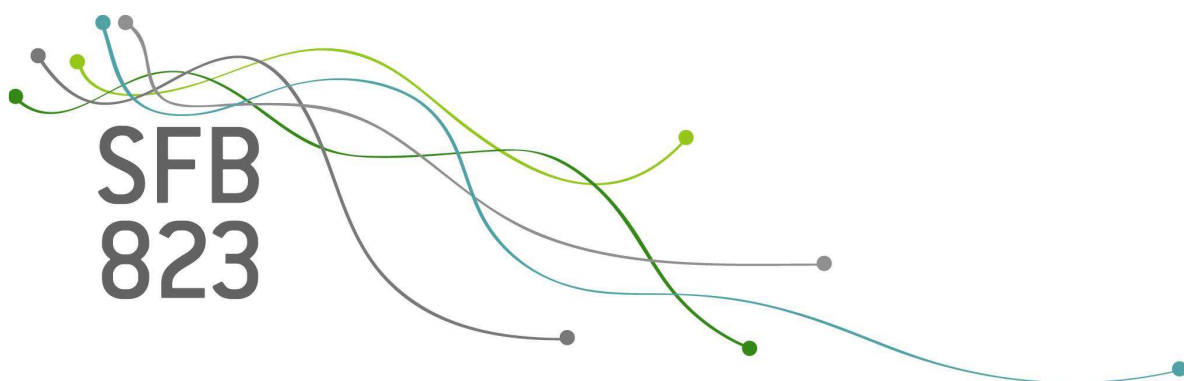


SFB
823

Disaggregate consumption feedback and energy conservation

Andreas Gerster, Mark A. Andor,
Lorenz Goette

Nr. 14/2021



Discussion Paper

Disaggregate Consumption Feedback and Energy Conservation

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

June 25, 2021

Abstract

Novel information technologies have the potential 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. Consumers with poor knowledge of appliance wattage respond most strongly to appliance-level feedback, consistent with the mechanism in our model. We estimate that a smart-meter rollout will yield much larger gains in consumer surplus if appliance-level feedback can be provided.

JEL codes: D12, D83, Q41, Q58.

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). In the context of resource conservation, information technologies have the potential to correct misperception of energy intensities. It is well-documented that individuals tend to underestimate the energy use of energy-intensive appliances, and overestimate the energy use of low-intensity applications (Attari et al., 2010). This 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 provide a new mechanism to foster energy conservation. We also develop a 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 these moderate effect sizes is that aggregate consumption data does not correct misperceptions of energy intensities of individual 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.,

¹For 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 Euro (EC, 2014a).

2018, 2019). Our study builds on a novel smart meter technology that leverages the potential of appliance-level feedback without installing additional costly infrastructure. The technology exploits the distinct signature that every appliance leaves in the electricity consumption data. These signatures 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 700 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 incentives to encourage the conservation of 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). 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 derive how these statistics can be identified as simple functions of the appliance-level treatment effects. To explore beliefs, we conduct surveys and find that households are poorly informed about the wattage of their appliances.² We also show that treatment effects are strongest for households with particularly low levels of knowledge about appliance wattage.

²By contrast, we find that perceptions of aggregate electricity use are very accurate, leaving little room for aggregate-level inattention to increase electricity use in our sample.

Our contribution to the literature is threefold. First, we disentangle the effects of appliance-level information from the effects of aggregate 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).³ 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 active usage. During these hours, we estimate large conservation effects of about 10%.

³Further 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.

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 amplified by monetary incentives and information about savings of other participants. We do not find that any of these strategies increases the response to 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 study, 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 12 – 15 Euro per household and annum, which scales to an increase in total consumer surplus of about 500 – 620 million Euro 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 and potentially extend into other applications such as improving the effectiveness of physical exercise and the nutritional quality of diets. 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 digitization 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 $\mathbf{x} = (x_1, \dots, x_J)'$ 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 α . For simplicity, we approximate consumers' utility by an additively separable utility function, which yields:

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

where u_{jj} denotes the second derivative of $u(\mathbf{x})$ with respect to x_j , while Δy_j and Δb_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 using 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 $x_1 e_1 + x_2 e_2 = (x_1 + x_2)e$. 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

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:

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

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:

$$\Delta \text{ConsSurplus} = \sum_j \frac{b_j}{e_j} \Delta E_j, \quad (3)$$

where $\Delta \text{ConsSurplus}$ 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 (ΔE_j). The intuition for Equation (3) is that a be-

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 (ΔE_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 η_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 from our data by 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 men-

tion 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 who own solar photovoltaic 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 shown by Table A1 in the Appendix, sociodemographic characteristics of study participants are comparable to 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 professionals, 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, when 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. We also conduct complementary analyses to estimate treatment effects in the subsequent months.

⁴When 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 A15 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. This results in a rich dataset of several billion observations over the entire study period. The high granularity of our data allows us to use load disaggregation techniques that disentangle the total electricity use into appliance-level uses. It sets our study apart from earlier attempts to analyze and disaggregate smart meter data, such as Google PowerMeter or Microsoft Hohm, which were discontinued in 2011 and 2012, respectively. A main reason for the failure of these services was insufficient access high-granular smart meter data (Donnal and Leeb, 2015).

The smart meter data is saved online in real-time and processed daily 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 allowing the laundry cool down and heating it up again. This produces a characteristic signature in the high-resolution smart meter data that the algorithm uses to detect appliance use events, i.e. the start/end date and the amount of electricity used by individual appliances (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 A5 in the Appendix gives descriptive statistics on the more than 300.000 appliance events that we observe during the core 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: $\text{Appliance Score} = 100 \times (\text{Monthly Appliance Consumption} - \text{Bench}_{\text{low}}) / (\text{Bench}_{\text{high}} - \text{Bench}_{\text{low}})$, where $\text{Bench}_{\text{low}}$ and $\text{Bench}_{\text{high}}$ correspond to pre-determined benchmark values for high and low appliance uses, respectively. We construct these benchmarks from survey data on typical appliance uses as well as product data sheets on the technical efficiency of appliances currently used in German households (for details, see Table A3 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. 2020b). 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 vacuum cleaners. By contrast, air conditioning is not prevalent in German households and thus cannot explain that increase.

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

Table 1: Overview of Experimental Conditions

	EC	T ₁	T ₂	T ₃	T ₄
AF: Aggregate Feedback	✓	✓	✓	✓	✓
D: Disaggregation	-	✓	✓	✓	✓
C: Challenges	-	-	✓	✓	✓
<i>Type of Challenge:</i>					
M: Monetary Incentives	-	-	✓	-	✓
R: Ranking	-	-	-	✓	✓

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, information on baseline electricity use is missing for 27 participants.⁶ Furthermore, 73 households experienced technical difficulties that prevented 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 assigned 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 frequencies.

Participants in our first treatment group T₁ have access to an additional app page that provides feedback on appliance-level usages, cost and appliance scores. In the three treatment

⁶For 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.

⁷In 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.

groups T₂-T₄, participants receive the same appliance-level feedback and are additionally invited to take part in appliance challenges. With these challenges, we test whether complementary interventions increase the effectiveness of appliance-level feedback (for research on multiple interventions, see e.g. Brandon et al., 2019). The challenges start at the beginning of the second month after installation of the app and require participants to increase one of the appliance scores 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 T₂, participants obtain 1 EUR per appliance score improvement, capped at a maximum of 20 EUR per monthly challenge. This treatment is motivated by studies that have found stronger conservation effects when monetary incentives are provided (e.g., Ito et al. 2018). In group T₃, participants receive a ranking that compares their score improvement with those of other study participants, but do not obtain a financial reward. A participant within the top percentile of monthly score improvements is classified as rank one, and similarly for all other percentiles. With this treatment, we test the impact of relative performance feedback, which has been found to be effective in educational (Azmat and Iriberry, 2010; Tran and Zeckhauser, 2012) and workplace settings (Mas and Moretti, 2009; Blanes i Vidal and Nossol, 2011). In group T₄, 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. This treatment allows us to explore the effectiveness of combining relative performance feedback with monetary incentives.

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 the same appliances more than once, given that all other appliances have been targeted already. At the end of our study, participants in T₂ and T₄ 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 A7 of the Appendix, the average realized payments per monthly efficiency challenge amount to 6.3 EUR in T₂ and to 4.5 EUR in T₄.⁸ The timing in our experiment is as follows: In the first month after app installation, participants in T₁-T₄ obtain appliance-level feedback but chal-

⁸The average payment in group T₄ does not exactly equal 5 EUR as we have less than 100 challenge participants in that group for some months.

Table 2: Descriptives

	EC	T ₁	T ₂	T ₃	T ₄	P-value
<i>a) Sociodemographics (n=700)</i>						
Baseline consumption, in kWh/day	10.2	10.5	10.1	10.5	10.5	0.90
No. of occupants	2.4	2.5	2.4	2.5	2.6	0.17
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
<i>b) Baseline Aggregate Consumption Beliefs (n=466)</i>						
Yearly consumption (belief), in kWh	3,643	3,650	3,616	3,496	3,727	0.81
Yearly consumption (actual), in kWh	3,524	3,551	3,497	3,467	3,589	0.97
<i>c) Wattage Ranking Task (n=598)</i>						
At least one mistake (0=no, 1=yes)	0.50	0.62	0.60	0.58	0.56	0.48
Share of mistakes in all comparisons	0.11	0.12	0.11	0.11	0.12	0.90
Number of households	140	136	143	143	138	$\Sigma=700$

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. The wattage ranking task refers to a task in the baseline survey that asked respondents to assess the wattage of a typical laptop, dish-washer, tumble dryer, and fanheater, relative to a 100 W lightbulb on a five point Likert scale (“much lower”, “lower”, “about the same”, “higher”, “much higher”). “Share of mistakes in all comparisons” gives the total number of mistakes, divided by the number of all binary comparisons implied by that task. Aggregate consumption beliefs are elicited in the same survey. We drop observations beyond the 95 percentile of the belief distribution as well as those with extreme relative errors (below the 2.5 and above the 97.5 percentile).

allenges have not yet started. In the months 2-6, participants in T₂-T₄ take part in challenges. After month 6, challenges do not occur any more, but T₁-T₄ participants continue to receive appliance-level feedback.

3.3 Data and Balancing

In Table 2, we show that sociodemographic variables are balanced across our experimental groups, as expected from randomization. When we conduct F-tests of mean equality across our five groups, we cannot reject the null hypothesis that the variable means are equal at the 5% level for any covariate, as shown in the last column of Table 2. The same holds true for various dwelling characteristics and the possession of household appliances (for details, see Appendix Table A2).

The survey that we conducted prior to the field test allows us to assess the extent to which households misjudge their aggregate and appliance-level electricity use. We elicited house-

holds' annual electricity consumption beliefs and compare them to their annual consumption values from the latest annual bill prior to the field test. This comparison does not provide any evidence that households underestimate their aggregate electricity consumption. On average, annual consumption beliefs are virtually on par with baseline consumption levels (Panel b of Table 2). In addition, we find that beliefs closely match the actual consumption values. A one-unit increase in baseline electricity use is associated with an increase in baseline beliefs by 0.93 units (Std. Err.: 0.026; see Table A6 and Figure A11 in the Appendix). As shown in Byrne et al. (2020), this estimate is a measure of the salience of aggregate consumption. Hence, our finding that it is only slightly below one indicates that lack of salience is unlikely to cause excessive resource use in our context. Taken together, this evidence suggests that correcting aggregate beliefs and increasing the salience of household-level electricity consumption through aggregate feedback may not lead to substantial behavioral change.

To assess appliance-level beliefs, our baseline survey contained a wattage ranking task that asked households to assess the wattage of a laptop, dish-washer, tumble dryer, and fanheater relative to a 100 watt lightbulb on a five point Likert scale ("much lower", "lower", "about the same", "higher", "much higher").⁹ We find that more than half of all respondents make at least one mistake in assessing appliance-level wattage. Furthermore, respondents make mistakes in about 11% of all the binary appliance-level wattage comparisons that are implicitly incorporated in the task. Both findings show that households' knowledge of appliance-level energy intensities is limited, as formalized by our model. In the following, we explore the implications of providing households with appliance-level feedback and introduce our empirical approach.

4 Empirical Strategy and Results

4.1 Treatment Effects on Total Electricity Use

To identify the average treatment effect (ATE) of our four treatment groups T_1 - T_4 , we estimate the following equation via ordinary least squares (OLS):

⁹Inspired by Attari et al. (2010), the wattage ranking task also included three further appliances: desktop computer, hifi system, and air conditioning. Depending on the configuration of these devices, their wattage varies to an extent that prevents us from establishing a clear wattage ranking. We thus exclude these devices from our analysis.

$$Y_{it}^{norm} = \alpha Y_i^b + \boldsymbol{\beta}' \mathbf{T}_i + \nu_t + \mu_m^b + \epsilon_{it}, \quad (4)$$

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 during the core study period (10.5 kWh). This normalization expresses treatment effects as a percentage of the average consumption level in the absence of treatment and is common in the literature (e.g., Allcott 2015).¹⁰ 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 (ν_t) and an error term ϵ_{it} . In order to increase the precision of our estimates, we also include a set of month-of-baseline fixed effects (μ_m^b) that equal one if the baseline metering period includes a given calendar month and zero otherwise. These fixed effects absorb variation in baseline electricity consumption that originates from seasonality and any month-specific shocks during the baseline period. We cluster standard errors at the household level, thereby accounting for serial correlation in the error terms. In Appendix A10, we demonstrate that our results remain unchanged when we estimate Equation 4 using hourly data and hour-by-day fixed effects.

As shown in Panel a) of Table 3, the ATE estimate for participants in the four treatment groups T_1 – T_4 amounts to -4.8% and is statistically significant at all conventional levels. When we estimate the ATEs separately for each treatment groups (Column 2), we do not find sizable differences in their magnitude, with point estimates ranging from -3.9% in T_2 to -6.6% in T_3 .¹¹ Our ATE estimates increase slightly when we consider only a balanced panel, i.e., when we include only participants with complete data transmission during the entire time span of the core study period (Table A11 of the Appendix).

¹⁰The treatment effects in kWh can be obtained by simply multiplying our estimates with the experimental control group mean (10.5 kWh).

¹¹When we conduct pairwise comparisons between the ATEs in the treatment groups $T_1 - T_4$, we cannot reject the null hypothesis that they are the same.

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

(a) Effect of Experimental Conditions			(b) Effects of App Elements			
	(1)	(2)		(3)	(4)	(5)
$D: T_1-T_4$	-0.048*** (0.016)		$D: \text{Disaggregation}$	-0.048*** (0.016)	-0.047*** (0.018)	-0.043** (0.019)
T_1		-0.044** (0.019)	$M: \text{Monetary incentives}$		0.014 (0.013)	0.003 (0.016)
T_2		-0.039** (0.019)	$R: \text{Ranking}$		-0.015 (0.013)	-0.026 (0.018)
T_3		-0.066*** (0.020)	$M: \text{Monet. inc.} \times R: \text{Rank.}$			0.023 (0.026)
T_4		-0.043** (0.021)				
$Y^b: \text{Baseline elec. use}$	0.895*** (0.022)	0.895*** (0.022)	$Y^b: \text{Baseline elec. use}$	0.895*** (0.022)	0.895*** (0.022)	0.895*** (0.022)
Day fixed effects (FE)	✓	✓	Day fixed effects (FE)	✓	✓	✓
Month-of-baseline FE	✓	✓	Month-of-baseline FE	✓	✓	✓
R^2	0.5586	0.5589	R^2	0.5720	0.5722	0.5722
Number of obs.	106,283	106,283	Number of obs.	106,283	106,283	106,283
Number of households	700	700	Number of households	700	700	700

Notes: ***, **, * denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parentheses and clustered at the household-match level. The outcome variable is daily electricity consumption, divided by the mean in the EC group (10.5 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:

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

where the scalar D_i equals one if a household is in any 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 β_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. Households that obtain such feedback reduce their consumption by almost 5 percent compared to households with aggregate feedback (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. Thus, 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).¹²

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 use this sample to construct a matched control group of households that did not receive any feedback. Our results are consistent with previous studies that have found only small savings from aggregate feedback (e.g., Degen et al. 2013). As shown in Appendix Table A12, our ATE estimate for households who obtain only aggregate feedback amounts to -0.9% (Std. Err.: 2.2%).

Furthermore, we use the additional observations after the core study period to estimate how treatment effects evolve in that time period. We estimate conservation effects between

¹²Given the standard error of 1.3% for the coefficient on monetary incentives in Table 3, we have sufficient power to detect interaction effects of 3.6%. Thus, we are sufficiently powered to reject negative interaction effects of monetary framings such as those found in Asensio and Delmas 2015, which are around 10%. Our study is also sufficiently powered to identify the overall disaggregation effect, where the minimal detectable effect is 4.4%.

−1.4% (Std. Err.: 1.7%) and −4.7% (Std. Err.: 2.2%), depending on the specification and estimation sample we use (for details, see Appendix Section A6 and Table A13). In particular, we cannot reject the null hypothesis that the average treatment effects during and after the core study period are equal at any conventional significance level (Difference: 0.038 percentage points, Std. Err.: 0.023 percentage points, see Appendix Section A6 and Table A14). Taken together, these findings indicate that the treatment effect from appliance-level feedback persists after the study period. Yet, we lack the statistical power to precisely pin down the evolution of treatment effects over time.

4.2 Treatment Effects by Hour-of-the-Day and Appliance Category

In this subsection, we investigate two sources of treatment effect heterogeneity. First, we exploit the high granularity of our data to investigate how treatment effects vary by hour-of-the-day. This analysis allows us to explore the timing of households’ responses to appliance-level feedback. We estimate the following equation:

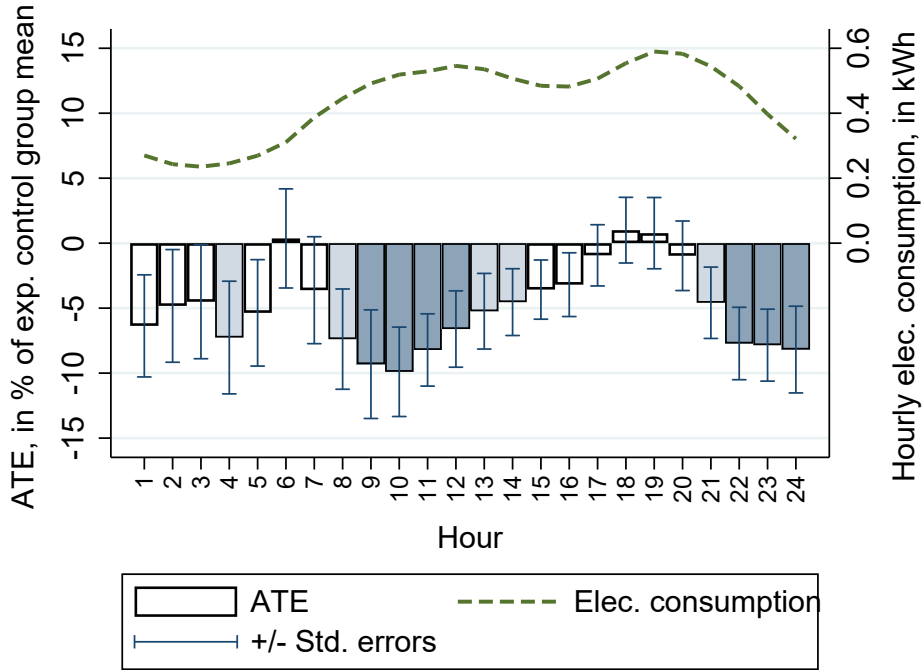
$$Y_{ith}^{norm} = \alpha Y_i^b + \sum_{h=1}^{24} \beta^h D_i + \nu_t + \mu_h + \mu_m^b + \epsilon_{ith}, \quad (6)$$

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 μ_h for every hour of the day. We normalize our outcome variable 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 hour. Again, we cluster standard errors at the household level.

Figure 1 shows that treatment effects are large during late morning hours and late evening hours. During these hours, they reach about 10% of the average control group consumption.¹³ Furthermore, we find 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 a.m. and 1 p.m., which coincide with large electricity consumption levels. However, during 4 and 8 p.m., consumption levels are similar, but households save considerably less. We also detect strong savings in late evening hours between 9 and 11 p.m., when consumption levels are rather low.

¹³The time pattern is slightly more pronounced when we consider absolute rather than relative electricity savings (Figure A9 of the Appendix).

Figure 1: 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 EC group mean in the respective hour. 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: $F(24, 699) = 1.88$, $p\text{-value} = 0.007$.

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 :

$$Y_{itj}^{norm} = \alpha_j Y_i^b + \beta_j D_i + \gamma_j C_i + v_t + \mu_m^b + \epsilon_{itj}, \quad (7)$$

which closely resembles 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

Table 4: ATE on Daily Appliance-Level Electricity Consumption

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Always-On	Refrigeration	Dish–Washer	Washing	Dryer	Oven	Other appl.
<i>D</i> : Disaggregation	0.031 (0.055)	0.001 (0.052)	−0.071 (0.100)	−0.026 (0.076)	−0.434** (0.173)	−0.095 (0.169)	−0.075** (0.031)
<i>C</i> : Challenges	−0.049 (0.045)	−0.012 (0.044)	−0.029 (0.078)	−0.003 (0.056)	−0.006 (0.117)	0.178 (0.130)	0.005 (0.022)
Υ^b : Baseline elec. use	1.156*** (0.065)	0.460*** (0.086)	0.746*** (0.103)	0.637*** (0.072)	1.064*** (0.166)	1.355*** (0.197)	0.851*** (0.033)
Day fixed effects (FE)	✓	✓	✓	✓	✓	✓	✓
Month-of-baseline FE	✓	✓	✓	✓	✓	✓	✓
R^2	0.368	0.152	0.046	0.028	0.035	0.040	0.356
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, 0.97, 0.29, 0.47, 0.16, 0.20, and 5.6 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).

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 1. In particular, we find that appliance-level feedback triggers a substantial reduction in the electricity consumption of dryers (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, yet these effects are not statistically significant at any conventional level. In Figure A10 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 1 can be attributed to these appliances.

In addition, we find that households reduce consumption for the category *Other Appliances*, an effect that is statistically significant at the 5% level. This category encompasses a variety of electric appliances, such as televisions, hi-fi systems, vacuum-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 entertainment purposes can explain the second peak in conservation effects in the late evening hours (Panel d of Figure A10 in the Appendix).

Next, we disentangle changes in the frequency of use from changes in utilization (Table A16 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, 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 and indicates that additional financial incentives or relative performance feedback did neither increase nor decrease the effectiveness of appliance-level information.

4.3 Appliance-Level Beliefs

To investigate the mechanisms that underlie households' responses to appliance-level feedback, we first explore whether treatment effects are particularly large for households who make mistakes in the wattage ranking task. This test is closely linked to our model, which predicts that households who misjudge the relative wattage of appliances should respond most strongly to the provision of appliance-level information. We define two groups of households: those who made at least one mistake in the wattage ranking task and those who did not. We then estimate our main specification (Equation 4) and interact the dummy variable D with both group indicators, which yields the the average treatment effect of appliance-level information in these groups. As shown in Column (1) of Table 5a, we find that the effect size amounts to -4.3% for households who make at least one mistake and is statistically significant at the 5% level. For households with no errors in the wattage ranking task, the effect size amounts to

Table 5: Effect Heterogeneity and Appliance-Level Beliefs**(a) Effect Heterogeneity by Wattage Ranking Task Results**

	(1)		(2)		(3)	
	Estimate	Std. err.	Estimate	Std. err.	Estimate	Std. err.
1(At least one mistake)	0.028	(0.021)	0.031	(0.022)		
1(No wattage ranking)			0.169**	(0.068)		
$D \times 1$ (No mistake)	-0.014	(0.019)	-0.010	(0.020)		
$D \times 1$ (At least one mistake)	-0.043**	(0.018)	-0.043**	(0.018)		
$D \times 1$ (No wattage ranking)			-0.155**	(0.069)		
D : Disaggregation					-0.012	(0.017)
Share of mistakes					0.125*	(0.067)
$D \times$ Share of mistakes					-0.156*	(0.085)
Y^b : Baseline elec. use	0.888***	(0.024)	0.893***	(0.022)	0.889***	(0.024)
Day fixed effects (FE)	✓	✓	✓	✓	✓	✓
Month-of-baseline FE	✓	✓	✓	✓	✓	✓
R^2	0.5613		0.5613		0.5615	
Number of obs.	91,638		106,283		91,638	
Number of participants	598		700		598	

(b) Appliance-Level Rank Differences between Beliefs and Measured Uses

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean rank diff. (B)	Mean rank diff. (E)	Change (E-B)	Std. err.	P-value	N
EC	1.81	1.79	-0.02	0.07	0.757	66
$T_1 - T_4$	1.76	1.56	-0.19	0.04	0.000	296
T_1	1.75	1.57	-0.18	0.10	0.077	61
T_2	1.75	1.42	-0.32	0.08	0.000	62
T_3	1.69	1.47	-0.22	0.10	0.031	54
T_4	1.79	1.54	-0.26	0.09	0.009	53

Notes: ***, **, * denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parentheses and clustered at the household level. “1(At least one mistake)” and “1(No wattage ranking)” denote dummy variables that equal one if a household made at least one mistake in the wattage ranking task or did not take part in that task, respectively. “Mean rank diff.” denotes the mean of the absolute rank differences between the rank of appliance consumptions (measured during the study period) and the rank of electricity consumption beliefs, which we elicited for every appliance category before (baseline, B) and during the study period (endline, E).

only -1.4% and is not statistically significant at any conventional level. In Column (2), we also consider a third group of households who did not indicate appliance-level beliefs in the baseline survey. The effect size that we estimate for this group of households is even larger than for those households who indicate beliefs and make at least one mistake. As not expressing beliefs may indicate particularly poor knowledge of appliance wattages, this finding is consistent with our previous results.

In addition, we explore the intensive margin of making mistakes by interacting the dummy variable D with the share of mistakes in all binary comparisons implied by the wattage ranking task (Column 3). The estimate of the main effect (D) shows that households who make no

mistakes save only about 1.2% in response to appliance-level feedback. By contrast, we find that an increase in the share of mistakes from 0 to 1 increases conservation effects by 15.6 percentage points (p-value: 0.069), which supports the idea that households with poor knowledge of appliance wattage have particularly large treatment effects. Furthermore, the regression results from Column 3 demonstrate that households who make mistakes in the wattage ranking task tend to have higher consumption levels overall. Taken together, these findings suggest that lack of energy-related knowledge among households with high consumption levels could explain why many previous feedback studies have detected particularly large treatment effects for these households (e.g., Allcott, 2011b; Andor et al., 2020a; Tiefenbeck et al., 2018).¹⁴

As a complementary test for the role of beliefs, we explore whether participants' appliance-level beliefs align more closely with the appliance-level uses that we measure during the experiment. We elicited appliance-level beliefs in a baseline survey that took place prior to the study period (November 2016) and an endline survey towards the end of the study period (July 2017). In particular, we asked participants to estimate their monthly electricity consumption for the appliance categories always-on, washing machine, dryer, refrigeration, and dish-washer. A limitation of our data is that we observe appliance-level uses only once our field experiment started. Hence, we cannot assess the accuracy of households' appliance-level beliefs before obtaining appliance-level feedback. Yet, we can test whether beliefs during the field tests align more closely with measured appliance-level uses for households in the treatment groups T_1 - T_4 than for households in the experimental control group.

To circumvent difficulties that may arise from noisy appliance-level belief measurements, we first translate all consumption beliefs into ranks.¹⁵ 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

¹⁴In our study, we also find evidence for such treatment effect heterogeneity (see Appendix Section A5 for details).

¹⁵The elicitation of appliance-level 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). For these reasons, we chose not to provide reference points.

ranks implied by participants' belief and those based on the appliance-level data during the study period. Averaging over these absolute differences across all appliances yields a measure of 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 5b 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 1.8 for all experimental groups. It reflects poor knowledge of energy intensities as it is only slightly lower than the expected rank difference of randomly determined ranks, which amounts to $70/36 \approx 1.94$. 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 with 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.

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 a quantitative measure of the relative bias b_j/e_j from our data. In particular, we exploit that $(b_j/e_j) = -(\Delta y_j/y_j)/\eta_j$, where η_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).

To approximate the price elasticities of appliance-level energy service demand η_j , we follow two strategies. First, we employ cross-sectional variation in our dataset that stems from the fact

that similar households in terms of observable characteristics pay different electricity prices, in particular owing to transmission charges that vary strongly by region. 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).¹⁶

To identify η_j from our data, we estimate the following regression separately for every appliance category 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 for identification 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 does not threaten the consistency of our estimates. Another concern is omitted variable bias. Electricity suppliers offer tariffs with lower marginal prices to households with higher consumption levels, 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, 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 -3.42 and -1.07 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 -0.29 to -0.55 . Our appliance-level elasticity estimates imply a household-level elasticity of -0.39 , which is close

¹⁶The 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 η_j			Uniform η_j	
	$\Delta y_j/y_j$	Avg. use in kWh/a	ΔE_j in EUR/a	η_j	b_j/e_j	$\Delta\text{ConsSurplus}$ in EUR/a	η_j	b_j/e_j	$\Delta\text{ConsSurplus}$ in EUR/a
Always-On	0.031	885.35	6.59	-0.31	0.10	0.66	-0.44	0.07	0.47
Refrigeration	0.001	372.88	0.06	-0.40	0.00	0.00	-0.44	0.00	0.00
Dish-Washer	-0.071	117.69	-2.00	-0.40	-0.18	0.36	-0.44	-0.16	0.32
Washing	-0.026	184.07	-1.13	-0.55	-0.05	0.05	-0.44	-0.06	0.07
Dryer	-0.434	69.10	-7.16	-3.42	-0.13	0.91	-0.44	-0.99	7.07
Oven	-0.095	85.47	-1.93	-1.07	-0.09	0.17	-0.44	-0.22	0.41
Other appl.	-0.075	2,145.96	-38.55	-0.29	-0.26	10.05	-0.44	-0.17	6.60
Total		3,860.51	-44.12	-0.39		12.20			14.94

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, ΔE_j , is calculated as the product of $\Delta y_j/y_j$ (Column 1), the average annual electricity use in the experimental control group (Column 2), and the average electricity price in our sample (0.238 EUR per kWh). η_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). Uniform η_j are drawn from Frondel et al. (2019).

to the estimate of -0.44 , taken from Frondel et al. (2019).¹⁷ This finding reduces concerns that our cross-sectional identification strategy yields strongly biased estimates.

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* and *Refrigeration*. Negative biases are most pronounced for the categories *Dish-Washer* and *Other Appliances*, where consumers underestimate energy intensities by 18% and 26%, respectively. For the categories *Washing*, *Dryer*, and *Oven*, we obtain less pronounced biases of -5% to -13% . 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 relatively 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),

¹⁷The elasticity of total consumption can be calculated as follows: $\eta = \prod_j \eta_j(y_j/y)$, where η_j and y_j denote the elasticity and the consumption level for appliance j , respectively, and $y = \sum_j y_j$ denotes total consumption.

and the average consumption level, y_j (Column 2). We then multiply the change in expenditures with the relative 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 12.2 EUR per annum (last row of Column 6), which is substantially less than the 44.1 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 rise when more accurate beliefs lead to higher 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 conclude that consumer surplus increases by 14.9 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).

For cost-benefit analyses, it is crucial to take into account that changes in consumer surplus are equal to the *weighted* sum of appliance-specific expenditures savings. The official cost-benefit analysis for Germany assumes that smart meter feedback achieves 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 28% (12.2 EUR / 44.1 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 2.9 EUR for a conservation effects of 1.2%. This back-of-the-envelope calculation illustrates that traditional cost-benefit analyses overestimate consumer benefits by a factor of about three. However,

our study also demonstrates that the provision of appliance-level feedback yields larger effect sizes. Using the effect sizes found in our study, we find that consumer surplus increases by about 12 EUR, which demonstrates 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 700 participants, we randomize information provision on a smartphone 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, i.e., feedback as provided by conventional smart meters. 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 a.m. – 1 p.m. and in the late evening hours between 9 – 11 p.m., when it reaches about 10%. 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.

Based on a conceptual model, we investigate the consequences of providing information to consumers who underestimate the energy use of energy-intensive appliances as documented in 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. We also use our model to 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 used by policy makers 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. Implicitly, this approach weights appliance-level savings by the relative consumption shares rather than the measure of consumers' bias. As the correct weights typically add up to less than one, current cost-benefit analyses tend to overestimate 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 12.2 – 14.9 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 500 – 620 million Euro. These estimates illustrate that appliance-level feedback is crucial for raising consumer surplus. By contrast, the additional cost of providing appliance-level feedback based on load decomposition, as in our study, are only small when smart meter infrastructure is available. Hence, 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 3.7 megatons per annum. Using the range of estimates for the social cost of carbon of 11 – 105 USD (10 – 93 EUR) per ton of CO₂ equivalents (IAWG, 2016), this reduction in emissions translates into a decrease in total external costs from carbon emissions by 42 – 401 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.¹⁸

Future research should investigate the external validity of our results along two dimensions. First, evaluating a roll-out of appliance-level feedback would allow to assess the scalability of this intervention. While we do not find large differences in observable characteristics between our study participants and the average German population, differences in unobservable characteristics or in program implementation might change the effectiveness of the intervention at scale (e.g., Al-Ubaydli et al. 2017b,a). Second, evidence on the effectiveness of appliance-level feedback in other countries would be valuable. The average electricity consumption of U.S. households, for example, exceeds the consumption of German households by more than a factor of three, which may imply larger conservation potentials (Andor et al.,

¹⁸A 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.

2020a). In particular, air conditioning use is widespread in the U.S., but not in Germany. It is easily captured by NALM algorithms and could be part of disaggregated feedback. Previous studies, measuring air conditioning use directly, have found that air conditioning use reduces in response to application-specific feedback (Bruelisauer et al., 2018). Hence, appliance-level feedback may in fact be more effective in countries such as the U.S.

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 digitization 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 provides a framework for policy makers to evaluate their benefits.

Appendix (For Online Publication)

A1 Effect of Appliance-Level Feedback on Energy Consumption	i
A2 Sufficient Statistics for Evaluating Consumer Surplus	iv
A3 Nonintrusive Appliance Load Monitoring	vii
A4 Non-Experimental Control Group	ix
A5 Heterogeneity in Treatment Effects by Baseline Consumption	xi
A6 Conservation Effects after Core Study Period	xii
A7 Appliance-Level Treatment Effects by Hour of the Day	xiii

Supplementary Figures

A1 Schematic Representation of the NALM Algorithm	viii
A2 Visualization of Screens I	xiv
A3 Visualization of Screens II	xv
A4 Appliance Signatures	xvi
A5 Distribution of Appliance Scores	xvii
A6 Decomposition of Electricity Uses, by Appliance Category	xviii
A7 Balancing between Study Participants and the Non-Experimental Sample	xix
A8 ATE of Experimental Control (EC) on Hourly Electricity Consumption (relative to Matched Control)	xx
A10 ATE of Disaggregation (D) on Hourly Electricity Consumption, by Appliance Category (Relative to Exp. Control)	xxii
A11 ATE of Experimental Control (EC) on Hourly Electricity Consumption (relative to Matched Control)	xxiii

Supplementary Tables

A1 Comparison of Experimental Sample with German Population	xxiv
A2 Descriptives: Appliance Possessions and House Characteristics	xxv
A3 Benchmarks for the Calculation of Efficiency Scores	xxvi
A4 Distribution of Daily Use, in kWh/day	xxvii
A5 Distribution of Use per Utilization, in kWh	xxvii
A7 Distribution of Payoffs from Appliance Challenges, in EUR	xxix
A8 Balance Experimental Sample vs. Matched Non-Experimental Sample	xxix
A9 Heterogeneity in ATE, by Baseline Consumption	xxx
A11 ATEs on Daily Electricity Consumption, Relative to Exp. Control (Balanced Panel)	xxxii
A12 ATEs on Daily Electricity Consumption, Relative to Matched Control	xxxiii

A13	ATE on Daily Electricity Consumption (After the Core Study Period), Only Main Effects	xxxiv
A14	Difference in ATEs During and After the Core Study Period	xxxv
A15	Difference in ATEs, by Subgroups	xxxvi
A16	ATE on Appliance-Level Consumption, Frequency of Use, and Electricity Use Intensity	xxxvii
A17	Appliance-Level Elasticity Estimates	xxxviii

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:

$$U(x_1, x_2, z) = u_1(x_1) + u_2(x_2) + z, \quad (8)$$

where $u_1(x_1)$ and $u_2(x_2)$ 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 $x_1 e_1 + x_2 e_2 = (x_1 + x_2)e$. 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:

$$\frac{b_1}{b_2} = -\frac{x_2}{x_1}. \quad (9)$$

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

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

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:

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

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

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

where $\eta_j(\tilde{e}_j)$ denotes the price elasticity of demand for energy service j under the naive belief e , i.e., $\eta_j(\tilde{e}_j) = (\partial x_j(e)/\partial p)/(x_j(e)/p)$. Using that:

$$\frac{\tilde{e}_2}{\tilde{e}_1} = \frac{\alpha e_2 + (1 - \alpha)e}{\alpha e_1 + (1 - \alpha)e'}$$

and that $e_1 > e_2$ by definition, we immediately obtain that:

$$\frac{e_1 \tilde{e}_2}{e_2 \tilde{e}_1} > 0.$$

One consequence of this finding is that 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., $\eta_1(\tilde{e}_1)/\eta_2(\tilde{e}_2) \geq 1$. Furthermore, 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:

$$\lim_{e_2 \rightarrow 0} \frac{e_1 \tilde{e}_2 \eta_1(\tilde{e}_1)}{e_2 \tilde{e}_1 \eta_2(\tilde{e}_2)} = \lim_{e_2 \rightarrow 0} \frac{\alpha + (1 - \alpha)(e/e_2) \eta_1(\tilde{e}_1)}{\alpha + (1 - \alpha)(e/e_1) \eta_2(\tilde{e}_2)} = \infty > 1.$$

As e_1 , η_1 , and η_2 are finite, and α weakly smaller than one, there always exists a sufficiently small e_2 so that Inequality (11) holds.

Extension to n appliances

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

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

From Equation 10, we obtain that:

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

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:

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

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

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

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

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

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 $k \in \{J+1, \dots, N\}$. Second, it holds when the demand response to a price change, $|\partial x_k(\tilde{e}_k) / \partial p|$, is sufficiently small for all $k \in \{J+1, \dots, N\}$.

A2 Sufficient Statistics for Evaluating Consumer Surplus

Let consumers have the following quasi-linear utility function:

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

where $u(\mathbf{x})$ denotes the utility from consuming J energy services denoted by the vector $\mathbf{x} = (x_1, \dots, x_J)'$, 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 α .

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

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

Utility maximization yields the FOCs:

$$u_j(\alpha) = p(e_j + b_j(\alpha)) \quad \forall j \in \{1, \dots, J\}. \quad (18)$$

Furthermore, normative utility is:

$$U^n(\alpha) = U^d(\alpha) + \sum_j p b_j(\alpha)x_j(\alpha).$$

We are interested in the welfare effect of a change in α . We find that:

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

We can approximate a change in consumer surplus by $\Delta CS \approx \frac{U^n}{\alpha} d\alpha$, 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 disaggregate 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 α yields the 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 of a diagonal Hessian, i.e., that utilities from different services are additively separable, we get:

$$\Delta x_j = \frac{p \Delta b_j}{u_{jj}}, \quad (19)$$

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:

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

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

$$\frac{\partial x_j}{\partial p} = \frac{e_j}{u_{jj}}. \quad (20)$$

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 rearrangements, 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

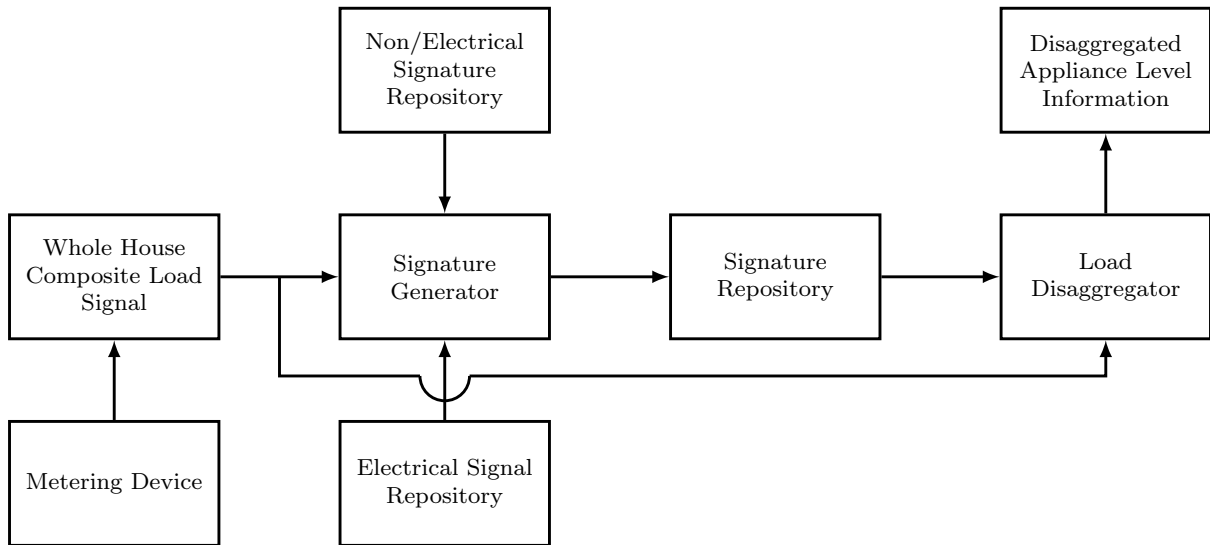
In this section, we describe the so-called nonintrusive appliance load monitoring (NALM) algorithm, which is employed in our study to determine appliance-level consumptions from high-frequency measurements of total electricity consumption of a household. Based on Hart (1992), we first introduce the general 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 appliances 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 algorithms 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 algorithm 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 algorithms is a signature repository, which collects appliance signatures. To construct this repository, the algorithm 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 Algorithm



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 algorithm 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 intervals 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 two covariates. First, we match on the average per day electricity consumption in the baseline period. Second, we control for differences in the timing of billing periods by matching on the end day of the billing period before the start of the field test. 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 have about the same baseline electricity use distribution. In Table A8, we additionally assess the balance in terms of further billing information. We find that the average end date of the bill is not statistically different for both groups, as expected from matching. Yet, the average start of the billing period starts about one month earlier in the experimental control group, compared to the matched non-experimental observations. To account for such differences in the baseline billing period, we include month-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):

$$Y_{it}^{norm} = \alpha Y_i^b + \beta EC_i + \gamma' \mathbf{T}_i + \nu_t + \mu_m^b + \epsilon_{it}, \quad (21)$$

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. ν_t and μ_m^b denote day and month-of-baseline fixed effects. 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 following specification, which allows us to identify the impact of the app elements:

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

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 A12 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: $F(24, 277) = 1.15$, p-value: 0.2938.

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 -5.5% 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 -4.7% in T_2 to -7.3% in T_3 . As shown in Panel b), the total conservation effects can be clearly attributed to disaggregate feedback, which yields a 4.7% reduction in electricity use, relative to the MC group. Again, we do not find 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 (10.5 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 A9). 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 A13 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 -1.4% , but is not statistically significant at any conventional level. When we restrict our analysis to a balanced panel, the treatment effect amounts to -3.3% , but is not statistically significant at any conventional level. Using the non-experimental sample to identify the total conservation effect (Panel b of Table A13), we find that disaggregate feedback yields a persistent reduction by -4.7% when using the full sample (Column 3) and of -2.7% when restricting our sample to a balanced panel of households (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:

$$Y_{it}^{norm} = \alpha Y_i^b + \tau \text{AfterCSP}_t + \beta_1 D_i + \beta_2 D_i \cdot \text{AfterCSP}_t + v_t + \mu_m^b + \epsilon_{it},$$

$$Y_{it}^{norm} = \alpha Y_i^b + \tau \text{AfterCSP}_t + \gamma_1 EC_i + \gamma_2 EC_i \text{AfterCSP}_t + \delta_1 D_i + \delta_2 D_i \text{AfterCSP}_t + v_t + \mu_m^b + \epsilon_{it},$$

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_{CSP} denotes the ATE during the core study period. The estimate δ_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 A14, 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:

$$Y_{it}^{norm} = \alpha Y_i^b + \sum_{h=1}^{24} \beta_j^h D_i + v_t + \mu_h + \epsilon_{itdj}.$$

In Figure A10, 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

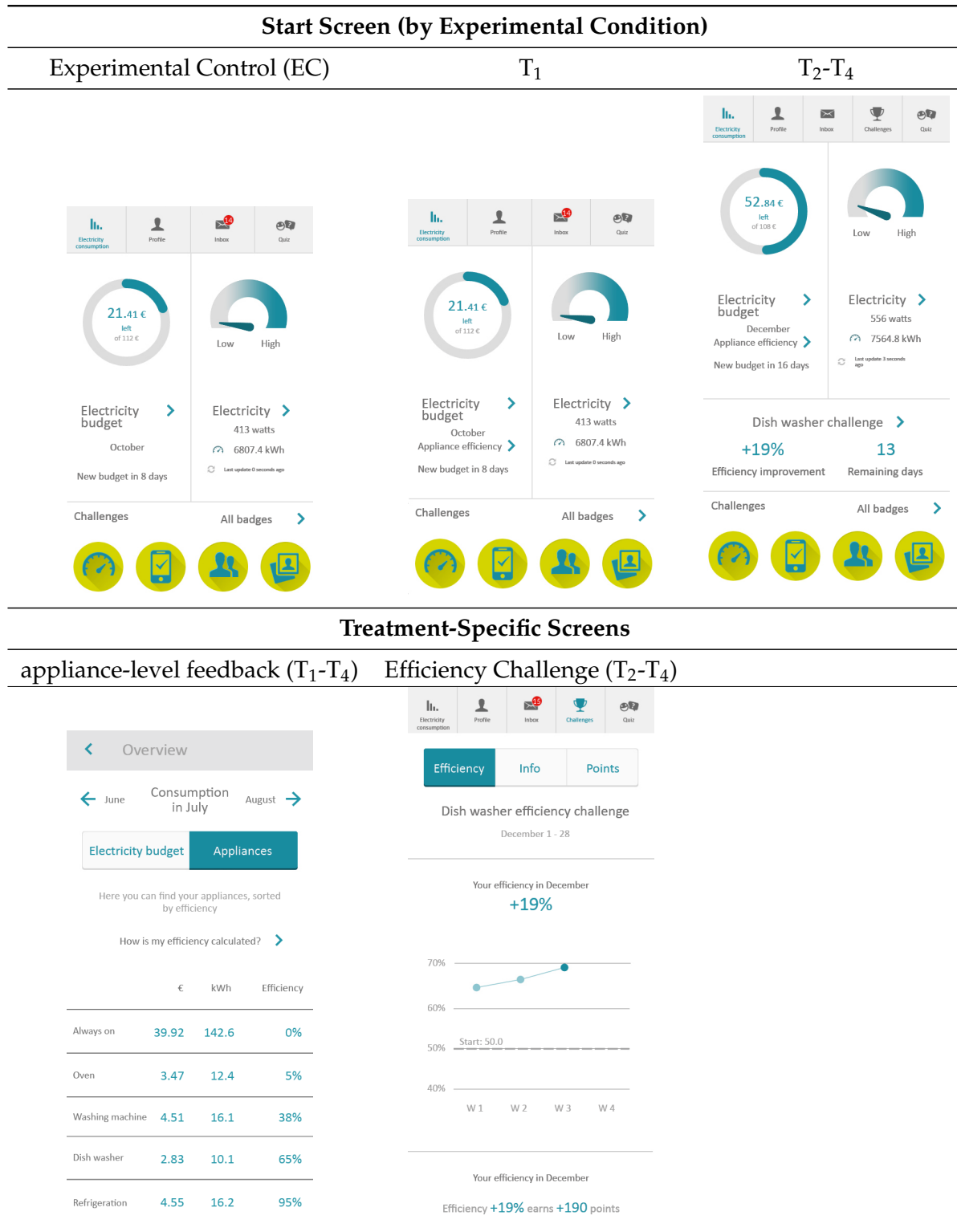


Figure A3: Visualization of Screens II

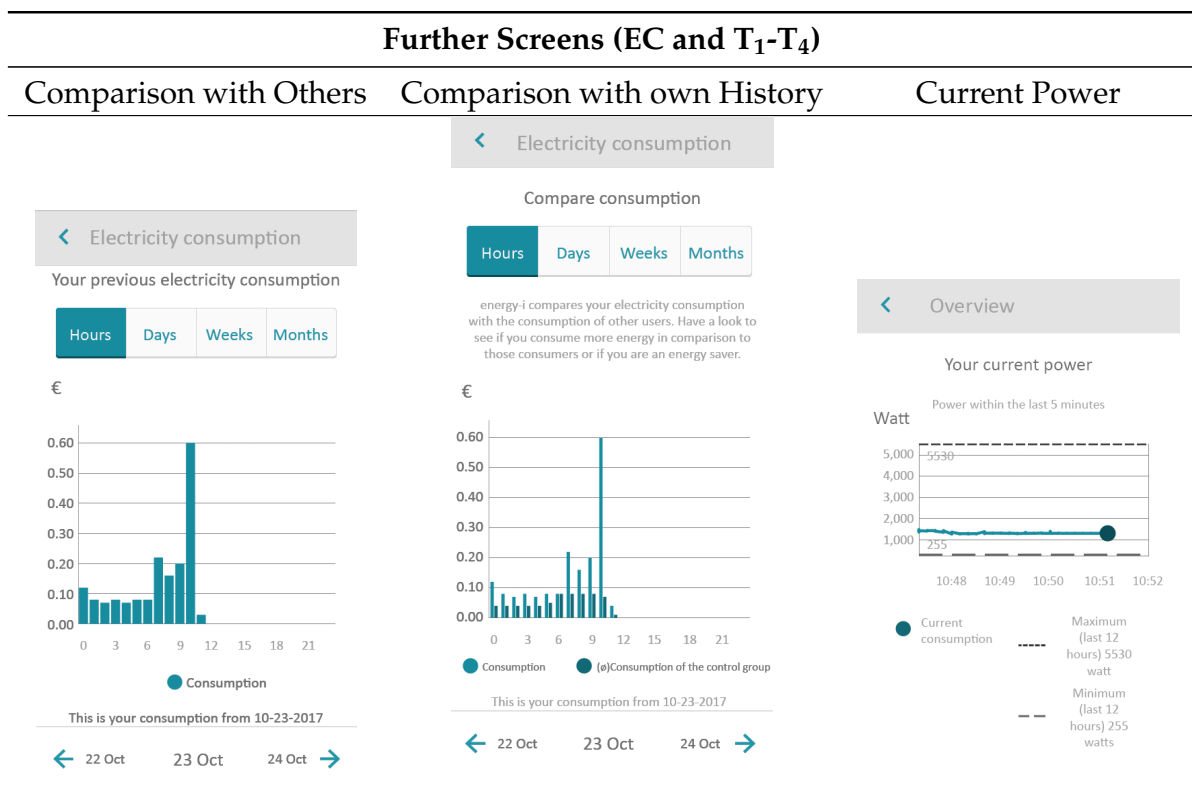
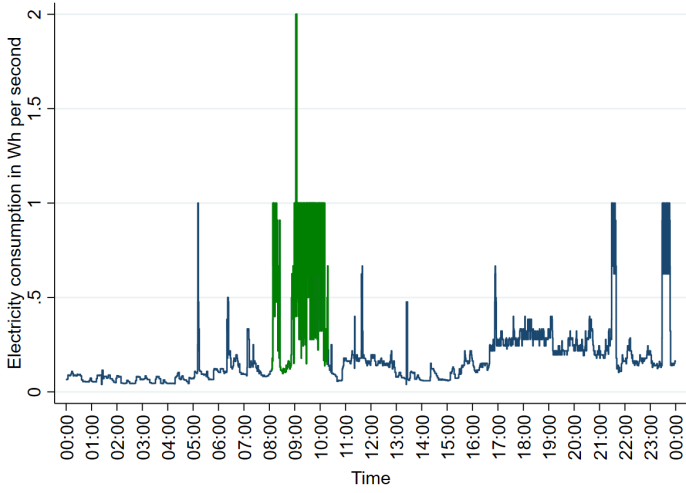
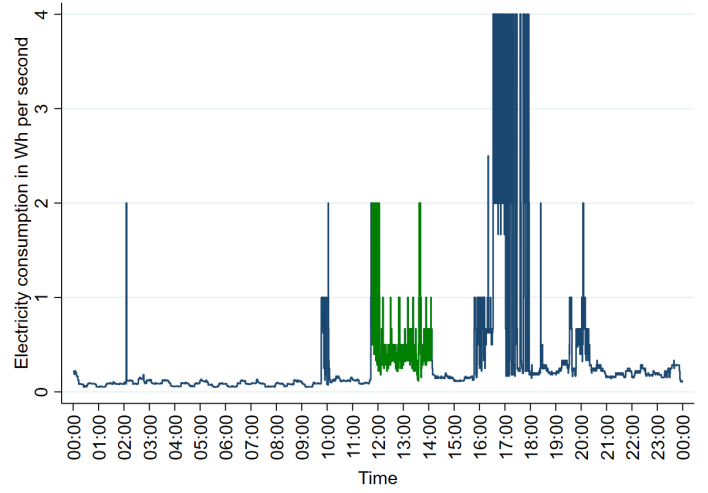


Figure A4: Appliance Signatures

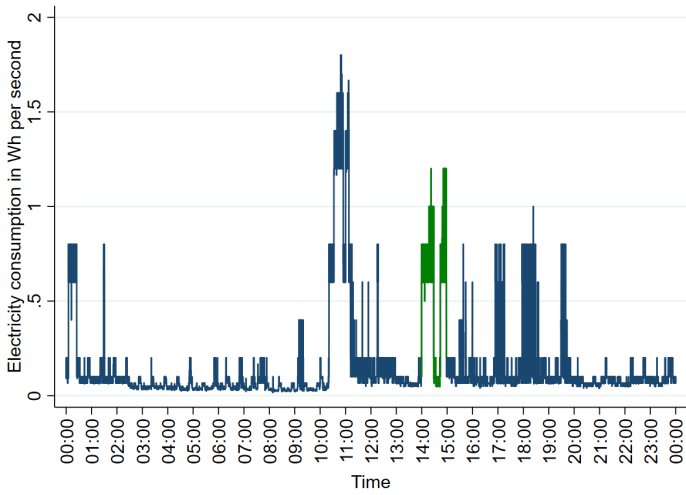
(a) Washing Machine



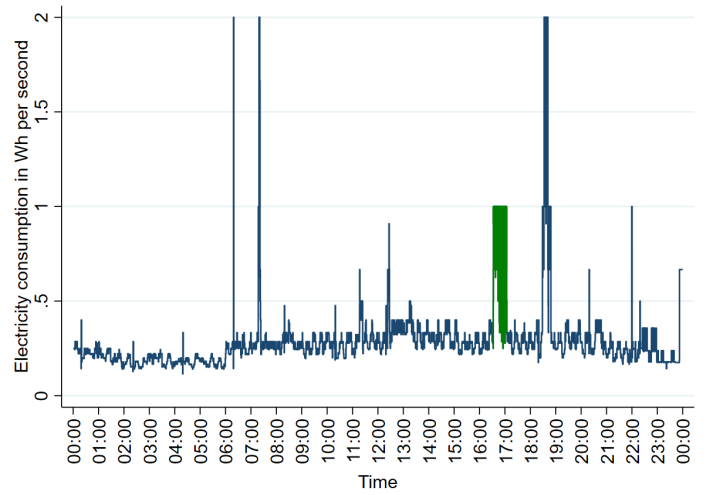
(b) Dryer



(c) Dish Washer

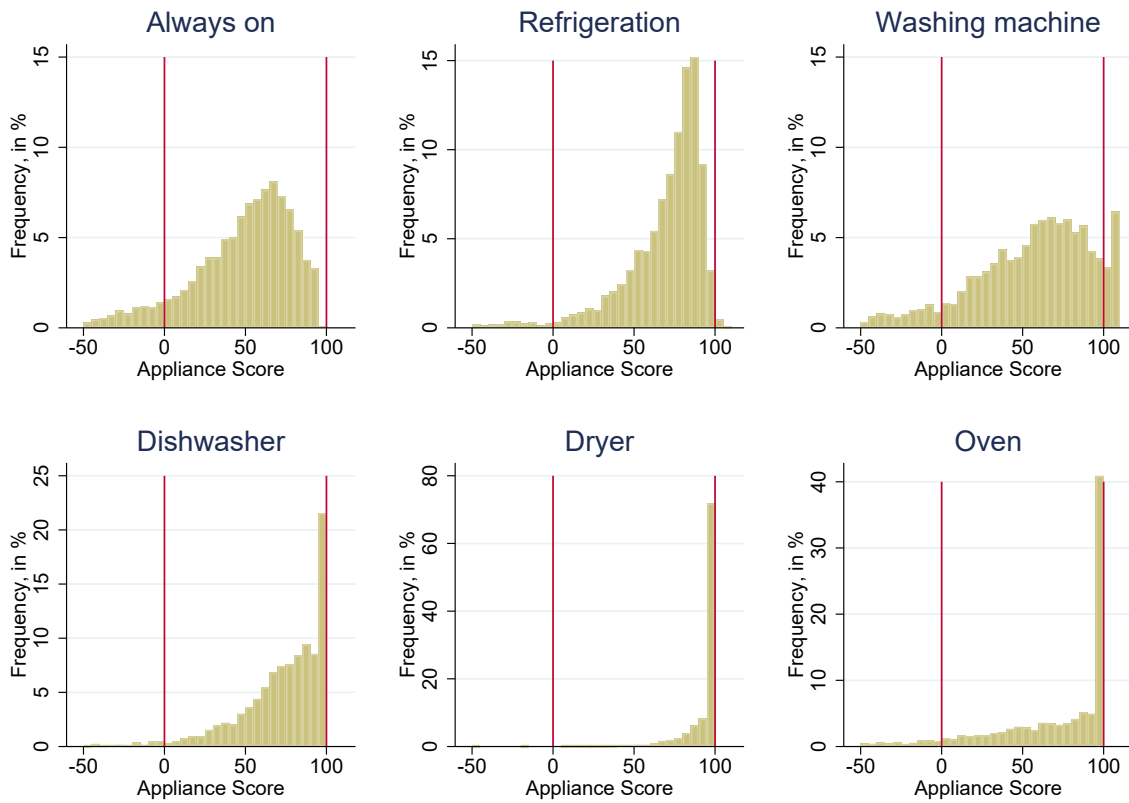


(d) Oven



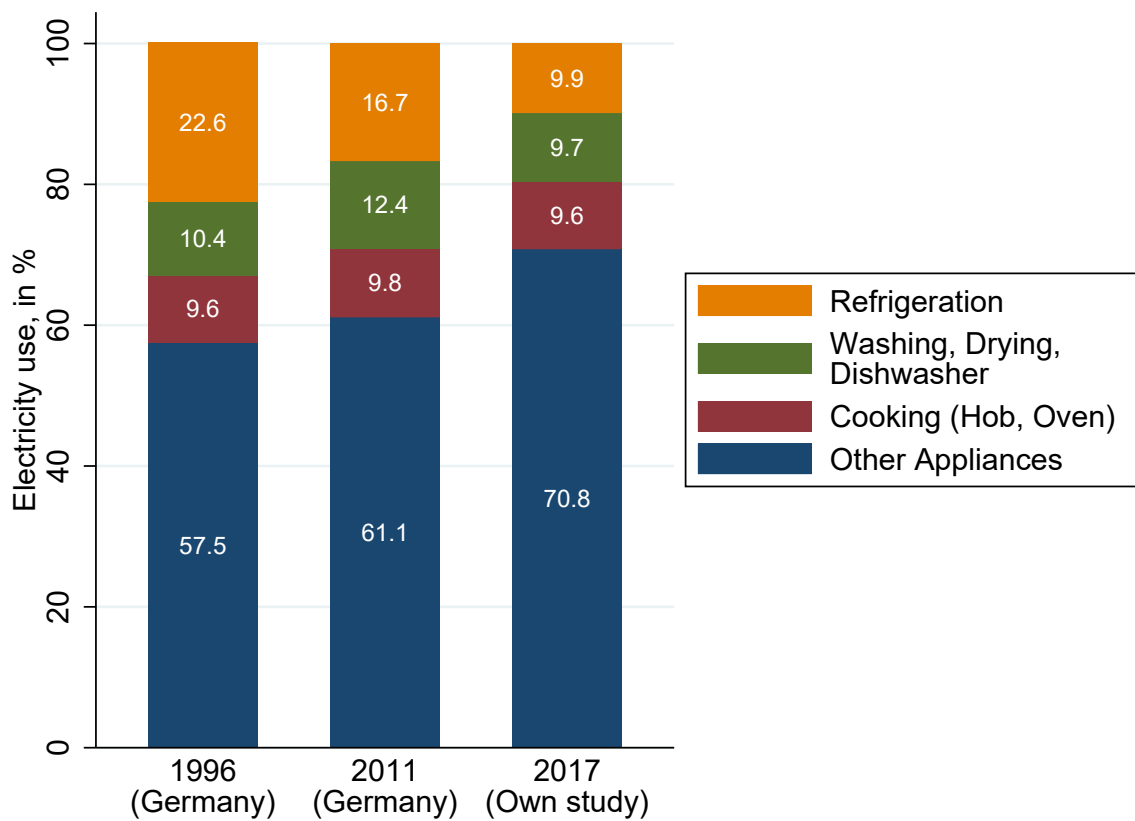
Notes: The graphs depict exemplary appliance signatures as detected by the nonintrusive appliance load monitoring (NALM) algorithm in our study. Periods highlighted in green correspond to a detected appliance use event for the respective appliance.

Figure A5: Distribution of Appliance Scores



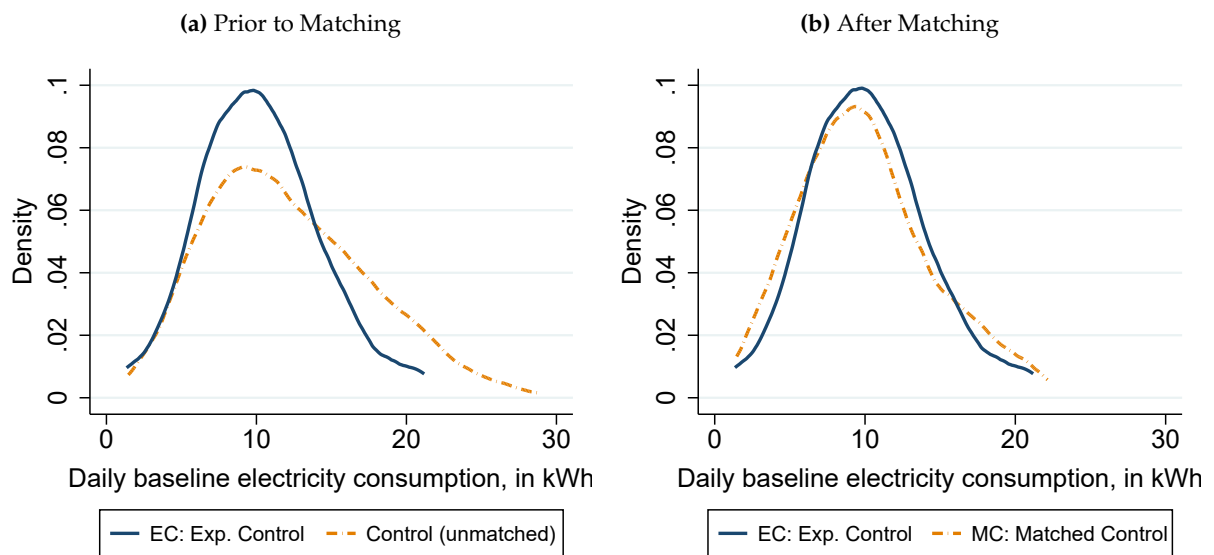
Notes: Appliance Scores are calculated as follows: $\text{Appliance Score} = 100 \times (\text{Monthly Appliance Consumption} - \text{Bench}_{\text{low}}) / (\text{Bench}_{\text{up}} - \text{Bench}_{\text{low}})$, where $\text{Bench}_{\text{low}}$ and $\text{Bench}_{\text{high}}$ correspond to pre-determined 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 A3).

Figure A6: Decomposition of Electricity Uses, by Appliance Category



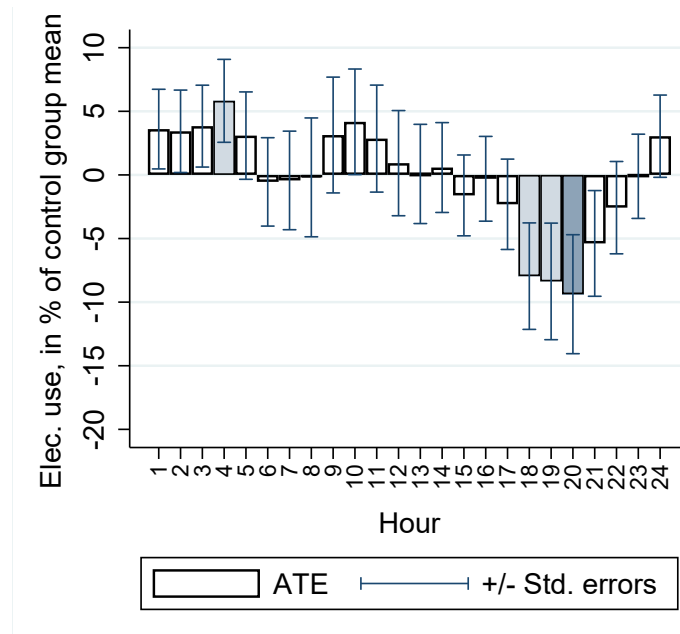
Notes: Values are expressed as percentages of total electricity consumption in a year. Values for 1996 and 2011 are drawn from BDEW (2016), Energie-Info - Stromverbrauch im Haushalt, Bundesverband der Energie- und Wasserwirtschaft. Values from our study are calculated for the Experimental Control group. 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-backofen/betriebswerte-und-energieverbrauchs-kennzeichnung, last access: February 27, 2020).

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



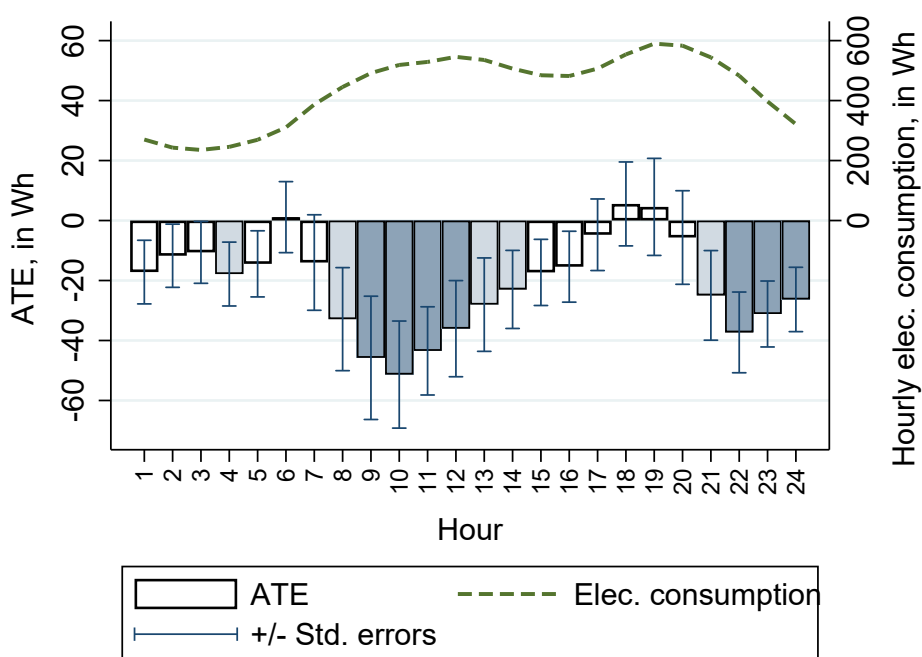
Notes: "Control (unmatched)" denotes all control group households, while "MC: Matched Control" denotes the group of households that have been matched to households in the Experimental Control (EC) condition.

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



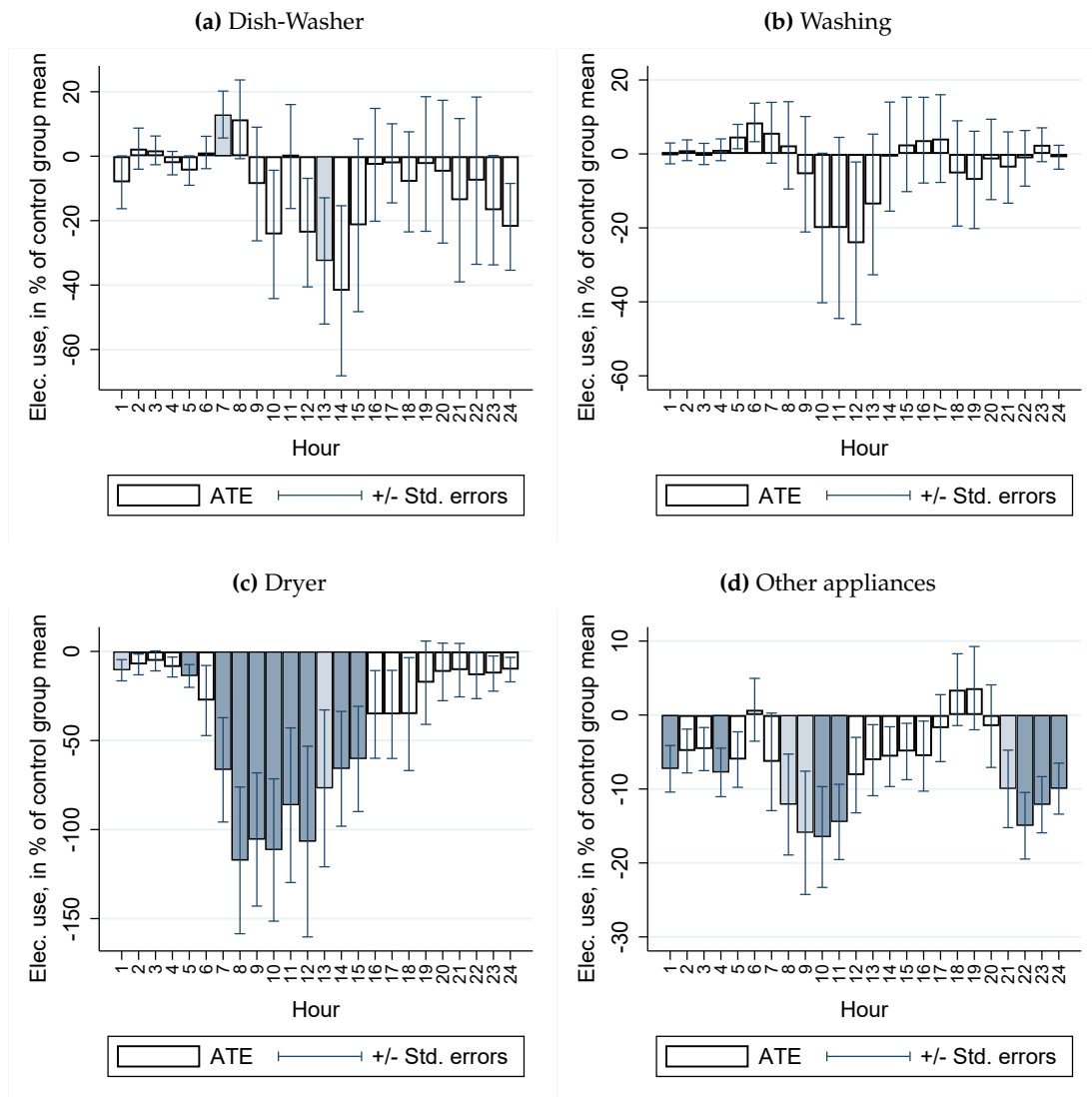
Notes: Shaded bars indicate that treatment effects are statistically significant at the 1% (blue shaded) or 5% (light blue shaded) level. Whiskers indicate a range of +/- 1 standard error (clustered at the household-match level). The outcome variable is daily electricity consumption, divided by the mean in the EC group. Using participants in the MC and the EC groups, we estimate the following equation: $Y_{i,t,h}^{norm} = \alpha Y_i^b + \sum_{h=1}^{24} \beta^h EC_i + v_t + \mu_h + \epsilon_{i,t,h}$. We cannot reject the null hypothesis that all hourly point estimates are zero: $F(24, 277) = 1.15$, p-value: 0.2938.

Figure A9: ATE of Disaggregation (*D*) relative to Experimental Control (*EC*), by Hour of the Day



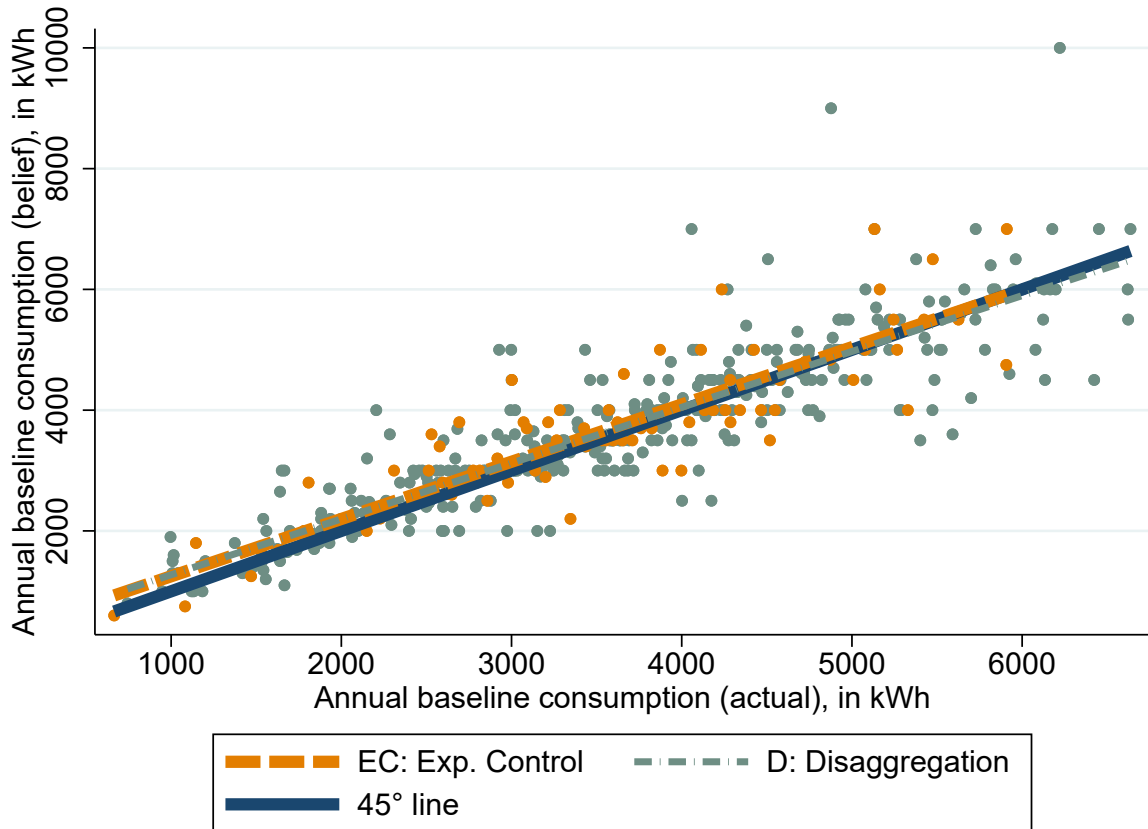
Notes: The ATE corresponds to the treatment effect of the groups T_1 - T_4 relative to the *EC* group (in kWh). Hourly elec. consumption corresponds to the consumption in the *EC* group. Shaded bars indicate that treatment effects are statistically significant at the 1% (blue shaded) or 5% (light blue 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.

Figure A10: 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 hourly mean in the EC group. The categories *Refrigeration* and *Always-On* are measured daily, so that we cannot estimate hourly treatment effects for them.

Figure A11: ATE of Experimental Control (*EC*) on Hourly Electricity Consumption (relative to Matched Control)



Notes: Dots represent data points, lines stem from linear regressions of beliefs on actual baseline consumption and a constant (separately for households in the experimental control group, *EC*, and households in the treatment groups $T_1 - T_4$, *D*). We elicited beliefs about the baseline electricity consumption in a survey that we conducted prior to the field experiment. Actual baseline consumption corresponds to the consumption of a household from the last (annual) bill prior to the field experiment. We drop outliers, defined as all observations above the 95 percentile of the actual baseline consumption distribution, as well as below the 2.5 or above the 97.5 percentile of the distribution of consumption beliefs, divided by the respective actual consumption.

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): Population 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/Materialflusse-Energieflusse/Tabellen/stromverbrauch-haushalte.html>, last access: March 9, 2020).

Table A2: Descriptives: Appliance Possessions and House Characteristics

	All	EC	T1	T2	T3	T4	P-value	N
Appliance Possession								
# of cooling appliances	2.27	2.23	2.24	2.27	2.27	2.33	0.82	700
# of dish washers	0.92	0.91	0.91	0.95	0.93	0.91	0.87	700
# of washing machines	1.02	1.00	1.04	1.00	1.01	1.04	0.61	700
# of tumble dryers	0.72	0.74	0.71	0.69	0.69	0.77	0.68	700
# of ovens	0.98	0.98	0.99	0.98	0.99	0.98	0.76	700
# of hobs	0.99	0.99	0.99	0.98	0.99	0.99	0.84	700
HH lives in detached house (1: yes, 0: no)	0.45	0.46	0.42	0.48	0.43	0.44	0.83	700
HH lives in semi-detached house (1: yes, 0: no)	0.14	0.13	0.16	0.10	0.16	0.16	0.37	700
HH lives in an apartment (1: yes, 0: no)	0.24	0.22	0.26	0.23	0.24	0.23	0.94	700
# of bedrooms	3.03	3.06	3.01	3.00	3.00	3.08	0.88	686
Water heating via gas (1: yes, 0: no)	0.62	0.59	0.63	0.64	0.60	0.64	0.83	700
Space heating via gas (1: yes, 0: no)	0.58	0.53	0.68	0.57	0.55	0.57	0.07	700

Notes: P-values are from F-tests of mean equality in all experimental conditions: EC, T₁, T₂, T₃, T₄ (heteroscedasticity-robust standard errors).

Table A3: 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-greecs>). 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, telephones, 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 A4: Distribution of Daily Use, in kWh/day

	Mean	Std. dev.	Min	Max	p1	p25	p50	p75	p99	N
Total electricity	10.16	5.36	0.00	100.07	1.72	6.56	9.27	12.75	27.03	106,283
Always on	2.47	1.97	0.00	47.49	0.14	1.10	2.01	3.30	9.33	93,187
Refrigeration	1.02	0.66	0.00	11.12	0.07	0.63	0.88	1.24	3.29	93,185
Dishwasher	0.30	0.63	0.00	14.15	0.00	0.00	0.00	0.00	2.64	84,511
Washing mach.	0.49	1.11	0.00	40.30	0.00	0.00	0.00	0.60	5.08	91,473
Dryer	0.14	0.59	0.00	36.06	0.00	0.00	0.00	0.00	2.88	65,852
Oven	0.24	0.81	0.00	20.57	0.00	0.00	0.00	0.00	3.86	93,187
Other appliances	5.56	3.89	0.00	56.58	0.00	3.01	4.74	7.15	19.07	93,187

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

Table A5: Distribution of Use per Utilization, in kWh

	Mean	Std. dev.	Min	Max	p1	p25	p50	p75	p99	N
Always-On	2.47	1.97	0.000	47.49	0.13	1.10	2.01	3.30	9.32	93,664
Refrigeration	1.02	0.66	0.000	11.12	0.07	0.63	0.88	1.24	3.29	93,659
Dishwasher	1.08	0.42	0.002	10.98	0.34	0.79	1.04	1.32	2.37	23,643
Washing mach.	1.00	0.75	0.001	18.54	0.05	0.53	0.81	1.27	3.65	45,270
Dryer	1.02	1.02	0.004	38.46	0.03	0.38	0.74	1.37	4.51	8,979
Oven	0.80	1.11	0.001	20.09	0.01	0.19	0.47	0.95	5.55	28,362

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 A6: Beliefs vs. Actual Baseline Consumption (at the Household-Level)**(a) Mean Annual Baseline Consumption, Beliefs vs. Actual**

	Belief (in kWh)	Actual (in kWh)	Mean Diff. (in kWh)	Std. Err.	P-value	n
Pooled	3,627.6	3,539.4	-88.29	(27.74)	0.002	497
<i>EC</i>	3,611.6	3,491.5	-120.06	(60.51)	0.048	92
T_1	3,675.8	3,614.0	-61.80	(63.01)	0.327	100
T_2	3,642.4	3,561.0	-81.34	(75.86)	0.284	99
T_3	3,531.8	3,483.0	-48.77	(44.84)	0.277	106
T_4	3,681.3	3,547.1	-134.30	(64.29)	0.037	100

(b) Alignment Between Beliefs and Actual Baseline Consumption

	(1)		(2)	
	Estimate	Std. Err.	Estimate	Std. Err.
Y^b : Baseline elec. use	0.928***	(0.026)	0.948***	(0.058)
D : T_1 - T_4			48.891	(205.672)
$D \times Y^b$			-0.024	(0.065)
Constant	342.923***	(82.470)	300.114	(184.225)
R^2	0.7878		0.7879	
Number of obs.	497		497	

Notes for both Panels: Beliefs about the electricity consumption in the baseline period were elicited in a survey that we conducted prior to the field experiment. Actual baseline consumptions were obtained from billing data. ***, **, * denote statistical significance at the 1%, 5%, 10% level, respectively. Robust standard errors are in parantheses. *EC* equals 1 for households in the experimental control group. D equals one for households in the groups $T_1 - T_4$. We drop outliers, defined as all observations above the 95 percentile of the actual baseline consumption, as well as below the 2.5 or above the 97.5 percentile of the consumption belief, divided by the actual consumption.

Notes for Panel b): In Column 1), we pool all experimental groups and estimate the model: $Y^{b,belief} = \alpha + \beta Y^b + \epsilon$, where $Y^{b,belief}$ and Y^b denotes the belief and the actual consumption of household i in the baseline period, respectively (and ϵ denotes an error term). In Column (2), we interact βY^b with the disaggregation dummy D that equals one for the groups T_1 - T_4 : $Y^{b,belief} = \alpha + \beta Y^b + \gamma D \cdot Y^b + \epsilon$. The estimate $\hat{\beta}$ gives the average change in beliefs as the actual consumption increases by one unit (for the *EC* group). The estimate $\hat{\gamma}$ gives the the change of this slope for the households in the treatment groups T_1 - T_4 (relative to the *EC* group).

Table A7: 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	4,554
T ₅	4.5	2.9	0.1	9.8	0.1	1.9	4.5	7.0	9.6	4,068

Notes: Statistics refer to monthly payoffs from the appliance challenges in the treatment groups T₃ and T₅. *p*1 denotes the first percentile, *p*25 the 25th percentile, etc. *N* denotes the number of month-appliance challenge observations.

Table A8: Balance Experimental Sample vs. Matched Non-Experimental Sample

	Difference EC-MC	Std. Err.	P-val.	N
Baseline elec. cons., in kWh/day	0.23	0.27	0.40	280
End of baseline billing, in days	-10.04	11.58	0.39	280
Start of baseline billing, in days	-33.75	13.28	0.01	280

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

Table A9: Heterogeneity in ATE, by Baseline Consumption

(a) Experimental Sample			(b) Sample Including Matched Controls		
	Estimate	Std. Err.		Estimate	Std. Err.
<i>D</i> : Disaggregation	-0.051***	(0.017)	<i>D</i> : Disaggregation	-0.047***	(0.015)
$D \times Y^{b,dm}$	-0.066	(0.052)	$D \times Y^{b,dm}$	-0.134***	(0.044)
$Y^{b,dm}$: Baseline elec. use (demeaned)	0.939***	(0.045)	$Y^{b,dm}$	1.007***	(0.035)
			<i>EC</i> : Exp. Control	0.008	(0.021)
			$EC \times Y^{b,dm}$	-0.068	(0.058)
Day fixed effects	✓		Day fixed effects	✓	
Month-of-baseline FE	✓		Month-of-baseline FE	✓	
R^2	0.5480		R^2	0.5542	
Number of obs.	106,283		Number of obs.	127,342	
Number of households	700		Number of households	840	

Notes: ***, **, * denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parentheses 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 A10: ATEs on Daily Electricity Consumption, Relative to Exp. Control

(a) Daily Aggregation, Day Fixed-Effects			(b) Hourly Aggregation, Hour-by-Day Fixed-Effects		
	(1)	(2)		(1)	(2)
$D: T_1-T_4$	-0.048*** (0.016)		$D: T_1-T_4$	-0.047*** (0.016)	
T_1		-0.044** (0.019)	T_1		-0.044** (0.019)
T_2		-0.039** (0.019)	T_2		-0.038** (0.019)
T_3		-0.066*** (0.020)	T_3		-0.066*** (0.020)
T_4		-0.043** (0.021)	T_4		-0.043** (0.021)
Day fixed effects (FE)	✓	✓	Hour-by-Day fixed effects (FE)	✓	✓
Month-of-baseline FE	✓	✓	Month-of-baseline FE	✓	✓
R^2	0.5586	0.5589	R^2	0.2217	0.2218
Number of obs.	106,283	106,283	Number of obs.	2,543,261	2,543,261
Number of households	700	700	Number of households	700	700

Notes: ***, **, * denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parentheses 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.

Table A11: ATEs on Daily Electricity Consumption, Relative to Exp. Control (Balanced Panel)

(a) Effect of Experimental Conditions			(b) Effects of App Elements			
	(1)	(2)		(3)	(4)	(5)
$D: T_1 - T_4$	-0.064*** (0.019)		$D: \text{Disaggregation}$	-0.064*** (0.019)	-0.068*** (0.021)	-0.064*** (0.022)
T_1		-0.065*** (0.023)	$M: \text{Monetary incentives}$		0.025 (0.015)	0.015 (0.017)
T_2		-0.048** (0.022)	$R: \text{Ranking}$		-0.016 (0.015)	-0.026 (0.020)
T_3		-0.087*** (0.024)	$M: \text{Monet. inc.} \times R: \text{Rank.}$			0.021 (0.030)
T_4		-0.056** (0.027)				
$Y^b: \text{Baseline elec. use}$	0.878*** (0.028)	0.879*** (0.028)	$Y^b: \text{Baseline elec. use}$	0.878*** (0.028)	0.880*** (0.028)	0.879*** (0.028)
Day fixed effects (FE)	✓	✓	Day fixed effects (FE)	✓	✓	✓
Month-of-baseline FE	✓	✓	Month-of-baseline FE	✓	✓	✓
R^2	0.5373	0.5380	R^2	0.5373	0.5379	0.5380
Number of obs.	79,562	79,562	Number of obs.	79,562	79,562	79,562
Number of households	460	460	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 A12: ATEs on Daily Electricity Consumption, Relative to Matched Control

(a) Effect of Experimental Conditions			(b) Effects of App Elements			
	(1)	(2)		(3)	(4)	(5)
<i>EC</i> : Exp. Control	-0.009 (0.022)	-0.009 (0.022)	<i>AF</i> : Aggregate feedback	-0.009 (0.022)	-0.009 (0.022)	-0.009 (0.022)
<i>D</i> : $T_1 - T_4$	-0.055*** (0.016)		<i>D</i> : Disaggregation	-0.047*** (0.016)	-0.049*** (0.018)	-0.045** (0.018)
T_1		-0.053*** (0.019)	<i>M</i> : Monetary incentives		0.016 (0.013)	0.005 (0.016)
T_2		-0.049** (0.020)	<i>R</i> : Ranking		-0.010 (0.013)	-0.021 (0.018)
T_3		-0.073*** (0.021)	<i>M</i> : Monet. inc. \times <i>R</i> : Rank.			0.023 (0.027)
T_4		-0.047** (0.021)				
Y^b : Baseline elec. use	0.910*** (0.020)	0.910*** (0.020)	Y^b : Baseline elec. use	0.910*** (0.020)	0.910*** (0.020)	0.910*** (0.020)
Day fixed effects	✓	✓	Day fixed effects	✓	✓	✓
Month-of-baseline FE	✓	✓	Month-of-baseline FE	✓	✓	✓
R^2	0.5687	0.5689	R^2	0.5687	0.5689	0.5690
Number of obs.	127,790	127,790	Number of obs.	127,790	127,790	127,790
Number of households	840	840	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 A13: ATE on Daily Electricity Consumption (After the Core Study Period), Only Main Effects

	(a) Experimental Sample		(b) Experimental and Matched Control Samples		
	(1) Full Sample	(2) Balanced Panel	(3) Full Sample	(4) Balanced Panel	
			<i>EC</i> : Experimental Control	-0.036 (0.026)	-0.026 (0.035)
<i>D</i> : Disaggregation	-0.014 (0.017)	-0.033 (0.022)	<i>D</i> : Disaggregation	-0.047** (0.022)	-0.027 (0.023)
Y^b : Baseline elec. use	0.759*** (0.022)	0.732*** (0.034)	Y^b : Baseline elec. use	0.781*** (0.024)	0.760*** (0.032)
Day fixed effects	✓	✓	Day fixed effects	✓	✓
Month-of-baseline FE	✓	✓	Month-of-baseline FE	✓	✓
R^2	0.4978	0.4909	R^2	0.5112	0.5004
Number of obs.	54,603	29,330	Number of obs.	66,084	35,765
Number of households	586	321	Number of households	708	391

Notes: ***, **, * denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parentheses 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 A14: Difference in ATEs During and After the Core Study Period

	(1) Experimental Sample	(2) Experimental and Matched Control Samples
1(After Core Study Period)	-0.052* (0.031)	-0.008 (0.034)
EC : Experimental Control		-0.010 (0.032)
EC : Exp. Control \times 1(After Core Study Period)		-0.025 (0.033)
D: Disaggregation	-0.063** (0.025)	-0.060** (0.026)
D: Disaggregation \times 1(After Core Study Period)	0.038 (0.023)	0.038 (0.023)
Y^b : Baseline elec. use	0.840*** (0.032)	0.866*** (0.028)
R^2	0.5280	0.5450
Number of obs.	84,891	103,411
Number of participants	321	321

Notes: ***, **, * denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parentheses 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 A15: Difference in ATEs, by Subgroups

	(1)	(2)
<i>D</i> : Disaggregation	-0.040 (0.028)	-0.019 (0.028)
1(Two Occupant)	0.071** (0.030)	
1(Three+ Occupants)	0.128*** (0.036)	
1(Own Property)		0.056* (0.029)
<i>D</i> : Disaggregation \times 1(Two Occupant)	0.003 (0.035)	
<i>D</i> : Disaggregation \times 1(Three+ Occupants)	-0.035 (0.039)	
<i>D</i> : Disaggregation \times 1(Own Property)		-0.040 (0.034)
γ^b : Baseline elec. use	0.864*** (0.024)	0.888*** (0.024)
R^2	0.5614	0.5598
Number of obs.	106,283	105,846
Number of participants	700	696

Notes: ***, **, * denote statistical significance at the 1%, 5%, 10% level, respectively. Standard errors are in parentheses 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 A16: ATE on Appliance-Level Consumption, Frequency of Use, and Electricity Use Intensity

(a) ATE on appliance-level electricity consumption, in % of control group							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Always-On	Refrigeration	Dish-Washer	Washing	Dryer	Oven	Other appl.
<i>D</i> : Disaggregation	0.031 (0.055)	0.001 (0.052)	-0.071 (0.100)	-0.026 (0.076)	-0.434** (0.173)	-0.095 (0.169)	-0.075** (0.031)
<i>C</i> : Challenges	-0.049 (0.045)	-0.012 (0.044)	-0.029 (0.078)	-0.003 (0.056)	-0.006 (0.117)	0.178 (0.130)	0.005 (0.022)
<i>Y^b</i> : Baseline elec. use	1.156*** (0.065)	0.460*** (0.086)	0.746*** (0.103)	0.637*** (0.072)	1.064*** (0.166)	1.355*** (0.197)	0.851*** (0.033)
Day fixed effects (FE)	✓	✓	✓	✓	✓	✓	✓
Month-of-baseline FE	✓	✓	✓	✓	✓	✓	✓
<i>R</i> ²	0.368	0.152	0.046	0.028	0.035	0.040	0.356
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.
<i>D</i> : Disaggregation	-	-	-0.023 (0.091)	-0.002 (0.058)	-0.398** (0.189)	-0.036 (0.156)	-
<i>C</i> : Challenges	-	-	-0.050 (0.070)	0.041 (0.046)	-0.010 (0.122)	0.307** (0.124)	-
<i>Y^b</i> : Baseline elec. use	-	-	0.723*** (0.094)	0.459*** (0.058)	1.097*** (0.181)	0.543*** (0.131)	-
Day fixed effects (FE)	✓	✓	✓	✓	✓	✓	✓
Month-of-baseline FE	✓	✓	✓	✓	✓	✓	✓
<i>R</i> ²	-	-	0.046	0.027	0.056	0.036	-
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.
<i>D</i> : Disaggregation	-	-	-0.058* (0.031)	-0.035 (0.043)	-0.038 (0.126)	-0.081 (0.087)	-
<i>C</i> : Challenges	-	-	0.032 (0.025)	-0.032 (0.029)	-0.020 (0.106)	-0.019 (0.071)	-
<i>Y^b</i> : Baseline elec. use	-	-	0.065** (0.027)	0.231*** (0.037)	0.052 (0.086)	0.769*** (0.093)	-
Day fixed effects (FE)	✓	✓	✓	✓	✓	✓	✓
Month-of-baseline FE	✓	✓	✓	✓	✓	✓	✓
<i>R</i> ²	-	-	0.056	0.042	0.082	0.106	-
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 parentheses, 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 A17: Appliance-Level Elasticity Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Always-On	Refrigeration	Dish-Washer	Washing	Dryer	Oven	Other appl.
ln p	-0.310 (0.355)	-0.400* (0.226)	-0.401 (0.722)	-0.552 (0.418)	-3.423*** (1.233)	-1.067 (1.024)	-0.289 (0.252)
# of occupants = 2	0.113 (0.069)	0.129*** (0.043)	0.872*** (0.192)	0.843*** (0.129)	-0.012 (0.359)	0.694*** (0.200)	0.278*** (0.055)
# of occupants = 3	0.038 (0.084)	0.116** (0.052)	1.099*** (0.201)	0.937*** (0.140)	0.104 (0.396)	0.970*** (0.229)	0.301*** (0.058)
# of occupants = 4	0.126 (0.087)	0.098* (0.058)	1.178*** (0.227)	1.019*** (0.144)	0.135 (0.406)	1.104*** (0.245)	0.321*** (0.062)
# of occupants = 5	-0.071 (0.106)	0.212*** (0.074)	1.460*** (0.287)	0.692*** (0.191)	0.508 (0.422)	0.947*** (0.325)	0.264*** (0.081)
Hh. net income	0.048*** (0.013)	-0.017* (0.009)	-0.001 (0.035)	0.026 (0.021)	-0.026 (0.052)	-0.042 (0.037)	-0.001 (0.008)
1(Hh. net income missing)	0.208** (0.091)	-0.100 (0.064)	-0.038 (0.241)	0.126 (0.140)	-0.419 (0.363)	-0.032 (0.252)	0.031 (0.061)
γ^b : Baseline elec. use	0.106*** (0.006)	0.030*** (0.004)	0.049*** (0.011)	0.059*** (0.009)	0.044*** (0.014)	0.072*** (0.013)	0.082*** (0.005)
Constant	-1.230** (0.507)	-0.986*** (0.326)	-3.728*** (1.066)	-3.338*** (0.616)	-7.236*** (1.879)	-4.725*** (1.504)	0.058 (0.356)
R^2	0.489	0.169	0.154	0.288	0.092	0.151	0.635
Number of participants	700	700	610	677	259	527	700

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>.
- Al-Ubaydli, Omar, John A. List, Danielle LoRe, and Dana Suskind.** 2017a. "Scaling for Economists: Lessons from the Non-Adherence Problem in the Medical Literature." *Journal of Economic Perspectives* 31 (4): 125–44.
- Al-Ubaydli, Omar, John A. List, and Dana L. Suskind.** 2017b. "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results." *American Economic Review* 107 (5): 282–86.
- 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.** 2015. "Site Selection Bias in Program Evaluation." *The Quarterly Journal of Economics* 130 (3): 1117–1165.
- 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, Jörg Peters, and Christoph M Schmidt.** 2020a. "Social Norms and Energy Conservation Beyond the US." *Journal of Environmental Economics and Management* 103 102351.
- Andor, Mark A., Andreas Gerster, and Stephan Sommer.** 2020b. "Consumer Inattention, Heuristic Thinking and the Role of Energy Labels." *The Energy Journal* 41 (1): 83–122.
- 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.
- Azmat, Ghazala, and Nagore Iriberry.** 2010. "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students." *Journal of Public Economics* 94 (7-8): 435–452.
- 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.
- Brandon, Alec, John A List, Robert D Metcalfe, Michael K Price, and Florian Rundhammer.** 2019. "Testing for Crowd Out in Social Nudges: Evidence from a Natural Field Experiment in the Market for Electricity." *Proceedings of the National Academy of Sciences* 116 (12): 5293–5298.
- 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 Koppalle.** 2019. "Experimental Evidence on the Effect of Information and Pricing on Residential Electricity Consumption." *NBER Working Paper* No. 25576.
- Byrne, David, Lorenz Goette, Leslie Martin, Samuel Schoeb, Verena Tiefenbeck, and Thorsten Staake.** 2020. "The Behavioral Mechanisms of Habit Formation: Evidence from a Field Experiment." Technical report, University of Melbourne.
- Byrne, David P., Andrea La Nauze, and Leslie A. Martin.** 2020. "An Experimental Study of Monthly Electricity Demand (In)Elasticity." *The Energy Journal* 42 (2): .
- Cadmus-Bertram, Lisa A., Bess H. Marcus, Ruth E. Patterson, Barbara A. Parker, and Brit-tany 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 Elek-trizitätsverbrauch." Final report to the Swiss Federal Office of Energy.
- destatis.** 2017. "Bevölkerung und Erwerbstätigkeit." Fachserie 1 Reihe 3. Federal Statistical Office of Germany, Wiesbaden.

- Donnal, J. S., and S. B. Leeb.** 2015. "Noncontact Power Meter." *IEEE Sensors Journal* 15 (2): 1161–1169.
- 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.
- Fronedel, 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.
- Mas, Alexandre, and Enrico Moretti.** 2009. "Peers at Work." *American Economic Review* 99 (1): 112–45.
- 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 Kimaio, 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"." *The Review of Economic Studies* 87 (5): 2399–2438.
- 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.
- Tran, Anh, and Richard Zeckhauser.** 2012. "Rank as an Inherent Incentive: Evidence from a Field Experiment." *Journal of Public Economics* 96 (9-10): 645–650.
- Blanes i Vidal, Jordi, and Mareike Nossol.** 2011. "Tournaments without Prizes: Evidence from Personnel Records." *Management Science* 57 (10): 1721–1736.
- Wichman, Casey J.** 2017. "Information Provision and Consumer Behavior: A Natural Experiment in Billing Frequency." *Journal of Public Economics* 152 13–33.

