
$$\equiv E[Y_{it}|D_{it} = 1, M_{it} = 1] - E[Y_{it}|D_{it} = 0, M_{it} = 1]$$

= *ATK*.

The *ATH* is inferred by netting out the effect attributable to technology (*ATK*) from the total effect (*ATE*). Next we describe our strategy for estimation and inference.

2.5 Estimation

We estimate our parameters of interest with the following linear model,

$$Y_{it} = \beta D_{it} (1 - M_{it}) + \delta D_{it} M_{it} + \gamma' X_{it} + U_{it},$$
(8)

where D_{it} is the treatment indicator in the long-term (i.e., $t > l^*$), X_{it} is a vector of observables, and U_{it} is the unobservable. Linking our parameters

of interest to the coefficients in equation 8 is straightforward. The pre-move effect of the HER is β , which corresponds to the *ATE*, and the post-move effect of the HER is δ , which corresponds to the *ATK*. *ATH* is then inferred from $\beta - \delta$. We conduct inference on these estimated parameters with heteroskedasticity-robust standard errors clustered by home.

2.6 Additional Notes

2.6.1 Timing of Moves

Over the course of an HER experiment, moves happen at different times and the timing of a move can influence the weight each home receives in the estimate of pre- and post-move HER effects (see, e.g., Goodman-Bacon, 2021). To evaluate whether our estimates are influenced by the timing of moves, we re-estimate the coefficients in equation 8 using the stacked difference-indifference procedure that Deshpande and Li (2019) developed. That is, we estimate equation 8 with panel datasets constructed to observe each home in the baseline, comparison, and move periods for the same amount of time.

2.6.2 Partial Identification

As described in Section 2.4, the validity of our identification strategy relies on three assumptions. Here, we describe how relaxing the second and third assumptions allows for the partial identification of the *ATH* and *ATK*. Recall that the second assumption requires the expected post-move habits to be equal across treated and control groups (i.e., no sorting based on habits) and the third assumption requires that the effect of the HER on technology adoption remains in the home after the initial resident moves.

As discussed in Section 2.4, the second assumption would likely hold if the HER did not alter pre-move technology adoption. If, however, the HER did alter pre-move technology and, in response, post-move residents sorted into homes based on their habits, then we believe that the most likely pattern would be that residents with a habit for higher levels of electricity consumption would sort away from control homes and towards treated homes because the return on energy efficiency investments is increasing in expected electricity consumption. In this case $E[H_{it}(1,1)|X_{it}] \ge E[H_{it}(0,1)|X_{it}]$ and our parameters of interest can be partially identified, with the pre-move effect still point identifying the *ATE*, the post-move effect identifying the lower bound of *ATK*, and the *ATE* net of *ATK* identifying the upper bound on *ATH*.

However, we cannot rule out the possibility that the HER altered premove technology in small but salient ways and that these alterations caused a different pattern of sorting on habits. For example, it is possible that environmental preferences are strong enough to cause post-move residents with a habit for lower levels of electricity consumption to sort away from control and towards treated homes. In Section 4.4.1 we consider whether this possibility is consistent with different sources of heterogeneity in the pre- and post-move effects of the HER.

The third assumption would be violated if moving causes the technology adopted in response to the HER to exit the home, depreciate, or spread to control group homes. However, all three of these possibilities suggest $E[K_{it}(1,0) - K_{it}(0,0)|X_{it}] \ge E[K_{it}(1,1) - K_{it}(0,1)|X_{it}]$, which allows for the partial identification of *ATH* and *ATK*. In other words, the pre-move effect of the HER would still point identify the *ATE*, the post-move effect would identify the lower bound of *ATK*. The *ATE* net of *ATK* yields an upper bound on *ATH*.



Figure 1: Example of Home Energy Report (HER)

Note: The figure presents the front and back of the Home Energy Report (HER). Before moving, treatment households receive HERs regularly (monthly, bi-monthly, or quarterly).

3. Background

In this section, we describe the administration of HER experiments and provide a statistical description of our mover sample.

3.1 Administration of Home Energy Report Experiments

Our analysis uses data from 38 natural field experiments administered by a company called Opower. These HER experiments were conducted between 2008 and 2013 with customers of 21 different residential energy providers across the United States. Figure 1 presents an example of an HER, which compared home and neighborhood electricity consumption, described conservation tips, and provided information on energy-efficient technology adop-

tion.

Each of the 38 HER experiments, or waves, used the same design, which is summarized in Figure 2. Homes were observed in the baseline period for twelve billing months and then randomly assigned to a treatment or control group. Homes then entered the comparison period, wherein Opower generated HERs for both groups, but only mailed the HER to treatment group households. Across the 38 waves, the HER was received monthly, bimonthly, or quarterly. We follow Allcott and Rogers (2014, pp. 3021) and pool across the HER frequency margin. Homes exited the comparison period and entered the move period when the initial resident deactivated their electricity service. Upon deactivation, generation of HERs ceased and the home was made ineligible for waves of HER experiments.

3.2 Description of Mover Sample

Our data were obtained via a data sharing agreement with Opower. These data allow us to observe: (i) the electricity bills of homes in each wave, (ii) treatment and control group assignment, (iii) the timeline of HER administration in each wave, (iv) the date on which a household deactivated their electricity service, and (v) household characteristics such as whether the home was a rental.

These data consist of 58,733,360 electricity bills for 1,810,096 homes. Each electricity bill includes the total consumption of electricity in kilowatt hours (kWh) and the length of the billing cycle. On average, an electricity bill covers 30 days, but this coverage varies. Our outcome measure adjusts for this variation by normalizing the electricity consumption by the length of the billing cycle, making average daily consumption over the course of a billing cycle our observed outcome.

To study the effect of the HER that remains in the home after the initial

ER Experiment	Move Period	If Assigned to Treatment: Initial Resident Moves Out, Home Ceases Receiving HER, & Subsequent Resident Moves In	If Assigned to Control: Initial Resident Moves Out & Subsequent Resident Moves In
f Homes in Mover Sample of HI	Comparison Period	<i>If Assigned to Treatment:</i> Home Periodically Receives HER	<i>If Assigned to Control:</i> Home Is Left Untouched
2: Timeline of		peq	
Figure	Baseline Period	Home Randomly Assign to Treatment or Control	Group

Note: This figure describes the three periods of an HER experiment for the mover sample. In the baseline period, homes are randomly assigned to a treatment or control group. In the comparison period, treatment group homes periodically receive the HER mailer and control group homes are left untouched. In the move then the subsequent resident moves. Control group homes in the move period see the initial resident move period, the receipt of the HER is ceased for treatment group homes once the initial resident moves out and out and then have the subsequent resident moves into the home.

Mover Sample (pp) (1)	Baseline Period kWh/day (2)	Frob. Home is Rental (pp) (3)	Prob. Heat is Elec. (pp) (4)	Months in Comparison Period (5)	Months in Move Period (6)	Comparison on 1st HER (% Diff.) (7)
(0.03)***	0.07)***	13.01	0.26	10.10	12.76	24.91
(0.03)***	(0.07)***	(0.14)***	(0.14)***	(0.04)***	(0.04)***	(0.28)***
0.01	0.11	-0.06	0.26	0.07	-0.08	-0.08
(0.04)	(0.09)	(0.18)	(0.18)	(0.05)	(0.05)*	(0.36)
Full	Mover	Mover	Mover	Mover	Mover	Mover
1,810,096	139,908	139,908	139,908	139,908	139,908	139,908

Table 1: Summary Statistics of Mover Sample

Notes: This table summarizes the characteristics of the mover sample. The first row reports the average value of each characteristic for homes assigned to the control group and the second row reports treatment group differences from the control group. Estimates regression-adjust for each HER wave. The first column reports the rate at which the full sample enters the mover sample. The subsequent columns report characteristics of the mover sample. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** p-value < 0.01, ** p-value < 0.05, * p-value < 0.10. resident moves, we construct a sample of movers from this data. This sample is comprised of homes that had a deactivation of the initial resident's account with their energy provider. Working with Opower, we eliminated homes where the deactivation was prompted by a name change or other changes unlikely to reflect a move by the initial resident.

We further restrict the mover sample to homes where deactivations occurred at or after the fourth HER had been received.³ We base this restriction on results in Allcott and Rogers (2014), which indicate that the effect of the HER plateaus around the receipt of the fourth HER. On average, the fourth report was generated 145 days, or approximately five months, after the start of the comparison period. We denote this subsample the "mover sample", which includes 5,768,148 electricity bills for 139,908 homes.

Table 1 provides a statistical summary of the mover sample. This summary presents averages of different features of the sample after regression adjusting with a dummy for each wave of an HER experiment. The first column shows that the mover sample is comprised of approximately 8 percent of the treatment and control group homes from the full sample. Subsequent columns show that, on average, mover sample homes consume about 38 kWh/day in the baseline period, nearly 14 percent of the mover sample homes are rental properties, and nearly 14 percent use electricity for heating. The average time spent in the comparison period by the mover sample is 16 months and nearly 13 months is spent on average in the move period. The first HER generated has the mover sample consuming an average of nearly 25 percent more energy than their neighbors ("Average neighbors" in Figure 1), likely because high-consumption homes were oversampled in early HER experiments (Allcott, 2015).

³This restriction can be applied to both treatment and control group homes, because, as noted above, Opower created HERs for both groups, but only sent out the mailers to treatment group homes.

Table 1 also provides evidence that supports the first assumption of our identification strategy. Treatment and control group homes select into the mover sample at statistically indistinguishable rates and these homes consume similar quantities of electricity in the baseline period. Furthermore, the two groups have a similar prevalence of rental arrangements and electric heating, spend similar amounts of time in the comparison and move periods, and on their first HER differ from the average electricity consumed by neighbors at similar levels (recall that treatment and control group HERs can be compared because they were generated for all homes, but only sent to treatment group homes).

We conclude this subsection by considering the extent to which informational frictions inhibited sorting into homes on the basis of whether a home once received the HER. The prevalence of these frictions lends credibility to the second assumption of our identification strategy. While mandates increasingly try to overcome informational frictions on energy consumption by requiring sellers to disclose energy bills to potential buyers (Palmer and Walls, 2017), only one mandate affected our experimental sample. Using Table 1 in Palmer and Walls (2017) and zip codes shared with us by Opower, we find that only 274 homes, or 0.02 percent of our mover sample, were affected by such a mandate. As a result, we conclude that movers were unlikely to have one important source of information that would have facilitated the type of sorting that would violate the second assumption of our identification strategy.

4. **Results**

This section presents estimates of the effects that decompose the long-term effectiveness of the HER into components attributable to habits and technology.

4.1 Event Study

Before we present estimates of the effects that decompose the long-term effectiveness of the HER, we investigate the underlying dynamics with an event study analysis of the mover sample. Our analysis divides time into six-month intervals in the baseline, comparison, and move periods. Each estimate is normalized by the level of control group electricity consumption in the baseline period (see column 2 in Table 1) and 95 percent confidence intervals are constructed with heteroskedasticity-robust standard errors clustered by home.

Figure 3 presents the results. Starting from the left of the figure we see an average difference between treatment and control group homes of approximately 0.3 percent in the baseline period. Scaling this difference by 38.1 kWh/day converts it to an estimated effect of 0.1 kWh/day. Such an effect is small: Approximately equivalent to treatment group homes using a 60-watt incandescent lightbulb for an extra two hours each day. Moreover, the confidence interval on this estimate shows it is statistically indistinguishable from an effect of zero. This balance in baseline period electricity consumption provides further support for the mean independence assumption discussed in Section 2.4.

Moving to the right of the first vertical line, which denotes the end of the baseline period and the start of the comparison period, the average effect starts by falling significantly, plateauing, and then rising modestly as the move period approaches. The negative sign on these estimates indicates the HER causes a reduction in household electricity consumption. Figure 3 reports an average effect that starts at -1.8 percent over the first six months of the comparison period, which then falls and plateaus at -2.5 to -2.7 over the subsequent year. This dynamic of an initial fall and subsequent plateau is consistent with the pattern documented in Allcott and Rogers (2014), where



Figure 3: Event Study of Average HER Effect on Mover Sample

Note: This figure reports estimated treatment effects on the mover sample. Each estimated effect is the average effect of treatment assignment at a given point in time. Each effect is presented in terms of percent changes relative to control group electricity consumption in the baseline period. Time is divided into six-month intervals. Observations that fall outside of the plotted intervals are assigned to an absorbing interval indicated on the figure with the superscript *a*. The omitted time period is the last six months of the baseline period. Brackets denote the 95 percent confidence interval. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with fixed effects for each six-month interval of event time, home, and year-by-season-by-wave. 95 percent confidence intervals are constructed with heteroskedasticity-robust standard errors clustered by home.

the average effect grows until the fourth HER and then plateaus. On average, our sample receives their fourth HER around the fifth month of the comparison period. Moving to the final year of the comparison period, the average effect of the HER starts to rise modestly, increasing to -2.2 and then -1.6 percent.

The average effects in the comparison period are -1.6 to -2.8 percent, which corresponds to approximately -0.6 to -1.0 kWh/day in levels. To put the magnitude of these estimates into perspective, such an effect is equivalent to treatment group homes using a 60-watt incandescent lightbulb for 10 to 17 fewer hours per day or replacing 2 to 4 60-watt incandescent lightbulbs that are used 5 hours per day with the CFL equivalent. The statistical significance of these effects can be seen by noting that the 95 percent confidence intervals do not overlap with zero in the comparison period. Economically, the average comparison period effects are also significant. Estimates of the price elasticity of electricity range from -0.07 to -0.30 (IEA, 2012), suggesting that utilities would have to increase the price of electricity by 5 to 39 percent to obtain the same effects reported over the comparison period in Figure 3.

Moving to the right of the second vertical line on Figure 3, we see that much of the average effect of the HER found in the comparison period persists in the move period. Over the first six months of the move period the HER continues to produce reductions in electricity consumption of -1.2 percent. The final estimate of Figure 3 shows that more than six months after moving, the estimated average effect is a -1.0 percent reduction in average electricity consumption.

The average effects in the move period are -1.0 to -1.2 percent, which equate to effects of approximately -0.4 to -0.5 kWh/day in levels. The 95 percent confidence intervals on these estimates show that the null hypothesis of no effect during the move period is rejected. Using estimates of the price elasticity of demand from IEA (2012), utilities would have to increase the

price of electricity by 3 to 17 percent to produce the same effects reported after the move period starts in Figure 3. The average effects in the move period are also significant when compared to the average effects in the comparison period, with 36 to 75 percent of the HER's average effect persisting in the home after the initial resident moves.

4.2 Parameters of Interest

Table 2 presents the estimates of the effects that decompose the long-term effectiveness of the HER. The first column presents the estimated pre-move effect of the HER for the full sample of homes. The second column presents the estimated pre- and post-move effects for the mover sample and the third column investigates heterogeneity in the pre- and post-move effects for the mover sample.

The estimated effects in the first two columns of Table 2 indicate that the majority of the HER's long-term effectiveness can be attributed to increases in technology adoption, with the remainder attributable to the formation of habits. To see how we reach this conclusion, recall that the pre-move effect of -2.1 percent in the first and second column of Table 2 identifies the long-term average treatment effect of the HER, i.e., the *ATE*.⁴ The post-move effect of -1.1 percent in the second column of Table 2 identifies the component of the long-term effect attributable to technology adoption, i.e., the *ATK*. Netting out the component attributable to technology identifies the component attributable to habits, which we call the *ATH*. For the mover sample the estimated component attributable to habits is -1.0 percent.

Normalizing these components by the *ATE* implies that 51.9 percent (*s.e.* = 13.1) of the long-term effectiveness is attributable to technology and

⁴Furthermore, the similarity of these estimates when estimated with the full and mover sample provides support for a stronger version of the mean independence assumption discussed in Section 2.4 that extends to selection into the mover sample.

	Electricity Con	s. (% of Contro	ol in Baseline)
	(1)	(2)	(3)
Pre-Move Effect	-2.14	-2.14	-2.55
	$(0.04)^{***}$	$(0.18)^{***}$	$(0.26)^{***}$
Post-Move Effect		-1.11	-1.01
		$(0.29)^{***}$	$(0.40)^{**}$
$Pre \times Elec.$ Heat			-3.04
			$(0.64)^{***}$
Post \times Elec. Heat			-2.35
			$(0.97)^{**}$
$Pre \times Rental$			0.29
			(0.51)
Post \times Rental			1.38
			$(0.77)^*$
Pre \times 1st Comp.			1.59
			$(0.34)^{***}$
Post \times 1st Comp.			0.50
			(0.54)
Sample	Full	Mover	Mover
Bills	58,733,360	5,768,148	5,768,148
Homes	1,810,096	139,908	139,908
R^2	0.63	0.54	0.59

Table 2: Average Effect of HER

Note: This table reports coefficients estimated with variants of equation 8. The coefficients respectively measure the average effect of treatment assignment after the fourth HER in the comparison period (pre-move effect) and in the move period (post-move effect). Each coefficient is presented in terms of percent changes relative to control group electricity consumption in the baseline period. Column 1 is estimated on the full sample and columns 2-3 on the mover sample. Column 3 interacts the pre- and post-move treatment indicators with whether a home has electric heating (elec. heat), whether a home is a rental (rental), and whether the first HER indicates that the home used less electricity than its neighborhood average (1st comp.). Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with fixed effects for each period of time, home, and year-by-season-by-wave. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

48.1 percent (s.e. = 13.1) is attributable to habits.

4.3 Heterogeneity

In this subsection we investigate heterogeneity in the components attributable to the long-term effectiveness of the HER. This investigation serves two goals. First, prior research finds significant heterogeneity in the effectiveness of the HER and applying our decomposition can help characterize the underlying causes of this heterogeneity. Second, under the identifying assumptions of our research design, some sources of heterogeneity should have predictable effects on our estimates. Finding these predicted effects would lend credibility to the identifying assumptions of our research design. In the remainder of this subsection, we focus on the first of these goals and discuss the second in the subsection that follows.

We investigate heterogeneity by estimating a variant of equation 8 that adds interactions between several covariates, the pre- and post-move treatment indicators, and the vector of observables that act as controls (i.e., we estimate a saturated regression). The covariates we interact measure whether a home uses electricity for heating, whether a home is a rental, and whether the first HER presents a favorable comparison of a home's electricity consumption to the neighborhood average. That is, whether the first HER reports that the home consumed less electricity than the neighborhood average. Prior research finds that electric heating increases the magnitude of the HER's effect, the effect is unaltered by the rental status of a home, and a favorable comparison on the first HER reduces the magnitude of the effect of the HER (Allcott, 2015). Unfortunately, our data do not allow us to address heterogeneity driven by whether a home has a pool, the square footage of the home, and the political ideology of the home's initial resident (Ayres et al., 2013; Costa and Kahn, 2013; Allcott, 2015).⁵

Column 3 of Table 2 presents the estimates of the coefficients. The first two respectively indicate average pre- and post-move effects of -2.6 and -1.0 percent for the omitted category, which is owner occupied homes with gas heating that used more electricity than the neighborhood average on the first HER. The subsequent coefficients report the extent to which changes to the covariates alter these effects. Table A1 presents estimates on each mover subsample separately.

The coefficients on the interaction terms with the electric heat covariate suggest that it increases the effectiveness of the HER because of technology adoption. As prior research has found, changing the energy source that a home uses for heating from natural gas to electricity significantly increases the magnitude of the pre-move effect. Column 3 of Table 2 shows that in homes with electric heating the pre-move effect drops by a statistically significant -3.0 percent and most of this drop persists after the initial resident moves, with a statistically significant post-move effect for homes with electric heating of -2.4 percent.

Interacting the rental status of a home with the pre- and post-move treatment indicators suggests that the effectiveness of the HER among rental properties is attributable to habit formation. Consistent with prior research, column 3 of Table 2 shows that going from an owner-occupied to a rental

⁵ In a separate analysis, we consider heterogeneity related to the total number of HERs generated before the initial resident moves. Figure A₂ plots the average effect of the HER by quartile of HERs generated. The top panel plots the effect of the first 4 HERs, the next panel plots the pre-move effect, and the middle panel plots the post-move effect. The post-move effect in this figure is limited to the six months after the initial resident moves so that similar lengths of time in the move period can be compared across the quartiles. The final two panels respectively plot the effect of the first 4 HERs and the pre-move effect divided by the post-move effect. Two sets of results stand out. First, the pre-move effect of the HER is statistically significantly larger for the second quartile of HERs generated than the other quartiles, as the 95 percent confidence intervals do not overlap. Second, the point estimate of the post-move effect for the first quartile is 0.06 percent, suggesting that technology was not adopted in response to the HER among the homes in the sample that received a total of 3 or fewer HERs.

home has an effect of 0.3 percent, which is statistically indistinguishable from zero. However, after the initial renter moves, the effect of the HER remaining in the home is statistically indistinguishable from zero. The coefficient on the interaction between the post-move treatment indicator and rental status is 1.4 percent, which is similar in magnitude to the post-move effects in columns 2 and 3 -1.0 to -1.1 percent. Furthermore, the estimate of 1.4 percent on the interaction term is on the margin of statistical significance, with a *p*-value = 0.07.

The estimates on the interaction terms with the first HER's comparison suggests that a favorable comparison reduces the effectiveness of the HER because habits are less likely to form. Column 3 of Table 2 shows that, consistent with prior research, moving from an unfavorable to a favorable comparison on the first HER reduces the effectiveness of the HER, with a statistically significant estimate of 1.6 percent for the coefficient on the interaction between the pre-move treatment indicator and a favorable comparison on the first HER. However, the favorability of the first comparison has a statistically null effect on the post-move effect, with an estimated coefficient of 0.5 percent on the corresponding interaction term.

Beyond offering an explanation for the underlying causes of heterogeneity in the effectiveness of the HER, our investigation of heterogeneity also lends credibility to our research design. We discuss this, as well as the the robustness of our estimates more generally, in the next subsection.

4.4 Robustness

4.4.1 Identifying Assumptions

The validity of our decomposition depends on three assumptions that we presented in Section 2.4: (1) mean independence between the potential outcomes, moving, and receipt of the HER; (2) treatment assignment did not

influence the habits of the residents who moved into a home ("balanced habits"); and (3) technology adoption in response to the HER was stable (i.e., remains in the home) after the initial resident moved ("technology stability"). In Section 3.2, we presented data consistent with the mean independence assumption. In Section 2.6.2, we argued that violations of the balanced habits and technology stability assumptions likely imply that our estimated effects identify a lower bound of the *ATK* and an upper bound of the *ATH*. In other words, by assuming that habits are balanced and technology adoption is stable, we likely obtain a conservative estimate of the contribution from technology adoption to the HER's long-term effectiveness, which reinforces our conclusion that the majority of the long-term effectiveness of the HER is due to technology.

Here, we argue that the results of our heterogeneity analysis in the third column of Table 2 are easily explained if the balanced habits and stable technology assumptions hold. Under these assumptions, our heterogeneity analysis finds that a larger proportion of the HER's long-term effectiveness is attributable to technology adoption in homes with electric heating than in homes with natural gas heating. This is predictable because homes with electric heating can conserve electricity by adopting technologies, like a door sweep on the home's front door, that would have a negligible effect on electricity consumption for homes with natural gas heating. Our heterogeneity analysis also finds the HER's long-term effectiveness is entirely attributable to habit formation for rental properties. This makes sense because rental arrangements constrain the technology that residents can install and diminish the incentive for residents to adopt the technology they are allowed to install (e.g., Davis, 2012).

The results of our heterogeneity analysis are also difficult to explain with alternatives to our identifying assumptions. The most concerning alternative would have the HER cause post-move residents with a habit for lower levels of electricity consumption to sort away from control homes and towards treated homes. This alternative is concerning because it would cause our identification strategy to overstate the role of technology and understate the role of habits in the HER's long-term effectiveness. However, under this alternative assumption, it is not clear why larger pre-move effects persist after the initial resident moves out of a home with electric heating, but not out of a home that has an unfavorable comparison generated on their first HER.

Collectively, the results of our heterogeneity analysis lend credibility to the assumptions underlying our identification strategy and the results of our decomposition. However, these results are not definitive. We conjecture that comparing the pre-move electricity consumption of the residents who move into treatment and control homes would help evaluate the possibility of sorting on habits, but our data does not allow us to make this comparison because it only identifies homes, not residents too.

4.4.2 Alternative Specifications of Controls

A basic concern with the estimates reported in Table 2 is that instead of capturing the effect of the HER, they reflect our decision to use fixed effects for each period of time, home, and year-by-season-by-wave as controls. Across a series of appendix tables we address this concern by showing that our results are robust to alternative control variable specifications. Table A4 demonstrates that the robustness of the results in the second column of Table 2. With respect to the results in the third column of Table 2, Tables A5, A6, and A7 report the same robustness for the subsamples of homes with electric heating, rentals, and homes where the comparison on the first HER was unfavorable. While different specifications of controls alter the level of the pre- and post-move effects, the same pattern of results reported in Table 2 are supported.

4.4.3 Timing of Moves

Another concern with the results in Table 2, which we touched on in Section 2.6.1 above, is that over the course of an HER experiment moves occur at different points in time and this can influence the weight that each home receives in the estimates of the pre- and post-move HER effects. To determine whether these weights influence our results, we estimate our parameters of interest with a stacked difference-in-difference approach (Deshpande and Li, 2019). That is, we estimate our parameters of interest with datasets that we construct where homes are observed for the same amount of time in the baseline, comparison, and move periods. Table A11 reports the pre- and post-move effects with samples where each home is observed for 365 days in the baseline period and 91, 192, 273, or 365 days in the comparison and move periods. Table A12 reports estimates with the same samples using the weighting procedure for stacked difference-in-difference models proposed by Wing et al. (2024). Across both tables, the same pattern of results found in Table 2 are obtained.

4.4.4 Mover Sample Construction

Yet another concern with the results in Table 2 is that instead of capturing our parameters of interest, they reflect our decision to limit the mover sample to homes where the initial resident moved after at least four HERs had been generated. Table A13 shows that alternative cutoff rules produce similar results. A related concern is that homes may sit idle at the start of the move period and if there happened to be slight imbalance in the likelihood of homes sitting idle, are results would be confounded. Table A14 shows that we still estimate statistically significant pre- and post-move effects when we drop homes that experience a 1, 2, or 3 standard deviation decrease in their move period electricity consumption.

5. Implications

Having presented our decomposition of the channels underlying the longterm effectiveness of the HER, we next consider broader implications of our findings for nudges.

5.1 Explaining the Persistent Effects of Social Comparison Nudges

The persistence of social comparison nudges in prior studies varies dramatically across contexts. Figure 4 presents the average effectiveness of these nudges one year after their discontinuation, with the estimates normalized by the average effect before discontinuation. The divergence in persistence across contexts can be seen by comparing the top and bottom panels of the figure. The top panel plots the average persistence when a social comparison nudge targets compliance, charitable giving, financial savings, or voter turnout. On average, just 4 percent of the initial effect of these social comparison nudges persists one year after discontinuation. In contrast, when a social comparison nudge targets water or energy conservation, 65 percent of the effectiveness, on average, remains a year after discontinuation.

Our decomposition results suggest a simple explanation for these divergent levels of persistence: The relative abundance of technologies for conserving energy and water. Recall that our decomposition of the HER's longterm effectiveness implies that 51.9 percent was attributable to technology adoption. We plot this estimated effect in Figure 4 and label it Mover Sample. As can be verified in the figure, this channel alone produces a level of persistence that is similar in magnitude to the total persistence arising from nudges to energy and water conservation and is much large in magnitude to the total persistence produced by nudging behaviors that are not easily



Figure 4: Effectiveness of Social Comparison Nudges After Discontinuation

Note: This figure presents the average effect of a social comparison nudge one year after it was discontinued. When such an estimate is not reported in a study, we fit an exponential decay model on the data presented in Online Appendix Figure A1. We then use the effect of the nudge one year after discontinuation that is predicted by the exponential decay model. Each effect is normalized by the average effect before discontinuation. The mover sample estimate includes the 95 percent confidence interval.

modified by technology adoption, such as voting. We interpret this pattern as indicative of a central role for technology adoption in the persistence of treatment effects after the discontinuation of a social comparison nudge.

This interpretation is, on the surface, at odds with prior research. Allcott and Rogers (2014) uses participation in utility sponsored energy efficiency programs as a proxy for technology adoption and find that technology adoption explains no more than 2 percent of the HER's long-term effectiveness. In the same vein, Bernedo et al. (2014) finds that, after the initial resident moves, the effect of a social comparison nudge does not lead to statistically significant savings in water consumption, and the authors conclude that technology adoption is not an important mechanism underlying persistence. We, however, reject these conclusions based on our decomposition results. Using conventional levels of statistical significance, the 51.9 percent that we attribute to technology adoption is estimated precisely enough to reject the 2 percent attributed to technology adoption by Allcott and Rogers (2014) and the null effect reported by Bernedo et al. (2014). We believe that the imperfect proxy for technology adoption used by Allcott and Rogers (2014) and the low statistical power of the analysis by Bernedo et al. (2014) can explain why their findings diverge from the results of our decomposition.

5.2 Net Benefits of Nudges

Our decomposition of the HER's long-term effectiveness also highlights a limitation of past evaluations of nudge-style interventions. These evaluations have compared the effectiveness of a nudge to the cost of their administration (Allcott and Mullainathan, 2010; Allcott and Rogers, 2014; Benartzi et al., 2017).⁶ This approach to calculating the costs of nudges implicitly assumes

⁶An additional approach implemented in Allcott and Kessler (2019) and Butera et al. (2022) elicits willingness to pay via incentivized surveys.

that there are no other financial costs created by the intervention. However, evaluations should also account for the indirect costs induced by an intervention (Heckman and Smith, 1997) and our analysis of the mover sample suggests that the HER induced costly adoption of energy efficient technology. While we have no data that allow us to infer the financial costs of the technology adopted in our mover sample, in Online Appendix Figure A3 we use different measures in the literature to illustrate the potential consequences of including such costs in net benefit calculations. This figure shows that, across the different costs of technology adoption reported in the literature (Billingsley et al., 2014; Gillingham et al., 2018), the net benefits drop by approximately 14 to 56 percent after accounting for the costs of HER-induced technology adoption.

6. Conclusion

Why do some nudges produce effects that persist and other nudges do not? This study develops a formal research design that addresses this question by decomposing the long-term effectiveness of a nudge into components attributable to habit formation and technology adoption. We apply our research design to the case of the HER, a nudge that is notable for its long-term effectiveness (see, e.g., Online Appendix Figure A1). We find that a majority of the HER effect stays in a home after the initial resident moves.

After assessing the plausibility of the identifying assumptions in our design and the robustness of our findings, we interpret our results as providing evidence for the primacy of technology adoption in the long-term effectiveness of the HER. This finding offers several contributions and points to new directions for future work.

First, our study provides a simple explanation for the divergent levels of persistence in treatment effects after social comparison nudges are discontin-

ued. The effect of a social comparison nudge is more likely to persist when the targeted behavior can be augmented by productive technologies, such as input efficient technologies to conserve energy and water. The effect is likely to persist when productive technologies are unavailable, such as in contexts where target behaviors are associated with compliance with rules, charitable giving, financial savings, tax evasion, and voting. Future work should explore the extent to which heterogeneity across experiments reflects differences in the costs or availability of productive technologies. For example, it would be fruitful to explore the extent to which differences in such costs and availability explain differences in persistence in multi-site experiments, such as Allcott and Rogers (2014) and Coppock and Green (2016).

Second, our study suggests that policymakers can replicate the long-term effectiveness of the HER in two ways. First, they can target behaviors that can be influenced by readily available technologies. Second, they can combine social comparison nudges with opportunities to adopt new technologies. For example, in the context of voting, our findings predict that the effects of social comparison nudges will persist in municipalities that provide an option to default into easier modes of voting in the future, such as mail-in voting. In the context of givings and savings, policymakers could pair social comparisons with an option for households to default to higher giving or savings rate in the future. Such defaults have been found to increase givings and savings (Madrian and Shea, 2001; Thaler and Benartzi, 2004; Goswami and Urminsky, 2016; Altmann et al., 2019), but our findings suggest combining these defaults with the framing of a social comparison will produce longer lived effects. Future work should explore this conjecture.

Third, our study illustrates the importance of accounting for the indirect costs induced by nudges. By isolating the mechanisms underlying the effectiveness of a nudge, we are able to infer one type of indirect cost, technology adoption, that is typically ignored in the evaluation of nudges. Using estimates in the literature of the financial cost of adopting energy efficient technology, we show that accounting for technology adoption attenuates previous estimates of HER net benefits by 14 to 56 percent. While this accounting exercise is highly stylized, it nonetheless illustrates how the application of our research design can isolate mechanisms that, in turn, can inform the economic evaluation of nudge-style interventions.

In addition to these three contributions, our study provides an important methodological contribution. To assess the mechanisms underlying behavioral responses to policies and programs, prior research has relied on survey measurements. However, relative to the cost of administering a nudge, a survey approach would be extraordinarily expensive. Our study thus complements previous work by developing a new research design that is well suited to isolate the mechanisms underlying the effectiveness of nudges. We imagine future research can build on this strategy. Potential applications include using the graduation of students or the separation of employees to understand the extent to which nudges, such as those respectively studied in Bettinger et al. (2012) and Earnhart and Ferraro (2021), produce human capital in the recipients of the nudge and in the organizations in which the recipients are nested.

References

- Acland, Dan and Matthew R. Levy, "Naiveté, Projection Bias, and Habit Formation in Gym Attendance," *Management Science*, 2015, *61* (1), 146–160.
- Allcott, Hunt, "Social norms and energy conservation," *Journal of Public Economics*, 2011, 95 (9-10), 1082–1095.
- _____, "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 2015, 130 (3), 1117–1165.

- and Judd B. Kessler, "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *American Economic Journal: Applied Economics*, 2019, 11 (1), 236–276.
- and Michael Greenstone, "Is There an Energy Efficiency Gap?," *Journal of Economic Perspectives*, 2012, 26 (1), 3–28.
- and Sendhil Mullainathan, "Behavior and Energy Policy," Science, 2010, 327 (5870), 1204–1205.
- and Todd Rogers, "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation," American Economic Review, 2014, 104 (10), 3003–3037.
- ____, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow, "The Welfare Effects of Social Media," *American Economic Review*, 2020, 110 (3), 629–676.
- _, Matthew Gentzkow, and Lena Song, "Digital Addiction," American Economic Review, 2022, 112 (7), 2424–2463.
- Altmann, Steffen, Armin Falk, Paul Heidhues, Rajshri Jayaraman, and Marrit Teirlinck, "Defaults and Donations: Evidence from a Field Experiment," *Review of Economics and Statistics*, 2019, 101 (5), 808–826.
- **Apesteguia, Jose, Patricia Funk, and Nagore Iriberri**, "Promoting rule compliance in daily-life: Evidence from a randomized field experiment in the public libraries of Barcelona," *European Economic Review*, 2013, *64*, 266–284.
- **Ayres, Ian, Sophie Raseman, and Alice Shih**, "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage," *The Journal of Law, Economics, and Organization*, 2013, 29 (5), 992–1022.

- Becker, Gary S., "Habits, Addictions, and Traditions," *Kyklos*, 1992, 45 (3), 327–345.
- and Kevin M. Murphy, "A Theory of Rational Addiction," Journal of Political Economy, 1988, 96 (4), 675–700.
- **Bell, Eric, Aimee Savage, John Ensley, and Robert Gottlieb**, "Evaluation of Southern California Edison's HER Persistence Pilot," Technical Report, Southern California Edison Co. 2020.
- Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing, "Should Governments Invest More in Nudging?," Psychological Science, 2017, 28 (8), 1041–1055.
- Bernedo, María, Paul J. Ferraro, and Michael Price, "The Persistent Impacts of Norm-Based Messaging and Their Implications for Water Conservation," *Journal of Consumer Policy*, 2014, 37, 437–452.
- Beshears, John, Hae Nim Lee, Katherine L. Milkman, Robert Mislavsky, and Jessica Wisdom, "Creating Exercise Habits Using Incentives: The Trade-off Between Flexibility and Routinization," *Management Science*, 2021, 67 (7), 3985–4642.
- Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment," *Quarterly Journal of Economics*, 2012, 127 (3), 1205–1242.
- Billingsley, Megan A., Ian M. Hoffman, Elizabeth Stuart, Steven R. Schiller, Charles A. Goldman, and Kristina LaCommare, "The Program Administrator Cost of Saved Energy for Utility Customer-Funded Energy

Efficiency Programs," Technical Report, Lawrence Berkeley National Lab 2014.

- Bursztyn, Leonardo, Davide Cantoni, David Y. Yang, Noam Yuchtman, and
 Y. Jane Zhang, "Persistent Political Engagement: Social Interactions and the Dynamics of Protests Movements," *American Economic Review: Insights*, 2021, 3 (2), 233–250.
- Butera, Luigi, Robert D. Metcalfe, William Morrison, and Dmitry Taubinsky, "Measuring the Welfare Effects of Shame and Pride," *American Economic Review*, 2022, 112 (1), 122–168.
- Charness, Gary and Uri Gneezy, "Incentives to Exercise," *Econometrica*, 2009, 77 (3), 909–931.
- **Choukhmane, Taha**, "Default Options and Retirement Savings Dynamics," *Working Paper*, 2021.
- **Coppock, Alexander and Donald P. Green**, "Is Voting Habit Forming? New Evidence from Experiments and Regression Discontinuities," *American Journal of Political Science*, 2016, 60 (4), 1044–1062.
- **Costa, Dora L. and Matthew E. Kahn**, "Energy Conservation "Nudges" and Environmentalist Ideology: Evidence From a Randomized Residential Electricity Field Experiment," *Journal of the European Economic Association*, 2013, 11 (3), 680–702.
- **Davis, Lucas W.**, "Evaluating the Slow Adoption of Energy Efficient Investments: Are Renters Less Likely to Have Eenergy Efficient Appliances?," in Donn Fullerton and Catherine Wolfram, eds., *The Design and Implementation of U.S. Climate Policy*, University of Chicago Press, 2012, pp. 301–316.

- **DellaVigna, Stefano and Elizabeth Linos**, "RCTs to Scale: Comprehensive Evidence from Two Nudge Units," *Econometrica*, 2022, *90* (1), 81–116.
- **Deshpande, Manasi and Yue Li**, "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy*, 2019, 11 (4), 213–248.
- **Earnhart, Dietrich and Paul J. Ferraro**, "The Effect of Peer Comparisons on Polluters: A Randomized Field Experiment among Wastewater Dischargers," *Environmental and Resource Economics*, 2021, 79, 627–652.
- **Ferraro, Paul J. and Juan José Miranda**, "Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment," *Resource and Energy Economics*, 2013, 35 (3), 356–379.
- Frey, Erin and Todd Rogers, "Persistence: How Treatment Effects Persist After Interventions Stop," Policy Insights from the Behavioral and Brain Sciences, 2014, 1 (1), 172–179.
- Fujiwara, Thomas, Kyle Meng, and Tom Vogl, "Habit Formation in Voting: Evidence from Rainy Elections," American Economic Journal: Applied Economics, 2016, 8 (4), 160–188.
- Gerarden, Todd D., Richard G. Newell, and Robert N. Stavins, "Assessing the Energy-Efficiency Gap," *Journal of Economic Literature*, 2017, 55 (4), 1486–1525.
- Gillingham, Kenneth, Amelia Keyes, and Karen Palmer, "Advances in Evaluating Energy Efficiency Policies and Programs," *Annural Review of Resource Economics*, 2018, 10, 511–532.

- **Goodman-Bacon, Andrew**, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, 225 (2), 254–277.
- **Goswami, Indranil and Oleg Urminsky**, "When should the Ask be a Nudge? The Effect of Default Amounts on Charitable Donations," *Journal of Marketing Research*, 2016, 53 (5), 829–846.
- Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev, "The behavioralist as tax collector: Using natural field experiments to enhance tax compliance," *Journal of Public Economics*, 2017, 148, 14–31.
- Heckman, James J. and Jeffrey Smith, "Evaluating the Welfare State," in Steinar Strøm, ed., *Econometrics and Economics in the 20th Century: The Ragnar Frisch Centenary*, Cambridge University Press, 1997, pp. 214–318.
- and Rodrigo Pinto, "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production Technologies with Unmeasured and Mismeasured Inputs," *Econometric Reviews*, 2015, 34 (1-2), 6–31.
- **Hummel, Dennis and Alexander Maedche**, "How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies," *Journal of Behavioral and Experimental Economics*, 2019, *80*, 47–58.
- **IEA**, "Understanding Electric Utility Consumers Summary Report: What We Know and What We Need to Know," Technical Report, Electric Power Research Institute 2012.
- _____, "The Potential of Behavioural Interventions for Optimising Energy Use at Home," Technical Report, International Energy Agency 2021.
- Jaffe, Adam B. and Robert N. Stavins, "The energy-efficiency gap: What does it mean?," *Energy Policy*, 1994, 22 (10), 804–810.

- John, Leslie K., George Loewenstein, Andrea B. Troxel, Laurie Norton, Jennifer E. Fassbender, and Kevin G. Volpp, "Financial Incentives for Extended Weight Loss: A Randomized, Controlled Trial," *Journal of General Internal Medicine*, 2011, 26, 621–626.
- Kast, Felipe, Stephan Meier, and Dina Pomeranz, "Saving more in groups: Field experimental evidence from Chile," *Journal of Development Economics*, 2018, 133, 275–294.
- Levitt, Steven D., John A. List, and Sally Sadoff, "The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment," *Working Paper*, 2016.
- Ludwig, Jens, Jeffrey R. Kling, and Sendhil Mullainathan, "Mechanism Experiments and Policy Evaluations," *Journal of Economic Perspectives*, 2011, 25 (3), 17–38.
- Madrian, Brigitte C. and Denns F. Shea, "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 2001, 116 (4), 1149–1187.
- **Palmer, Karen and Margaret Walls**, "Using information to close the energy efficiency gap: A review of benchmarking and disclosure ordinances," *Energy Efficiency*, 2017, *10*, 673–691.
- **Pollak, Robert A.**, "Habit Formation and Dynamic Demand Functions," *Journal of Political Economy*, 1970, 78 (4), 745–763.
- **Rogers, Todd and Erin Frey**, "Changing Behavior Beyond the Here and Now," in Gideon Keren and George Wu, eds., *Wiley Blackwell Handbook of Judgement and Decision Making*, John Wiley and Sons, 2016, pp. 725–748.

- _ , Donald P. Green, John Ternovski, and Carolina Ferrerosa Young, "Social pressure and voting: A field experiment conducted in a high-salience election," *Electoral Studies*, 2017, 46, 87–100.
- **Royer, Heather, Mark Stehr, and Justin Sydnor**, "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company," *American Economic Journal: Applied Economics*, 2105, 7 (3), 51–84.
- Shang, Jen and Rachel Croson, "A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods," *Economic Journal*, 2009, 119 (540), 1422–1439.
- Thaler, Richard H. and Cass R. Sunstein, *Nudge*, Yale University Press, 2008.
- and Shlomo Benartzi, "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, 2004, 112 (S1), S164–S187.
- **Vollaard, Ben and Daan van Soest**, "Punishment to promote prosocial behavior: a field experiment," *Journal of Environmental Economics and Management*, 2024, p. 102899.
- Wing, Coady, Seth M. Freedman, and Alex Hollingsworth, "Stacked Difference-in-Differences," *NBER Working Paper No.* 32054, 2024.

A. Online Appendix



Figure A1: Persistence of Social Nudge Effects

Note: This figure reports the proportion of the effect of a social comparison nudge that persists after it is discontinued. Estimated effects that are not statistically significant at the five percent level are set to zero.



Figure A2: Average HER Effects by HERs Generated

Note: This figure reports different average effects of the HER by the quartile of total number of HERs generated for a given home (see footnote 5 for more). Brackets indicate the 95 percent confidence interval. Confidence intervals are constructed with heteroskedasticity-robust standard errors clustered by home.



assumes a different cost of technology. The first panel follows prior work by assuming no cost. uses the cost of technology per kWh saved from Gillingham et al. (2018) of \$0.03 and the last panel uses the cost of technology per kWh saved from Billingsley et al. (2014) of \$0.12 per kWh saved. The calculations assume the HER is administered monthly for a period of one year and benefits accrue undiscounted for two years. The cost of electricity is assumed to be \$0.10 per kWh and the cost of administering the HER is assumed to be \$1 *Note:* This figure presents the benefits, costs, and net benefits of the HER for the mover sample. Each panel per HER.

Figure A3: Net Benefits by Cost of Technology per kWh Saved

)			•	
		Electricity Con	is. (% of Contre	ol in Baseline)	
	(1)	(2)	(3)	(4)	(5)
Pre-Move Effect	-2.14 $(0.04)^{***}$	-2.14 $(0.18)^{***}$	-1.78 (0.48)***	-4.72 $(0.62)^{***}$	-3.08 (0.25)***
Post-Move Effect	~	-1.11 (0.29)***	0.68 (0.73)	-3.24 (0.95)***	-1.14 (0.38)***
Sample	Full	Mover	Mover & Renter	Mover & Elec. Heat	Mover & Comparison on 1st HER
Bills Homes R ²	58,733,360 1,810,096 0.63	5,768,148 139,908 0.54	718,129 19,270 0.49	782,283 19,334 0.59	> 0 3,446,889 84,434 0.52

Table A1: Average Effect of HER Estimated with Mover Subsamples

tricity for heating, and column 5 on mover sample homes that used more electricity than the average of their sure the average effect after the fourth HER in the comparison period and in the move period. Each coefficient is presented in terms of percent changes to the full or mover sample's control group electricity consumption in the baseline period. Estimates normalized by each mover subsample's control group baseline period electricity consumption are presented in A2. Column 1 is estimated on the full sample, column 2 on the mover sample, column 3 on mover sample homes that were rentals, column 4 on mover sample homes that used elecneighbors reported on their first HER. Estimates are obtained by weighting by the duration of each electricity *Note*: This table reports coefficients estimated with variants of equation 8. The coefficients respectively meabill and are regression-adjusted with fixed effects for each period of time, home, and year-by-season-by-wave. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

1			I		
		Electricity Con	s. (% of Contro	ol in Baseline)	
	(1)	(2)	(3)	(4)	(5)
Pre-Move Effect	-2.14	-2.14	-1.91	-3.39	-2.60
Post-Move Effect	(0.04)	(0.1δ) - 1.11	0.73	(0.44)	-0.96
		$(0.29)^{***}$	(0.79)	$(0.68)^{***}$	$(0.32)^{***}$
Sample	Full	Mover	Mover &	Mover &	Mover &
			Renter	Elec. Heat	Elec. Cons. on 1st HER
					Exceeded
					Neighbors
Bills	58,733,360	5,768,148	718,129	782,283	3,446,889
Homes	1,810,096	139,908	19,270	19,334	84,434
R^2	0.63	0.54	0.49	0.50	0.52

Table A2: Average Effect of HER Estimated with Mover Subsamples: Alternative Normalization

group electricity consumption in the baseline period. Table A3 present the baseline period control group electricity consumption used to normalize each estimate. Column 1 is estimated on the full sample, column 2 on the mover sample, column 3 on mover sample homes that were rentals, column 4 on mover sample homes that used electricity for heating, and column 5 on mover sample homes that used more electricity than the sure the average effect after the fourth HER in the comparison period and in the move period. Each coefficient average of their neighbors reported on their first HER. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with fixed effects for each period of time, home, and year-*Note*: This table reports coefficients estimated with variants of equation 8. The coefficients respectively meaoy-season-by-wave. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses is presented in terms of percent changes to the full sample's, mover sample's, and mover subsample's control below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

0.52

0.59

0.49

0.54

0.63

		Electricit	y Cons. in	Baseline (kW	/h/day)
	(1)	(2)	(3)	(4)	(5)
Control	39.80	38.10	38.10	38.10	38.10
Sample	Full	Mover	Mover & Renter	Mover & Elec. Heat	Mover & Comparison on 1st HER
					> 0

Table A3: Baseline Electricity Consumption by Mover Subsample

Note: This table reports baseline period average electricity consumption per day (kWh/day) for different samples in the control group.

	Electricity	trol in Baseline)	
	(1)	(2)	(3)
Pre-Move Effect	-4.62	-2.39	-2.18
	$(0.25)^{***}$	$(0.25)^{***}$	$(0.18)^{***}$
Post-Move Effect	-2.60	-0.71	-0.90
	$(0.30)^{***}$	$(0.31)^{**}$	$(0.30)^{***}$
Controls	Treatment,	Treatment,	Treatment,
	Period,	Period,	Period,
	Wave	Year-by-Season	Year-by-Season
		of Bill-by-Wave	of Move-by-Wave,
			Year-by-Season
			of Bill-by-Wave,
			Avg. Elec.
			Consby-Baseline
			Season-by-Wave
Sample	Mover	Mover	Mover
Bills	5,768,148	5,768,148	5,768,148
Homes	139,908	139,908	139,908
R^2	0.16	0.22	0.47

Table A4: Robustness of HER Effects to Specification of Control Variables: Mover Sample

Note: This table reports coefficients estimated with equation 8 on the mover sample with different specifications of control variables. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to control group electricity consumption in the baseline period. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with the controls denoted. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity	Cons. (% of Con	trol in Baseline)
	(1)	(2)	(3)
Pre-Move Effect	-3.39	-2.14	-1.78
	$(0.65)^{***}$	$(0.65)^{***}$	$(0.42)^{***}$
Post-Move Effect	0.60	1.36	0.81
	(0.77)	$(0.78)^*$	(0.75)
Controls	Treatment,	Treatment,	Treatment,
	Period,	Period,	Period,
	Wave	Year-by-Season	Year-by-Season
		of Bill-by-Wave	of Move-by-Wave,
			Year-by-Season
			of Bill-by-Wave,
			Avg. Elec.
			Consby-Baseline
			Season-by-Wave
Sample	Mover &	Mover &	Mover &
-	Renter	Renter	Renter
Bills	718,129	718,129	718,129
Homes	19,270	19,270	19,270
R^2	0.15	0.22	0.46

Table A5: Robustness of HER Effects to Specification of Control Variables: Mover and Renter Sample

Note: This table reports coefficients estimated with equation 8 on the mover sample homes that were rentals with different specifications of control variables. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to mover sample control group electricity consumption in the baseline period. Estimates normalized by mover and renter control group consumption in the baseline period are presented in Table A8. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with the controls denoted. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity	Cons. (% of Con	trol in Baseline)
	(1)	(2)	(3)
Pre-Move Effect	-6.99	-4.23	-4.26
	$(0.81)^{***}$	$(0.82)^{***}$	$(0.53)^{***}$
Post-Move Effect	-2.24	-2.97	-3.69
	$(0.96)^{**}$	$(1.02)^{***}$	$(0.98)^{***}$
Controls	Treatment,	Treatment,	Treatment,
	Period,	Period,	Period,
	Wave	Year-by-Season	Year-by-Season
		of Bill-by-Wave	of Move-by-Wave,
			Year-by-Season
			of Bill-by-Wave,
			Avg. Elec.
			Consby-Baseline
			Season-by-Wave
Sample	Mover &	Mover &	Mover &
	Elec. Heat	Elec. Heat	Elec. Heat
Bills	782,283	782,283	782,283
Homes	19,334	19,334	19,334
R^2	0.10	0.34	0.56

Table A6: Robustness of HER Effects to Specification of Control Variables: Mover and Electric Heating Sample

Note: This table reports coefficients estimated with equation 8 on the mover sample homes that used electricity for heating with different specifications of control variables. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to mover sample control group electricity consumption in the baseline period. Estimates normalized by mover and electric heating control group consumption in the baseline period are presented in Table A9. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with the controls denoted. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity	Cons. (% of Con	trol in Baseline)
	(1)	(2)	(3)
Pre-Move Effect	-5.43	-3.32	-2.63
	$(0.35)^{***}$	$(0.35)^{***}$	$(0.21)^{***}$
Post-Move Effect	-2.23	-0.68	-1.09
	$(0.41)^{***}$	(0.42)	$(0.40)^{***}$
Controls	Treatment,	Treatment,	Treatment,
	Period,	Period,	Period,
	Wave	Year-by-Season	Year-by-Season
		of Bill-by-Wave	of Move-by-Wave,
			Year-by-Season
			of Bill-by-Wave,
			Avg. Elec.
			Consby-Baseline
			Season-by-Wave
Sample	Mover &	Mover &	Mover &
	Comp. on	Comp. on	Comp. on
	1st HER > 0	1st HER > 0	1st HER > 0
Bills	3,446,889	3,446,889	3,446,889
Homes	84,434	84,434	84,434
R^2	0.13	0.19	0.49

Table A7: Robustness of HER Effects to Specification of Control Variables: Mover and Electricity Consumption Exceeded Neighbors on 1st HER Sample

Note: This table reports coefficients estimated with equation 8 on the mover sample homes that used more than the average of their neighbors reported on the first HER with different specifications of control variables. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to mover sample control group electricity consumption in the baseline period. Estimates normalized by mover sample homes that used more than the average of their neighbors reported on the first HER control group consumption in the baseline period are presented in Table A10. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with the controls denoted. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity	Cons. (% of Con	trol in Baseline)
	(1)	(2)	(3)
Pre-Move Effect	-3.64	-2.30	-2.11
	$(0.70)^{***}$	$(0.70)^{***}$	$(0.52)^{***}$
Post-Move Effect	0.64	1.46	0.94
	(0.82)	$(0.84)^*$	(0.81)
Controls	Treatment,	Treatment,	Treatment,
	Period,	Period,	Period,
	Wave	Year-by-Season	Year-by-Season
		of Bill-by-Wave	of Move-by-Wave,
		2	Year-by-Season
			of Bill-by-Wave,
			Avg. Elec.
			Consby-Baseline
			Season-by-Wave
Sample		Mover & Rent	er
Bills	718,129	718,129	718,129
Homes	19,270	19,270	19,270
R ²	0.15	0.22	0.44

Table A8: Robustness of HER Effects to Specification of Control Variables:Mover and Renter Sample, Alternative Normalization

Note: This table reports coefficients estimated with equation 8 on the mover sample homes that were rentals with different specifications of control variables. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to mover and renter sample control group electricity consumption in the baseline period. Table A₃ present the baseline period control group electricity consumption used to normalize the estimates. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with the controls denoted. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity	Cons. (% of Con	trol in Baseline)
	(1)	(2)	(3)
Pre-Move Effect	-5.01	-3.04	-3.07
	$(0.58)^{***}$	$(0.59)^{***}$	$(0.43)^{***}$
Post-Move Effect	-1.61	-2.13	-2.51
	$(0.69)^{**}$	$(0.73)^{***}$	$(0.70)^{***}$
Controls	Treatment,	Treatment,	Treatment,
	Period,	Period,	Period,
	Wave	Year-by-Season	Year-by-Season
		of Bill-by-Wave	of Move-by-Wave,
		-	Year-by-Season
			of Bill-by-Wave,
			Avg. Elec.
			Consby-Baseline
			Season-by-Wave
Sample		Mover & Elec. H	Heat
Bills	782,283	782,283	782,283
Homes	19,334	19,334	19,334
R^2	0.10	0.34	0.54

Table A9: Robustness of HER Effects to Specification of Control Variables: Mover and Electric Heating Sample, Alternative Normalization

Note: This table reports coefficients estimated with equation 8 on the mover sample homes that used electricity for heating with different specifications of control variables. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to mover and electric heating sample control group electricity consumption in the baseline period. Table A₃ present the baseline period control group electricity consumption used to normalize the estimates. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with the controls denoted. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity Cons. (% of Control in Baseline)			
	(1)	(2)	(3)	
Pre-Move Effect	-4.60	-2.81	-2.22	
	$(0.30)^{***}$	$(0.30)^{***}$	$(0.18)^{***}$	
Post-Move Effect	-1.88	-0.58	-0.93	
	$(0.35)^{***}$	(0.35)	$(0.34)^{***}$	
Controls	Treatment,	Treatment,	Treatment,	
	Period,	Period,	Period,	
	Wave	Year-by-Season	Year-by-Season	
		of Bill-by-Wave	of Move-by-Wave,	
		-	Year-by-Season	
			of Bill-by-Wave,	
		Avg. Elec		
		Consby-Basel		
			Season-by-Wave	
Sample	Mover &	Mover &	Mover &	
	Comp. on	Comp. on	Comp. on	
	1st HER > 0	1st HER > 0	1st HER > 0	
Bills	3,446,889	3,446,889	3,446,889	
Homes	84,434	84,434	84,434	
R^2	0.13	0.19	0.49	

Table A10: Robustness of HER Effects to Specification of Control Variables: Mover and Electricity Consumption Exceeded Neighbors on 1st HER Sample, Alternative Normalization

Note: This table reports coefficients estimated with equation 8 on the mover sample homes that used more than the average of their neighbors reported on the first HER with different specifications of control variables. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to mover sample homes that used more than the average of their neighbors reported on the first HER control group electricity consumption in the baseline period. Table A3 present the baseline period control group electricity consumption used to normalize the estimates. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with the controls denoted. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity Cons. (% of Control in Baseline)			
	(1)	(2)	(3)	(4)
Pre-Move Effect	-3.17	-3.72	-3.31	-1.82
	$(0.29)^{***}$	$(0.26)^{***}$	$(0.29)^{***}$	$(0.36)^{***}$
Post-Move Effect	-2.02	-4.09	-2.87	-2.69
	$(0.40)^{***}$	$(0.39)^{***}$	$(0.45)^{***}$	$(0.61)^{***}$
Days Cutoff	91	182	273	365
Reweighted	No	No	No	No
Obs.	64472155	63494442	49197644	30859290
Homes	117,865	87,098	54,004	28,182
R^2	0.49	0.49	0.47	0.47

Table A11: Robustness of HER Effects to Staggered Moves: Estimates from Stacked Model

Note: This table reports coefficients that measure the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to control group electricity consumption in the baseline period. Columns 1-4 are estimated on a mover sample that is observed in the baseline period for 365 days and respectively observed in the comparison and move periods for at least 91, 182, 273, and 365 days. The unit of observation is average daily electricity consumption in the final 365 days of the baseline period and in the final 91, 182, 273, or 365 days of the comparison and move periods. Estimates are regression-adjusted with two-way fixed effects, i.e., a fixed effect for each day in event time and each home. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity Cons. (% of Control in Baseline)			
	(1)	(2)	(3)	(4)
Pre-Move Effect	-2.90	-3.46	-2.95	-1.81
	$(0.29)^{***}$	$(0.27)^{***}$	$(0.29)^{***}$	$(0.36)^{***}$
Post-Move Effect	-1.53	-3.36	-2.18	-1.96
	$(0.41)^{***}$	$(0.39)^{***}$	$(0.46)^{***}$	$(0.61)^{***}$
Days Cutoff	91	182	273	365
Reweighted	Yes	Yes	Yes	Yes
Obs.	64472155	63494442	49197644	30859290
Homes	117,865	87,098	54,004	28,182
R^2	0.49	0.49	0.47	0.47

Table A12: Robustness of HER Effects to Staggered Moves: Estimates from Stacked and Reweighted Model

Note: This table reports coefficients that measure the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to control group electricity consumption in the baseline period. Columns 1-4 are estimated on a mover sample that is observed in the baseline period for 365 days and respectively observed in the comparison and move periods for at least 91, 182, 273, and 365 days. The unit of observation is average daily electricity consumption in the final 365 days of the baseline period and in the final 91, 182, 273, or 365 days of the comparison and move periods. Estimates are obtained by reweighting according to the procedure proposed in Wing et al. (2024) and regressionadjusted with two-way fixed effects, i.e., a fixed effect for each day in event time and each home. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity Cons. (% of Control in Baseline)			
	(1)	(2)	(3)	(4)
Pre-Move Effect	-2.08	-2.07	-2.08	-2.08
	$(0.18)^{***}$	$(0.18)^{***}$	$(0.18)^{***}$	$(0.18)^{***}$
Post-Move Effect	-0.93	-0.96	-1.04	-1.03
	$(0.25)^{***}$	$(0.26)^{***}$	$(0.27)^{***}$	$(0.31)^{***}$
HER Cutoff	1	2	3	5
Bills	7,334,722	7,017,358	6,486,682	5,282,567
Homes	182,559	173,105	157,415	121,980
R^2	0.53	0.53	0.53	0.53

Table A13: Robustness of HER Effects to Mover Sample Cutoff

Note: This table reports coefficients estimated with equation 8 on different constructions of the mover sample. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to control group electricity consumption in the baseline period. Columns 1-4 are estimated on a mover sample that receives at least 1, 2, 3, and 5 HERs before moving. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with fixed effects for event time, home, and year-by-season-by-wave. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.

	Electricity Cons. (% of Control in Baseline)		
	(1)	(2)	(3)
Pre-Move Effect	-2.37	-2.08	-2.11
	$(0.22)^{***}$	$(0.19)^{***}$	$(0.18)^{***}$
Post-Move Effect	-1.46	-1.41	-1.23
	$(0.29)^{***}$	$(0.28)^{***}$	$(0.28)^{***}$
Change	Drop Homes with	Drop Homes with	Drop Homes with
	1σ Decrease	2σ Decrease	3σ Decrease
	in Move Period	in Move Period	in Move Period
Bills	3,925,965	5,162,824	5,571,764
Homes	94,372	124,881	134,987
R^2	0.54	0.53	0.54

Table A14: Robustness of HER Effects to Dropping Low Use Move Period Homes

Note: This table reports coefficients estimated with equation 8 on different constructions of the mover sample. The coefficients measures the average effect of treatment assignment in the comparison and move periods. Each effect is presented in terms of percent changes to control group electricity consumption in the baseline period. Columns 1-3 are estimated on a mover sample that has move period electricity consumption within 1, 2, or 3 standard deviations of baseline and comparison period electricity consumption. Estimates are obtained by weighting by the duration of each electricity bill and are regression-adjusted with fixed effects for event time, home, and year-by-season-by-wave. Heteroskedasticity-robust standard errors clustered by home are reported in parentheses below each estimate. *** *p*-value < 0.01, ** *p*-value < 0.05, * *p*-value < 0.10.