**Discussion Paper Series – CRC TR 224**

Discussion Paper No. 182
Project B 07

# Disaggregate Consumption Feedback and Energy Conservation

Andreas Gerster *
Mark A. Andor **
Lorenz Goette ***

June 2020

*Andreas Gerster, University of Mannheim, L7, 3-5, 68161 Mannheim, Germany; Tel.: +49 621 181 1791,
E-Mail: gerster@uni-mannheim.de.
** RWI – Leibniz Institute for Economic Research
*** University of Bonn; National University of Singapore

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

# Disaggregate Consumption Feedback and Energy Conservation

Andreas Gerster[*], Mark A. Andor[†], Lorenz Goette[‡]

June 16, 2020

**Abstract**

Novel information technologies hold the promise to improve decision making. In the context of smart metering, we investigate the impact of providing households with appliance-level electricity feedback. In a randomized controlled trial, we find that the provision of appliance-level feedback creates a conservation effect of an additional 5% relative to a group receiving standard (aggregate) feedback. These conservation effects are largely driven by reductions in electricity use of 10% to 15% during peak hours. Consumers with appliance-level feedback hold more accurate beliefs about the energy consumption of different appliances, consistent with the mechanism in our accompanying model. Our result suggests that conservation effects from a smart-meter rollout will be much larger if appliance-level feedback can be provided. Based on a sufficient statistics approach, we estimate that appliance-level feedback could raise consumer surplus by about 570 to 600 million Euro per annum for German households.

*JEL codes*: D12, D83, L94, Q41.

*Keywords*: Randomized controlled trial, disaggregation, consumption feedback, energy conservation.

*Correspondence*: Andreas Gerster, University of Mannheim, L7, 3-5, 68161 Mannheim, Germany; Tel.: +49 621 181 1791, E-Mail: gerster@uni-mannheim.de.

*Acknowledgements*: We gratefully acknowledge financial support by the Swiss National Science Foundation (SNF) under grant 407140_153675 (NFP 71). Furthermore, this work has been partly supported by the German Research Foundation (DFG) through CRC TR 224 (Project B07) and the Collaborative Research Center "Statistical Modeling of Nonlinear Dynamic Processes" (SFB 823), within the framework of Project A3, "Dynamic Technology Modeling". We are grateful for valuable comments and suggestions by Ken Gillingham, Timo Goeschl, Sébastien Houde, Katrina Jessoe, Andreas Lange, Mike Price, David Rapson, Ulrich J. Wagner as well as by participants of the 8th Atlantic Workshop, the 24th Annual Conference of the EAERE, the 1st RWI EEE Workshop, and seminars at RWTH Aachen, Technical University Berlin, UC Davis, University of Hamburg, University of Mannheim, Toulouse School of Economics, and Yale University. We thank Thomas Jaletzky, Gerhard Radtke, Lara-Franziska Rogowski, and Peter Schulten for excellent cooperation in conducting the field experiment as well as Kathrin Kaestner and Yero Ndiaye for excellent research assistance.

[*]University of Mannheim

[†]RWI – Leibniz Institute for Economic Research

[‡]University of Bonn; National University of Singapore

# 1 Introduction

Novel information technologies enable consumers to make better informed decisions. For example, navigation systems assist car drivers in improving their travel planning (Chorus et al., 2006), text message reminders increase patients' adherence to medical treatments (Pop-Eleches et al., 2011), and fitness trackers help athletes to sustain higher levels of physical exercise (Cadmus-Bertram et al., 2015). Another context where information technologies have the potential to overcome problems from imperfect information is resource use. It is well-documented that individuals underestimate the energy use of energy-intensive appliances (Attari et al., 2010). A bias in the perception of energy intensity can increase the energy use beyond the level a perfectly informed consumer would choose. If so, it would also exacerbate externalities associated with the use of energy, such as carbon emissions and local air pollution.

In this paper, we explore the potential of providing appliance-level feedback to correct households' misperceptions of energy intensities and to unlock a new mechanism to foster energy conservation. We also develop a simple framework to investigate the impact of disaggregate feedback on consumer welfare and derive sufficient statistics to quantify it, thus providing a novel tool for regulatory impact analysis. The context of our study is the deployment of advanced metering infrastructure throughout the world, which involves multi-billion investments into so-called smart electricity meters.[1]

A core rationale for the deployment of smart meters is that these devices can provide households with consumption feedback. Feedback may foster awareness about the cost and environmental impact of electricity usage, and hence lead to energy conservation (EC, 2014b). To date, smart meters typically provide information about household-level electricity consumptions. Previous evidence suggests that such feedback leads to modest conservation effects of 2% to 5% (Degen et al., 2013; McKerracher and Torriti, 2013; Schleich et al., 2017; Houde et al., 2013). One potential reason for why these effects are small is that aggregate consumption data does not correct misperceptions of energy intensities of different appliances. Indeed, studies that installed additional hardware to provide appliance-specific feedback found much larger effects of 10% to 20% (Asensio and Delmas, 2015; Bruelisauer et al., 2018; Tiefenbeck et al., 2018, 2019). Our study builds on a novel smart meter technology that leverages the potential

[1] For example, 79 million households in the United States and 472 million households in China have been outfitted with smart meters by 2017 (IEA, 2017). In addition, the European Union has committed to installing 200 million smart meters at an estimated cost of 45 billion Euros (EC, 2014a).

of appliance-level feedback without installing additional costly infrastructure. The technology exploits that every appliance leaves a distinct signature in the electricity consumption data, which can be used to estimate appliance-level electricity use from aggregate, high-frequency electricity data (for details, see Hart 1992; Gupta et al. 2017, and Appendix A3).

We conduct a field experiment with some 800 participants and provide electricity use feedback through a smartphone app. Participants in the experimental control group obtain aggregate feedback as traditionally provided by smart metering interventions. Participants in our first treatment group additionally receive information on their appliance-level electricity use. We also implement three further treatment groups that complement appliance-level feedback with different forms of incentives to conserve energy. Participants in the second treatment group receive a bonus based on their absolute reduction in electricity use for a given appliance. In the third group, we inform participants how their electricity savings compare to the savings of other study households. In the fourth group, participants receive a bonus if their savings are high relative to the savings of others.

Based on a conceptual model of household service demand, we characterize the conditions under which appliance-level feedback triggers reductions in consumers' total electricity use. In our model, consumers correctly assess their overall electricity consumption, but inaccurately perceive the energy intensity of different appliances, as found by Attari et al. (2010) and recently formalized by Graeber and Enke (2019), for example. Appliance-specific feedback corrects this misperception. As we show, if energy intensities are sufficiently different or price elasticities of energy demand sufficiently similar across appliances, correcting the misperceptions leads to lower overall electricity use. Beyond providing hypotheses on the sign of the average treatment effect of appliance-level feedback, we use our model to derive sufficient statistics for evaluating its impact on consumer surplus. We quantify these statistics by exploiting our high-frequency data on appliance-level electricity uses. In addition, we measure participants' beliefs about the electricity use of their appliances in a survey to test whether appliance-level feedback shifts them as predicted by our model.

Our contribution to the literature is threefold. First, we disentangle the effects of aggregate and appliance-level smart meter information in a tailored randomized controlled trial, which allows for a clean identification of the additional savings induced by appliance-level feedback. Our results show that the provision of appliance-level feedback reduces electricity consumption strongly by around 5%, compared to an experimental condition in which individuals receive 'standard' aggregate smart metering feedback. Previous studies on the effectiveness of standard smart metering find conservation effects ranging from 2% to 3% (Carroll et al., 2014; Degen et al., 2013; Martin and Rivers, 2017), up to 5% (Schleich et al., 2017; Houde et al., 2013).[2] Thus, our estimates suggest that providing disaggregate feedback adds considerably to the effectiveness of smart meter interventions. Furthermore, Asensio and Delmas (2015) find that appliance-level feedback is only effective when combined with a health frame that highlights the adverse impacts of electricity generation on air pollution. Our results broaden that perspective on appliance-level feedback by showing that information provision without a health frame also substantially changes consumer behavior.

Second, we contribute to a literature that has investigated how smart meter technologies can reduce electricity consumption during times of high demand. Previous studies have shown that households reduce their electricity consumption during peak hours when facing time-of-use pricing (e.g., Allcott 2011a and Burkhardt et al. 2019). Consumers' price responsiveness to time-of-use prices increases when consumers obtain aggregate real-time feedback (Jessoe and Rapson, 2014) and automation technology such as programmable thermostats (Bollinger and Hartmann, 2019). Furthermore, Ito et al. (2018) find that monetary incentives trigger persistent reductions in consumption during peak period, whereas the effect of moral appeals is only short-lived.

We complement this literature by investigating the time-of-use pattern of electricity savings from appliance-level feedback. We detect a conservation pattern that is closely, but not perfectly, aligned with peak periods. In particular, the overall treatment effect is driven by a reduction in electricity consumption during hours where individuals are at home. This finding suggests that the conservation effects in our study stem primarily from appliances that also require time as an input. During these hours, we estimate large conservation effects of 10% to 15%. These magnitudes are similar to findings from feedback provision in the context of showering (e.g., Tiefenbeck et al., 2018). We also test whether the effectiveness of appliance-level feedback can be fostered by additional monetary incentives and information about savings by other participants. We do not find that any of these strategies increases the response to

[2] Further studies have shown that feedback via more frequent billing can even lead to an increase in resource use (Wichman, 2017). With respect to consumption feedback, Gosnell et al. (2019) find that providing information on social comparisons and demand disaggregation via an advanced app is more effective than providing aggregate information via inhome displays and a basic app, but only for one of two smart meter installers.

appliance-level feedback. This result demonstrates that additional monetary incentives do not necessarily improve the degree to which consumers engage with informational interventions.

Third, our study relates to the literature on welfare analysis under optimization errors by consumers (see Farhi and Gabaix 2020 for an overview). One strand of this literature has primarily focused on the implications of tax misperceptions (e.g., Chetty et al. 2009; Rees-Jones and Taubinsky 2019). Another strand has derived optimal corrective taxes and subsidies for behaviorally biased consumers (e.g., Allcott and Taubinsky 2015; Gerster and Kramm 2019; O'Donoghue and Rabin 2006). By contrast, we analyze informational instruments when consumers misperceive product attributes and the cost of household services. We derive formulas for evaluating consumer surplus that can be used for policy analysis of informational interventions to overcome such misperceptions.

The formulas are simple variants of those currently used in regulatory cost benefit assessments. In particular, we find that changes in consumer surplus can be calculated as the *weighted* sum over appliance-level cost savings. The weights are given by consumers' relative bias, i.e., the perceived energy intensity divided by the actual intensity. Because these weights are typically smaller than one (in absolute value), approximating changes in consumer surplus by the realized energy cost savings – as currently done in regulatory impact analyses in the U.S. and the EU, for example (Faruqui et al., 2011; Giordano et al., 2012) – strongly overestimates the welfare gains of smart meters. In our application, we find that consumer surplus is overestimated by a factor of about three. Nevertheless, the welfare gains we calculate are substantial: we find that appliance-level feedback increases consumer surplus by about 14 – 15 Euro per household and annum, which scales to an increase in total consumer surplus of about 570 – 600 million Euros for all German households every year.

The relevance of our findings extends beyond the context of smart meter feedback. Consumers hold misperceptions not only about the electricity consumption of appliances, but also about the effectiveness of fitness activities, the caloric content of foods, and benefits from schooling returns, for example (e.g., Attari et al. 2010; Bollinger et al. 2011; Jensen 2010). Our theoretical model and our empirical estimates indicate that eliminating biases via feedback will reduce energy spending, improve the effectiveness of physical exercise as well as the nutritional quality of foods, and foster investments into education. Our paper also develops a method for evaluating how misperceptions affect consumer surplus, which provides policy makers with a tool for weighting the cost of policy interventions against their benefits. Such tools are particularly important as progresses in digitilization will likely raise the policy relevance of informational interventions in the future.

The remainder of the paper is structured as follows. Section 2 introduces a conceptual model that clarifies how disaggregate consumption feedback reduces the total use of an input for household services, such as electricity. Based on that model, we also derive the sufficient statistics needed for quantifying changes in consumer surplus. Section 3 presents the experimental design and the data. In Section 4, we estimate the impact of appliance-level feedback on energy consumption and analyze treatment effect heterogeneity. Section 5 quantifies the impact of appliance-level feedback on consumer surplus. Section 6 concludes.

## 2 Conceptual Model

We start by investigating the effects of providing appliance-level feedback to consumers based on the following household services model (Becker, 1965). Let consumers have the following quasi-linear utility function:

$$U(\mathbf{x}, z) = u(\mathbf{x}) + z,$$

where $u(\mathbf{x})$ is quasi-concave and denotes the utility from consuming $J$ household services denoted by the vector $\mathbf{x} = (x_1, \ldots, x_J)'$ and $z$ represents the numeraire good, whose price is normalized to 1. The consumption of household service $j$ requires inputs of $y_j = x_j e_j$, where $e_j$ denotes the input intensity of service $x_j$. In our application, households consume energy services by using a particular appliance, such as a dish-washer or dryer, and the input intensity refers to the amount of electricity that is needed to operate an appliance. Consumers maximize their utility subject to the budget constraint $w = z + \sum_j y_j p$, where $w$ denotes their exogenous income and $p$ denotes the price of the input, in our case electricity.

In line with the literature (Attari et al., 2010), let consumers have biased perceptions of input intensities $\tilde{e}_j = e_j + b_j$, where $b_j$ is a bias term that decreases in absolute terms when beliefs become more accurate. Consumers maximize their utility under the beliefs $\tilde{e}_j$, which yields the following first-order condition:

$$u_j = p(e_j + b_j),$$

where $u_j$ denotes the first derivative of $u(\mathbf{x})$ with respect to the service $x_j$. This condition shows that, in optimum, consumers equate the marginal benefits of consuming a service $j$ with its perceived marginal cost, which differs from the actual marginal cost $(pe_j)$ by the term $pb_j$.

We now assess the impact of a technology that increases the accuracy of beliefs by, for example, providing appliance-level feedback. To provide intuition, we totally differentiate the first-order condition of consumers' utility maximization problem with respect to $\alpha$. For simplicity, we approximate consumers' utility by an additively separable utility function, which yields:

$$\Delta y_j = e_j \Delta x_j = e_j \frac{p}{u_{jj}} \Delta b_j, \tag{1}$$

where $u_{jj}$ denotes the second derivative of $u(\mathbf{x})$ with respect to $x_j$, while $\Delta y_j$ and $\Delta b_j$ denote the change in the use of the input for appliance $j$ and the change in the belief biases $b_j$, respectively, in response to the increase in belief accuracy. As concavity of $u(\mathbf{x})$ implies that $u_{jj}$ is negative, Equation (1) shows that consumption of an energy service $j$ decreases ($\Delta y_j < 0$) when consumers underestimate $e_j$ prior to the intervention ($\Delta b_j > 0$). Equation (1) clarifies that learning that the marginal cost of a household service is higher than anticipated has the same effects as a price increase and reduces consumption of that service. The opposite holds true when consumers learn about lower cost.

When does appliance-level feedback reduce total input use? For the purpose of illustration, consider the case of two appliances and let subscript 1 denote the more input intensive appliance, i.e., $e_1 > e_2$. Following the literature on cognitive uncertainty and noisy Bayesian cognition models (Gabaix, 2017; Graeber and Enke, 2019), let consumers have input intensity beliefs $\tilde{e}_j = \alpha e_j + (1-\alpha)e$, where $\alpha \in [0,1]$ is the weight that individuals put on the correct intensity and $e$ denotes a naive belief that is consistent with the total input use, but does not distinguish between different appliances, thus satisfying $x_1 e_1 + x_2 e_2 = (x_1 + x_2)e$.[3] Furthermore, let consumers' bias, defined by $b_j = \tilde{e}_j - e_j$, be zero when appliance-level feedback is provided. In Appendix A1, we derive sufficient conditions for a decrease in the total input use

[3] In Graeber and Enke (2019), consumers make decisions under uncertainty and form beliefs about probabilities, where $e$ is interpreted as an average probabililty. We apply their rationale to our non-stochastic setting of biased beliefs.

in response to the provision of appliance-level feedback ($\Delta y = \Delta y_1 + \Delta y_2 < 0$). We find that input use decreases if the following inequality holds:

$$\frac{e_1}{e_2}\frac{\eta_1(\tilde{e}_1)}{\eta_2(\tilde{e}_2)} > 1, \tag{2}$$

where $\eta_j(\tilde{e}_j)$ denotes the price elasticity of demand for service $j$ under the belief $\tilde{e}_j$, i.e, $\eta_j(\tilde{e}_j) = (\partial x_j(\tilde{e}_j)/\partial p)/(x_j(\tilde{e}_j)/p)$. A sufficient condition for Inequality (2) to hold is that $e_2$ is sufficiently small, which follows from letting $e_2$ approach zero. Another sufficient condition is that the price elasticity of the more input intensive service 1 is at least as large as the price elasticity of service 2, i.e., $(\eta_1(\tilde{e}_1)/(\eta_2(\tilde{e}_2)) \geq 1$. In Appendix A1, we show that the same intuition holds true when we analyze the case of $n$ appliances.

To understand the main mechanism for reductions in total input use, let us consider the special case where preferences for both services are identical and consumers hold fully naive beliefs ($\tilde{e_j} = e$). In that case, consumption choices are the same for both appliances $(x_1(\tilde{e}_1) = x_2(\tilde{e}_2))$. Furthermore, the change in the bias in response to a higher belief accuracy has the same magnitude for both appliances as the naive belief $e$ lies at the midpoint between $e_1$ and $e_2$, which implies that $\Delta b_1 = -\Delta b_2$. As preferences for both services are identical in our example, eliminating the bias increases and decreases household service demand for appliance 1 and 2 by the same margin ($\Delta x_1 = -\Delta x_2$). Yet, total input use reduces $(\Delta y = e_1 \Delta x_1 + e_2 \Delta x_2 < 0)$, as appliance 1 is more input intensive than appliance 2 ($e_1 > e_2$) by definition. When consumers' beliefs are only partly biased ($\alpha > 0$), the same intuition holds true as all behavioral responses are scaled by a common factor, which reduces the change in total input use, but not its sign (for details, see Appendix A1).

As derived in Appendix A2, we can express the response of consumer surplus to an intervention that increases belief accuracy as follows:

$$\Delta ConsSurplus = \sum_j \frac{b_j}{e_j} \Delta E_j, \tag{3}$$

where $\Delta ConsSurplus$ denotes the change in consumer surplus and $\Delta E_j = p\Delta y_j$ gives the average treatment effect of the intervention on the input expenditures for household service $j$. Equation (3) shows that we can estimate changes in consumer surplus based on few sufficient statistics: the relative bias $(b_j/e_j)$, as well as the appliance-specific average treatment effects of the intervention on electricity expenditures ($\Delta E_j$). The intuition for Equation (3) is that a belief bias $b_j > 0$ creates a wedge between consumers' perceived and actual marginal cost from consuming an energy service $j$, which induces first-order distortions of consumption choices. Consumer surplus increases when beliefs become more accurate. The increase is proportional to the size of the relative belief bias $(b_j/e_j)$ and the behavioral response to an increase in belief accuracy, as captured by the change in input expenditures ($\Delta E_j$).

Equation (3) demonstrates a fundamental flaw of current cost benefit analyses that approximate changes in consumer surplus by the sum of expenditure savings (e.g., Faruqui et al. 2011; Giordano et al. 2012). In fact, the change in consumer surplus equals the *weighted* sum of changes in expenditures, where the weights are given by the relative biases $b_j/e_j$. These weights are typically less than one in absolute value, and thus influence consumer surplus estimates. Equation (3) also clarifies that expenditure savings are not a necessary condition for increases in consumer surplus, as implicitly assumed in current cost benefit analyses. If consumers overestimate the cost of using an energy service $j$ ($b_j/e_j > 0$), their welfare increases if they use that service more, thereby increasing expenditures ($\Delta E_j > 0$).

All sufficient statistics for evaluating consumer surplus can be estimated from our data. First, our appliance-level consumption measures allow us to estimate the average treatment effects of disaggregate feedback for each appliance category. Second, under additive separability and a pure nudge assumption (e.g. Allcott and Taubinsky 2015), we can identify the relative bias from our data as $(b_j/e_j) = -(\Delta y_j/y_j)/\eta_j$, where $(\Delta y_j/y_j)$ denotes the ATE on energy consumption of appliance $j$, expressed as a percentage of the input use $y_j$, and $\eta_j$ is the price elasticity of energy service demand $j$ (for derivations, see Appendix A2). To obtain estimates for price elasticities, we follow a double-pronged strategy. First, we draw upon estimates from the literature. Second, we estimate them based on our data, exploiting cross-sectional price variation across German regions (details are provided in Section 5). This variation partly stems from differences in grid surcharges, which are higher in regions with substantial electricity generation from renewable energy sources.

# 3 Experimental Design and Data

## 3.1 Study Implementation

We conducted the randomized controlled trial with customers of a large German utility. Customers were invited to take part in a smart meter study via an email that did not mention appliance-level feedback as the purpose of the study. To be eligible for participation in the experiment, customers had to have a smartphone and wireless internet access. Furthermore, participants with own electricity generation by solar panels were excluded. Out of around 50,000 customers we invited, 800 participants agreed to take part in the study and met our eligibility criteria.

As Table A1 in the Appendix shows, sociodemographic characteristics of study participants roughly match German averages. In terms of age, employment status, sex, and household net income, our experimental sample is similar to the German population. Participating households consist of slightly more occupants (2.5 vs. 2.0 in Germany), which is mostly driven by a smaller percentage of single-person households (12% vs. 42% in Germany). For households of a given size, electricity consumption levels are similar in our study and the German population. The larger average number of occupants per household in our sample translates into larger average electricity consumption levels compared to the German average (10.4 vs. 8.6 kWh per day). Furthermore, households in our sample live more often in their own property than the average German household (76% vs. 44% in Germany).[4]

All participating households obtained a high-resolution smart meter, an internet gateway that connected the smart meter with the internet, and access to a smartphone app. After the smart meter had been installed by professional plumbers, the utility sent participants the internet gateway along with instructions how to install the app. As soon as participants had activated the gateway and installed the app, they shared their smart meter data and our study started. More than 90% of our study participants entered the field test between November, 2016, and January, 2017, and the remaining few participants joined afterwards. The core study period extended for 6 months, where consumers had access to the full functionality of the app in their respective treatment group. From month 7 onwards, households were free to continue to use the app for another three months, but add-on functionalities beyond disaggregation were stopped in three treatment groups. As the number of participants declined considerably during that period, our analysis focuses on the core study period and conducts complementary analyses to estimate treatment effects in the months thereafter.

[4]When we test for differences in the effectiveness of treatments by the number of occupants in a household and the ownership status of the house, we cannot detect statistically significant differences (Table A12 in the Appendix). The point estimates indicate that the response is somewhat larger for single-person households, but lower for households who do not own their apartment or house.

The smart meters measure electricity consumption at a high frequency, typically every second, which results in an rich dataset of more than 10 billion observations over the entire study period. The smart meter data is saved online in real-time and processed every day to detect appliance usages based on a so-called nonintrusive appliance load monitoring (NALM) algorithm, which employs machine learning techniques for load disaggregation. The algorithm exploits the fact that appliances have characteristic electricity use signatures. These signatures can be used to disaggregate high-resolution smart meter data into appliance-specific electricity uses (see Appendix A3 for details). For example, a dryer first heats up the wet laundry, thereby consuming much electricity, and then iterates between periods of letting the laundry cool down and heating it up again. This produces a characteristic signature in the high-resolution smart meter data that the algorithm exploits to detect so-called appliance use events, i.e. the start/end date and the respective electricity use (see Figure A4 in the Appendix for exemplary signatures). Detection of appliance events is possible for the major appliances of a typical household, including the categories *Dishwasher*, *Washing Machine*, *Dryer*, and *Oven*, as well as a *Refrigeration* category that captures refrigerators and freezers. The algorithm also identifies an *Always-on* category as the typical consumption at 3 a.m. In addition to the appliance categories that we can directly measure, we construct a residual category *Other Appliances*, which captures the electricity consumptions of all other appliances. Table A4 in the Appendix gives descriptive statistics on the more than 600.000 appliance events that we observe during the study period.

Based on the appliance use events, we determine participants' monthly electricity consumption for every appliance category. To facilitate the assessment of appliance consumptions, we also use participants' electricity prices to calculate monthly operating cost by appliance. Furthermore, we transform monthly appliance-level consumptions into an appliance score that informs households about their usage intensity.[5] In the app, households can click on a button that provides a detailed description of the meaning of the appliance scores. The score is 100 if a household's appliance use is very low and 0 if it is very high, compared to typical usage behaviors and energy intensities of the respective appliance. More specifically, the score is calculated as follows: Appliance Score = 100 $\times$ (Monthly Appliance Consumption - $\text{Bench}_{\text{low}}$)/($\text{Bench}_{\text{up}}$ - $\text{Bench}_{\text{low}}$), where $\text{Bench}_{\text{low}}$ and $\text{Bench}_{\text{high}}$ correspond to pre-determined

[5] After consulting with the app designers, we denoted this score as an "efficiency score", as this term has intuitive appeal to an average household.

benchmark values for high and low appliance uses, respectively. These benchmarks are based on survey data on typical appliance uses and product data sheets on the technical efficiency of appliances currently used in German households (for details, see Table A2 in the Appendix).

In Figure A5 in the Appendix, we display the distributions of the appliance scores by appliance category. As we determined the appliance score benchmarks prior to the experiment, assessing the range of the appliance scores is a good test to evaluate the plausibility of the detected appliance use events. For all appliance categories, the vast majority of appliance scores lies between 0 and 100, which supports the credibility of the disaggregation. For the categories *Dishwasher*, *Dryer* and *Oven*, there is bunching at indices of 100, which indicates that some participants have not used these appliances at all in some months.

In addition, we benchmark the appliance-level measurements by comparing our data from 2017 with the average appliance uses in Germany, which is only available for 1996 and 2011. As Figure A6 in the Appendix shows, the percentage of electricity used for cooking (9.6%) and for washing, drying, and dish-washing (9.7%) aligns with German averages in 1996 and 2011 (about 9.6 – 9.8% and 10.4 – 12.4%, respectively). In our study, refrigeration accounts for 9.9%, which is less than German averages for 1996 and 2011 (22.6% in 1996 and 16.7% in 2011). This divergence likely reflects a gradual increase in energy efficiency of refrigerators and freezers over time, not least owing to ever increasing minimum standards (see, e.g., Andor et al. 2020). The percentage of the category *Other Appliances* amounts to 71%, which is slightly larger than the German averages (57.5% in 1996 and 61.1% in 2011). This deviation is likely driven by the general trend that households use more electric devices, such as smart TVs, computers, smartphones, and robotic vaccuum cleaners. By contrast, air conditioning is not prevalent in German households and thus cannot explain that increase.

We conducted three surveys to elicit participants' sociodemographic characteristics, their household characteristics and beliefs. The first survey took place in November 2016 prior to the start of the field test, followed by two additional surveys in March 2017 and July 2017. Furthermore, we obtained participants' most recent yearly electricity use before the field test started, which serves as our baseline. Of the 800 participants, we cannot observe baseline electricity use for 27 participants.[6] Furthermore, 73 households experienced technical difficulties

[6]For 18 participants, baseline electricity use is missing. In addition, we set electricity baseline to missing when the difference between the baseline and the experimental period is in the top or bottom percentile of its distribution within each treatment group (e.g. above +126% and -64% for the experimental control group) or larger than 25 kWh per day in absolute terms, which concerns 9 participants.

**Table 1:** Overview of Experimental Conditions

| | EC | $T_1$ | $T_2$ | $T_3$ | $T_4$ |
|---|---|---|---|---|---|
| AF: Aggregate Feedback | ✓ | ✓ | ✓ | ✓ | ✓ |
| D: Disaggregation | - | ✓ | ✓ | ✓ | ✓ |
| C: Challenges | - | - | ✓ | ✓ | ✓ |
| *Type of Challenge*: | | | | | |
| M: Monetary Incentives | - | - | ✓ | - | ✓ |
| R: Ranking | - | - | - | ✓ | ✓ |

that prevent them from connecting their smart meter to the internet. As a result, the final sample used for our analyses consists of 700 participants.

### 3.2 Experimental Design

We randomly assign participants into one of five experimental conditions, which are summarized in Table 1 (for screenshots of the smartphone app, see Figures A2 and A3 in the Appendix). Participants in our Experimental Control (EC) group get access to an app that provides information about their household-level electricity use. On the start screen, participants see whether the cost of their current monthly electricity consumption exceeds the value of their monthly advance payment.[7] Furthermore, participants can observe a real-time power meter that visualizes their current wattage. Beyond that, participants can compare their electricity consumption with their own history, as well as with other study participants, at a monthly, weekly, daily, and hourly frequency.

Participants in our first treatment group $T_1$ have access to an additional app page that provides feedback on appliance-level usages, cost and appliance scores. The three treatment groups $T_2$-$T_4$ provide the same appliance-level feedback and additionally invite participants to take part in so-called appliance challenges. These challenges start at the beginning of the second month after installation of the app and ask participants to increase the appliance score of an appliance by as much as they can within a month. At the end of the month, the change in the appliance score relative to the previous month is evaluated as follows. In group $T_2$, participants obtain 1 EUR per appliance score improvement, capped at a maximum of 20 EUR per monthly challenge. In group $T_3$, participants see a ranking that compares their score improvement with those of other study participants, but do not obtain a financial reward. A

[7] In Germany, typically, billing occurs annually and monthly advance payments are intended to smooth electricity costs over the year. Exceeding the monthly advance payment has no financial consequence, but indicates that households may face additional payments when the next yearly billing occurs.

**Table 2:** Descriptives

| | EC | $T_1$ | $T_2$ | $T_3$ | $T_4$ | P-value |
|---|---|---|---|---|---|---|
| Baseline cons., in kWh/day | 10.2 | 10.5 | 10.1 | 10.5 | 10.5 | 0.90 |
| Basel. period length, in days | 363.6 | 361.1 | 361.6 | 363.3 | 362.2 | 0.90 |
| # of occupants | 2.4 | 2.5 | 2.4 | 2.5 | 2.6 | 0.17 |
| # of appliances for refrigeration | 2.2 | 2.3 | 2.3 | 2.3 | 2.2 | 0.74 |
| Net income, in EUR | 3,004 | 3,188 | 3,030 | 3,194 | 3,091 | 0.63 |
| Own property, in % | 73.6 | 79.7 | 73.8 | 77.5 | 73.2 | 0.67 |
| Employed, in % | 50.2 | 53.1 | 50.9 | 55.4 | 46.0 | 0.23 |
| Share of females, in % | 44.8 | 49.0 | 48.7 | 46.9 | 47.5 | 0.40 |
| Age, in years | 47.6 | 44.8 | 47.6 | 45.9 | 43.4 | 0.29 |
| Number of households | 140 | 136 | 143 | 143 | 138 | |

*Notes:* P-values are from F-tests of mean equality in all experimental conditions (clustered at the household level). Variables are measured at the household level, except for employed, share of females, and age, which we measure at the household member level.

participant within the top percentile of monthly score improvements is classified as rank one, and similarly for all other percentiles. In group $T_4$, we implement the same ranking, but reward participants according to their rank: rank one translates into 10 EUR, rank two into 9.9 EUR, etc., and rank 100 into 0 EUR. At the end of our study, participants in $T_2$ and $T_4$ receive an Amazon voucher to the amount of their earnings from taking part in the challenges. We designed our reward scheme to yield similar average payments in both treatment groups. As shown in Table A5 of the Appendix, the average realized payments per monthly efficiency challenge amount to 6.3 EUR in $T_2$ and to 4.5 EUR in $T_4$.[8]

Participants take part in a maximum of five challenges. The first challenge is always targeted towards the appliance with the lowest appliance score, followed by the appliance with the second-lowest score, etc. If less than five appliances are detected for a household, challenges can target an appliances more than once, given that all other appliances have been targeted already.

The timing in our experiment is as follows: In the first month after app installation, participants in $T_1$–$T_4$ obtain appliance-level feedback but challenges have not yet started. In the months 2-6, participants in $T_2$-$T_4$ take part in challenges. After month 6, challenges do not occur any more, but $T_1$–$T_4$ participants continue to receive appliance-level feedback. In Table 2, we show that central sociodemographics are balanced across experimental conditions, as

[8]The average payment in group $T_4$ does not exactly equal 5 EUR as we have less than 100 challenge participants in that group for some months.

expected from randomization. When we conduct F-tests of mean equality in all experimental conditions, we cannot reject the null hypothesis of no difference for any of the variables at the 5% level, as shown in the last column of Table 2.

# 4 Empirical Strategy and Results

## 4.1 Treatment Effects on Total Electricity Use

We start by identifying the average treatment effect (ATE) of our four treatment groups $T_1$-$T_4$. We estimate the following equation via ordinary least squares (OLS):

$$Y_{it}^{norm} = \alpha Y_i^b + \boldsymbol{\beta}'\mathbf{T_i} + \nu_t + \mu_w^b + \epsilon_{it}, \tag{4}$$

where $Y_{it}^{norm}$ denotes the electricity use of participant $i$ at day $t$, divided by the average daily electricity use in the experimental control group (9.86 kWh). Dividing by a constant does not change our results, but expresses all treatment effects as percentage of the average consumption, which facilitates interpretation. $Y_i^b$ denotes the average daily consumption during the baseline period, which we normalize in the same way as the outcome variable. Households in our four treatment groups are denoted by the vector $\mathbf{T}_i = (T_{1i}, T_{2i}, T_{3i}, T_{4i})$. The ATEs of the respective treatment group are thus identified by the parameter estimates $\hat{\boldsymbol{\beta}}$. We also estimate a specification with a disaggregation dummy variable $D_i$ that equals one if a household is in one of the four treatment groups ($T_1$–$T_4$) that obtained appliance-level feedback. Our model includes fixed effects for every day ($\nu_t$) and an error term $\epsilon_{it}$. In order to increase the precision of our estimates, we also include a set of week-of-baseline fixed effects ($\mu_w^b$) that equal one if the baseline metering period includes a particular week and zero otherwise. We cluster standard errors at the household level, thereby accounting for serial correlation in the error terms.

As shown in Panel a) of Table 3, the ATE estimate for participants in the four treatment groups $T_1 - T_4$ amounts to $-5.2\%$ and is statistically significant at all conventional levels. When we estimate the ATEs separately for all treatment groups (Column 2), we do not find sizable differences in their magnitude, with point estimates ranging from $-4.1\%$ in $T_2$ to $-6.0\%$ in $T_3$. Our ATE estimates increase slightly when we consider only a balanced panel, i.e., when we exclude participants that do not share data for the entire time span of the core study period (Table A8 of the Appendix).

**Table 3:** ATEs on Daily Electricity Consumption, Relative to Exp. Control

(a) Effect of Experimental Conditions

| | (1) | (2) |
|---|---|---|
| $D$: $T_1$-$T_4$ | $-0.051^{***}$ | |
| | (0.018) | |
| $T_1$ | | $-0.060^{***}$ |
| | | (0.021) |
| $T_2$ | | $-0.041^{*}$ |
| | | (0.022) |
| $T_3$ | | $-0.060^{***}$ |
| | | (0.022) |
| $T_4$ | | $-0.044^{*}$ |
| | | (0.023) |
| $Y^b$: Baseline elec. use | $0.899^{***}$ | $0.900^{***}$ |
| | (0.021) | (0.021) |
| Day fixed effects (FE) | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ |
| $R^2$ | 0.5720 | 0.5722 |
| Number of obs. | 106,283 | 106,283 |
| Number of households | 700 | 700 |

(b) Effects of App Elements

| | (3) | (4) | (5) |
|---|---|---|---|
| $D$: Disaggregation | $-0.051^{***}$ | $-0.056^{***}$ | $-0.054^{***}$ |
| | (0.018) | (0.020) | (0.020) |
| $M$: Monetary incentives | | 0.016 | 0.013 |
| | | (0.014) | (0.017) |
| $R$: Ranking | | $-0.006$ | $-0.009$ |
| | | (0.013) | (0.019) |
| $M$: Monet. inc. $\times$ $R$: Rank. | | | 0.007 |
| | | | (0.028) |
| $Y^b$: Baseline elec. use | $0.899^{***}$ | $0.900^{***}$ | $0.900^{***}$ |
| | (0.021) | (0.021) | (0.021) |
| Day fixed effects (FE) | ✓ | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ | ✓ |
| $R^2$ | 0.5720 | 0.5722 | 0.5722 |
| Number of obs. | 106,283 | 106,283 | 106,283 |
| Number of households | 700 | 700 | 700 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parantheses and clustered at the household-match level. The outcome variable is daily electricity consumption, divided by the mean in the $EC$ group (9.86 kWh). $AF$ equals one for the households in the groups $EC$ and $T_1 - T_4$, while being zero for households in $MC$. $D$ equals one for households in the groups $T_1 - T_4$, $M$ equals one for households in the groups $T_2$ and $T_4$, and $R$ equals one for households in the groups $T_3$ and $T_4$, while being zero for the other participants, respectively.

Next, we disentangle the differential impact of the app elements by estimating the following equation:

$$Y_{it}^{norm} = \alpha Y_i^b + \beta_1 D_i + \beta_2 M_i + \beta_3 R_i + \beta_4 M_i R_i + \nu_t + \mu_w^b + \epsilon_{it}, \tag{5}$$

where the scalar $D_i$ equals one if a household is in one of the four treatment groups $(T_{1i}, T_{2i}, T_{3i}, T_{4i})$, so that $\hat{\beta}_1$ identifies the conservation effect of providing appliance-level feedback, compared to providing aggregate feedback only. Furthermore, $M_i$ equals one for the treatment groups $T_{2i}$ and $T_{4i}$, where participants receive monetary rewards for saving electricity. Similarly, $R_i$ equals one for the treatment groups $T_{3i}$ and $T_{4i}$, where participants obtain information on their savings relative to those of other participants. All three groups $T_2$-$T_4$ also receive appliance-level feedback. Hence, the parameter estimates $\hat{\beta}_2$ and $\hat{\beta}_3$ identify by how much the effectiveness of appliance-level feedback changes when monetary incentives and rank information are provided additionally. We also interact $M_i$ and $R_i$ to test whether the effectiveness of monetary incentives increases when they are tied to a relative ranking rather than an absolute appliance score improvement. This interaction effect is identified by parameter $\beta_4$.

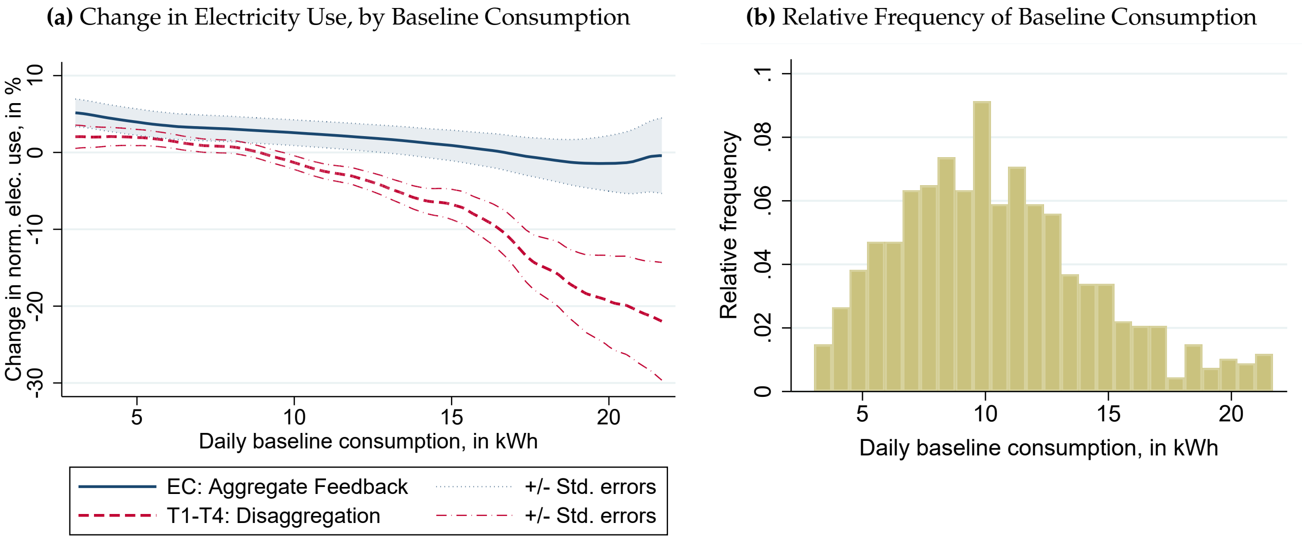
Our results from Panel b) of Table 3 show that the provision of appliance-level feedback in $T_1$-$T_4$ is the main driver of the electricity conservation we observe. When consumers obtain such feedback, their conservation effect increases by about 5 percentage points, relative to aggregate feedback only (Column 3). In Column (4), we test whether the provision of monetary incentives and rankings intensify the response to appliance-level feedback. We find that the point estimates are close to zero and not statistically significant at any conventional level. Hence, neither monetary incentives nor rankings trigger higher electricity savings compared to appliance-feedback alone. Furthermore, we do not find support for the conjecture that monetary incentives become more effective when information about participants' rank is also provided, as shown by the small and statistically insignificant interaction effect between $M$ and $R$ in Column (5).

At least two explanations can rationalize our finding that the additional challenges did not increase conservation effects. First, the rank information and the monthly financial incentives of about 5 EUR on average may not have sparked households' interest to actively take part in the challenges. Second, the high effectiveness of providing appliance-level feedback may not have left much further potential for conserving electricity without large utility losses. As we cannot detect statistically significant differences between the ATEs of the treatment groups $T_1 - T_4$, we pool them and use the disaggregation dummy $D$ for our subsequent analyses.

To identify the overall conservation effect, we additionally obtain data from a sample of households with smart meters that are served by the same utility. In Appendix A4, we explain how we construct a matched control group of households that did not receive any feedback from this sample. The results, presented in Table A9 in the Appendix, show that households in the experimental control group do not consume less electricity than households in the matched control group ($\text{ATE}_{\text{EC}}$: 0.001, Std. Err.: 0.02). This finding indicates that giving aggregate real-time feedback does not induce an energy conservation effect in our study population. Hence, we interpret the average treatment effects from Table 3 as total electricity savings.

Furthermore, we use the additional observations after the core study period to estimate how treatment effects evolve after the core study period. We cannot reject the null hypothesis that the difference between the average treatment effects during and after the core study period are equal to zero at any conventional significance level ($\text{ATE}_{\text{AfterCoreStudyP}} - \text{ATE}_{\text{CoreStudyP}}$:

**Figure 1:** Change in Electricity Use, by Baseline Consumption




*Notes for Panel a):* "Change in norm. elec. use" denotes the difference in daily electricity consumption between the core study period and the baseline electricity use from the previous billing year, divided by the mean in the EC group (9.86 kWh). For local mean smoothing, we use an Epanechikov kernel with rule-of-thumb bandwidths 1.47 and 4.07 for the *D* and *EC* group, respectively.

0.039, Std. Err.: 0.025, see Appendix Section A6 and Table A11). While the point estimate indicates a slight decrease in the effectiveness, we do not have sufficient power to precisely pin down how treatment effects evolve over time.

## 4.2 Treatment Effect Heterogeneity

In this subsection, we investigate two sources of treatment effect heterogeneity. First, we analyze how treatment effects vary with households' baseline consumption levels. Our model implies that households with large biases should have high baseline consumption levels. Hence, they should reduce their electricity consumption the most after receiving appliance-level feedback. Similar conservation patterns have been found in other studies on energy conservation (see e.g. Allcott 2011b; Andor et al. 2018; Tiefenbeck et al. 2018). Second, we exploit the high granularity of our data to investigate how treatment effects vary by hour-of-the-day. This analysis is motivated by the fact that baseline electricity levels also vary greatly within a household over the course of a day. In addition, conservation effects during peak periods can imply co-benefits from avoiding high production cost of electricity and reducing the likelihood of critical peak events during these times, for example.

To explore how effect sizes differ by baseline consumption, Figure 1 visualizes the change in electricity consumption between the baseline and the study period for households in the experimental control group, $EC$, and the disaggregation group, $D$. For each group, we estimate the relationship between consumption changes and baseline consumption levels by local mean smoothing. The difference between both lines can be interpreted as a difference-in-differences estimate of the ATE for households with a particular baseline consumption level. Figure 1 shows that households with appliance-level feedback ($D$) reduce their electricity consumption more strongly than households in the $EC$ group. These reductions are particularly large for participants with large baseline consumptions.

In Appendix A5, we quantify this treatment effect heterogeneity by estimating a model that interacts the $D$ dummy with baseline electricity use. Our point estimate amounts to $-0.066$ (Table A7 of the Appendix), which implies that one-unit increase in normalized baseline electricity use (about 10 kWh per day) is associated with an increase of the ATE by 6.6 percentage points. While this estimate is not statistically significant, we find a similar (yet, statistically significant) interaction effect for the $D$ group when using additional data from our non-experimental matched control group (Table A7).

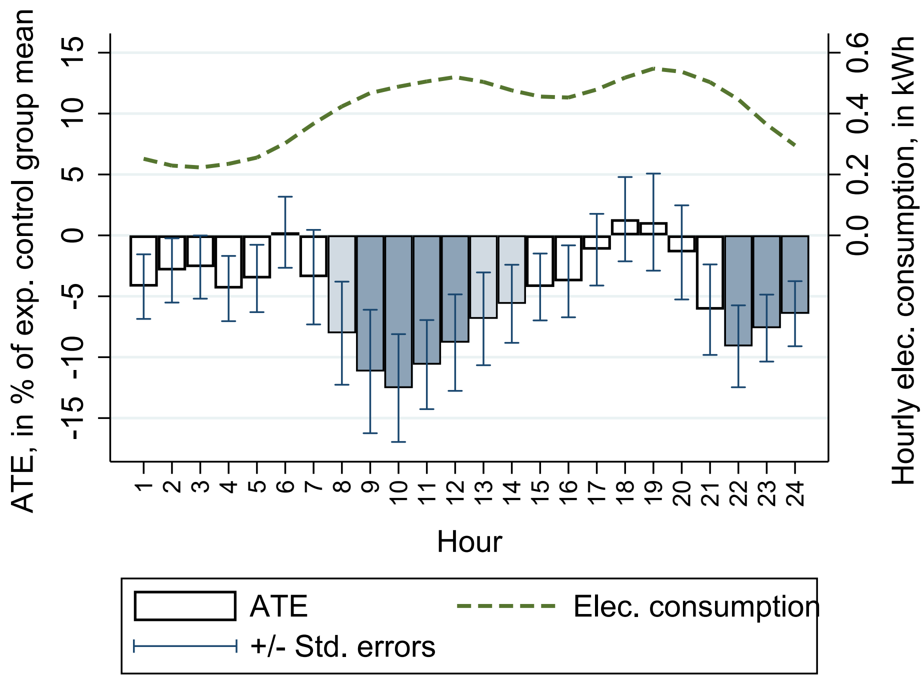
Next, we investigate how the conservation effect of appliance-level feedback varies by hour-of-the-day. We exploit the rich time dimension of our data set and estimate the following equation:

$$Y_{ith}^{norm} = \alpha Y_i^b + \sum_{h=1}^{24} \beta^h D_i + \nu_t + \mu_h + \mu_w^b + \epsilon_{ith}, \tag{6}$$

where all variables are defined as in Equation (1), except that we now investigate the hourly electricity consumption of participant $i$ on day $t$ and hour $h$, and additionally include 24 fixed effects $\mu_h$ for every hour of the day. We normalize hourly consumptions by the average hourly consumption in the $EC$ group, so that our estimates $\hat{\beta}^h$ capture the ATE in hour $h$, expressed as a percentage of the average consumption in that group. Again, we cluster standard errors at the household level.

Figure 2 shows that treatment effects are large during late morning hours and late evening hours, reaching about 10 – 15% of the average hourly consumption. This finding demonstrates that appliance-level feedback achieves considerable conservation effects that align with evidence from other contexts, such as showering (Tiefenbeck et al., 2018). Furthermore, we find

**Figure 2:** ATE of Disaggregation ($D$) relative to Experimental Control ($EC$), by Hour of the Day




*Notes:* The outcome variable is hourly electricity consumption, divided by the mean over all hours in the *EC* group (0.41 kWh). Shaded bars indicate that treatment effects are statistically significant at the 1% (blue shaded) or 5% (light blue shaded) level. Whiskers indicate a range of +/- 1 standard error (clustered at the household level). Based on conducting an F-test, we can reject the null hypothesis that all hourly point estimates are zero: $F(24, 699) = 1.82$, p-value $= 0.0096$.

that the magnitude of electricity conservation cannot be predicted by baseline electricity consumption levels alone. Electricity reductions are particularly strong in the late morning hours between 8 – 13 a.m., which coincide with large electricity consumption levels. During 4 and 8 p.m., consumption levels are similar, but households save considerably less. Furthermore, we detect strong savings of about 6–9% in late evening hours between 9 – 11 p.m., when consumption levels are rather low.

### 4.3 Treatment Effects on Appliance-Level Electricity Consumptions

In the following, we investigate whether the time pattern of conservation effects originates from heterogeneous appliance-level responses. To determine the appliance-specific ATEs, we estimate the following equation separately for every appliance $j$:

$$Y_{itj}^{norm} = \alpha_j Y_i^b + \beta_j D_i + \gamma_j C_i + \nu_t + \epsilon_{itj}, \tag{7}$$

which closely mimicks Equation (4), but substitutes the outcome variable by the average daily consumption of each appliance category, normalized by the respective control group average. To analyze how our challenges affect appliance-level consumptions, we also include a variable $C$, which equals one if a household is in one of the treatment groups $T_2$-$T_4$ and zero otherwise. In this specification, $\hat{\beta}_j$ captures the average treatment effect of disaggregate feedback, while $\hat{\gamma}_j$ identifies by how much this effect changes when challenges are implemented in addition.

As shown in Table 4, we find that the conservation effects from appliance-level feedback are close to zero for appliance categories which are typically used throughout the day, such as *Refrigeration* and *Always-On*. The low response may partly reflect that consumers' demand for refrigeration is largely constant, irrespective of (perceived) cost. By contrast, we detect considerable conservation effects for appliances that are only used during daytime, which can explain the time pattern of conservation effects from Figure 2. In particular, we detect that appliance-level feedback triggers a substantial reduction in the electricity consumption of dryers by around 42% (Column 5 of Table 4). As dryers are an electricity intensive appliance, it is plausible that consumers underestimate it (Attari et al., 2010) and hence reduce their energy consumption after receiving appliance-level feedback. In addition, substitutes for using the dryer are often available as dry-hanging clothes is common for German households. We also find some evidence that participants have reduced their use of the washing machine, the dish-washer, and the oven by 3.6 to 8.6%, yet these effects are not statistically significant at any conventional level. In Figure A9 of the Appendix, we present conservation effects for every hour-of-the-day and show that energy conservation for dryers, washing-machines, and dish-washers primarily occurs around midday. This finding suggests that the first peak in conservation effects from Figure 2 can be attributed to these appliances.

In addition, we find that households reduce consumption for the category *Other Appliances* by around 8%, an effect that is statistically significant at the 5% level. This category encompasses a variety of electric appliances, such as televisions, hi-fi systems, vaccuum-cleaners, computers, as well as lighting. While no direct feedback is given for these appliances, participants can update their beliefs by exploiting the feedback on other appliances to learn about residual consumption. For example, if a consumer learns that all large appliances are energy-efficient, he or she may also figure out which of the remaining appliances are particularly electricity-intensive. The fact that many of the *Other Appliances*, such as televisions, serve en-

**Table 4:** ATE on Daily Appliance-Level Electricity Consumption

| | (1) Always-On | (2) Refrigeration | (3) Dish−Washer | (4) Washing | (5) Dryer | (6) Oven | (7) Other appl. |
|---|---|---|---|---|---|---|---|
| *D*: Disaggregation | 0.031 | −0.003 | −0.082 | −0.036 | −0.423** | −0.086 | −0.080** |
| | (0.058) | (0.054) | (0.108) | (0.081) | (0.183) | (0.181) | (0.034) |
| *C*: Challenges | −0.047 | −0.007 | −0.023 | 0.006 | −0.051 | 0.165 | −0.001 |
| | (0.047) | (0.047) | (0.084) | (0.059) | (0.126) | (0.140) | (0.023) |
| $Y^b$: Baseline elec. use | 1.127*** | 0.434*** | 0.715*** | 0.612*** | 1.014*** | 1.328*** | 0.850*** |
| | (0.062) | (0.103) | (0.098) | (0.075) | (0.165) | (0.198) | (0.034) |
| Day fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $R^2$ | 0.353 | 0.123 | 0.039 | 0.025 | 0.028 | 0.035 | 0.346 |
| Number of obs. | 93,187 | 93,185 | 84,511 | 91,473 | 65,852 | 93,187 | 93,187 |
| Number of households | 700 | 700 | 635 | 686 | 499 | 700 | 700 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors in parantheses, clustered at the household level. The outcome variable is daily electricity consumption of an appliance, divided by the mean for the same appliance in the *EC* group (2.30, 0.98, 0.31, 0.48, 0.18, 0.22, and 5.49 kWh for Columns 1 to 7, respectively). The dummy variable *C* equals one if participants are invited to take part in challenges ($T_2$, $T_3$, $T_4$).

tertainment purposes can explain the second peak in conservation effects in the late evening hours (Panel d of Figure A9 in the Appendix).

Next, we disentangle changes in the frequency of use from changes in utilization (Table A13 in the Appendix). We find that the reduction in electricity use for dryers stems from a lower frequency of use rather than a decrease in the electricity intensity per utilization. Reducing the frequency of use can for example be achieved by dry-hanging clothes or by washing only full loads. For dish-washers, we detect a decrease in the electricity intensity for dish-washers, which typically stems from choosing washing programs that use lower temperatures and thus energy, but need more time.

As shown by the small and statistically insignificant estimates for the dummy variable *C* in Table 4, we cannot detect that the appliance challenges in our treatment groups 3-5 triggered additional savings beyond the effects of appliance-level feedback. This finding is consistent with our results from Section 4.1, where we did not find evidence that challenges increase the conservation effect from appliance-level feedback.

### 4.4 Treatment Effects on Beliefs

To investigate the mechanisms behind the reduction in overall electricity use, we test whether the accuracy of participants' beliefs improves over the course of the experiment. The elicitation of energy consumption beliefs is subject to vivid controversy and a methodological consensus has not been reached so far (Frederick et al., 2011; Attari et al., 2010, 2011). To help consumers who are not familiar with energy consumption units, some researchers provide reference points and inform households about the energy consumption of a reference appliance prior to asking participants about energy consumption beliefs (Attari et al., 2010). While this approach can reduce excessive variance in participants' answers, providing a reference point has been shown to also bias belief estimates towards that reference point. In addition, changing the unit of measurement from watts to kilowatts, for example, can induce framing effects that also bias belief elicitation (Frederick et al., 2011).

In our study, we elicited beliefs in a baseline survey that took place prior to the core study period and an endline survey in July, i.e., after most consumers had finished the core study period. We asked participants to guess their monthly electricity consumption for the appliance categories always-on, washing machine, dryer, refrigeration, and dish-washer without providing reference points. As a consequence of not providing a reference point, participant's belief estimates are noisy. For example, some participants appear to express their beliefs in other unit of measurement than the ones we asked for, such as yearly consumptions.

To circumvent difficulties that may arise from noisy belief measurements, we translate all consumption beliefs into ranks. We assign rank 1 if a participant believes that the monthly consumption of an appliance was highest among all of his or her appliances. Similarly, rank 2 corresponds to a belief that the appliance consumption occupies the second rank, etc. We also calculate the same ranks based on the appliance-level data that we can observe. This allows us to calculate a rank difference as the absolute difference between the ranks implied by participants' belief and those based on our data. Averaging about these absolute difference over all appliances gives us a measure for the accuracy of beliefs that is robust to differences in the unit of measurement used by participants when expressing their beliefs. For this measure to be comparable, we drop all participants who have not stated beliefs for all appliances in one of the two surveys. The mean rank difference is zero if a consumer is correct about the rank of all appliances and can reach up to 3 if the estimated consumption ranks are exactly opposite to the ranks from our appliance-level measurements.

Table 5 displays the average absolute rank difference for the baseline survey that we conducted prior to the experiment (B) and the survey after the core study period (E), as well as the change in the rank difference (E-B). The baseline rank difference amounts to about 2 for all experimental groups. This value coincides with the expected rank difference of ran-

**Table 5:** Analysis of Rank Differences between Beliefs and Measured Appliance Uses

| | Mean rank diff. (B) | Mean rank diff. (E) | Change (E−B) | Std. err. | P-value | N |
|---|---|---|---|---|---|---|
| EC | 1.97 | 1.92 | −0.05 | 0.06 | 0.450 | 65 |
| $T_1$ | 1.95 | 1.71 | −0.24 | 0.10 | 0.025 | 58 |
| $T_2$ | 1.98 | 1.65 | −0.33 | 0.09 | 0.000 | 59 |
| $T_3$ | 1.96 | 1.63 | −0.33 | 0.10 | 0.001 | 52 |
| $T_4$ | 2.02 | 1.70 | −0.31 | 0.10 | 0.002 | 55 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parantheses and clustered at the household level. "Mean rank diff." denotes the mean of the absolute rank differences between the rank of measured appliance consumptions and the rank of electricity consumption beliefs, calculated for every appliance category.

domly determined ranks and thus reflects poor knowledge of energy intensities. For the *EC* group, we cannot reject the null hypothesis that the mean differences in the rank difference (E-B) equals zero at any conventional significance level. This finding is consistent with the absence of an conservation effect for that group (for details, see Section A4 in the Appendix). By contrast, for each of the treatment groups $T_1$-$T_4$ that have received appliance-level feedback, we observe a decrease in the absolute rank difference that is statistically significant at the 5% level. Accordingly, participants in these groups adjusted their beliefs in response to obtaining appliance-level feedback, which our model has identified as a prerequisite for a change in consumption.

## 5 Consumer Surplus

In this section, we go beyond estimating conservation effects and quantify the impact of appliance-level feedback on consumer surplus. In Section 2, we have identified the relative bias $(b_j/e_j)$ and the appliance-level treatment effects $(\Delta y_j/y_j)$ as sufficient statistics to estimate changes in consumer surplus. We retrieve the relative bias $b_j/e_j$ from our data by estimating the price elasticities of appliance-level energy consumption $\eta_j$, using that $(b_j/e_j) = -(\Delta y_j/y_j)/\eta_j$, where $\eta_j$ is a price elasticity of energy service demand $j$ and $(\Delta y_j/y_j)$ denotes the ATE on electricity consumption of appliance $j$, expressed as a percentage of baseline electricity use $y_j$ (for derivations, see Section 2).

We follow two strategies to approximate $\eta_j$ in our setting. First, we employ cross-sectional variation in our dataset that stems from the fact that similar households in terms of income, household size, and annual consumption pay different electricity prices, in particular owing to transmission charges that vary strongly by region in Germany. Second, we draw upon estimates by Frondel et al. (2019), who use instrumental variable techniques to estimate a price elasticity of aggregate electricity consumption of German households, and proxy our appliance-level elasticities by their estimate (-0.44).[9]

To identify $\eta_j$ from our data, we estimate the following regression separately for every appliance $j$:

$$ln\, y_{ij} = \eta_j\, ln\, p_i + \beta_j' \mathbf{X}_i + \epsilon_{ij},$$

where $\mathbf{X}$ denotes a vector of socio-demographic variables for household $i$ such as its income and size, as well as the annual consumption level in the year prior to the experiment. Cross-sectional identification faces some challenges that we discuss in the following. A typical concern is non-linear pricing, where marginal prices change with the level of electricity consumption. As households in our sample face constant marginal prices, non-linear pricing cannot threaten the validity of our approach. Another concern is omitted variable bias. Electricity suppliers offer tariffs with lower marginal prices to households who consume more, which could negatively bias our elasticity estimates. To circumvent such bias, we control for baseline electricity consumption, as well as for household income and size. While controlling for additional covariates reduces concerns from omitted variable bias, we cannot rule out that it may still be present to some degree. As an indirect test of its magnitude, we calculate the household-level elasticity implied by our appliance-level estimates and compare it to the findings by Frondel et al. (2019). If our appliance-level elasticity estimates systematically suffered from omitted variable bias, we would expect to find that the implied household-level elasticity is biased as well.

Our point estimates, depicted in Column (4) of Table 6, show that appliance-level consumptions are particularly elastic for the categories dryer and oven, where the elasticity estimates reach $-4.35$ and $-1.18$, respectively. This finding reflects that consumers can easily substitute these energy services by, for example, dry-hanging clothes. For the categories always-on, washing, and refrigeration, we obtain much smaller elasticities of $-0.20$ to $-0.38$. Our appliance-level elasticity estimates imply a household-level elasticity of $-0.36$, which is close

[9]The mean (short-run) price elasticities reported in the literature amounts to -0.35, with estimates ranging up to -2.01 (Espey and Espey, 2004). As a lower bound, some studies find that households do not respond to varying prices at all (e.g., (Byrne et al., 2020)).

**Table 6:** Changes in Consumer Surplus from Appliance-Level Feedback

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Estimated $\eta_j$ | | Uniform $\eta_j$ (Frondel et al., 2019) | | |
| | $\Delta y_j/y_j$ | Avg. use in kWh/a | $\Delta E_j$ in EUR/a | $\eta_j$ | $b_j/e_j$ | $\Delta$ConsSurplus in EUR/a | $\eta_j$ | $b_j/e_j$ | $\Delta$ConsSurplus in EUR/a |
| Always−On | 0.031 | 838.68 | 6.19 | −0.20 | 0.16 | 0.97 | −0.44 | 0.07 | 0.44 |
| Refrigeration | −0.003 | 359.48 | −0.22 | −0.38 | −0.01 | 0.00 | −0.44 | −0.01 | 0.00 |
| Dish−Washer | −0.082 | 111.90 | −2.20 | −0.36 | −0.23 | 0.50 | −0.44 | −0.19 | 0.41 |
| Washing | −0.036 | 176.64 | −1.51 | −0.31 | −0.12 | 0.18 | −0.44 | −0.08 | 0.12 |
| Dryer | −0.423 | 65.46 | −6.60 | −4.35 | −0.10 | 0.64 | −0.44 | −0.96 | 6.35 |
| Oven | −0.086 | 78.59 | −1.62 | −1.18 | −0.07 | 0.12 | −0.44 | −0.20 | 0.32 |
| Other appl. | −0.080 | 2,005.63 | −38.11 | −0.27 | −0.30 | 11.46 | −0.44 | −0.18 | 6.90 |
| Total | | 3,636.39 | −44.07 | −0.36 | | 13.87 | | | 14.54 |

*Notes:* $\Delta y_j/y_j$ correspond to the point estimate for *D: Disaggregation* from Table 4, which gives the ATE of providing appliance-level feedback, divided by the respective *EC* group mean of daily electricity consumption. The change in Expenditures, $\Delta E_j$, is calculated as the product of $\Delta y_j/y_j$ (Column 1), the average experimental control group mean (Column 2), and the average electricity price in our sample (0.238 EUR per kWh). $\eta_j$ denotes the price elasticity of energy service demand with respect to the electricity price, which we estimate as described in Section 5. $b_j/e_j$ denotes the relative bias, which we calculate as $b_j/e_j = -(\Delta y_j/y_j)/\eta_j$ (for derivations, see Appendix A2). Changes in consumer surplus are calculated as $\Delta\text{ConsSurplus} = \sum_j (b_j/e_j)\Delta E_j$ (see Equation 3).

to the estimate of $-0.44$, taken from Frondel et al. (2019).[10] This finding reduces concerns that our cross-sectional identification strategy is biased.

Based on the appliance-level elasticities and ATEs, we estimate how consumer surplus responds to the provision of appliance-level feedback. Column (1) of Table 6 reproduces the appliance-level ATEs in response to disaggregate feedback from Table 4, which correspond to $\Delta y_j/y_j$ in our model. Using these ATEs, as well as the estimated appliance-level price elasticities from Columns (4), we estimate the relative bias as $b_j/e_j = -(\Delta y_j/y_j)/\eta_j$, which is depicted in Column (5). We find that relative biases are negative for all appliance categories, except for *Always-On*. Negative biases are most pronounced for the categories *Dish-Washer* and *Other Appliances*, where consumers underestimate energy intensities by 23% and 30%, respectively. For the categories *Washing*, *Dryer*, and *Oven*, we obtain less pronounced biases of $-7\%$ to $-12\%$. The bias for the *Refrigeration* is close to zero, which may reflect that the electricity use of fridges and freezers is largely independent of usage behaviors and thus rather easy to assess.

To determine the change in consumer surplus, we calculate the appliance-level change in expenditures ($\Delta E_j = p\Delta y_j$) as the product of the relative ATE, $\Delta y_j/y_j$ (Column 1 of Table 6), and the average consumption level, $y_j$ (Column 2). We then multiply the change in expen-

[10]The elasticity of total consumption can be calculated as follows: $\eta = \prod_j \eta_j(y_j/y)$, where $\eta_j$ and $y_j$ denote the elasticity and the consumption level for appliance $j$, respectively, and $y = \sum_j y_j$ denotes total consumption.

ditures with the the relativ bias ($b_j/e_j$) to obtain the change in consumer surplus that can be attributed to every appliance category (Column 6). Summing over all categories, we find that total consumer surplus increases by 13.9 EUR per annum (last row of Column 6), which is substantially less than the 44 EUR decrease in expenditures (last row of Column 3). There are two main reasons why changes in total expenditures are an incorrect measure for changes in consumer surplus. First, less consumption of an energy service not only reduce expenditures, but also utility. The reduction in utility is proportional to the relative bias $b_j/e_j$, which explains why this term is crucial for determining consumer welfare. Second, consumer surplus can also raise when more accurate beliefs lead to more consumption of an energy service, despite the fact that expenditures increase. In our setting, this occurs for the category *Always-on*, where our estimates imply that consumers overestimate energy intensity.

As a robustness check, Columns (7)-(9) present our findings when we estimate changes in consumer surplus based on a uniform price elasticity of $-0.44$. In that case, we find that consumer surplus increases by 14.5 EUR per annum, which closely matches our findings from Column (6). While the total change in consumer surplus is almost identical, we obtain slight differences in the contributions of the different appliances. Based on our elasticities estimates, we find that the largest increase in consumer surplus stems from the category *Other Appliances* (Column 6). The contribution of the category *Dryer* to the change in consumer surplus increases, owing to a larger implied bias in that category, while the reverse holds true for the category *Other Appliances* (Column 9).

Taking into account that changes in consumer surplus are equal to the *weighted* sum of appliance-specific expenditures savings is crucial for cost-benefit analyses. The official cost-benefit analysis for Germany assumes that smart meter feedback can achieve annual electricity savings of 1.2% (EC, 2014c), which translates into annual expenditure savings of 10.4 EUR for the average household in our sample. Following the EU methodology (Giordano et al., 2012), all expenditure savings are considered as a measure for how much consumers benefit from smart metering. Yet, our results have shown that only 32% (13.9 EUR / 44 EUR) of the estimated expenditure savings translate into changes in consumer surplus (Column 6 of Table 6). Hence, consumer benefits are considerably smaller and would reach only 3.3 EUR for an conservation effects of 1.2%. However, our study also demonstrates that larger effect sizes are possible when appliance-level feedback is provided. Using the effect sizes found in our study, we find that consumer surplus increases by about 14 EUR, which underlines the potential of appliance-level feedback.

## 6 Discussion and Conclusion

In this paper, we investigate the effects of providing households with appliance-specific feedback. Based on a field experiment among some 800 participants, we randomize information provision on a smart phone app. Participants in our control group obtain aggregate electricity consumption feedback, while treatment group participants additionally receive information on appliance-level consumptions. Furthermore, we test whether monetary incentives or information about the electricity savings relative to other households foster the effectiveness of appliance-level feedback.

Our results show that the provision of appliance-level feedback reduces electricity consumption by around 5% compared to a control group that receives aggregate feedback only. We find that appliance-level feedback helps consumers to reduce the electricity consumption of energy-intensive appliances, such as tumble dryers. The behavioral response to appliance-level feedback is particularly large in the late morning hours between 8 – 13 a.m. and in the late evening hours between 9 – 11 p.m., when it reaches about 10% to 15%. Our experimental evidence also demonstrates that the effectiveness of appliance-level feedback is not further increased by monetary incentives or information about the conservation efforts of others.

Theoretically, we investigate the consequences of providing information to consumers who underestimate the energy use of energy-intensive appliances as shown by Attari et al. (2010). Our model demonstrates that correcting consumers' beliefs via appliance-level feedback leads to lower energy consumption if energy intensities of different appliances are sufficiently different, or the respective demand elasticities sufficiently similar. Based on this model, we also develop sufficient statistics to assess the gains in consumer surplus from correcting this bias. We find that such gains can be calculated as the *weighted* sum of appliance-level energy cost savings, where the weights are given by a measure of consumers' bias. Hence, current cost-benefit analyses in the U.S. and the EU (Giordano et al., 2012; Faruqui et al., 2011) that approximate changes in consumer surplus by the realized aggregate energy cost savings are fundamentally flawed. As the weights tend to be smaller than one in absolute value, current cost-benefit analyses tend to overerestimate the gains in consumer surplus substantially.

When we estimate the sufficient statistics from our experimental data, we find that appliance-level feedback increases consumer surplus by 13.9 – 14.5 Euro per household and annum. Extrapolating to the entire German population of 41.3 million households (destatis, 2017), we obtain an annual increase in consumer surplus by about 570 – 600 million Euro. These estimates illustrate that appliance-level feedback is crucial for raising consumer surplus. The installation of smart meters without providing such feedback would forgo large welfare gains.

By reducing electricity consumption, appliance-level feedback also mitigates externalities associated with the generation of electricity. Using the German average carbon intensity of electricity generation of about 486 g per kWh (Icha and Kuhs, 2019), we extrapolate that carbon emissions from residential electricity consumption in Germany would reduce by 3.9 megatons per annum. Using the range of estimates for the social cost of carbon of 11 – 105 USD (10 – 93 EUR) per ton of $CO_2$ equivalents (IAWG, 2016), this reduction in emissions translates into a decrease in total external costs from carbon emissions by 44 – 422 Mio. EUR per year. While this range reflects the uncertainties in estimating the social cost of carbon, it clearly demonstrates that co-benefits from emission reductions are non-negligible. A caveat of these extrapolations is that the presence of cap-and-trade schemes, such as the European Union Emissions Trading Scheme, could undermine the effectiveness of additional conservation efforts. Yet, policy makers could circumvent such offsetting effects by adjusting the emission cap accordingly.

More broadly, our study advances the understanding of how disaggregate consumption feedback changes consumer behavior and welfare. Our experiment used smart meter feedback to correct consumer misperceptions about the energy intensity of household appliances, but the relevance of our findings extends beyond that context. For example, biased beliefs about the caloric content of food (Bollinger et al., 2011) compromise consumers' ability to choose a healthy diet. Providing disaggregate feedback solves this problem by informing consumers about the relative benefits and costs of their food choices. Similar applications arise when consumers are unaware of the returns to education (Jensen, 2010), the carbon footprint of consumption alternatives, and the effectiveness of protective measures against natural disasters or a pandemic, for example. As digitilization progresses, novel interventions to reduce such misperceptions will likely develop. The conservation effects we detect in our experiment showcase the potential of such interventions to change behaviors. In addition, our method for evaluating consumer surplus allows policy makers to assess their benefits.

# Appendix



## Supplementary Figures



## Supplementary Tables

## A1 Effect of Appliance-Level Feedback on Energy Consumption

We illustrate the impact of providing appliance-level feedback based on the following conceptual model. Let consumers have the following quasi-linear utility function:

$$U(x_1, x_2, z) = u_1(x_1) + u_2(x_2) + z, \tag{8}$$

where $u_1(x_1)$ and $u(x_2)$ denote the utility from consuming two energy services $x_1$ and $x_2$, which we assume to be additively separable (for simplicity). The variable $z$ denotes the numeraire good, whose price is normalized to 1. The consumption of energy services requires energy inputs of $x_j e_j$, $\forall j \in \{1, 2\}$, where $e_j$ denotes the energy intensity of energy service $x_j$. Without loss of generality, let the subscript 1 denote the more energy intensive appliance, i.e. $e_1 > e_2$. Consumers maximize their utility subject to the budget constraint $w = z + x_1\tilde{e}_1 p + x_2\tilde{e}_2 p$, where $w$ denotes their endowment, $p$ denotes the price of energy, and $\tilde{e}_j$ denotes the energy intensity belief for appliance $j$.

Substituting the budget constraint into Equation (8), consumers' utility function can be written as:

$$U(x_1, x_2; e_1, e_2) = u_1(x_1) + u_2(x_2) + w - p(x_1\tilde{e}_1 + x_2\tilde{e}_2).$$

Let consumers' naive belief on the energy intensity of both appliances be denoted by $e$. It is consistent with the available aggregate information on total energy consumption and thus implicitly defined by $x_1e_1 + x_2e_2 = (x_1 + x_2)e$. Rewriting this equation yields:

$$e = \frac{x_1}{x_1 + x_2} e_1 + \frac{x_2}{x_1 + x_2} e_2.$$

In the absence of appliance-level feedback, households' energy intensity beliefs are biased towards the naive belief $e$, i.e. $\tilde{e}_j = \alpha e_j + (1 - \alpha)e$, where $\alpha \in [0, 1]$ denotes the belief accuracy. Hence, we obtain that:

$$b_1 = \tilde{e}_1 - e_1 = \alpha e_1 + (1 - \alpha)e - e_1 = (1 - \alpha)(e - e_1) = (1 - \alpha)\frac{x_2}{x_1 + x_2}(e_2 - e_1),$$

which implies that:

$$\frac{b_1}{b_2} = -\frac{x_2}{x_1}. \tag{9}$$

Under belief $\tilde{e}_j$, the first-order condition of utility maximization yields $u_j = pe$ for $j \in \{1,2\}$. Totally differentiating it with respect to the price of energy $p$ gives:

$$\begin{aligned} u_{jj} \frac{\partial x_j(\tilde{e}_j)}{\partial p} &= \tilde{e}_j \\ \Leftrightarrow u_{jj} &= \frac{\tilde{e}_j}{\partial x_j(\tilde{e}_j)/\partial p}. \end{aligned} \tag{10}$$

We now investigate how feedback that reduces belief biases from $b_j$ to zero for $j \in \{1,2\}$ changes total energy consumption. Consumers reduce total energy consumption if:

$$\Delta y = \Delta y_1 + \Delta y_2 < 0.$$

Under a pure nudge assumption ($\Delta b_j = -b_j$) and using that $\Delta y_j = \frac{p\, e_j}{u_{jj}} \Delta b_j$, we obtain:

$$\Delta y = -\frac{p\, e_1}{u_{11}} b_1 - \frac{p\, e_2}{u_{22}} b_2 < 0.$$

After some rearrangements (taking into account that $u_{22} < 0$, $b_2 > 0$, and $e_2 > 0$), this inequality holds if:

$$-\frac{e_1}{e_2} \frac{u_{22}}{u_{11}} \frac{b_1}{b_2} > 1.$$

Using Equations (9) and (10), and expanding by $p$, we can rewrite this inequality as:

$$\begin{aligned} \frac{e_1}{e_2} \frac{\tilde{e}_2}{\tilde{e}_1} \frac{\partial x_1(\tilde{e}_1)/\partial p}{\partial x_2(\tilde{e}_2)/\partial p} \frac{(x_2(\tilde{e}_2)/p)}{(x_1(\tilde{e}_1)/p)} &> 1 \\ \Leftrightarrow \frac{e_1}{e_2} \frac{\tilde{e}_2}{\tilde{e}_1} \frac{\eta_1(\tilde{e}_1)}{\eta_2(\tilde{e}_2)} &> 1, \end{aligned} \tag{11}$$

where $\eta_j(\tilde{e}_j)$ denotes the price elasticity of demand for energy service $j$ under the naive belief $e$, i.e, $\eta_j(\tilde{e}_j) = (\partial x_j(e)/\partial p)/(x_j(e)/p)$. As $\alpha \in [0,1]$, we can bound the expression:

$$\frac{\tilde{e}_2}{\tilde{e}_1} = \frac{\alpha e_2 + (1-\alpha)e}{\alpha e_1 + (1-\alpha)e} \in \left[\frac{e_2}{e_1}, 1\right].$$

Using these bounds, we obtain:

$$\frac{e_1}{e_2} \frac{\eta_1(\tilde{e}_1)}{\eta_2(\tilde{e}_2)} \geq \frac{e_1}{e_2} \frac{\tilde{e}_2}{\tilde{e}_1} \frac{\eta_1(\tilde{e}_1)}{\eta_2(\tilde{e}_2)} > 1,$$

so that a sufficient condition for $\Delta y < 0$ is that:

$$\frac{e_1}{e_2}\frac{\eta_1(\tilde{e}_1)}{\eta_2(\tilde{e}_2)} > 1.$$

First, as $e_1 > e_2$ by assumption, Inequality (11) holds when the price elasticity of energy service demand of the more energy intensive appliance 1 is at least as large as the price elasticity of the less energy intensive appliance 2, i.e., $\eta_1(\tilde{e}_1)/\eta_2(\tilde{e}_2) \geq 1$. Second, Inequality (11) holds when $e_2$ is sufficiently small. This finding immediately follows from taking the limits to both sides of the inequality, which yields:

$$\lim_{e_2 \to 0} \frac{e_1}{e_2}\frac{\eta_1(\tilde{e}_1)}{\eta_2(\tilde{e}_2)} = \infty > 1.$$

Hence, as $e_1$, $\eta_1$, and $\eta_2$ are finite, there always exists a sufficiently small $e_2$ so that Inequality (11) holds.

### Extension to n appliances

Let $i \in \{1, ..., N\}$ denote a particular appliance and assume that $e_1 > e_2 > ... > e_n$ without loss of generality. Furthermore, let $j \in \{1, ..., J\}$ denote appliances where consumers hold negative biases ($b_j < 0$), i.e. underestimate energy intensities, while the $k \in \{J+1, ..., N\}$ denotes appliances where consumers hold positive biases.

Using that $\Delta b_j = -b_j$ and that $\Delta y_j = \frac{p\,e_j}{u_{jj}}\Delta b_j$, we obtain:

$$\begin{aligned} & \Delta y < 0 \\ \Leftrightarrow & \sum_j \Delta y_j + \sum_k \Delta y_j < 0 \\ \Leftrightarrow & -\sum_j \frac{p\,e_j}{u_{jj}} b_j - \sum_k \frac{p\,e_k}{u_{kk}} b_k < 0. \end{aligned}$$

We can rearrange as follows (using that $b_N > 0$, $u_{NN} > 0$, $e_n > 0$):

$$-\sum_j \frac{e_j}{e_N}\frac{u_{NN}}{u_{jj}}\frac{b_j}{b_N} - \sum_{k \neq N} \frac{e_k}{e_N}\frac{u_{NN}}{u_{kk}}\frac{b_k}{b_N} > 1. \tag{12}$$

From Equation 10, we obtain that:

$$\frac{u_{NN}}{u_{jj}} = \frac{\tilde{e}_N}{\tilde{e}_j}\frac{\eta_j(\tilde{e}_j)}{\eta_N(\tilde{e}_N)}\frac{x_j}{x_N}. \tag{13}$$

Furthermore, given that $e = \sum_i (x_i / \sum_k x_k) e_i = \sum_i (x_i/\bar{x}) e_i$, we have that:

$$\begin{aligned}
b_l &= \tilde{e}_l - e = (1-\alpha)(e - e_l) \\
\Leftrightarrow e - e_l &= (1-\alpha)\left(e_l\left(\frac{x_l}{\bar{x}} - 1\right) + \sum_{i\neq l}\frac{x_i}{\bar{x}}e_i\right) \\
\Leftrightarrow &= (1-\alpha)\left(-e_l\frac{\sum_{i\neq l} x_i}{\bar{x}} + \sum_{i\neq l}\frac{x_i}{\bar{x}}e_i\right) \\
\Leftrightarrow &= (1-\alpha)\left(\sum_{i\neq l}\frac{x_i}{\bar{x}}(e_i - e_l)\right),
\end{aligned}$$

where $\bar{x} = \sum_i x_i$. Accordingly:

$$\frac{b_l}{b_N} = \frac{\sum_{i\neq l} x_i(e_i - e_l)}{\sum_{i\neq N} x_i(e_i - e_N)}. \tag{14}$$

Using Equation (13), we can rewrite Equation (12) as follows:

$$-\sum_j \frac{e_j}{e_N}\frac{\tilde{e}_N}{\tilde{e}_j}\frac{\eta_j(\tilde{e}_j)}{\eta_N(\tilde{e}_N)}\frac{x_j}{x_N}\frac{b_j}{b_N} > 1 + \sum_{k\neq N}\frac{e_k}{e_N}\frac{\tilde{e}_N}{\tilde{e}_k}\frac{\eta_k(\tilde{e}_k)}{\eta_N(\tilde{e}_N)}\frac{x_k}{x_N}\frac{b_k}{b_N}, \tag{15}$$

where we obtain the two appliance-case as a special case after inserting Equation (14). Rearranging Equation (15) yields:

$$\begin{aligned}
-\sum_j \frac{e_j}{\tilde{e}_j}\eta_j(\tilde{e}_j)x_j b_j &< \sum_k \frac{e_k}{\tilde{e}_k}\eta_k(\tilde{e}_k)x_k b_k \\
\Leftrightarrow -\sum_j \frac{e_j}{\tilde{e}_j}\frac{\partial x_j(\tilde{e}_j)}{\partial p}b_j &< \sum_k \frac{e_k}{\tilde{e}_k}\frac{\partial x_k(\tilde{e}_k)}{\partial p}b_k.
\end{aligned} \tag{16}$$

The sufficient conditions for Inequality (16) to hold closely resemble the two-appliance case. First, the inequality holds when all $e_k$ are sufficiently small for all $k \in \{J+1, ..., N\}$. Second, it holds when the demand response to a price change, $|\partial x_k(\tilde{e}_k)/\partial p|$, is sufficiently small for all $k \in \{J+1, ..., N\}$.

## A2 Sufficient Statistics for Evaluating Consumer Surplus

Let consumers have the following quasi-linear utility function:

$$U(\mathbf{x}, z) = u(\mathbf{x}) + z, \tag{17}$$

where $u(\mathbf{x})$ denotes the utility from consuming $J$ energy services denoted by the vector $\mathbf{x} = (x_1, \ldots, x_J)'$, and $z$ denotes the numeraire good, whose price is normalized to 1. The consumption of energy service $j$ requires energy inputs of $y_j = x_j e_j$, where $e_j$ denotes the energy intensity of energy service $x_j$. Consumers maximize their utility subject to the budget constraint $w = z + \sum_j y_j p$, where $w$ denotes their exogenous income and $p$ denotes the price of energy. Let consumers have biased perceptions of energy intensities $\tilde{e}_j = e_j + b_j(\alpha)$, where $b_j(\alpha)$ is a bias term that decreases (in absolute terms) in the accuracy of beliefs, denoted by $\alpha$.

Following the reduced-form approach to behavioral public finance (Mullainathan et al., 2012), we can write decision utility as:

$$U^d = u(\mathbf{x}) - \sum_j p(e_j + b_j)x_j.$$

Utility maximization yields the FOCs:

$$u_j = p(e_j + b_j) \;\; \forall\, j \in \{1, \ldots, J\}. \tag{18}$$

Furthermore, normative utility is:

$$U^n = U^d + \sum_j p\, b_j x_j.$$

We are interested in the welfare effect of a change in $\alpha$. We find that:

$$\begin{aligned}
\frac{\partial U^n}{\partial \alpha} &= \frac{\partial U^d}{\partial \alpha} + \sum_j p x_j \frac{\partial b_j}{\partial \alpha} + \sum_j b_j p \frac{\partial x_j}{\partial \alpha} \\
&= \sum_j \underbrace{\frac{\partial U^d}{\partial x_j}}_{=0} \frac{\partial x_j}{\alpha} - \sum_j p x_j \frac{\partial b_j}{\partial \alpha} + \sum_j p x_j \frac{\partial b_j}{\partial \alpha} + \sum_j b_j p \frac{\partial x_j}{\partial \alpha} \\
&= \sum_j b_j p \frac{\partial x_j}{\partial \alpha}.
\end{aligned}$$

We can approximate a change in consumer surplus by $\Delta CS \approx \frac{U^n}{\partial \alpha} d\alpha$, which gives:

$$\begin{aligned}\Delta CS &\approx \sum_j b_j p \frac{\partial x_j}{\partial \alpha} d\alpha \\ &= \sum_j \frac{b_j}{e_j} \cdot p \cdot \Delta y_j,\end{aligned}$$

where $\Delta y_j = \Delta x_j e_j$ is the average treatment effect of providing disaggegate information on the electricity consumption of appliance category $j$, and $b_j / e_j$ denotes the relative bias.

Totally differentiating Equation 18 with respect to $\alpha$ yields the following system of equations:

$$\begin{bmatrix} \frac{\partial x_1}{\partial \alpha} \\ \vdots \\ \frac{\partial x_J}{\partial \alpha} \end{bmatrix} d\alpha = p\, H^{-1} \begin{bmatrix} \frac{\partial b_1}{\partial \alpha} \\ \vdots \\ \frac{\partial b_J}{\partial \alpha} \end{bmatrix} d\alpha,$$

where H is the Hessian matrix of consumers' maximization problem. Under the assumption that the Hessian is diagonal, i.e., that utilities from different services are additively separable, we obtain:

$$\Delta x_j = \frac{p\ \Delta b_j}{u_{jj}}, \tag{19}$$

where $\Delta x_j = (\partial x_j / \partial \alpha) d\alpha$ and $\Delta b_j = (\partial b_j / \partial \alpha) d\alpha$.

Totally differentiating Equation (18) with respect to p (when $b_j = 0$) yields the following system of equations:

$$\begin{bmatrix} \frac{\partial x_1}{\partial p} \\ \vdots \\ \frac{\partial x_J}{\partial p} \end{bmatrix} = H^{-1} \begin{bmatrix} e_1 \\ \vdots \\ e_J \end{bmatrix}$$

Under additive separability, we obtain for all $j \in \{1, \dots, J\}$:

$$\frac{\partial x_j}{\partial p_j} = \frac{e_j}{u_{jj}}. \tag{20}$$

Substituting for $u_{jj}$ in Equation (19) and (20) and rearranging yields:

$$\frac{\Delta b_j}{e_j} = \frac{\Delta x_j}{\frac{\partial x_j}{\partial p} p}.$$

Under a pure nudge assumption, $\Delta b_j = -b_j$, and after some expansions, we get:

$$\frac{b_j}{e_j} = -\frac{\frac{\Delta x_j e_j}{x_j e_j}}{\frac{\partial x_j e_j}{\partial p} \frac{p}{x_j e_j}} = -\frac{\frac{\Delta y_j}{y_j}}{\frac{\partial y_j}{\partial p} \frac{p}{y_j}} = -\frac{(\Delta y_j / y_j)}{\eta_j},$$

where $(\Delta y_j / y_j)$ is the ATE, expressed as a percentage of the mean of actual consumption, and $\eta_j = \frac{\partial y_j / \partial p}{y_j / p}$ denotes the price elasticity of energy consumption for appliance category $j$.

## A3 Nonintrusive Appliance Load Monitoring

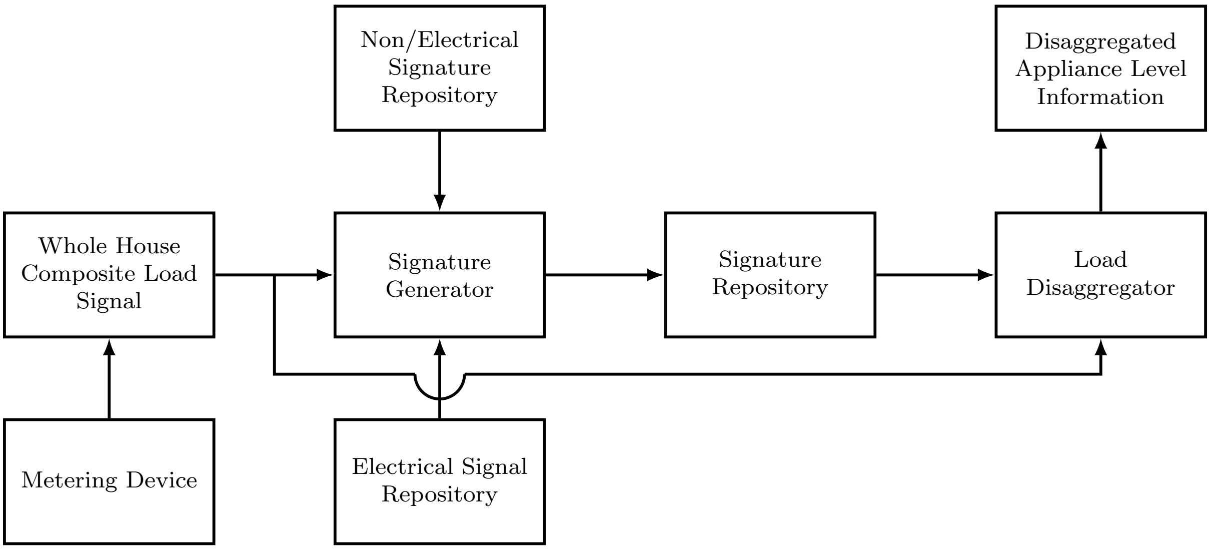
In this section, we describe the so-called nonintrusive appliance load monitoring (NALM) algorithm, which is employed in our study to determine appliance-level consumptions from high-frequency measurements of total electricity consumption of a household. Based on Hart (1992), we first introduce the general approach of NALM algorithms and then present the structure of the algorithm employed in our study (see Gupta et al. 2017 for details).

NALM algorithms exploit that applianced are typically wired in parallel, so that the power they consume is additive. The fact that appliances are switched on and off creates distinct patterns in high frequency data, which can be used to decode appliance-level consumptions. This decoding process is simplified by the fact that every appliance has a distinct signature, i.e. a characteristic pattern of the power it consumes, during use. For example, washing machines use different amounts of power when they heat water, wash, and spin. NALM algorighms represent appliances as so-called finite state machines (FSMs), i.e., model appliances as having a finite set of states (e.g. off, heating, washing, spinning) and transitions between states (e.g. off $\rightarrow$ heating $\rightarrow$ washing $\rightarrow$ spinning $\rightarrow$ off). These FSM models are then mapped with observable shifts in electricity usage to determine appliance-level consumptions. While the methodology has already been proposed almost 30 years ago (e.g. Hart 1992), the mapping between FSMs and empirical transitions has been facilitated by recent advances in machine learning.

The structure of the NALM algorighm used in our study is depicted in Figure A1. A metering device records both the electric power consumed and the voltage at a high-frequency (in our case, every few seconds), thus measuring the "whole house composite load signal". This signal is analyzed in order to detect so-called transitions in the data, i.e., changes in consumption levels.

A core element of NALM algoritms is a signature repository, which collects appliance signatures. To construct this repository, the algorighm uses a comprehensive collection of electrical load signature patterns of common appliances. For example, the load signature of an electric clothes dryer typically consists of three states (off, high heat, cool down) and of typical power consumptions for each of these states (e.g. 0 W, 4500-6000 W, 200-300 W, respectively). Another input is the non-electric signature repository which includes typical behavioral parameters of appliance usages (e.g. that a clothes dryer is typically used for 30-75 min). Based

**Figure A1:** Schematic Representation of the NALM Algorighm




*Notes*: The representation is based on Gupta et al. (2017).

on these inputs, the household specific signature repository is constructed as follows. First, the NALM algorighm uses methods from cluster analysis to define clusters of shifts in electricity consumption. In a subsequent step, it classifies these clusters by comparing them to the typical states and transitions of a particular appliance. This classification step is typically performed via supervised machine learning techniques based on training data.

In a subsequent step, a load disaggregator uses the whole house composite load signal as well as the signal repository to decompose the entire signal into appliance-specific consumptions. In our case, load disaggregation was performed once a day, so that households could access appliance-level information always on day following appliance usage.

We depict exemplary signatures from our data for the appliances dish washer, washing machine, dryer, and oven in Figure A4. The figures show that appliances leave a distinct pattern in high-frequency electricity consumption data, which allows to determine the start and end date of an appliance, as well as the electricity consumed by it. For the dryer, for example, it is easy to spot the pattern of a long heating period after switching on the appliance, followed by an iteration between periods for letting cool down the laundry and heating it up again.

## A4 Non-Experimental Control Group

To identify the overall conservation effect, we additionally obtain data from a non-experimental sample of smart meter households that are served by the same utility. These households have agreed to report their electricity consumption to the grid operator who uses the data to forecast load profiles. We obtain smart meter data in 15 minute invervals for 577 households, starting from November 1, 2016, which is when our field test started. This data allows us to identify the effect sizes for all experimental conditions relative to obtaining no feedback at all. To ensure that observable household characteristics are balanced across our experimental and non-experimental sample, we select a control group using a propensity matching method.

As the left panel of Figure A7 in the Appendix shows, the baseline consumption in the non-experimental sample is slightly larger than in the experimental sample. To account for such differences, we follow a matching approach to determine the subset of control households that we use in our analyses, denoted henceforth as matched control (MC) households. For every participant in the EC group, we determine the nearest neighbor in the non-experimental sample by implementing a 1:1 matching algorithm without replacement, based on three covariates. First, we match on the average per day electricity consumption in the baseline period. Second, to control for possible differences between both populations in the presence at the start of the experiment, we match on the electricity consumption in November 2016, when the meter was installed. Third, we control for different lengths of the billing period. As a result, we obtain a group of 140 matched households.

The right panel of Figure A7 shows that, after matching, participants in the EC and in the non-experimental MC group are balanced in terms of their baseline use. In Table A6, we additionally assess the balance in terms of further baseline billing information. We find that the average billing period starts and ends around 20 days earlier in our experimental sample, compared to the non-experimental sample. To account for such differences in the baseline billing period, we include a full set of week-of-baseline fixed effects in our post-matching regressions.

Based on the Matched Control (MC) group and our experimental sample, we estimate the average total conservation effects of our five experimental conditions by estimating the following equation via ordinary least squares (OLS):

$$Y_{it}^{norm} = \alpha Y_i^b + \beta EC_i + \boldsymbol{\gamma}' \mathbf{T_i} + \nu_t + \mu_w^b + \epsilon_{it}, \tag{21}$$

where $Y_{it}^{norm}$ denotes the electricity use of participant i at day t, divided by the average daily electricity use in the experimental control group, and $EC_i$ denotes households in the experimental control group. To account for serial correlation in the error terms, we cluster standard errors at the household-match level, where households in the Matched Control group obtain the same id as their matched counterparts (Abadie and Spiess, 2019). In addition, we estimate the followinig specification, which allows us to identify the impact of the app elements:

$$Y_{it}^{norm} = \alpha Y_i^b + \beta AF_i + \beta_1 D_i + \beta_2 M_i + \beta_3 R_i + \beta_4 M_i R_i + \nu_t + \mu_w^b + \epsilon_{it}, \tag{22}$$

where $AF_i$ denotes that household $i$ has access to aggregate feedback, which is the case for all households in our experimental sample.

Panel a) of Table A9 presents the average conservation effects in the treatment groups, relative to the $MC$ group. It shows that we do not detect statistically significant electricity savings for households in the $EC$ group, which obtain only aggregate feedback (Column 1). This finding is supported when we investigate hourly treatment effects (see Figure A8 in the Appendix). Using our hourly estimates, we cannot reject the null hypothesis that all hourly point estimates are zero: $F(24, 277) = 1.24$, p-value: 0.2027.

Hence, the total conservation effects that we estimate align closely with the electricity conservation effects relative to the experimental control group from Table 3. The ATE estimate for participants in the four treatment conditions $T_1 - T_4$ amounts to $-5.2\%$ and is statistically significant at all conventional levels. When we estimate the ATEs separately for all treatment groups (Column 2), we do not find sizable differences, with point estimates ranging from $-4.3\%$ in $T_2$ to $-6.0\%$ in $T_3$. As shown in Panel b), the total conservation effects can be clearly attributed to disaggregate feedback, which yields a 5.3% reduction in electricity use, relative to the $MC$ group. Again, we cannot detect that additional app features significantly increase conservation effects.

## A5 Heterogeneity in Treatment Effects by Baseline Consumption

To investigate treatment effect heterogeneity in more detail, we estimate two equations. First, we estimate the following equation by OLS, using only our experimental sample:

$$Y_{it}^{norm} = \alpha Y_i^{b,dm} + \beta EC_i Y_i^{b,dm} + \nu_t + \mu_w^b + \epsilon_{it},$$

where $Y_i^{b,dm}$ denotes the baseline consumption of household $i$, expressed as a percentage of the average daily consumption in the EC group. We also demean this variable, so that we can interpret $\hat{\beta}$ as the average treatment effect at the mean of baseline consumption (10.4 kWh). In addition, we estimate the following equation using the experimental and the matched control groups:

$$Y_{it}^{norm} = \alpha Y_i^{b,dm} + \beta_0 EC_i + \beta_1 EC_i Y_i^{b,dm} + \gamma_0 D_i + \gamma_1 D_i Y_i^{b,dm} + \nu_t + \mu_w^b + \epsilon_{it}.$$

Using the experimental sample, the estimates for the interaction term between the disaggregation dummy $D$ and baseline electricity consumption amounts to $-0.66$, but is not statistically significant (Panel a of Table A7). When using data from the experimental sample and matched control observations, we find that the interaction effect reaches -0.124 and is statistically significant at the 1% level. Furthermore, we cannot reject the null hypothesis that both the main effect of $EC$ and its interaction with the baseline electricity use are zero (F-test stat.: 1.35, p-value: 0.26), while we can reject the corresponding null hypothesis for the disaggregation groups $D$ (F-test stat: 7.47, p-value: 0.001).

## A6 Conservation Effects after Core Study Period

After our core study period of 6 months, participants continued to have access to the app, but with limited functionality in treatment group $T_2 - T_4$. In particular, participants were not invited to take part in efficiency challenges any longer, but still received appliance-level consumption feedback. To investigate the treatment effects after the core study period, we estimate Equation (4) and (5), but restrict the sample to the time period from month 7 of the field test onwards.

Panel a) of Table A10 gives the average conservation effect relative to the $EC$ group and shows that the point estimate for the treatment effect in the disaggregation groups $T_1 - T_4$ amounts to $-1.2\%$, but is not statistically significant at any conventional level. Using the non-experimental sample to identify the total conservation effect (Panel b of Table A10), we find that disaggregate feedback yields a persistent reduction by $-1.5\%$ when using the full sample (Column 3) and $-5.0\%$ when restricting our sample to a balanced panel of households that we can observe for the full study period (Column 4).

We test for differences in the average treatment effect between the core study period and the period thereafter by estimating the following regressions based on our experimental sample, as well as on our experimental sample and our matched control group, respectively:

$$Y_{it}^{norm} = \alpha Y_i^b + \tau\ AfterCSP_t + \beta_1 D_i + \beta_2 D_i \cdot AfterCSP_t + \nu_t + \mu_w^b + \epsilon_{it},$$

$$Y_{it}^{norm} = \alpha Y_i^b + \tau\ AfterCSP_t + \gamma_1 EC_i + \gamma_2 EC_i\ AfterCSP_t + \delta_1 D_i + \delta_2 D_i\ AfterCSP_t + \nu_t + \mu_w^b + \epsilon_{it},$$

where $AfterCSP_t$ equals one if day $t$ occurs after the beginning of study month 7 and zero otherwise. To avoid that differences in average treatment effects arise from changes in the sample composition over time, we use a balanced panel for both regressions. The estimate $\hat{\beta}_2$ identifies the difference in the average treatment effect from appliance-level feedback relative to the experimental control group: $ATE_{\text{AfterCSP}}$-$ATE_{\text{CSP}}$, where $ATE_{\text{CSP}}$ denotes the ATE during the core study period. The estimate $\delta_2$ has the same interpretation, but gives the change in the ATEs relative to the matched control group, i.e., relative to obtaining no feedback at all. As shown in Table A11, we find that that $\hat{\beta}_2$ and $\hat{\delta}_2$ are positive, but not statistically significant from zero at any conventional level.

## A7 Appliance-Level Treatment Effects by Hour of the Day

To identify the hourly average treatment effects at the appliance-level, we estimate the following model separately for every appliance:

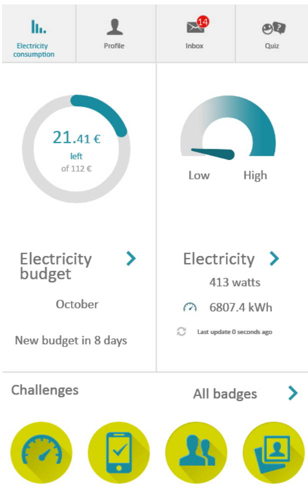
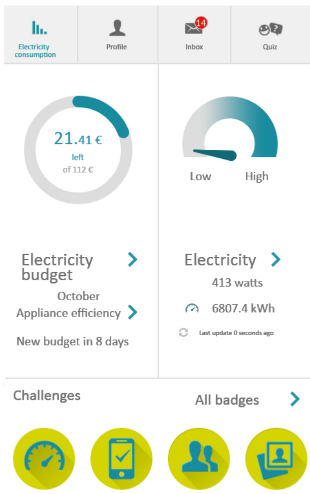
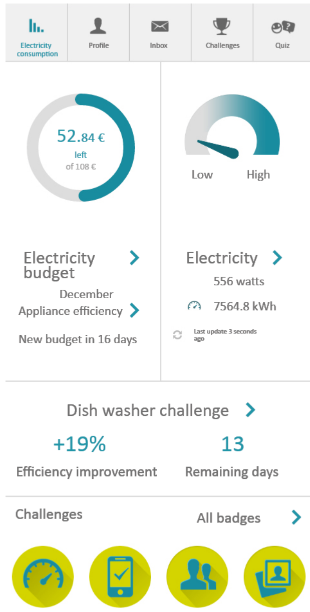
$$Y_{ithj}^{norm} = \alpha Y_i^b + \sum_{h=1}^{24} \beta_j^h D_i + \nu_t + \mu_h + \epsilon_{ihdj}.$$

In Figure A9, we show how appliance-level consumptions change in response to appliance-level feedback over the hours of a day. For dish-washers, dryers, and washing machines, we find a distinct pattern that savings occur only during the day, between 7 a.m. and 15 p.m., which coincides with typical usage patterns of these appliances. By contrast, consumption reductions in the category *Other Appliances* occur particularly during late morning hours, as well as during late evening hours, between 8 p.m. and 4 a.m. As the categories *Refrigeration* and *Always-On* are measured daily, we cannot estimate hourly treatment effects for them.

# Supplementary Figures

**Figure A2:** Visualization of Screens I

**Start Screen (by Experimental Condition)**

| Experimental Control (EC) | $T_1$ | $T_2$-$T_4$ |
|---|---|---|
|  |  |  |







**Treatment-Specific Screens**

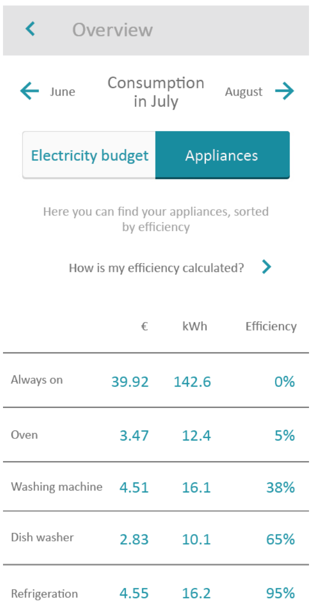
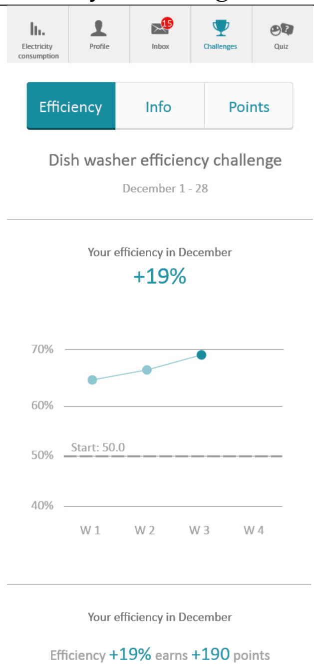
| appliance-level feedback ($T_1$-$T_4$) | Efficiency Challenge ($T_2$-$T_4$) |
|---|---|
|  |  |

**Figure A3:** Visualization of Screens II

**Further Screens (EC and $T_1$-$T_4$)**

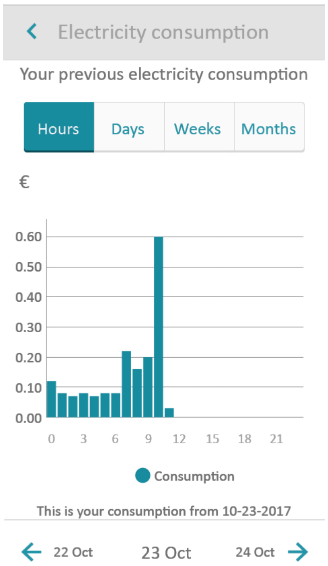
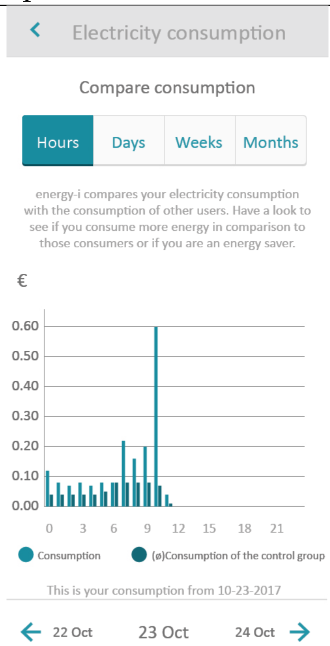
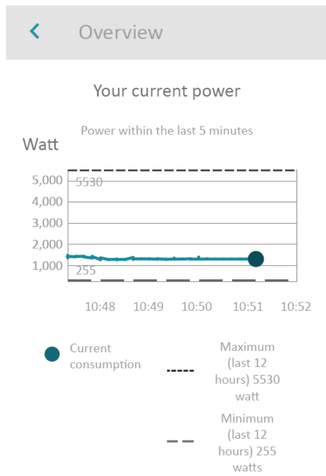
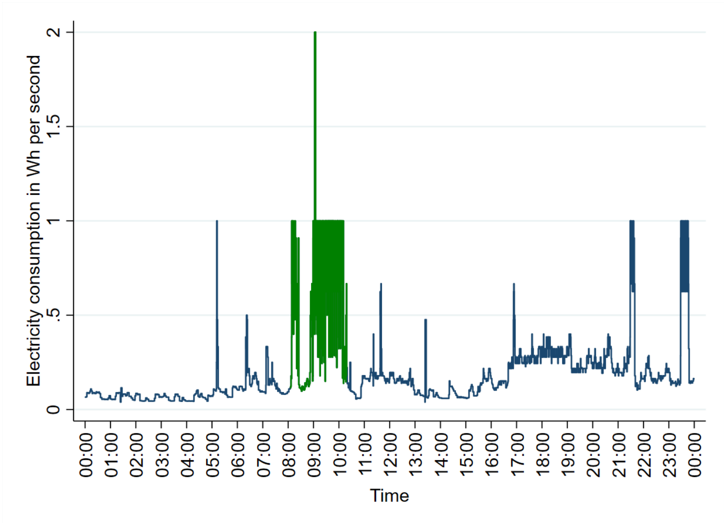
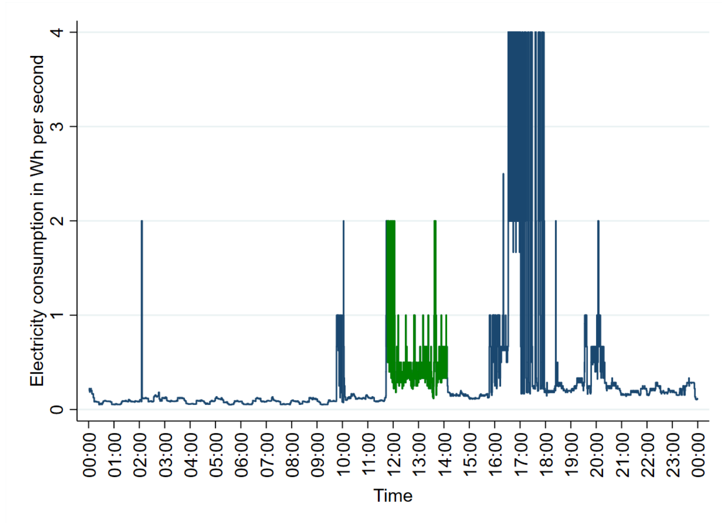
| Comparison with Others | Comparison with own History | Current Power |
|---|---|---|
|  |  |  |

**Figure A4:** Appliance Signatures

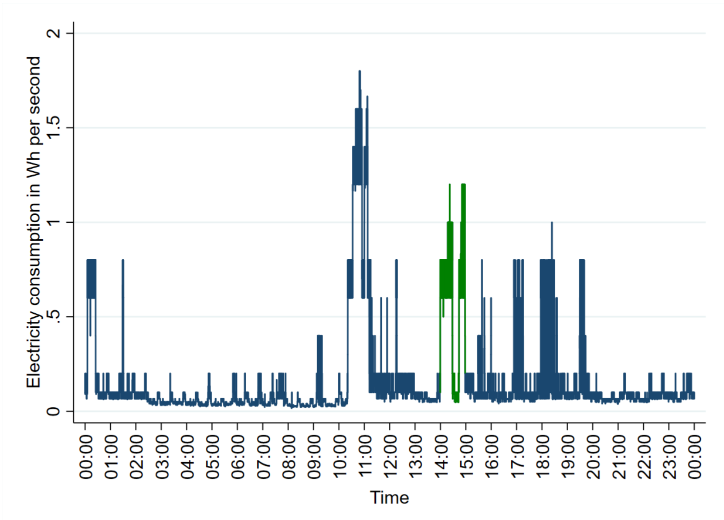
**(a)** Washing Machine



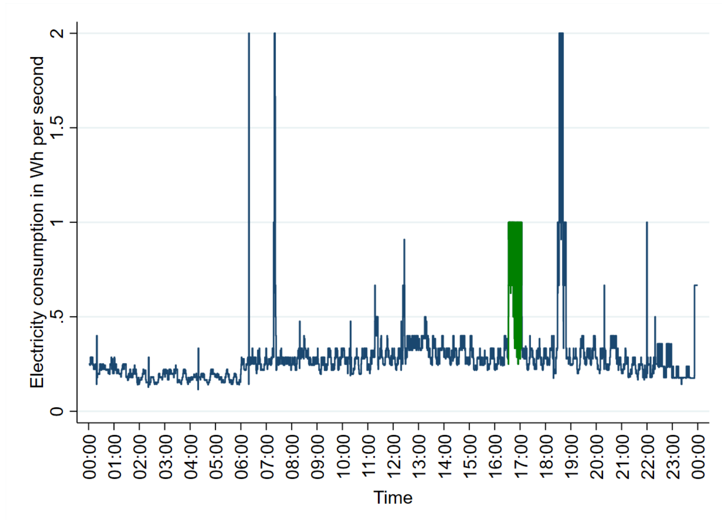

**(b)** Dryer




**(c)** Dish Washer

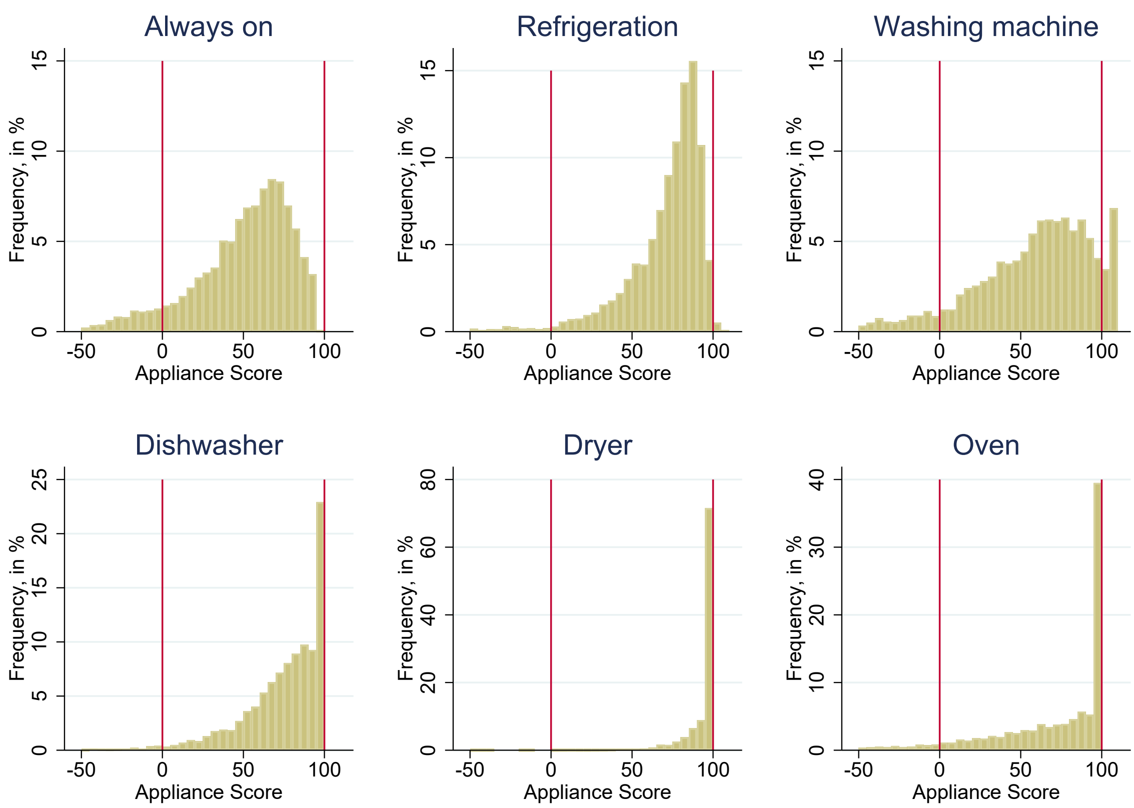



**(d)** Oven



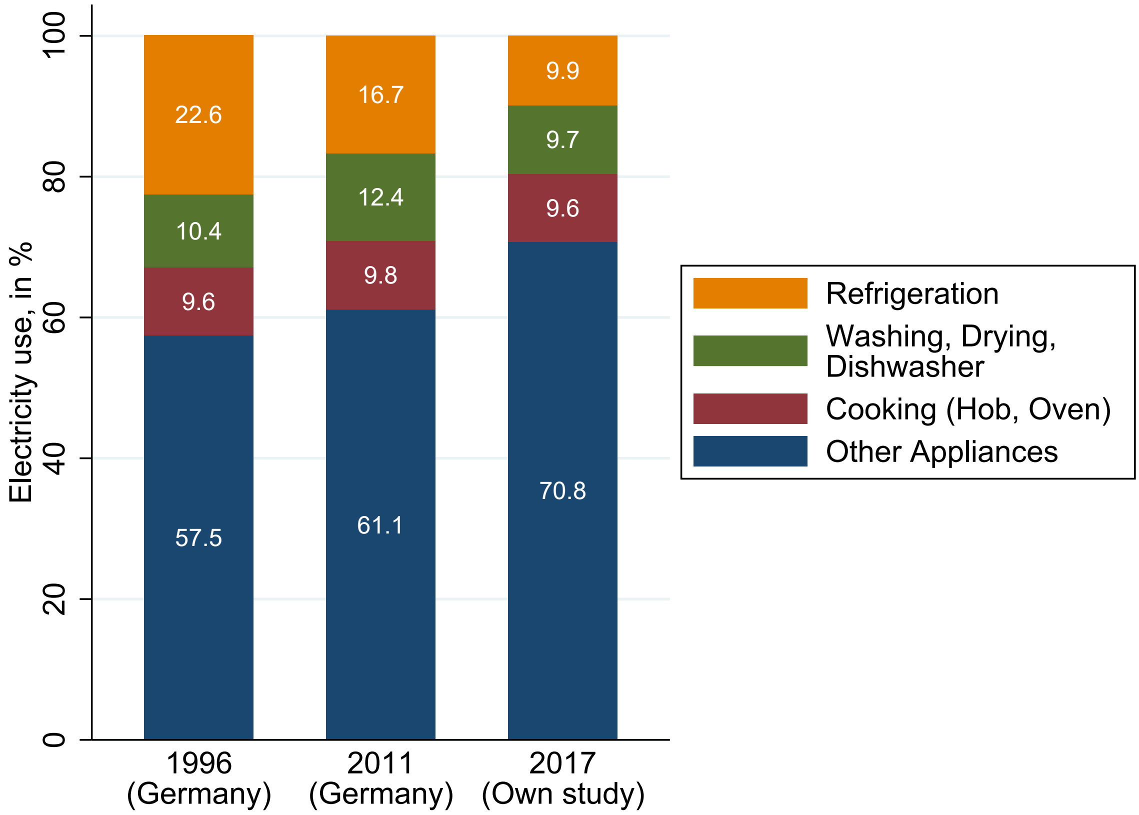

*Notes:* The graphs depict exemplary appliance signatures as detected by the nonintrusive appliance load monitoring (NALM) algorithm in our study. Periods highlighted in green correspond to a detected appliance use event for the respective appliance.

**Figure A5:** Distribution of Appliance Scores



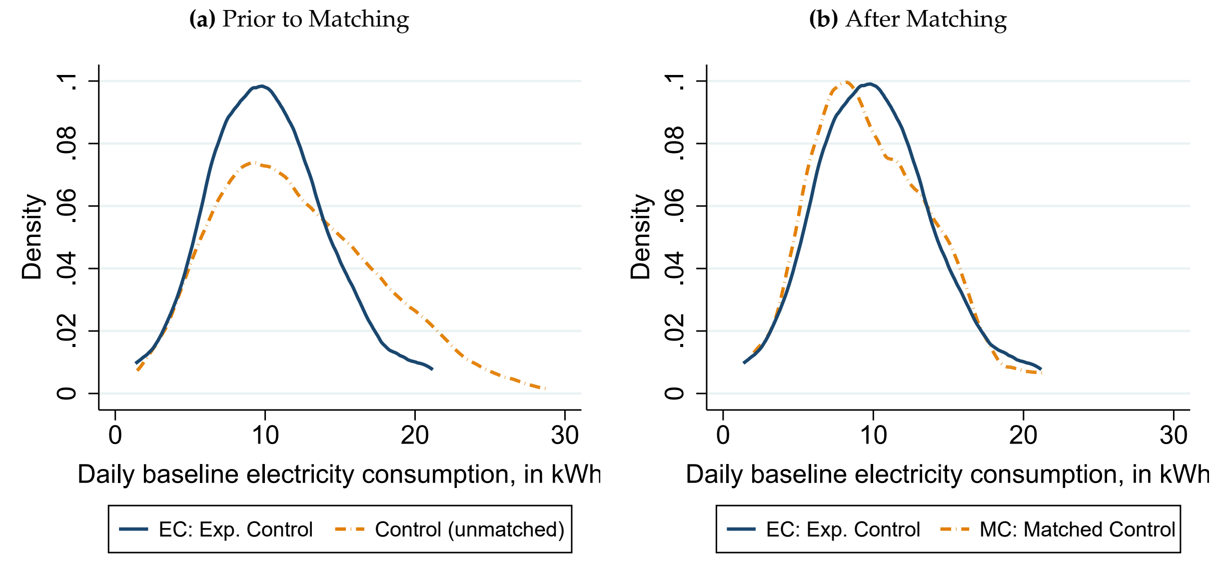

*Notes:* Appliance Scores are calculated as follows: Appliance Score = 100 × (Monthly Appliance Consumption - $\text{Bench}_{\text{low}}$)/($\text{Bench}_{\text{up}}$ - $\text{Bench}_{\text{low}}$), where $\text{Bench}_{\text{low}}$ and $\text{Bench}_{\text{high}}$ correspond to pre-determined benchmark values for high and low appliance uses, respectively. These benchmarks are based on survey data on typical appliance uses and product data sheets on the technical efficiency of appliances currently used in German households (for details, see Table A2).

**Figure A6:** Decomposition of Electricity Uses, by Appliance Category



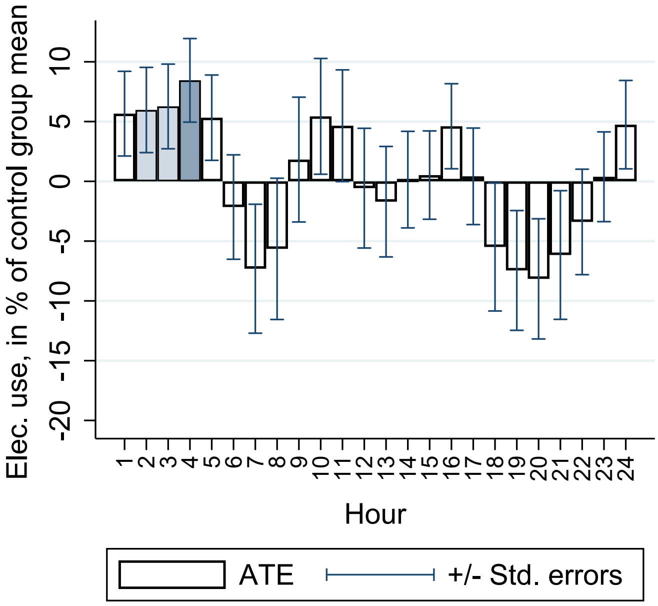

*Notes:* Values are expressed as percentages of total electricity consumption in a year. Values for 1996 and 2011 are drawn from BDEW (2016), Energie-Info - Stromverbrauch im Haushalt, Bundesverband der Energie- und Wasserwirtschaft. Values from our study are calculated for the Experimental Control group. As we do not have data on hobs, we extrapolate their consumption based on the rule-of-thumb that a hob accounts for 77.5% (75-80%) of total electricity consumption for cooking, as stated by the energy efficiency advocacy HEA (`www.hea.de/fachwissen/herde-backoefen/betriebswerte-und-energieverbrauchskennzeichnung`, last access: February 27, 2020).

**Figure A7:** Balancing in Terms of Basline Electricity Consumption between Study Participants and the Non-Experimental Sample



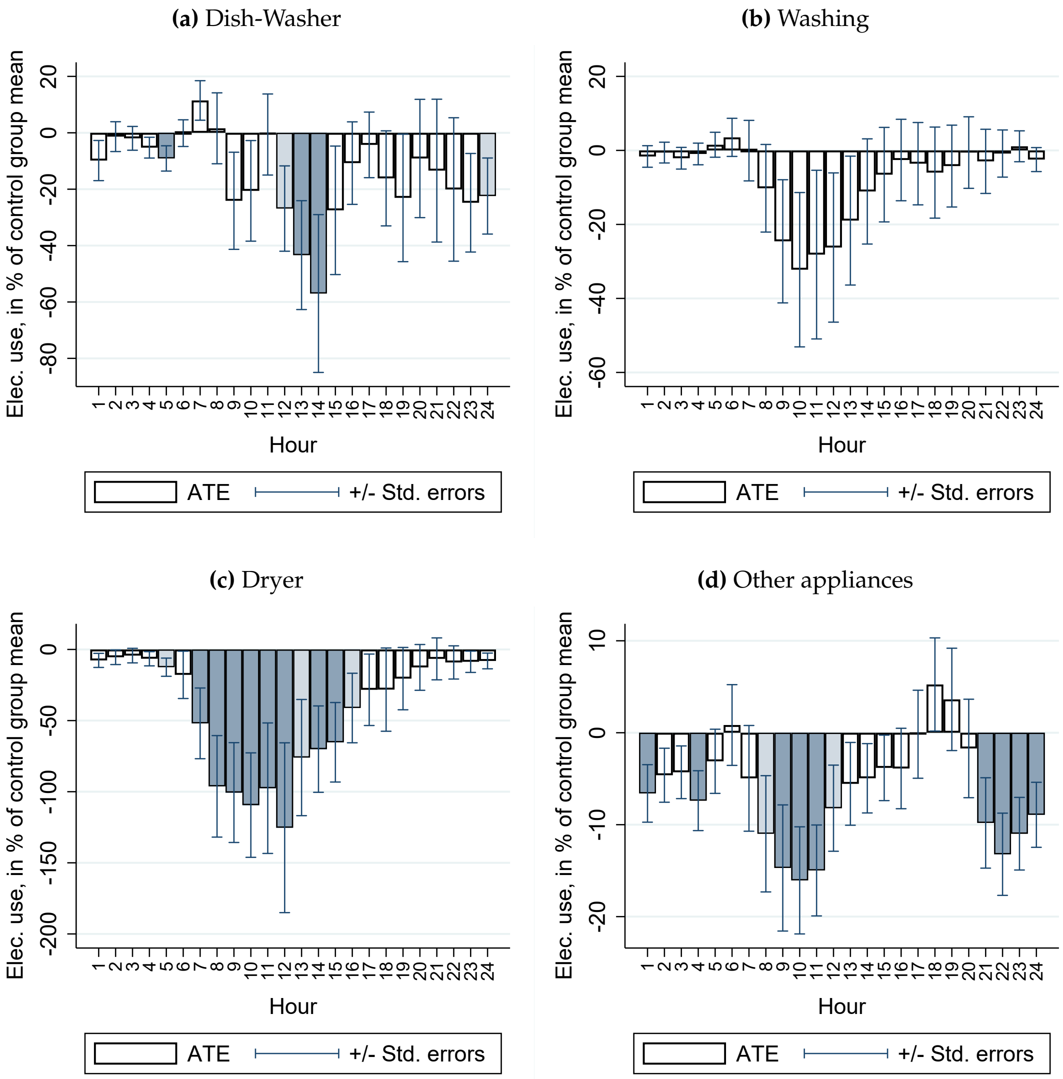

*Notes:* "Control (unmatched)" denotes all control group households, while "MC: Matched Control" denotes the group of households that have been matched to households in the Experimental Control (*EC*) condition.

**Figure A8:** ATE of Experimental Control (*EC*) on Hourly Electricity Consumption (relative to Matched Control)




*Notes:* Shaded bars indicate that treatment effects are statistically significant at the 1% (blue shaded) or 5% (light blue shaded) level. Whiskers indicate a range of +/- 1 standard error (clustered at the household-match level). The outcome variable is daily electricity consumption, divided by the mean in the *EC* group. Using participants in the *MC* and the *EC* groups, we estimate the following equation: $Y_{ith}^{norm} = \alpha Y_i^b + \sum_{h=1}^{24} \beta^h EC_i + \nu_t + \mu_h + \epsilon_{ihd}$. We cannot reject the null hypothesis that all hourly point estimates are zero: $F(24, 277) = 1.24$, p-value: 0.2027.

**Figure A9:** ATE of Disaggregation (*D*) on Hourly Electricity Consumption, by Appliance Category (Relative to Exp. Control)




*Notes:* Shaded bars indicate that treatment effects are statistically significant at the 1% (blue shaded) or 5% (light blue shaded) level. Whiskers indicate a range of +/- 1 standard error (clustered at the household level). The outcome variable is hourly electricity consumption at the appliance level, divided by the respective mean in the *EC* group. The categories *Refrigeration* and *Always-On* are measured daily, so that we cannot estimate hourly treatment effects for them.

## Supplementary Tables

**Table A1:** Comparison of Experimental Sample with German Population

| | Experimental sample | German population |
|---|---|---|
| Baseline cons., in kWh/day | 10.4 | 8.6 |
| # of occupants | 2.5 | 2.0 |
| # of refrigeration appliances | 2.2 | 2.4 |
| Net income, in EUR per month | 3,103 | 3,314 |
| Own property, in % | 75.6 | 44.0 |
| Employed, in % | 51.2 | 57.8 |
| Share of females, in % | 47.4 | 50.7 |
| Age, in years | 45.8 | 44.3 |
| Baseline cons., in kWh/day (1 person household) | 6.2 | 5.5 |
| Baseline cons., in kWh/day (2 person household) | 9.8 | 8.8 |
| Baseline cons., in kWh/day (3+ person household) | 11.6 | 13.3 |
| 1 person household, in % | 12.0 | 41.8 |
| 2 person household, in % | 51.1 | 33.5 |
| 3+ person household, in % | 36.9 | 24.7 |

*Notes:* German averages are taken from the following German Statistical Office publications (for the year 2017): Populalation Statistics (Mikrozensus); Environmental-Economic Accounting; Income, Receipts, and Expenditures; Consumption Expenditures. The average electricity consumption is for the baseline year 2016 (`https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/Materialfluesse-Energiefluesse/Tabellen/stromverbrauch-haushalte.html`, last access: March 9, 2020).

**Table A2:** Benchmarks for the Calculation of Efficiency Scores

| | 1 person | 2 persons | 3 persons | 4 persons | 5 persons |
|---|---|---|---|---|---|
| Always on | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 90.7 | 125.7 | 158.7 | 174.1 | 190.5 |
| Refrigeration (1 appliance) | 4.8 | 4.8 | 5.7 | 5.9 | 6.3 |
| | 39.9 | 44.3 | 49.3 | 54.2 | 59.6 |
| Refrigeration (2 appliances) | 8.3 | 11.5 | 11.5 | 12.4 | 12.5 |
| | 66.5 | 73.9 | 82.1 | 90.3 | 99.3 |
| Refrigeration (3+ appliances) | 8.3 | 11.5 | 11.5 | 12.4 | 12.5 |
| | 94.1 | 103.1 | 112.1 | 121.6 | 132.0 |
| Washing machine | 1.2 | 1.8 | 2.3 | 2.8 | 3.4 |
| | 14.0 | 23.9 | 36.4 | 43.6 | 44.6 |
| Dishwasher | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 17.7 | 27.7 | 34.0 | 43.6 | 54.6 |
| Dryer | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 31.9 | 62.7 | 68.8 | 70.5 | 72.2 |
| Oven | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| | 6.3 | 12.6 | 19.0 | 25.3 | 31.6 |

*Notes:* The benchmarks were calculated taking the technical energy efficiencies of appliances on the market into account(energy efficient – energy inefficient), as well as typical usage behaviors (rare user – heavy user). The main sources for technical efficiency are product data sheets for efficient appliances from EcoTopTen, an online platform for energy efficient products (URL: `https://www.ecotopten.de/`), as well as product data sheets of inefficient appliances from product tests by Stiftung Warentest, a renowned German consumer organisation (URL: `www.test.de`). In addition, we use information on typical usage behaviors from surveys such as the German Residential Energy Consumption Survey (RWI-GRECS, URL: `http://www.rwi-essen.de/forschung-und-beratung/fdz-ruhr/datenangebot/mikrodaten/rwi-grecs`). For the category *Always-On*, we calculate the upper benchmark based on the stand-by electricity use for a range of appliances, including TVs, hifi systems, PCs, routers, telefones, coffee machines, washing machines, and microwaves (using energy inefficient appliance varieties). These calculations take typical appliance possessions by household size into account. We set the lower benchmark to zero, assuming that always-on consumption can be avoided. For the category *Refrigeration*, we calculate benchmarks based on the number of appliances (1, 2, and more than 2). For each of them, we consider the most energy efficient and inefficient appliances available on the market, whose cooling volume is as recommended for the respective household size. For the categories *Washing machine*, we use data on the energy consumption per use for energy efficient and inefficient appliances and consider the typical frequency of use for heavy users and rare users (for every household size). We proceed in the same manner for the categories *Dish washer*, *Dryer*, and *Oven*, but assume that the lower benchmark is zero as households can substitute these energy services with hand-washing, dry-hanging and eating-out, for example. Details on the calculations can be obtained from the authors upon request.

**Table A3:** Distribution of Daily Use, in kWh/day

| | Mean | Std. dev. | Min | Max | p1 | p25 | p50 | p75 | p99 | N |
|---|---|---|---|---|---|---|---|---|---|---|
| Total electricity | 9.64 | 5.16 | 0.00 | 100.07 | 1.64 | 6.15 | 8.81 | 12.12 | 25.90 | 160,886 |
| Always on | 2.38 | 1.89 | 0.00 | 47.49 | 0.16 | 1.08 | 1.91 | 3.16 | 8.89 | 140,538 |
| Refrigeration | 0.98 | 0.59 | 0.00 | 11.12 | 0.08 | 0.61 | 0.85 | 1.19 | 2.95 | 140,536 |
| Dishwasher | 0.28 | 0.60 | 0.00 | 14.15 | 0.00 | 0.00 | 0.00 | 0.00 | 2.53 | 127,657 |
| Washing mach. | 0.46 | 1.05 | 0.00 | 40.30 | 0.00 | 0.00 | 0.00 | 0.54 | 4.88 | 137,897 |
| Dryer | 0.13 | 0.57 | 0.00 | 36.06 | 0.00 | 0.00 | 0.00 | 0.00 | 2.79 | 99,201 |
| Oven | 0.22 | 0.75 | 0.00 | 20.57 | 0.00 | 0.00 | 0.00 | 0.00 | 3.54 | 140,538 |
| Other appliances | 5.27 | 3.70 | 0.00 | 56.58 | 0.02 | 2.84 | 4.51 | 6.79 | 18.08 | 140,538 |

*Notes:* $p1$ denotes the first percentile, $p25$ the 25th percentile, etc. $N$ denotes the number of daily appliance-level observations.

**Table A4:** Distribution of Use per Utilization, in kWh

| | Mean | Std. dev. | Min | Max | p1 | p25 | p50 | p75 | p99 | N |
|---|---|---|---|---|---|---|---|---|---|---|
| Always-On | 2.37 | 2.13 | 0.000 | 200.40 | 0.16 | 1.04 | 1.86 | 3.16 | 8.94 | 225,568 |
| Refrigeration | 0.96 | 0.57 | 0.000 | 11.12 | 0.08 | 0.61 | 0.84 | 1.18 | 2.88 | 224,809 |
| Dishwasher | 1.06 | 0.42 | 0.002 | 10.98 | 0.34 | 0.77 | 1.00 | 1.28 | 2.36 | 54,365 |
| Washing mach. | 0.97 | 0.74 | 0.001 | 23.97 | 0.05 | 0.51 | 0.78 | 1.22 | 3.58 | 105,488 |
| Dryer | 1.05 | 1.11 | 0.003 | 38.46 | 0.03 | 0.37 | 0.76 | 1.38 | 4.98 | 21,362 |
| Oven | 0.80 | 1.08 | 0.001 | 26.12 | 0.01 | 0.20 | 0.48 | 0.98 | 5.25 | 63,152 |

*Notes:* In the categories *Always-On* and *Refrigeration*, the unit of utilization is one day. For all other appliances, average use is given per utilization event. $p1$ denotes the first percentile, $p25$ the 25th percentile, etc. $N$ denotes the number of appliance-use events.

**Table A5:** Distribution of Payoffs from Appliance Challenges, in EUR

| | Mean | Std. dev. | Min | Max | p1 | p25 | p50 | p75 | p99 | N |
|---|---|---|---|---|---|---|---|---|---|---|
| $T_3$ | 6.3 | 7.6 | 0.0 | 20.0 | 0.0 | 0.0 | 2.0 | 12.0 | 20.0 | 4,048 |
| $T_5$ | 4.5 | 2.9 | 0.1 | 9.8 | 0.1 | 1.9 | 4.5 | 7.0 | 9.6 | 3,616 |

*Notes:* Statistics refer to monthly payoffs from the appliance challenges in the treatment groups $T_3$ and $T_5$. *p*1 denotes the first percentile, *p*25 the 25th percentile, etc. *N* denotes the number of month-appliance challenge observations.

**Table A6:** Balance Experimental Sample vs. Matched Non-Experimental Sample

| | Difference EC-MC | Std. Err. | P-val. | N |
|---|---|---|---|---|
| Baseline elec. cons., in kWh/day | 0.23 | 0.25 | 0.36 | 280 |
| Meter baseline elec. cons., in kWh/day | 0.34 | 0.35 | 0.33 | 280 |
| Length of baseline period, in days | 0.81 | 1.11 | 0.47 | 280 |
| Start of baseline billing, in days | −19.14 | 13.34 | 0.15 | 280 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors in parantheses, clustered at the household level.

**Table A7:** Heterogeneity in ATE, by Baseline Consumption

**(a)** Experimental Sample

| | Estimate | Std. Err. |
|---|---|---|
| $D$: Disaggregation | -0.054*** | (0.018) |
| $D \times Y^{b,dm}$ | -0.066 | (0.052) |
| $Y^{b,dm}$: Baseline elec. use (demeaned) | 0.939*** | (0.045) |
| Day fixed effects | ✓ | |
| Week of baseline FE | ✓ | |
| $R^2$ | 0.5480 | |
| Number of obs. | 106,283 | |
| Number of households | 700 | |

**(b)** Sample Including Matched Controls

| | Estimate | Std. Err. |
|---|---|---|
| $D$: Disaggregation | -0.057*** | (0.016) |
| $D \times Y^{b,dm}$ | -0.124*** | (0.050) |
| $Y^{b,dm}$ | 0.997*** | (0.042) |
| $EC$: Exp. Control | -0.002 | (0.021) |
| $EC \times Y^{b,dm}$ | -0.058 | (0.059) |
| Day fixed effects | ✓ | |
| Week of baseline FE | ✓ | |
| $R^2$ | 0.5542 | |
| Number of obs. | 127,342 | |
| Number of households | 840 | |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parantheses and clustered at the household and the household-match level for Panel a) and b), respectively. The outcome variable is daily electricity consumption, divided by the mean in the $EC$ group. Baseline electricity use $Y^{b,dm}$ is demeaned.

**Table A8:** ATEs on Daily Electricity Consumption, Relative to Exp. Control (Balanced Panel)

(a) Effect of Experimental Conditions

| | (1) | (2) |
|---|---|---|
| $D$: $T_1 - T_4$ | -0.064*** | |
| | (0.023) | |
| $T_1$ | | -0.078*** |
| | | (0.026) |
| $T_2$ | | -0.052* |
| | | (0.027) |
| $T_3$ | | -0.073*** |
| | | (0.028) |
| $T_4$ | | -0.055** |
| | | (0.028) |
| $Y^b$: Baseline elec. use | 0.889*** | 0.890*** |
| | (0.027) | (0.027) |
| Day fixed effects (FE) | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ |
| $R^2$ | 0.5578 | 0.5580 |
| Number of obs. | 79,562 | 79,562 |
| Number of households | 460 | 460 |

(b) Effects of App Elements

| | (3) | (4) | (5) |
|---|---|---|---|
| $D$: Disaggregation | -0.064*** | -0.074*** | -0.074*** |
| | (0.023) | (0.025) | (0.025) |
| $M$: Monetary incentives | | 0.022 | 0.022 |
| | | (0.016) | (0.019) |
| $R$: Ranking | | 0.001 | 0.001 |
| | | (0.015) | (0.022) |
| $M$: Monet. inc. $\times$ $R$: Rank. | | | -0.000 |
| | | | (0.033) |
| $Y^b$: Baseline elec. use | 0.889*** | 0.889*** | 0.889*** |
| | (0.027) | (0.027) | (0.027) |
| Day fixed effects (FE) | ✓ | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ | ✓ |
| $R^2$ | 0.5578 | 0.5580 | 0.5580 |
| Number of obs. | 79,562 | 79,562 | 79,562 |
| Number of households | 460 | 460 | 460 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. The regressions are based on a balanced panel for the core study period (months 1-6). Standard errors are in parantheses and clustered at the household-match level. The outcome variable is daily electricity consumption, divided by the mean in the *EC* group. $D$ equals one for households in the groups $T_1 - T_4$, $M$ equals one for households in the groups $T_2$ and $T_4$, and $R$ equals one for households in the groups $T_3$ and $T_4$, while being zero for the other participants, respectively.

**Table A9:** ATEs on Daily Electricity Consumption, Relative to Matched Control

(a) Effect of Experimental Conditions

| | (1) | (2) |
|---|---|---|
| *EC*: Exp. Control | 0.001 | 0.001 |
| | (0.020) | (0.020) |
| $D$: $T_1 - T_4$ | -0.052*** | |
| | (0.016) | |
| $T_1$ | | -0.060*** |
| | | (0.020) |
| $T_2$ | | -0.043** |
| | | (0.020) |
| $T_3$ | | -0.064*** |
| | | (0.020) |
| $T_4$ | | -0.044** |
| | | (0.022) |
| $Y^b$: Baseline elec. use | 0.915*** | 0.915*** |
| | (0.019) | (0.019) |
| Day fixed effects | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ |
| $R^2$ | 0.5713 | 0.5715 |
| Number of obs. | 127,342 | 127,342 |
| Number of households | 840 | 840 |

(b) Effects of App Elements

| | (3) | (4) | (5) |
|---|---|---|---|
| *AF*: Aggregate feedback | 0.001 | 0.001 | 0.001 |
| | (0.020) | (0.020) | (0.020) |
| $D$: Disaggregation | -0.053*** | -0.059*** | -0.058*** |
| | (0.018) | (0.020) | (0.020) |
| $M$: Monetary incentives | | 0.019 | 0.015 |
| | | (0.014) | (0.017) |
| $R$: Ranking | | -0.005 | -0.009 |
| | | (0.013) | (0.018) |
| $M$: Monet. inc. $\times$ $R$: Rank. | | | 0.009 |
| | | | (0.028) |
| $Y^b$: Baseline elec. use | 0.915*** | 0.915*** | 0.915*** |
| | (0.019) | (0.019) | (0.019) |
| Day fixed effects | ✓ | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ | ✓ |
| $R^2$ | 0.5713 | 0.5715 | 0.5715 |
| Number of obs. | 127,342 | 127,342 | 127,342 |
| Number of households | 840 | 840 | 840 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. These regressions include observations from our matched control group, as described in Section A4. Standard errors are in parantheses and clustered at the household-match level. The outcome variable is daily electricity consumption, divided by the mean in the *EC* group. *AF* equals one for the households in the groups *EC* and $T_1 - T_4$, while being zero for households in *MC*. $D$ equals one for households in the groups $T_1 - T_4$, $M$ equals one for households in the groups $T_2$ and $T_4$, and $R$ equals one for households in the groups $T_3$ and $T_4$, while being zero for the other participants, respectively.

**Table A10:** ATE on Daily Electricity Consumption (After the Core Study Period), Only Main Effects

**(a)** Experimental Sample

| | (1) Full Sample | (2) Balanced Panel |
|---|---|---|
| $D$: Disaggregation | -0.012 | -0.013 |
| | (0.019) | (0.027) |
| $Y^b$: Baseline elec. use | 0.759*** | 0.754*** |
| | (0.022) | (0.031) |
| Day fixed effects | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ |
| $R^2$ | 0.5282 | 0.5414 |
| Number of obs. | 54,603 | 29,330 |
| Number of households | 586 | 321 |

**(b)** Experimental and Matched Control Samples

| | (3) Full Sample | (4) Balanced Panel |
|---|---|---|
| $EC$: Experimental Control | -0.003 | -0.034 |
| | (0.023) | (0.029) |
| $D$: Disaggregation | -0.015 | -0.050* |
| | (0.019) | (0.025) |
| $Y^b$: Baseline elec. use | 0.762*** | 0.758*** |
| | (0.019) | (0.028) |
| Day fixed effects | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ |
| $R^2$ | 0.5257 | 0.5363 |
| Number of obs. | 65,782 | 35,580 |
| Number of households | 704 | 389 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parantheses and clustered at the household level in Columns (1) and (2), and at the household-match level for Columns (3) and (4). The outcome variable is daily electricity consumption, divided by the mean in the *EC* group. $D$ equals one for households in the groups $T_1 - T_4$, while being zero for other participants.

**Table A11:** Difference in ATEs During and After the Core Study Period

| | (1) Experimental Sample | (2) Experimental and Matched Control Samples |
|---|---|---|
| 1(After Core Study Period) | -0.073* | -0.032 |
| | (0.044) | (0.040) |
| $EC$ : Experimental Control | | -0.012 |
| | | (0.057) |
| $EC$ : Exp. Control $\times$ 1(After Core Study Period) | | -0.005 |
| | | (0.027) |
| $D$: Disaggregation | -0.029 | -0.048 |
| | (0.060) | (0.060) |
| $D$: Disaggregation $\times$ 1(After Core Study Period) | 0.039 | 0.039 |
| | (0.025) | (0.025) |
| $R^2$ | 0.2434 | 0.2339 |
| Number of obs. | 84,891 | 102,800 |
| Number of participants | 321 | 389 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parantheses and clustered at the household and at the household-match level for Columns (1) and (2), respectively. The outcome variable is daily electricity consumption, divided by the mean in the $EC$ group. 1(After Core Study Period) is a dummy variable that equals one when an observation occurs from month 7 onwards, while being zero otherwise. $D$ equals one for households in the groups $T_1 - T_4$, while being zero for other participants.

**Table A12:** Difference in ATEs, by Subgroups

| | (1) | (2) |
|---|---|---|
| $D$: Disaggregation | -0.070** | -0.005 |
| | (0.036) | (0.030) |
| 1(Two Occupant) | 0.051 | |
| | (0.038) | |
| 1(Three+ Occupants) | 0.101** | |
| | (0.041) | |
| $D$: Disaggregation $\times$ 1(Two Occupant) | 0.026 | |
| | (0.045) | |
| $D$: Disaggregation $\times$ 1(Three+ Occupants) | 0.003 | |
| | (0.045) | |
| 1(Own Property) | | 0.052 |
| | | (0.036) |
| $D$: Disaggregation $\times$ 1(Own Property) | | -0.040 |
| | | (0.040) |
| $Y^b$: Baseline elec. use | 0.868*** | 0.903*** |
| | (0.023) | (0.024) |
| $R^2$ | 0.5741 | 0.5765 |
| Number of obs. | 106,283 | 95,431 |
| Number of participants | 700 | 622 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parantheses and clustered at the household level. The outcome variable is daily electricity consumption, divided by the mean in the *EC* group. 1(Two Occupants) is a dummy variable that equals one when a household consists of two occupants, while 1(Three+ Occupants) is one for households with at least three occupants. 1(Own Property) equals one if a household lives in his own property. $D$ equals one for households in the groups $T_1 - T_4$, while being zero for other participants.

**Table A13:** ATE on Appliance-Level Consumption, Frequency of Use, and Electricity Use Intensity

**(a)** ATE on appliance-level electricity consumption, in % of control group

| | Always-On | Refrigeration | Dish-Washer | Washing | Dryer | Oven | Other appl. |
|---|---|---|---|---|---|---|---|
| $D$: Disaggregation | 0.031 | -0.003 | -0.082 | -0.036 | -0.423** | -0.086 | -0.080** |
| | (0.058) | (0.054) | (0.108) | (0.081) | (0.183) | (0.181) | (0.034) |
| $C$: Challenges | -0.047 | -0.007 | -0.023 | 0.006 | -0.051 | 0.165 | -0.001 |
| | (0.047) | (0.047) | (0.084) | (0.059) | (0.126) | (0.140) | (0.023) |
| $Y^b$: Baseline elec. use | 1.127*** | 0.434*** | 0.715*** | 0.612*** | 1.014*** | 1.328*** | 0.850*** |
| | (0.062) | (0.103) | (0.098) | (0.075) | (0.165) | (0.198) | (0.034) |
| Day fixed effects | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $R^2$ | 0.353 | 0.123 | 0.039 | 0.025 | 0.028 | 0.035 | 0.346 |
| Number of obs. | 93,187 | 93,185 | 84,511 | 91,473 | 65,852 | 93,187 | 93,187 |
| Number of households | 700 | 700 | 635 | 686 | 499 | 700 | 700 |

**(b)** ATE on appliance-level frequency of use, in % of control group

| | Always-On | Refrigeration | Dish-Washer | Washing | Dryer | Oven | Other appl. |
|---|---|---|---|---|---|---|---|
| $D$: Disaggregation | - | - | -0.024 | -0.007 | -0.396* | -0.005 | - |
| | - | - | (0.096) | (0.059) | (0.203) | (0.156) | - |
| $C$: Challenges | - | - | -0.041 | 0.054 | -0.026 | 0.309** | - |
| | - | - | (0.074) | (0.047) | (0.127) | (0.125) | - |
| $Y^b$: Baseline elec. use | - | - | 0.684*** | 0.425*** | 1.028*** | 0.487*** | - |
| | - | - | (0.090) | (0.059) | (0.164) | (0.125) | - |
| Day fixed effects | - | - | ✓ | ✓ | ✓ | ✓ | - |
| $R^2$ | - | - | 0.039 | 0.024 | 0.049 | 0.024 | - |
| Number of obs. | - | - | 84,511 | 91,473 | 65,852 | 93,187 | - |
| Number of participants | - | - | 635 | 686 | 499 | 700 | - |

**(c)** Appliance-level electricity consumption per use, in % of control group

| | Always-On | Refrigeration | Dish-Washer | Washing | Dryer | Oven | Other appl. |
|---|---|---|---|---|---|---|---|
| $D$: Disaggregation | - | - | -0.076** | -0.034 | -0.015 | -0.103 | - |
| | - | - | (0.033) | (0.047) | (0.140) | (0.096) | - |
| $C$: Challenges | - | - | 0.033 | -0.040 | -0.052 | -0.051 | - |
| | - | - | (0.026) | (0.030) | (0.102) | (0.079) | - |
| $Y^b$: Baseline elec. use | - | - | 0.055** | 0.226*** | 0.004 | 0.707*** | - |
| | - | - | (0.028) | (0.036) | (0.083) | (0.103) | - |
| Day fixed effects | - | - | ✓ | ✓ | ✓ | ✓ | - |
| $R^2$ | - | - | 0.028 | 0.033 | 0.050 | 0.089 | - |
| Number of obs. | - | - | 20,314 | 29,290 | 6,721 | 21,088 | - |
| Number of households | - | - | 599 | 673 | 257 | 525 | - |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors in parantheses, clustered at the household level. The outcome variable is daily electricity consumption (Panel a), frequency of use (Panel b), and energy use per utilization (Panel c) at the appliance level, divided by the respective mean in the *EC* group. We do not present estimates for *Always-On* and *Refrigeration* as consumptions are measured by day, so that there is no extensive margin effect.

**Table A14:** ATE on Daily Electricity Consumption (After the Core Study Period), Relative to Experimental Control

**(a)** Effects of Experimental Conditions

| | (1) | (2) |
|---|---|---|
| $D$: $T_1 - T_4$ | -0.012 | |
| | (0.019) | |
| $T_1$ | | 0.003 |
| | | (0.023) |
| $T_2$ | | 0.017 |
| | | (0.023) |
| $T_3$ | | -0.018 |
| | | (0.024) |
| $T_4$ | | -0.051** |
| | | (0.024) |
| $Y^b$: Baseline elec. use | 0.759*** | 0.762*** |
| | (0.022) | (0.022) |
| Day fixed effects | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ |
| $R^2$ | 0.5282 | 0.5300 |
| Number of obs. | 54,603 | 54,603 |
| Number of households. | 586 | 586 |

**(b)** Effects of App Elements

| | (3) | (4) | (5) |
|---|---|---|---|
| $D$: Disaggregation | -0.012 | 0.015 | 0.003 |
| | (0.019) | (0.022) | (0.023) |
| $M$: Monetary incentives | | -0.009 | 0.014 |
| | | (0.017) | (0.023) |
| $R$: Ranking | | -0.044*** | -0.020 |
| | | (0.016) | (0.024) |
| $M$: Monet. inc. # $R$: Rank. | | | -0.047 |
| | | | (0.034) |
| $Y^b$: Baseline elec. use | 0.759*** | 0.760*** | 0.762*** |
| | (0.022) | (0.022) | (0.022) |
| Day fixed effects | ✓ | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ | ✓ |
| $R^2$ | 0.5282 | 0.5296 | 0.5300 |
| Number of obs. | 54,603 | 54,603 | 54,603 |
| Number of households | 586 | 586 | 586 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parantheses and clustered at the household-match level. The outcome variable is daily electricity consumption, divided by the mean in the $EC$ group. $AF$ equals one for the households in the groups $EC$ and $T_1 - T_4$, while being zero for households in $MC$. $D$ equals one for households in the groups $T_1 - T_4$, $M$ equals one for households in the groups $T_2$ and $T_4$, and $R$ equals one for households in the groups $T_3$ and $T_4$, while being zero for the other participants, respectively.

**Table A15:** ATE on Daily Electricity Consumption (After the Core Study Period), Relative to Matched Control

**(a)** Effects of Experimental Conditions

| | (1) | (2) |
|---|---|---|
| $EC$: Exp. Control | -0.003 | -0.004 |
| | (0.023) | (0.023) |
| $D$: $T_1 - T_4$ | -0.018 | |
| | (0.017) | |
| $T_1$ | | -0.002 |
| | | (0.022) |
| $T_2$ | | 0.004 |
| | | (0.021) |
| $T_3$ | | -0.022 |
| | | (0.022) |
| $T_4$ | | -0.057** |
| | | (0.023) |
| $Y^b$: Baseline elec. use | 0.762*** | 0.764*** |
| | (0.019) | (0.019) |
| Day fixed effects | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ |
| $R^2$ | 0.5257 | 0.5271 |
| Number of obs. | 65,782 | 65,782 |
| Number of households | 704 | 704 |

**(b)** Effects of App Elements

| | (3) | (4) | (5) |
|---|---|---|---|
| $AF$: Aggregate feedback | -0.003 | -0.004 | -0.004 |
| | (0.023) | (0.023) | (0.023) |
| $D$: Disaggregation | -0.015 | 0.013 | 0.002 |
| | (0.019) | (0.022) | (0.024) |
| $M$: Monetary incentives | | -0.015 | 0.006 |
| | | (0.016) | (0.023) |
| $R$: Ranking | | -0.041** | -0.020 |
| | | (0.016) | (0.024) |
| $M$: Monet. inc. $\times$ $R$: Rank. | | | -0.042 |
| | | | (0.034) |
| $Y^b$: Baseline elec. use | 0.762*** | 0.762*** | 0.764*** |
| | (0.019) | (0.019) | (0.019) |
| Day fixed effects | ✓ | ✓ | ✓ |
| Week of baseline FE | ✓ | ✓ | ✓ |
| $R^2$ | 0.5257 | 0.5269 | 0.5271 |
| Number of obs. | 65,782 | 65,782 | 65,782 |
| Number of households | 704 | 704 | 704 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parantheses and clustered at the household-match level. The outcome variable is daily electricity consumption, divided by the mean in the $EC$ group. $AF$ equals one for the households in the groups $EC$ and $T_1 - T_4$, while being zero for households in $MC$. $D$ equals one for households in the groups $T_1 - T_4$, $M$ equals one for households in the groups $T_2$ and $T_4$, and $R$ equals one for households in the groups $T_3$ and $T_4$, while being zero for the other participants, respectively.

**Table A16:** Appliance-Level Elasticity Estimates

| | (1) Always-On | (2) Refrigeration | (3) Dish-Washer | (4) Washing | (5) Dryer | (6) Oven | (7) Other appl. |
|---|---|---|---|---|---|---|---|
| ln p | -0.197 | -0.380 | -0.361 | -0.308 | -4.348*** | -1.180 | -0.265 |
| | (0.363) | (0.235) | (0.775) | (0.449) | (1.355) | (1.130) | (0.265) |
| # of occupants = 2 | 0.082 | 0.135*** | 1.023*** | 0.842*** | -0.169 | 0.614*** | 0.309*** |
| | (0.074) | (0.050) | (0.222) | (0.145) | (0.415) | (0.223) | (0.056) |
| # of occupants = 3 | 0.013 | 0.118** | 1.220*** | 0.953*** | -0.219 | 1.018*** | 0.323*** |
| | (0.087) | (0.057) | (0.231) | (0.156) | (0.454) | (0.242) | (0.059) |
| # of occupants = 4 | 0.108 | 0.112* | 1.294*** | 0.965*** | -0.121 | 0.928*** | 0.372*** |
| | (0.098) | (0.067) | (0.254) | (0.160) | (0.444) | (0.269) | (0.064) |
| # of occupants = 5 | -0.190* | 0.220*** | 1.722*** | 0.660*** | 0.005 | 0.859** | 0.324*** |
| | (0.115) | (0.074) | (0.271) | (0.201) | (0.475) | (0.361) | (0.081) |
| Hh. net income | 0.048*** | -0.009 | -0.028 | 0.027 | -0.018 | -0.034 | 0.003 |
| | (0.014) | (0.010) | (0.034) | (0.021) | (0.060) | (0.040) | (0.009) |
| 1(Hh. net income missing) | 0.178* | -0.060 | -0.186 | 0.172 | -0.258 | 0.078 | 0.086 |
| | (0.099) | (0.072) | (0.234) | (0.143) | (0.411) | (0.275) | (0.063) |
| $Y^b$: Baseline elec. use | 0.107*** | 0.030*** | 0.049*** | 0.055*** | 0.047*** | 0.070*** | 0.078*** |
| | (0.007) | (0.005) | (0.011) | (0.009) | (0.015) | (0.014) | (0.005) |
| Constant | -1.049** | -1.005*** | -3.663*** | -2.944*** | -8.466*** | -4.872*** | 0.064 |
| | (0.521) | (0.340) | (1.119) | (0.654) | (2.055) | (1.660) | (0.369) |
| $R^2$ | 0.503 | 0.181 | 0.176 | 0.303 | 0.074 | 0.152 | 0.648 |
| Number of participants | 560 | 560 | 490 | 539 | 197 | 429 | 560 |

*Notes:* ***, **,* denote statistical significance at the 1%, 5%, 10% level, respectively. The outcome variable is the average daily electricity consumption in the study period (in logs). The regressors are net income, baseline electricity use, as well as a set of dummy variables for the number of occupants and missings for the variable net income. Heteroskedasticity robust standard errors in parantheses.

## References


**Abadie, Alberto, and Jann Spiess.** 2019. "Robust Post-Matching Inference." `https://scholar.harvard.edu/spiess/publications/robust-post-matching-inference`.

**Allcott, Hunt.** 2011a. "Rethinking Real-Time Electricity Pricing." *Resource and Energy Economics* 33 (4): 820–842.

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

**Allcott, Hunt, and Dmitry Taubinsky.** 2015. "Evaluating Behaviorally Motivated Policy: Experimental Evidence from the Lightbulb Market." *American Economic Review* 105 (8): 2501–2538.

**Andor, Mark A., Andreas Gerster, and Stephan Sommer.** 2020. "Consumer Inattention, Heuristic Thinking and the Role of Energy Labels." *The Energy Journal* 41 (1): 83–122.

**Andor, Mark, Andreas Gerster, Jörg Peters, and Christoph M. Schmidt.** 2018. "Social Norms and Energy Conservation Beyond the US." *USAEE Working Paper* No. 18-356.

**Asensio, Omar I., and Magali A. Delmas.** 2015. "Nonprice Incentives and Energy Conservation." *Proceedings of the National Academy of Sciences* 112 (6): 510–515.

**Attari, Shahzeen Z., Michael L. DeKay, Cliff I. Davidson, and Wändi Bruine de Bruin.** 2011. "Reply to Frederick et al.: Anchoring Effects on Energy Perceptions." *Proceedings of the National Academy of Sciences* 108 (8): E24–E24.

**Attari, Shahzeen Z., Michael L. DeKay, Cliff I. Davidson, and de Bruin, Wändi Bruine.** 2010. "Public Perceptions of Energy Consumption and Savings." *Proceedings of the National Academy of Sciences* 107 (37): 16054–16059.

**Becker, Gary S.** 1965. "A Theory of the Allocation of Time." *Economic Journal* 75 (299): 493–517.

**Bollinger, Bryan K., and Wesley R. Hartmann.** 2019. "Information vs. Automation and Implications for Dynamic Pricing." *Management Science* 66 (1): 290–314.

**Bollinger, Bryan, Phillip Leslie, and Alan Sorensen.** 2011. "Calorie Posting in Chain Restaurants." *American Economic Journal: Economic Policy* 3 (1): 91–128.

**Bruelisauer, Marcel, Lorenz Goette, Zhengyi Jiang, Jan Schmitz, and Renate Schubert.** 2018. "Appliance Specific Feedback and Social Comparisons: Evidence From a Field Experiment on Electricity Saving." *SSRN Working Paper*. 10.2139/ssrn.3299500.

**Burkhardt, Jesse, Kenneth Gillingham, and Praveen Kopalle.** 2019. "Experimental Evidence on the Effect of Information and Pricing on Residential Electricity Consumption." *NBER Working Paper* No. 25576.

**Byrne, David P., Andrea La Nauze, and Leslie A. Martin.** 2020. "An Experimental Study of Monthly Electricity Demand (In)Elasticity." *The Energy Journal* (forthcoming).

**Cadmus-Bertram, Lisa A., Bess H. Marcus, Ruth E. Patterson, Barbara A. Parker, and Brittany L. Morey.** 2015. "Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women." *American Journal of Preventive Medicine* 49 (3): 414–418.

**Carroll, James, Seán Lyons, and Eleanor Denny.** 2014. "Reducing Household Electricity Demand Through Smart Metering: The Role of Improved Information about Energy Saving." *Energy Economics* 45 234–243.

**Chetty, Raj, Adam Looney, and Kory Kroft.** 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–1177.

**Chorus, Caspar G., Eric J. E. Molin, and Bert van Wee.** 2006. "Use and Effects of Advanced Traveller Information Services (ATIS): A Review of the Literature." *Transport Reviews* 26 (2): 127–149.

**Degen, Kathrin, Charles Efferson, Fabian Frei, Lorenz Goette, and Rafael Lalive.** 2013. "Smart Metering, Beratung oder Sozialer Vergleich: Was Beeinflusst den Elektrizitätsverbrauch." Final report to the Swiss Federal Office of Energy.

**destatis.** 2017. "Bevölkerung und Erwerbstätigkeit." Fachserie 1 Reihe 3. Federal Statistical Office of Germany, Wiesbaden.

**EC.** 2014a. "Benchmarking Smart Metering Deployment in the EU-27 with a Focus on Electricity." `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2014:356:FIN`, Report by the European Commission.

**EC.** 2014b. "Cost-Benefit Analyses & State of Play of Smart Metering Deployment in the EU-27.." `https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=SWD:2014:189:FIN`, European Commission Staff Working Document.

**EC.** 2014c. "Country Fiches for Electricity Smart Metering." `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=SWD:2014:188:FIN`, European Commission Staff Working Document.

**Espey, James A., and Molly Espey.** 2004. "Turning on the Lights: A Meta-Analysis of Residential Electricity Demand Elasticities." *Journal of Agricultural and Applied Economics* 36 (1): 65–81.

**Farhi, Emmanuel, and Xavier Gabaix.** 2020. "Optimal Taxation with Behavioral Agents." *American Economic Review* 110 (1): 298–336.

**Faruqui, Ahmad, Doug Mitarotonda, Lisa Wood, Adam Cooper, and Judith Schwartz.** 2011. "The Costs and Benefits of Smart Meters for Residential Customers." *IAEE Whitepaper*, `https://www.edisonfoundation.net/iee/Documents/IEE_BenefitsofSmartMeters_Final.pdf`.

**Frederick, Shane W., Andrew B. Meyer, and Daniel Mochon.** 2011. "Characterizing Perceptions of Energy Consumption." *Proceedings of the National Academy of Sciences* 108 (8): E23–E23.

**Frondel, Manuel, Gerhard Kussel, and Stephan Sommer.** 2019. "Heterogeneity in the Price Response of Residential Electricity Demand: A Dynamic Approach for Germany." *Resource and Energy Economics* 57 119–134.

**Gabaix, Xavier.** 2017. "Behavioral Inattention." *National Bureau of Economic Research Working Paper* No. 24096.

**Gerster, Andreas, and Michael Kramm.** 2019. "Correct Me If You Can - Optimal Non-Linear Taxation of Internalities." *CESifo Working Paper* No. 7732.

**Giordano, Vincenzo, Ijeoma Onyeji, Gianluca Fulli, M. Jiménez, and Constantina Filiou.** 2012. "Guidelines for Cost Benefit Analysis of Smart Metering Deployment." 10.2790/39435, European Commission Joint Research Centre.

**Gosnell, Greer, Ralf Martin, Mirabelle Muûls, Quentin Coutellier, Goran Strbac, Mingyang Sun, and Simon Tindermans.** 2019. "Making Smart Meters Smarter the Smart Way." *CEP Discussion Paper* No. 1602.

**Graeber, Thomas, and Benjamin Enke.** 2019. "Cognitive Uncertainty." `http://thomasgraeber.com//assets/papers/Graeber_JMP1_CognitiveUncertainty.pdf`.

**Gupta, Abahy, Vivek Garud, Uri Rodney, and Subodh Bapat.** 2017. "Systems and Methods for Improving the Accuracy of Appliance Level Disaggregation in Non-Intrusive Appliance Load Monitoring Techniques.." US Patent No: US 9,612.286 B2.

**Hart, George William.** 1992. "Nonintrusive Appliance Load Monitoring." *Proceedings of the IEEE* 80 (12): 1870–1891.

**Houde, Sébastien, Annika Todd, Anant Sudarshan, June A. Flora, and K. Carrie Armel.** 2013. "Real-Time Feedback and Electricity Consumption: A Field Experiment Assessing the Potential for Savings and Persistence." *The Energy Journal* 34 (1): 87–102.

**IAWG.** 2016. "Technical Support Document: Technical Update of the Social Cost of Carbon for Regulatory Impact Analysis." `https://www.epa.gov/sites/production/files/2016-12/documents/sc_co2_tsd_august_2016.pdf`, U.S. Interagency Working Group on Social Cost of Greenhouse Gases.

**Icha, Petra, and Gunter Kuhs.** 2019. "Entwicklung der spezifischen Kohlendioxid-Emissionen des deutschen Strommix in den Jahren 1990 – 2018." *Climate Change* 10 4–28.

**IEA.** 2017. "Smart Meter Deployment in Key Markets, 2013-2017." `https://www.iea.org/data-and-statistics/charts/smart-meter-deployment-in-key-markets-2013-2017`, International Energy Agency.

**Ito, Koichiro, Takanori Ida, and Makoto Tanaka.** 2018. "Moral Suasion and Economic Incentives: Field Experimental Evidence from Energy Demand." *American Economic Journal: Economic Policy* 10 (1): 240–267.

**Jensen, Robert.** 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125 (2): 515–548.

**Jessoe, Katrina, and David Rapson.** 2014. "Knowledge is (Less) Power: Experimental Evidence from Residential Energy Use." *American Economic Review* 104 (4): 1417–1438.

**Martin, Steve, and Nicholas Rivers.** 2017. "Information Provision, Market Incentives, and Household Electricity Consumption: Evidence from a Large-Scale Field Deployment." *Journal of the Association of Environmental and Resource Economists* 5 (1): 207–231.

**McKerracher, Colin, and Jacopo Torriti.** 2013. "Energy Consumption Feedback in Perspective: Integrating Australian Data to Meta-Analyses on In-Home Displays." *Energy Efficiency* 6 (2): 387–405.

**Mullainathan, Sendhil, Joshua Schwartzstein, and William J. Congdon.** 2012. "A Reduced-Form Approach to Behavioral Public Finance." *Annual Review of Economics* 4 511–540.

**O'Donoghue, Ted, and Matthew Rabin.** 2006. "Optimal Sin Taxes." *Journal of Public Economics* 90 (10-11): 1825–1849.

**Pop-Eleches, Cristian, Harsha Thirumurthy, James P. Habyarimana, Joshua G. Zivin, Markus P. Goldstein, Damien de Walque, Leslie MacKeen, Jessica Haberer, Sylvester Kimaiyo, John Sidle, Duncan Ngare, and David R. Bangsberg.** 2011. "Mobile Phone Technologies Improve Adherence to Antiretroviral Treatment in a Resource-Limited Setting: a Randomized Controlled Trial of Text Message Reminders." *AIDS* 25 (6): 825–834.

**Rees-Jones, Alex, and Dmitry Taubinsky.** 2019. "Measuring 'Schmeduling'." *NBER Working Paper* No. 22884.

**Schleich, Joachim, Corinne Faure, and Marian Klobasa.** 2017. "Persistence of the Effects of Providing Feedback Alongside Smart Metering Devices on Household Electricity Demand." *Energy Policy* 107 225–233.

**Tiefenbeck, Verena, Lorenz Goette, Kathrin Degen, Vojkan Tasic, Elgar Fleisch, Rafael Lalive, and Thorsten Staake.** 2018. "Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation." *Management Science* 64 (3): 983–1476.

**Tiefenbeck, Verena, Anselma Wörner, Samuel Schöb, Elgar Fleisch, and Thorsten Staake.** 2019. "Real-Time Feedback Promotes Energy Conservation in the Absence of Volunteer Selection Bias and Monetary Incentives." *Nature Energy* 4 (1): 35–41.

**Wichman, Casey J.** 2017. "Information Provision and Consumer Behavior: A Natural Experiment in Billing Frequency." *Journal of Public Economics* 152 13–33.