Discussion Paper Series – CRC TR 224

Discussion Paper No. 149
Project B 07

Complementarities in Behavioral Interventions: Evidence from a Field Experiment on Energy Conservation

Ximeng Fang*
Lorenz Goette**
Bettina Rockenbach***
Matthias Sutter****
Verena Tiefenbeck†
Samuel Schoeb††
Thorsten Staake†††

January 2020

*University of Bonn (email: x.fang@uni-bonn.de)
**University of Bonn, National University of Singapore
***University of Cologne
****Max Planck Institute for Research into Collective Goods, University of Cologne, University of Innsbruck
†Friedrich-Alexander Universität Nürnberg-Erlangen, ETH Zurich
††University of Bamberg
†††University of Bamberg

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

Complementarities in Behavioral Interventions: Evidence from a Field Experiment on Energy Conservation*

Ximeng Fang Lorenz Goette Bettina Rockenbach Matthias Sutter
Verena Tiefenbeck Samuel Schoeb Thorsten Staake ‡

23 January 2020

Abstract

Behavioral policy often aims at overcoming barriers like imperfect information and limited attention that contribute to suboptimal consumer decisions. When multiple barriers are present, a single intervention that does not overcome all barriers simultaneously may fail to unfold its full potential. We conduct a three-month randomized field experiment on energy conservation in a resource-intensive everyday activity, using two different interventions. Home energy reports fail to reduce energy use despite achieving significant knowledge gains; real-time feedback induces considerable conservation effects. Strikingly, combining both interventions boosts these effects by over 50%. This showcases how barrier multiplicity can generate complementarities in behavioral interventions.

JEL classification: D12, D83, Q41

Keywords: behavioral interventions, energy conservation, inattention, real-time feedback, home energy reports, policy interactions, randomized controlled trials

*We are indebted to our research assistant team (in particular Benedikt Kauf and Kristina Steinbrecher) for their indefatigable support in running the field experiment. Further, we would like to thank audiences in Bonn, at the ESA Dijon, the IAREP-SABE Dublin, the EAERE Manchester, the 3rd CRC TR 224 Conference, the PCBS 2019 Prague, the 12th RGS Doctoral Conference, and the 2nd BoMa PhD Workshop. Financial support by the University of Cologne Forum "Energy", the SUN Institute Environment & Sustainability, and by the German Research Foundation (DFG) through CRC TR 224 (Project B07) and ECONtribute (under Germany´s Excellence Strategy – EXC 2126/1– 390838866) is gratefully acknowledged. This study is registered in the AEA RCT Registry under the identification number "AEARCTR-0004865".

‡Fang: University of Bonn (email: x.fang@uni-bonn.de). Goette: University of Bonn, National University of Singapore. Rockenbach: University of Cologne. Sutter: Max Planck Institute for Research into Collective Goods, University of Cologne, University of Innsbruck. Tiefenbeck: Friedrich-Alexander Universität Nürnberg-Erlangen, ETH Zurich. Schoeb: University of Bamberg. Staake: University of Bamberg.
Author contributions: Authors 1 to 5 designed and planned the experiment, author 1 coordinated the implementation of the experimental design. Authors 6 and 7 provided invaluable technical support. Author 1 analyzed the data and drafted the manuscript; feedback by authors 2 to 7 helped improve it significantly.

1. Introduction

Individuals regularly make decisions that are not in line with their own values and intentions. For example, gym members want to stay in shape and healthy, yet exercise less often than they initially plan to (DellaVigna and Malmendier, 2006); entrepreneurs want to manage their businesses effectively, yet fail to follow simple rules for good financial practice (Drexler, Fischer and Schoar, 2014). One particularly relevant domain is pro-environmental behavior. Amidst growing public concern about climate change, many people are willing to make personal sacrifices in order to protect the environment, yet often fail to act pro-environmentally in their everyday lives (Kollmuss and Agyeman, 2002; Frederiks, Stenner and Hobman, 2015). This causes negative externalities, as the ever-rising demand for energy and natural resources fuels existential societal challenges like global warming and water scarcity (IPCC, 2014).

The gap between consumers' intentions and actions could be partly driven by behavioral barriers such as imperfect information, limited attention, self-control problems, status quo bias, etc. (Allcott, 2016; Handel and Schwartzstein, 2018). These informational or cognitive frictions and biases can also mute responses to monetary incentives and thereby dampen the effectiveness of traditional price-based policies (Jessoe and Rapson, 2014; Madrian, 2014). Behavioral interventions aimed at overcoming these types of barriers have been used to facilitate behavioral change in a variety of contexts such as retirement savings or public health (Thaler and Sunstein, 2008; Madrian, 2014), and are also regularly advocated as promising policy tool for fostering more environmentally sustainable household consumption behavior (Dietz et al., 2009; Allcott and Mullainathan, 2010; Reddy et al., 2017). However, the existing evidence on their effectiveness remains mixed.[1]

In this paper, we emphasize that multiple behavioral barriers can be present at the same time. For example, people tend to underestimate the energy-intensity of room heating (Attari et al., 2010), and additionally they may be inattentive and forget to turn off the heating system when leaving the house, or fail to resist the temptation of turning up temperature on a cold winter's day. Presence of multiple barriers in turn implies that a single intervention that does not overcome all relevant barriers to a sufficient degree may stay below its potential, because its effect would be dampened by the remaining barriers. For example, information provision about environmental consequences could fall flat if individuals do not activate their newly acquired knowledge when making decisions. Thus, we hypothesize that interventions that each overcome a different set of barriers can be complements, in the sense that each intervention is more effective in conjunction with the other(s) than in isolation; put differently, the whole is greater than the sum of its parts. If such complementarities indeed exist, then clever policy bundling may increase the cost-effectiveness of behavioral interventions beyond what can be achieved

[1] Pro-environmental interventions have drawn from a broad set of instruments such as information campaigns, feedback, social norms, goal-setting, etc. For reviews, see e.g. Abrahamse et al. (2005), Fischer (2008), Delmas, Fischlein and Asensio (2013), Karlin, Zinger and Ford (2015), and Andor and Fels (2018).

with piecemeal approaches. To the best of our knowledge, this hypothesis has not been formulated explicitly and tested empirically yet. Our paper aims to fill this gap.

We first introduce a stylized conceptual framework to demonstrate the hypothesized complementarity theoretically. In the framework, multiple barriers (e.g. lack of information, inattention, present bias) distort agents' choices by preventing them from fully incorporating the marginal costs of consumption. Importantly, the different barriers enter in a multiplicative way, which implies that behavioral interventions will only unfold their full effect if all relevant barriers are eliminated at the same time. We then report evidence from a three-month randomized field experiment in which we use two behavioral policy tools, home energy reports and real-time feedback, to overcome various barriers to resource conservation in a specific energy- and water-intensive everyday activity. Crucially, we evaluate both the combined intervention as well as *each* intervention in isolation, as this is necessary for empirical identification of policy interaction effects. The home energy reports in our setting are primarily aimed at closing knowledge gaps about environmental impacts while real-time feedback is targeted to channel attention in the exact moment individuals are engaged in the resource-consuming activity.[2]

Lack of knowledge about resource-intensity of everyday activities is regarded as a major barrier to pro-environmental behavior (Gardner and Stern, 2008); indeed, consumers often underestimate the impact of highly energy- or water-intensive activities (Attari et al., 2010; Attari, 2014).[3] Many existing intervention strategies aim at providing information, yet conservation effects tend to be modest in methodologically more rigorous studies (Abrahamse et al., 2005; Delmas, Fischlein and Asensio, 2013; Karlin, Zinger and Ford, 2015). One possible explanation is that the existence of further barriers prevents information interventions from taking full effect (Dietz et al., 2009). Limited attention is a prime candidate, as knowledge and good intentions are insufficient if not acted on in the moment of decision-making.[4] For example, in a consumer goods context, Chetty, Looney and Kroft (2009) show that explicitly posting the sales tax on price tags in U.S. supermarkets decreases product demand, although shoppers are in principle perfectly knowledgeable about the tax, even if it is not shown. There is also evidence pointing

[2]Even if there are alternative channels at work, complementarity can still arise as long as the set of barriers targeted by the two interventions, respectively, are sufficiently distinct.

[3]Similarly, consumers may misperceive gasoline costs (Allcott, 2011*a*) or underestimate the emissions impact associated with food choice (Camilleri et al., 2019). Knowledge gaps can give rise to subtimal behavior, e.g. if an agent has misspecified priors that are never corrected due to lack of feedback (Hanna, Mullainathan and Schwartzstein, 2014; Schwartzstein, 2014). Hence, households often engage in conservation measures that are relatively ineffective (Gardner and Stern, 2008; Tonke, 2019).

[4]Of course, there can be other important barriers. Most prominently, agents might suffer from self-control problems, e.g. due to present bias. Werthschulte and Loeschel (2019) find correlational evidence for a positive association between present bias and energy consumption. Similarly, individuals might be biased towards concentrated rather than dispersed payoffs (Koszegi and Szeidl, 2013; Dertwinkel-Kalt et al., 2019). Status quo bias and inertia can also prevent consumers from adopting more sustainable behavior, which is why changing defaults can be effective (Pichert and Katsikopoulos, 2008; Fowlie et al., 2018). Furthermore, literature in the psychology of risk perception argues that the complex and uncertain nature of climate change breeds overoptimism and makes it difficult for people to internalize it as threat that requires immediate action (Gifford, 2011; Markowitz and Shariff, 2012).

towards the role of (in)attention in environmental behavior (Allcott and Rogers, 2014; Keefer and Rustamov, 2018; Langenbach et al., 2019).[5] Making resource use and environmental costs more salient can thus help to bridge the gap between attitudes and actions (Tiefenbeck et al., 2018).

We borrow the terminology for our "home energy reports" from a prominent program by the company Opower, in which millions of U.S. households received information mails containing feedback on their own as well as their neighbors' energy use.[6] The original mails reduced total household energy use by about two percent (Allcott, 2011*b*), although home energy reports have been found to be ineffective in other samples (Andor et al., 2017; Myers and Souza, 2019). Our real-time feedback intervention is based on Tiefenbeck et al. (2018), who find that providing salient feedback about resource use in real-time through smart meters results in 22% lower energy use in the shower, which translated into a 5% reduction in total household energy use.

We conducted our field experiment in student dormitories in the cities of Bonn and Cologne, Germany, and following Tiefenbeck et al. (2018), we focused on resource consumption behavior in the shower. Showering is an interesting paradigm for several reasons. First, it is resource-intensive: an average shower in our sample requires 2.2 kWh of energy to heat up 38 liters of water.[7] Second, individuals tend to neglect the energy-intensity of warm water use. Third, showering is prone to behavioral biases like inattention and self-control problems, as the pleasure of a warm shower is salient and immediate, whereas the cost of resource use seems abstract and distant. Fourth, showering is a routine activity that occurs within a household context.

A total of 351 students participated in our experiment, with all of them living in single dorm apartments with a private bathroom. It is noteworthy that subjects in our sample had no monetary incentives for conserving energy or water, because utilities are completely included in the flat monthly rent. For the duration of our study, from early December 2016 until late February/early March 2017, each subject was equipped with a smart shower meter that could be installed directly below the shower head and recorded detailed data on each water extraction. We then implemented two types of interventions, home energy reports and real-time feedback, both tailored specifically to resource use in the shower. Subjects were randomly assigned into one of four experimental conditions in a 2×2 design: no intervention (CON group), home energy reports only (HER group), real-time feedback only (RTF group), or both interventions combined (DUAL group).

[5]For example, Langenbach et al. (2019) document a moderating effect of cognitive capacity on the relation between pro-environmental intentions and actions. Allcott and Rogers (2014) observe "action-and-backsliding" patterns after receiving home energy reports. Also, the positive relation between feedback frequency and the effectiveness of feedback interventions seems consistent with inattention playing a role (Fischer, 2008; Karlin, Zinger and Ford, 2015).

[6]There are some major differences between the Opower reports and our intervention. We provide feedback only to one specific energy-intensive activity, and we also put less focus on social norms but more on additional information about environmental impacts. Details will follow later in the paper.

[7]Total household use per capita in Germany was on average 22.4 kWh of energy and 123 liters of water in 2016 (Source: German Federal Statistical Office).

Home energy reports were sent twice via email and provided feedback on the subject's water and energy use in the shower as well as additional information on environmental impacts in terms of CO_2 emission. In contrast to the Opower home energy reports, there is little focus on social norms in our experimental intervention. We do, however, include a social comparison element in the second home energy report. Real-time feedback provided instant measurement of water use through a local display on the smart meter, thus drawing immediate attention while showering, but there was no extra information on energy use or environmental impacts. Apart from evaluating the effectiveness of home energy reports and real-time feedback in isolation, respectively, we further evaluate the interaction effect between the two interventions. Our main hypothesis is that there could be considerable complementarities, as agents may only engage in resource conservation if they believe that their behavior has sufficient environmental impact *and* pay immediate attention to resource use in the shower.

Our empirical results show that subjects in the RTF group reduce their energy (water) consumption by about 0.4 kWh (6.3 liters) per shower compared to subjects in the CON group, which corresponds to 17–18% of baseline resource use — a conservation effect that is consistent with results in previous studies using the same smart shower meter (Tiefenbeck et al., 2018, 2019). The treatment effect remains fairly stable over the entire 3-month duration of the study. In contrast, home energy reports in isolation (HER group) do not lead to any discernible conservation effects at all. Crucially, in line with our hypothesis, we observe a striking complementarity between the two interventions. Combining home energy reports with real-time feedback (DUAL group) further increases the treatment effect of real-time feedback by 0.22 kWh of energy (3.8 liters of water) per shower, more than 50% of the original effect in the RTF group. This additional resource use reduction in the DUAL group is not driven by short-lived boosts directly after receiving a home energy report, but rather seems to unfold over time, which speaks in favor of learning effects and against cueing or Hawthorne effects as the underlying mechanism.

Additional survey analyses show that both interventions help subjects form more precise beliefs about their water use in the shower. Additionally, information included in home energy reports induces drastic (upwards) updates in beliefs about CO_2 emissions due to warm water use in the shower. Hence, the zero conservation effect of home energy reports in isolation is not due to lack of learning. There is also no evidence that subjects in the DUAL group read their reports more carefully than subjects in the HER group. Instead, it seems that, in the absence of real-time feedback, inattention and lack of immediate visibility have prevented knowledge gains about environmental impacts from translating into effective conservation behavior. Overall, our findings showcase that accounting for the presence of multiple behavioral barriers can guide the design of an appropriate policy mix that unlocks the full potential of behavioral policy.

A number of other studies on pro-environmental behavior also explore the idea that some policy measures become more successful when accompanied by other interventions.[8] For example, Jessoe and Rapson (2014) find that pricing schemes that incentivize lower peak electricity consumption fall flat if consumers do not know how to effectively adjust electricity usage; only households who have been outfitted with in-home-displays reduce electricity use significantly in response to price hikes. List et al. (2017) study the combination between Opower home energy reports and financial reward programs and find that adding the latter encourages energy conservation in households that do not respond strongly to the reports alone. Note that these studies do not have a complete 2×2 experimental design, and thus, strictly speaking, cannot cleanly identify complementarities. Brandon et al. (2018) evaluate both the combined and the individual effects of two very similar interventions on household energy conservation, home energy reports and "peak energy reports", which are basically home energy reports for peak electricity periods, and find neither strong evidence for complementarity nor for substitutability. The novel contribution of our paper is that we demonstrate how barrier multiplicity can generate complementarities in behavioral interventions that each target a different set of barriers.

Our paper also relates to other studies on information provision or the role of attention in economic decision making that do not focus on policy interaction effects. Information provision and disclosure are traditional policy instruments which are often proposed when there are strong informational frictions, e.g. asymmetries, search costs, or complexity. They have been shown to influence behavior in a wide variety of other contexts, such as energy efficiency investments (Newell and Siikamäki, 2014), financial decision making (Bertrand and Morse, 2011), educational choice (Jensen, 2010; Hastings and Weinstein, 2008), job search (Altmann et al., 2018), social program/benefit take-up (Bhargava and Manoli, 2015; Liebman and Luttmer, 2015), and food choice (Bollinger, Leslie and Sorensen, 2011; Camilleri et al., 2019). In particular, information can have strong impacts when agents' prior beliefs are severely misspecified, for example when students underestimate returns to higher education (Jensen, 2010), or when farmers neglect the importance of certain input dimensions (Hanna, Mullainathan and Schwartzstein, 2014).

There has also been increasing interest in the role of limited attention in economic decision-making, both in theoretical and empirical literature (Chetty, Looney and Kroft, 2009; Gabaix, 2017). When agents are partially inattentive, sending reminders (Karlan et al., 2016), simplification (Bhargava and Manoli, 2015), or making information more salient (Chetty, Looney and Kroft, 2009; Taubinsky and Rees-Jones, 2018) can help bring intentions and relevant choice dimensions to the top of mind.

The remainder of this paper is structured as follows: Section 2 introduces the theoretical framework for policy interactions under multiple barriers. Section 3 describes the experimental setup and derives behavioral predictions. Section 4 presents our data as

[8]Policy interaction have also been considered in development economics, where a number of studies experimentally test the effect of combined interventions on financial savings (Dupas and Robinson, 2013; Jamison, Karlan and Zinman, 2014), education (Mbiti et al., 2019), risky sexual behavior (Duflo, Dupas and Kremer, 2015), or demand for health products (Ashraf, Jack and Kamenica, 2013).

well as some descriptive statistics. Section 5 explains our empirical approach and Section 6 presents our main empirical results. In Section 7, we study the potential mechanisms underlying the results and provide robustness checks. Section 8 concludes.

2. Conceptual framework

We begin by introducing a stylized conceptual framework to understand how complementarities in behavioral interventions can arise in settings where multiple behavioral barriers are present, e.g. imperfect information, limited attention and present bias. Even if marginal returns to behavioral policy are decreasing, complementarities can arise if interventions target different barriers that operate (to some degree) independently of each other, because interventions that overcome only one barrier in isolation may still leave intact other barriers that prevent behavioral change.

2.1. Policy interactions in a theoretical framework

Behavioral policy can help to bridge the gap between consumers' intentions and actions by mitigating relevant barriers. In this framework, we highlight that the presence of multiple behavioural barriers can generate complementarities between different interventions, in the sense that each intervention becomes more effective in the presence of the other interventions.

Basic setup.— We consider a simple setup in which an agent engages in an energy-intensive activity, say showering, and the policy objective is to reduce energy use. The agent's consumption level is determined by a trade-off between the consumption utility (pleasure, instrumental benefits, etc.) and the perceived costs of resource use (monetary costs, environmental concern, etc.). She chooses energy use level $e \geq 0$ to maximize

$$U(e) = V(e) - B \cdot C(e)\,, \tag{1}$$

where $V(e)$ is the instantaneous consumption utility and $C(e)$ is the cost of energy consumption.[9] In addition to standard smoothness conditions, we assume that V is hump-shaped (locally increasing at 0, strictly concave, unique maximum) and that C is strictly monotonically increasing and weakly convex. For simplicity, we abstract from uncertainty or dynamics. In the absence of monetary motives, as in our empirical setting, $C(e)$ is the "moral" cost the agent perceives in face of the negative externalities from energy use. However, the cost function is attenuated by an aggregate bias/barrier factor B, and energy use is biased upwards if $B \in [0, 1)$.

[9] The agent may not explicitly optimize with regard to energy use, but as long as the mapping from actual decision variable (e.g. shower duration) to resource use is injective, we can represent the problem as if the agent was optimizing over energy use.

Multiple barriers.— This aggregate barrier can be the product of a collection of separate frictions and biases, so

$$B = b_1 \cdot b_2 \cdot \ldots b_K \,. \tag{2}$$

The various barriers b_k enter multiplicatively and thus jointly prevent the agent from fully implementing her conservation motive.[10] For example, the agent could underestimate energy-intensity, so the first bias factor is $b_1 = \phi \in [0, 1)$. In addition, she may be partially inattentive and consider environmental costs only with weight $b_2 = \theta \in [0, 1)$. As cream topping, the agent may be present-biased and like to save her good intentions for another day, so that $b_3 = \beta \in [0, 1)$. Put together, $B = \phi \cdot \theta \cdot \beta$. Note that we assume for simplicity that b_k does not change with energy use e for any barrier k.

Consumption behavior.— The agent's consumption level is defined by equating marginal utility and marginal costs, but with the latter being diminished by the aggregate barrier:

$$V'(e) = \prod_{k=1}^{K} b_k \cdot C'(e) = B \cdot C'(e) \,. \tag{3}$$

If $B < 1$, than the true marginal cost is underweighted and energy use is thus biased upwards. Defining f such that $f(e) = \frac{V'(e)}{C'(e)}$ for all $e \in [0, \infty)$, we can directly map the relation between implemented energy use and aggregate bias as

$$e(B) = f^{-1}(B) \,. \tag{4}$$

Notice that f^{-1} is a strictly decreasing function, so the weaker the aggregate bias, i.e. B closer to 1, the lower the energy use.[11]

Behavioral interventions.— In this setup, we define behavioral interventions as policies that (only) work through manipulating B. In contrast, price-based policies are aimed at increasing the marginal costs of energy use, $C'(e)$ that the agent faces. Equation (4) shows that any behavioral policy P that mitigates the aggregate bias compared to no-intervention state o will induce the agent to conserve energy. Hence,

$$\Delta B^P = B^P - B^o > 0 \implies \Delta e^P = e(B^P) - e(B^o) < 0 \,, \tag{5}$$

and the more successful an intervention in mitigating the aggregate bias the larger its effect on the outcome of interest. Recall that the policy objective here is to reduce energy use, so $\Delta e^P < 0$ moves in the desired direction.

[10] An alternative interpretation, more from a social planner's point of view, is that the agent should internalize the full social cost $C^s(e)$, and the ratio of private to social cost $C(e)/C^s(e)$ would then be another factor entering into the aggregate bias B^s, so decision utility is $U(e) = V(e) - B^s \cdot C^s(e)$. This interpretation highlights the overarching policy objective of reducing externalities instead of "internalities". Efforts to increase the privately perceived cost can include carbon pricing, social norms, goal-setting, etc.

[11] This is because marginal consumption utility $V'(e)$ is strictly decreasing and marginal cost $C'(e)$ is non-decreasing. Hence, f is strictly increasing, so the inverse function f^{-1} exists and is strictly decreasing.

Policy interactions.— Behavioral interventions I and II are complements if their combination reduces consumption by more than the sum of their individual effects, i.e $\Delta e^{\mathrm{I+II}} < \Delta e^{\mathrm{I}} + \Delta e^{\mathrm{II}}$. If they are substitutes, the inequality sign is reversed. Notice that even under substitutability, it can be the case that $\mathrm{I} + \mathrm{II}$ is more effective than both I and II in isolation, i.e. $\Delta e^{\mathrm{I+II}} < \Delta e^{\mathrm{I}}$ and $\Delta e^{\mathrm{I+II}} < \Delta e^{\mathrm{II}}$, which is exactly the reason why empirical identification of policy interaction effects requires evaluation of the combined intervention as well as of *each* intervention in isolation.

In a single-barrier setting, where $B = b_1$, there is little room for complementarity between interventions I and II that both target b_1. First, there will likely be some degree of redundancy or crowding out, so $\Delta B^{\mathrm{I+II}} < \Delta B^{\mathrm{I}} + \Delta B^{\mathrm{II}}$. Second, one would typically expect decreasing marginal returns, so the scope for further conservation effects diminishes with every intervention that is piled upon another. In our framework, this corresponds to function f^{-1} being convex.[12] Intuitively, resource consumption is more inelastic at lower levels, e.g. due to desire for satisfying basic needs like hygiene; an analogy could be that concave utility of money implies decreasing marginal returns to monetary incentives. In principle, there could be complementarities if the introduction of one policy makes another policy more salient, or if there are foot-in-the-door effects, but we are abstracting from such sources of policy interaction in this paper.[13]

When moving to a multiple-barrier setting as described in equation (2), the same factors can still play a role. In particular, marginal returns to bias mitigation tend to be decreasing, and there could also be redundancy or crowding out if interventions do not target entirely disjunct sets of barriers.[14] However, barrier multiplicity gives rise to a new force that potentially generates complementarities: behavioral policy may need to overcome all barriers simultaneously to unfold its full potential. Mathematically, this statement is based on a simple fact:

$$\frac{\partial^K B}{\partial b_1 \ldots \partial b_K} > 0 \,. \tag{6}$$

For example, even a highly motivated agent needs to be knowledgeable about the impacts of her actions, pay attention at the right moment, and have sufficient willpower,

[12] For example, if V has a positive third derivative and the cost function C is linear or quadratic, then f^{-1} is strictly decreasing and convex. A positive third derivative is often labelled prudence and implies a desire for precautionary saving in choice under risk. Of course, f^{-1} could in principle also be concave, so marginal returns are increasing, but this seems implausible. For example, it can imply that resource conservation interventions have larger effects for low-baseline households, although the opposite is usually true.

[13] Furthermore, it is unclear whether these effects will be substantial, especially since there could also be counteracting forces like competition for attention in a policy-crowded environment or moral licensing effects. In Section 7, we present robustness checks showing that these effects are unlikely to explain our empirical results.

[14] Crowding out can also be the indirect result of interactions across barriers, i.e. if the strength of one barrier is endogenously linked to another barrier. For example, information provision may incidentally induce agents to pay more attention (Hanna, Mullainathan and Schwartzstein, 2014; Gabaix, 2017) or set up commitment devices against their self-control problems.

etc., in order to act pro-environmentally. An informational intervention may correct misperceptions about environmental impacts, but the agent can still be present-biased and inattentive, and that would need to be addressed as well. Formally, a policy intervention P_l that overcomes only a single barrier l still leaves the other barriers intact, so

$$\Delta B^{P_l} = B^{P_l} - B^o = (1 - b_l^o) \prod_{k \neq l} b_k^o . \tag{7}$$

The more severe these remaining barriers the more the effectiveness of P_j is attenuated. Thus, it can be more successful when combined with other interventions that raise $\prod_{k \neq j} b_k$, and vice versa. In the extreme case where $b_k^o = 0$ for all k, it becomes necessary to eliminate every single barrier in one fell swoop, in order to achieve any substantial behavioral change.

To show how superadditivity in barrier mitigation carries over to the outcome of interest, we can look at a Taylor series approximation. The effect of a behavioral policy P on the agent's energy consumption level is as follows:

$$\Delta e^P \;=\; e(B^P) - e(B^o) \;=\; \Delta B^P \,\frac{1}{f'} \;-\; \frac{1}{2}\left(\Delta B^P\right)^2 \frac{f''}{(f')^3} \;+\; \mathcal{O}\left(\left(\Delta B^P\right)^3\right), \tag{8}$$

where $f' = \frac{V''C' - V'C''}{(C')^2} < 0$ and evaluated at baseline level $e(B^o)$. The first-order effect is unambiguously consumption-reducing and directly proportional to barrier manipulation ΔB^P, and it becomes stronger when marginal costs C' are high.[15] In contrast, the sign of the second-order effect depends on the sign of f''. Typically, one would expect $f'' \geq 0$, e.g. due to desire for basic needs satisfaction, which pushes policy interaction effects in the direction of substitutability. Nevertheless, complementarity can arise between interventions I and II that each target a different set of barriers if $\Delta B^{\mathrm{I+II}} > \Delta B^{\mathrm{I}} + \Delta B^{\mathrm{II}}$, which translates directly into a superadditive first-order energy conservation effect.

2.2. Illustration with two barriers

To demonstrate the potential complementarity more clearly, let us consider a simple case with two barriers, b_1 and b_2. For example, b_1 could be an awareness parameter, where $b_1 < 1$ indicates that the agent is unaware of the full environmental impact of her behavior, e.g. if she underestimates the amount of energy required for heating up water for showering. Moreover, b_2 could be an (exogenous) attention parameter, so $b_2 < 1$ indicates partial inattention to environmental costs in the moment of decision making, e.g. while showering. The aggregate bias factor prior to any intervention is therefore $B = b_1 \, b_2$, where we dropped the superscript o for ease of notation. Crucially, the agent has to hold sufficiently high beliefs about environmental impacts as well as be sufficiently attentive

[15]Therefore, our framework also predicts that behavioral interventions can be complementary to traditional price-based policies such as Pigouvian taxation. See also footnote 10.

toward it in order to implement her pro-environmental intentions.

Suppose that we have two interventions. Intervention I only provides information ($b_1^{\text{I}} > b_1$) but does not address limited attention, so that $B^{\text{I}} = b_1^{\text{I}} b_2$ and $\Delta B^{\text{I}} = (b_1^{\text{I}} - b_1)b_2$. Intervention II only channels attention ($b_2^{\text{II}} > b_2$) but does not close knowledge gaps, so that $B^{\text{II}} = b_1 b_2^{\text{II}}$ and $\Delta B^{\text{II}} = (b_2^{\text{II}} - b_2)b_1$. Limited attention dampens the effect of intervention I, and lack of information dampens the effect of intervention II. When the two interventions are introduced jointly, however, both information and attention problems are mitigated, so $\Delta B^{\text{I+II}} = (b_1^{\text{I}} b_2^{\text{II}} - b_1 b_2)$.[16] The joint bias mitigation effect is not only larger than that of any individual intervention, which is easy to see, but it is also larger than the sum of the two individual mitigation effects, because

$$\begin{aligned}\Delta B^{\text{I+II}} - \Big(\Delta B^{\text{I}} + \Delta B^{\text{II}}\Big) &= (b_1^{\text{I}} b_2^{\text{II}} - b_1 b_2) - (b_1^{\text{I}} - b_1)b_2 - (b_2^{\text{II}} - b_2)b_1 \\ &= (b_1^{\text{I}} - b_1)(b_2^{\text{II}} - b_2) > 0\,. \end{aligned} \tag{9}$$

This additional mitigation effect is stronger the more severe each barrier is initially. For example, if the agent is completely inattentive in the absence of intervention II, i.e. $b_2 = 0$, then information intervention I in isolation will not have any effect at all.

The interventions I and II are complements if their efficacy in reducing energy consumption is also superadditive, so $\Delta e^{\text{I+II}} < \Delta e^{\text{I}} + \Delta e^{\text{II}}$. Within this framework, an equivalent way to write the condition is that

$$f^{-1}\Big(B^{\text{II}} + (B^{\text{I+II}} - B^{\text{II}})\Big) - f^{-1}(B^{\text{II}}) < f^{-1}\Big(B + (B^{\text{I}} - B)\Big) - f^{-1}(B)\,, \tag{10}$$

so intervention I must be more effective together with intervention II than in isolation, where I and II are interchangeable. The level differences due to additional intervention II can work to the disadvantage of the combined intervention if f^{-1} is convex, but equation (9) shows that $B^{\text{I+II}} - B^{\text{II}} > B^{\text{I}} - B$, which works to its advantage. Hence, the key takeaway is that despite opposing forces like decreasing marginal returns, the presence of multiple distinct barriers potentially gives rise to complementarities that can be exploited to enhance the effectiveness of behavioral interventions — in particular if some biases are extremely severe. Whether or not this source of policy complementarity is practically relevant in real-world settings is then an empirical question.

3. Experimental setup

We conducted our field experiment in student dormitories in the cities of Bonn and Cologne, Germany, from early December 2016 to late February/early March 2017. Each

[16]Note that we have implicitly assumed that there is no crowding out, i.e. $b_2^{\text{I}} = b_2$ and $b_1^{\text{II}} = b_1$, which rules out endogenous attention or induced information search, although the complementarity property may even be retained when allowing for a limited degree of redundancy. In the setting here, $\Delta B^{\text{I+II}} - \left(\Delta B^{\text{I}} + \Delta B^{\text{II}}\right) > 0$ as long as $b_2^{\text{I}} + b_1^{\text{II}} < 1 + b_1 b_2$.

participant was outfitted with a smart shower meter that measured individual resource consumption in the shower over the three-month experimental period. Subjects were randomly assigned to receive real-time feedback on their water use through the smart meter, or home energy reports with detailed information on water use, energy use, and CO_2 emissions via email, or both. Students in our dormitories have no direct monetary incentives to conserve energy or water, because they pay a flat monthly rent that includes all utility bills, which effectively creates a zero-marginal-cost environment.

3.1. Recruitment of participants

We selected six student dormitory sites in Bonn and Cologne for our study, based on the number of single-person apartments and ease of access for our local research teams. All residents of these dorms are students at the University of Bonn or the University of Cologne, or at various smaller universities in the cities. We recruited our subjects from the pool of dorm tenants living in single-person apartments with private bathroom, as this allows us to precisely measure resource use of each individual. To participate in the study, residents had to actively agree based on the principle of informed consent. Two additional criteria were levied: subject should not have lengthy absences planned within the intended study period (except during Christmas vacation), and they should own a smartphone compatible with Bluetooth 4.0, which was necessary for implementing home energy report treatments.

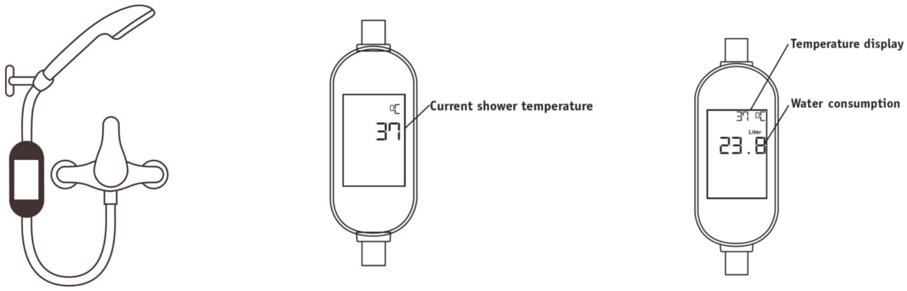
The recruiting process started around mid-October 2016. Posters and flyers informed residents of the selected dorms about the upcoming study, and our local research assistant teams engaged in door-to-door recruiting. Interested students had to complete an online registration survey to provide required information and to give their consent for us collecting and analyzing data on their showering behavior. It was explicitly (and truthfully) stated that we would treat any collected data confidentially and not share it with the dormitory administration. As remuneration, each participant received 20 Euros after completing the study, and ten participants were randomly drawn to receive a 300 Euro cash prize. In total, 406 students registered for the study, out of which 361 met our participation criteria.[17] Ten students subsequently dropped out of the study, either because they had to move out unexpectedly or we were not able to contact them again. This leaves us with a sample of 351 participants.

3.2. Smart shower meters and smartphone app

At the beginning of the study, starting from 5th Dec 2016, each participant was outfitted with an Amphiro b1 smart shower meter that measures and records data on every water

[17] The total number of all single apartments in the selected dorms is 1380, thus our recruitment rate was almost 30%. More than half of the dorm resident pool was lost to us because they were never at home when we knocked, so out of the students we actually managed to talk to, the majority registered for the study.

Figure 1: Amphiro b1 smart shower meter



(a) Position in the shower (b) Display in control mode (c) Display in feedback mode

extraction in the shower. The device can be easily attached between the shower head and shower hose, and features a smartphone-sized liquid crystal display, which can be programmed to display various types of information. The display faces directly towards the user while under the shower, as illustrated in Figure 1a. The smart meter is designed such that showering convenience should not be inhibited, as it is small, lightweight, and needs no battery; power is generated through an integrated hydro turbine, without noticeably affecting water flow in the process. One drawback of the lack of battery is that the device is unaware of the global time: showers can only be recorded in temporal order, but without time stamps. Once water flow in the shower starts, the smart meter is powered on and begins to measure, among others, the amount of water flowing through, water temperature, and the time passed since beginning of water flow. After water flow stops, the device remains powered on for three minutes. If water is turned on again within this time frame, it will continue measurement from the point where it had previously stopped. Once water flow stops for more than three minutes, the device terminates measurement and stores the recorded data as most recent observation point.

We programmed the shower meters to display select pieces of information to participants in real time, i.e. while they are taking their showers. Depending on the study progress and assigned experimental condition, the device was either in control or feedback mode. In control mode, the display only showed information about current water temperature in degree Celsius (see Figure 1b), whereas in feedback mode, it would additionally display the amount of water used (in liter) since the start of the shower (see Figure 1c). In addition, we asked all participants to install the Amphiro smartphone app around week 5 of the experiment, shortly after the end of the Christmas break. The participants could use the app to upload data from their shower meters via Bluetooth connection.[18] We were then able to access the data uploaded through the app and use it to create personalized home energy reports. The original Amphiro smartphone app also calculates summary statistics about users' resource use in the shower, but we deactivated this feature for our study participants, so the its only functionality was data uploading.

[18]The process was quite simple. After installing the smartphone app, subjects created an account, which was then paired to their shower meter through manual entry of a device-specific code. After successful pairing, the meter automatically transmitted all stored data to the app via Bluetooth whenever it was powered on and the smartphone was within range.

One ancillary benefit of the app was that it stored time and date of each data upload, which allows us to construct approximate time windows for each shower. Unfortunately, 28% of all participants did not succeed in uploading any data to the app in time, mostly due to technical problems. The most common sources of failure were problems with the Bluetooth connection or unexpected incompatibility between smartphone and app. We will explain how to deal with this issue later.

3.3. Implementation of real-time feedback

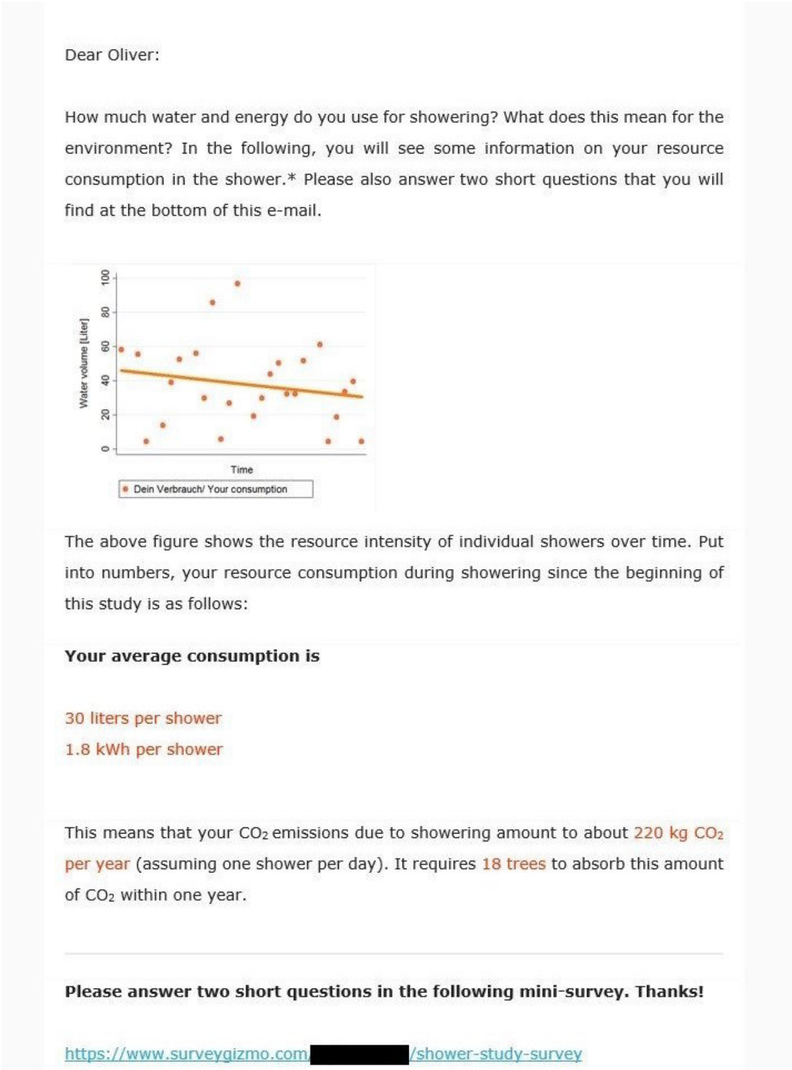
The live tracking of water use on the shower meter display in feedback mode is what we refer to as real-time feedback, our first type of intervention. We programmed half of the smart meters as control devices and the other half as treatment devices. Control devices only displayed the current water temperature throughout the entire study. Treatment devices also started in control mode for the first ten showers, which we use to measure baseline behavior, but switched permanently to feedback mode starting from shower number eleven.

3.4. Implementation of home energy reports

Our second type of intervention consists of two personalized home energy reports. These reports were sent via e-mail and showed descriptive statistics about the subject's water and energy use in the shower, as well as information about environmental impacts. To allow for learning about outcomes of single showers, a graphical representation of the subject's history of water use per shower is included. The reports were constructed based on data that was uploaded by subjects through the smartphone app. We sent out additional reminders to upload data before each planned delivery of the home energy report, but the reports themselves were not explicitly announced. Subjects who did not manage to upload any data only received empty report templates instead of statistical figures and graphs.

Figure 3 shows the screenshot of an example home energy report (Oliver is a a fictitious person). After a short introductory text, subjects see a scatter plot of their history of water use per shower since beginning of the study, including a fitted regression line to help recognize trends and averages. Below the graph, average water use (in liters) and energy use (in kWh) per shower are stated numerically. Furthermore, there is an information panel on projected CO_2 emissions per year and the number of trees required to absorb the corresponding amount of CO_2. The whole report is formulated concisely in neutral language, to avoid any normative or moral suasion elements. In the second report, we added a social comparison component in the spirit of the original Opower home energy reports (Allcott, 2011*b*). Specifically, we assigned a random anonymous peer to each

Figure 3: Screenshot (partial) of a typical home energy report


Dear Oliver:

How much water and energy do you use for showering? What does this mean for the environment? In the following, you will see some information on your resource consumption in the shower.* Please also answer two short questions that you will find at the bottom of this e-mail.

The above figure shows the resource intensity of individual showers over time. Put into numbers, your resource consumption during showering since the beginning of this study is as follows:

Your average consumption is

30 liters per shower
1.8 kWh per shower

This means that your CO_2 emissions due to showering amount to about 220 kg CO_2 per year (assuming one shower per day). It requires 18 trees to absorb this amount of CO_2 within one year.

Please answer two short questions in the following mini-survey. Thanks!

https://www.surveygizmo.com[redacted]/shower-study-survey

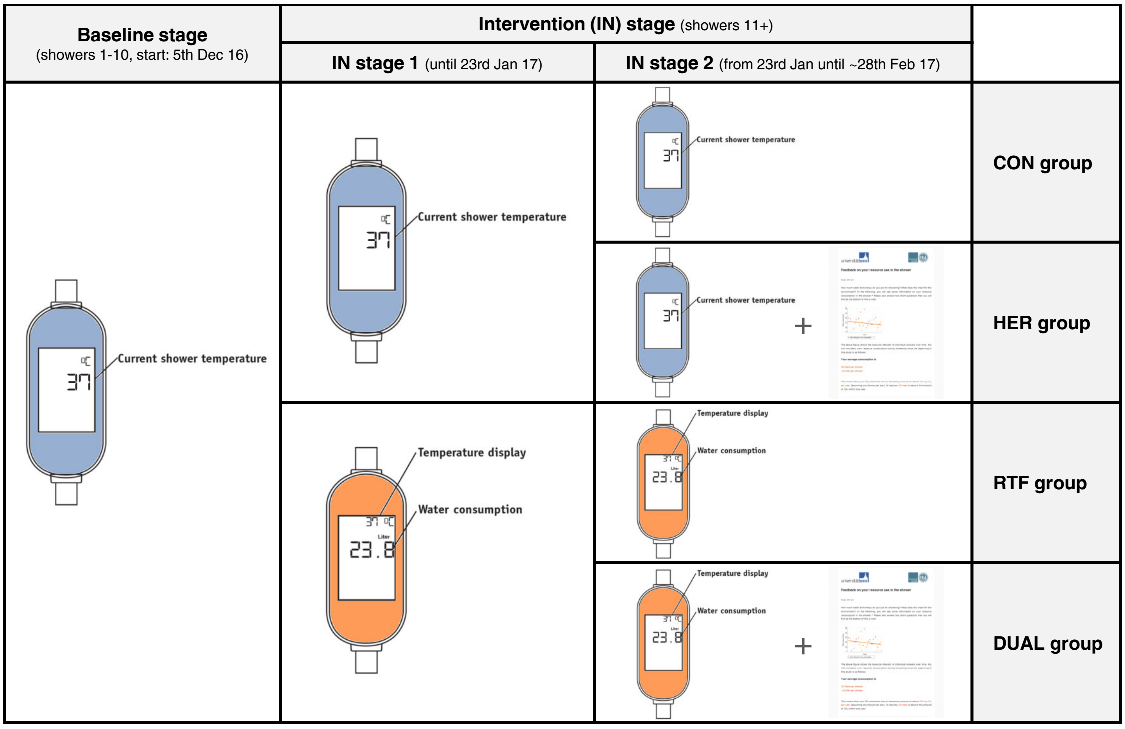
subject and displayed statistics on the peer's energy and water use.[19] At the bottom of each report, there was a personalized link to a mini-survey that we asked subjects to fill out. We can use this information to verify if, and how closely, the email has been read.

3.5. Experimental design

We implemented a complete 2×2 design with four experimental conditions. Subjects in the control (CON) group received no intervention at all; subjects in the RTF group only received real-time feedback through the smart shower meters; subjects in the HER only received home energy reports; and subjects in the DUAL group received both real-time feedback and home energy reports. Treatment assignment was randomized and the

[19]For a screenshot, see Figure A1 in Appendix A. The matching procedure was one-sided and ensured that each subject (except the most and the least efficient) was equally likely to see a peer with lower or higher energy use per shower.

Figure 4: Experimental design and timing of interventions



group sizes are as follows: 82 in CON, 88 in HER, 90 in RTF, 91 in DUAL.[20]

Figure 4 illustrates the experimental design in detail. Each shower meter goes through a baseline stage of ten showers, in which it only displays the current water temperature, regardless of the experimental condition. We use these showers to measure baseline consumption behavior. Starting from the eleventh shower (intervention stage), devices in RTF and DUAL additionally displayed water use in real time, whereas devices in CON and HER permanently stayed in control mode. About halfway into the study, we started sending home energy reports to each subject in the HER or DUAL group; the first report was sent on 23 January 2017 and the second report on 8 February 2017, about two weeks later. We distinguish between intervention (IN) stage 1, in which real-time feedback is switched on but there were no reports yet, and intervention (IN) stage 2, which is the period beginning after the first home energy report was sent out.[21] In order to hold interaction with experimenters constant, subjects in CON and RTF groups received placebo emails at the exact same time the home energy reports were sent out, which simply asked them to fill out a mini-survey — the same that came along with the actual reports.

This staggered experimental design allows us to exploit both between- and within-subject variation to cleanly identify treatment effects of interest. The effect of real-time feedback in isolation is identified by the comparison between the RTF and CON groups in the (entire) intervention stage, or alternatively by the comparison between the pooled

[20]For the exact randomization protocol, see Appendix B.

[21]In practice, the distinction between IN stage 1 and 2 is not perfect, as a small fraction of subjects had yet to complete 10 showers when the first report was sent out. If anything, this generates measurement error in our treatment indicators and thus biases estimates toward zero.

RTF/DUAL group and the pooled CON/HER group in IN stage 1. The effect of home energy reports in isolation is identified by the comparison between the HER and CON groups in IN stage 2. The additional effect of home energy reports, when combined with real-time feedback is identified by the comparison between the DUAL and RTF groups in IN stage 2. Differences between the effects of home energy reports with and without real-time feedback are indicative of policy interaction effects, i.e. whether the two interventions are substitutes or complements.

3.6. Behavioral predictions

In order to derive behavioral predictions for each of our experimental groups, we first briefly discuss the channels through which each of the two interventions work. Recall from our theoretical framework that the energy conservation effect of a behavioral policy P is $\Delta e^P = f^{-1}(B^P) - f^{-1}(B^o)$, where B is the aggregate behavioral barrier parameter, f is the ratio of marginal benefits to marginal costs, and its inverse f^{-1} is strictly decreasing. Assuming that the function f remains unaffected — we evaluate this empirically in Section 7 —, the effectiveness of the three intervention regimes therefore depends on the degree to which they succeed in overcoming the aggregate barrier.

Real-time feedback visually displays live measurement of water use in the shower. This is likely to significantly reduce inattention problems, as users are constantly facing the smart meter display, and the previously abstract and elusive notion of resource use suddenly becomes salient and palpable through the steadily upward moving liter count. Of course, water volume information as such may affect behavior if individuals previously underestimated the amount of water they use. However, there is no additional information on energy use or carbon emissions, so there may still be severe knowledge gaps about the environmental relevance of showering remaining. Therefore, our leading interpretation is that real-time feedback mainly affects behavior by targeting limited attention. As the RTF condition in our experiment is essentially a replication of the intervention by Tiefenbeck et al. (2018), albeit more minimalistic and in a sample without monetary incentives, we also expect to find significant conservation effects.

Prediction 1. *Providing real-time feedback through the smart shower meter display in treatment RTF leads to a reduction in water and energy consumption in the shower.*

Home energy reports provide information on subjects' water use in the shower as well as additional information about energy use and CO_2 emissions. The second report also included a social comparison element, but it seems unlikely to us that this is very important, as Tiefenbeck et al. (2018) find no effect for including comparisons with the co-resident in a two-person household and social norms are even less likely to arise in our setting, with the peer being random and anonymous. In contrast to real-time feedback, the reports are not immediately salient while showering and therefore fail to target barriers like limited attention or self-control problems.[22] We therefore interpret home energy reports as information intervention and expect that they induce conservation behavior mainly by closing knowledge gaps about water- and energy-intensity (and implied environmental impacts) of showering.

Prediction 2. *Providing information through home energy reports in treatment HER leads to a reduction in water and energy consumption in the shower.*

Finally, the combined intervention should reduce all barriers that are either targeted by home energy reports or real-time feedback. In particular, both the problem of limited knowledge about environmental impacts as well as the problem of limited attention should be mitigated. This should therefore result in a larger conservation effect than can be achieved by either intervention in isolation.

Prediction 3a. *The combined intervention that adds home energy reports to real-time feedback (treatment DUAL) leads to a larger overall reduction in water and energy consumption in the shower than either intervention in isolation (treatments RTF or HER).*

The combined effect of home energy reports and real-time feedback could be smaller or larger than the sum of their individual effects. In general, returns to behavioral policy may be expected to decrease, as there is less room for further conservation efforts. However, the presence of multiple barriers can generate complementarities. Recall that the aggregate bias B is then the product of many individual barriers, so $B = b_1 \cdot b_2 \cdot ...b_K$. This implies that the effect of providing information through home energy reports could be mitigated by strong remaining barriers like limited attention or self-control problems. Analogously, the effect of channelling attention through real-time feedback could be mitigated by lack of awareness about energy use and carbon emissions due to warm water use in the shower. Combining both interventions could therefore have a conservation effects that is larger than the sum of the effects of each intervention in isolation.

Prediction 3b. *Home energy reports may lead to a larger reduction in water and energy consumption in the shower for subjects who receive real-time feedback (treatment DUAL) than for subjects who do not receive real-time feedback (treatment HER).*

This means that our two interventions would be complements, whereby improved attentional focus enables knowledge gains to induce a larger treatment effect and vice versa.

[22]In principle, it is possible that participants also become more attentive about resource use even without visual aid through the smart meter, as would be predicted by rational inattention models when updates in beliefs about environmental impacts are sufficiently large. This would be an instance of indirect redundancies between interventions. However, if there is such an effect, it may prove short-lived once reports fade out of memory and resolutions cool off (Allcott and Rogers, 2014; Schwartz and Loewenstein, 2017).

4. Data and descriptive statistics

In this section, we describe our experimental data and show some summary statistics about our sample and their baseline resource use behavior. We also perform randomization checks to affirm the validity of our assignment into experimental conditions. Furthermore, we offer some suggestive evidence of the presence of potentially severe information and attention barriers to resource conservation in our setting.

4.1. Measurement data on resource use behavior

The smart shower meters measured and stored data on, among others, water volume, average temperature, and average flow rate of every water extraction. The amount of energy used was then calculated based on volume and temperature data, using the standard engineering formula for heat energy.[23] Every subject had a shower meter installed for the whole duration of the study, starting from early-December 2016. At the end of the study, from late February to early March 2017, we retrieved the devices and read out the data manually. In this way, we were able to extract an initial data set of 21,469 showers by 327 participants. Unfortunately, no data could be obtained in 24 cases, either because the smart meter was defective or because subjects never used it, or because subjects simply disappeared without a trace (and their shower meters with them).

A number of data cleaning steps are performed before running the empirical analyses. We briefly describe the most important steps here, as a more detailed documentation can be found in Appendix C. First, we drop the very first data point of each participant, as they usually started with a test run to check if the device was working. Following Tiefenbeck et al. (2018), we further drop any water extraction with volume below 4.5 liters (in total $2,942$ extractions), as these are unlikely to be actual showers but rather minor extractions for other purposes such as cleaning. We further remove 37 extreme outlier points, defined as such by being more than 4.5 times the subject-specific interquartile range away from the closest quartile.[24] We further exclude 1 device with erratic data, 5 devices with fewer than 10 recorded extractions, as well as 3 devices with an abnormally large baseline consumption of on average 168 liters or more per shower, which is about 40 liters (1.5 standard deviations) away from the rest of the field. In 8 cases, the device's temperature sensor broke at some point, and we impute missing information with the average temperature of showers taken while the sensor was still intact. The final data set used for our empirical analyses includes $17,942$ showers by 318 participants.

[23]The formula for energy use of water heating is $Q \times m \times c_p \times \Delta T$, with heat energy Q, mass of water m, heat capacity c_p , and ΔT the difference between the measured water temperature and cold water temperature (assumed to be 12 degrees Celsius). Following Tiefenbeck et al. (2018), we also assume boiling efficiency losses of 35% and distribution losses of 24%.

[24]We are particularly strict in only excluding the most implausible data points here. Conventionally, 1.5 or 3 times the interquartile range (IQR) are used as criterion for outliers. For a normal distribution, 4.5 times the IQR away from the nearest quartile corresponds to 6.745 standard deviation away from the mean.

The shower meter stores the temporal order of showers, so we can easily classify each shower into baseline or intervention stage, as real-time feedback (in the RTF and DUAL groups) started from shower number 11 onwards. Assigning showers into intervention stage 1 (pre-reports) or stage 2 (post-reports) is slightly more tricky, as the device has no counter for global time. Fortunately, the smartphone app stores the date and time of when data about a shower was uploaded, which allows us to construct bounds for when a shower took place. We instructed subjects to use the smartphone app regularly starting from 11 January 2017, and send additional reminders before each home energy report was to be sent out. Using this timing information, we classify observations into pre-report showers (IN stage 1) or post-report showers (IN stage 2). If there are multiple showers within the range of uncertainty around report dates, we use the switching point implied by constant shower frequency. One complication is that we do not know the timing of showers by the subjects who did not manage to upload any data to the app. Therefore, we impute the timing of showers for these non-uploaders based on the assumption that timing of home energy reports follows the same distribution for uploaders and non-uploaders. To operationalize this, we use timing information from uploaders to estimate the probability that a shower took place after a home energy report, and then assign the implied post-report probabilities to showers of non-uploaders. Figure A2 in Appendix A plots the estimated CDFs.[25]

4.2. Questionnaire data

To supplement our behavioral data on resource use in the shower, we administered several questionnaires. In the baseline survey, we collected information on individual characteristics (i.e. age, gender, etc.), self-assessed resource use in the shower, shower comfort (i.e. how much they enjoy showering), environmental attitudes and beliefs, as well as a number of personality attributes (i.e. Big Five, patience, etc). In the post-intervention survey, we again collected self-reported data on self-assessed resource use, shower comfort, and environmental attitudes. Furthermore, we administered mini-surveys with each home energy report, which also asked subjects to estimate their resource use in the shower.

We mainly make use of information on self-assessed resource use, shower comfort, and environmental attitudes, and how they change in response to our interventions. Environmental attitude is elicited using four items about pro-environmental behavior and identity, e.g. "I do what is right for the environment, even when it costs more money or takes more time".[26] Shower comfort is elicited using five items on how much subjects

[25] For details of the imputation procedure, see Appendix D.

[26] The other items are "Environmental friendliness is part of my personal identity", "How often do you try to conserve water?", and "How often do you try to conserve energy?". We also include a set of questions adapted from Nolan et al. (2008), but only in the baseline questionnaire.

Table 1: Descriptive statistics – baseline showers

	Mean	Std. dev.	10th pctile	Median	90th pctile	Obs.
Energy use [kWh]	2.21	1.91	0.43	1.71	4.58	2489
Volume [liter]	37.82	30.45	9.20	29.60	76.00	2489
Duration [min]	7.00	5.01	1.96	5.83	13.01	2489
Temperature [Celsius]	36.16	5.22	32.00	37.00	40.00	2463
Flow rate [l/min]	5.71	2.45	2.80	5.40	9.10	2489

Includes only showers taken in the baseline stage, i.e. first 10 showers and before home energy reports were sent out. For temperature statistics, devices with broken temperature sensors are excluded. Duration is net of any breaks and calculated by dividing water volume by flow rate.

enjoy showering, e.g. "I find it relaxing to take a shower".[27] We create indices for shower comfort and environmental attitude, respectively, by taking the simple average of the individual's responses to the relevant items (rated on a 4- or 5-point Likert scale) and then normalizing to mean 0 and standard deviation 1. For self-assessments, we asked participants to estimate how many liters of water they typically use when taking a shower. These estimates can then be directly compared to their actual water use as measured by the smart meter. Note that we refrained from eliciting subjects' beliefs about energy use and environmental impacts, because we did not want to raise awareness about these issues before the intervention and potentially undermine the home energy report treatments.

4.3. Sample characteristics and baseline behavior

All participants in the field experiment are students at universities in Bonn or Cologne living in one-person dorm apartments, so our sample is rather homogeneous. From the 318 participants represented in our main data set, 203 lived in a dorm in Bonn and 115 lived in a dorm in Cologne at the time of our study. The share of females is 61 percent.[28] Average age was 23.8 years (median 23 years), with students from all stages of their studies being represented in our sample. About 34 percent are non-German students, reflecting the over-representation of international students in dorms. The distribution of majors is as follows: 38 percent mathematics or natural sciences, 31 percent social sciences, 16 percent business/economics, 10 percent agricultural/environmental sciences, and 5 percent arts/performance.

Using the nine showers (the first being excluded) in the baseline stage, where only the current water temperature was displayed, we can measure baseline resource use behavior of each subject. Table 1 presents descriptive statistics about baseline energy and water use per shower, as well as shower duration (net of breaks), water temperature, and flow

[27] The other items are "I like showering", "For me, taking a shower is just a means to an end", "I like to let my mind wander when I shower", and "I try to shower as quickly as possible".

[28] In 2016/17, the share of female students was 55% at the University of Bonn and 60% at the University of Cologne, suggesting that there was no substantial gender-based selection into our study.

Table 2: Randomization checks

	Panel A. Baseline averages by individual					Panel B.
	Energy use [kWh]	Volume [liter]	Duration [min]	Temperature [Celsius]	Flow rate [l/min]	Number of showers
HER group	-0.066	-1.901	0.181	0.959	-0.435	3.393
	(0.220)	(3.468)	(0.548)	(0.608)	(0.320)	(5.226)
RTF group	-0.111	-1.253	0.284	0.086	-0.124	-2.312
	(0.215)	(3.427)	(0.597)	(0.595)	(0.370)	(5.183)
DUAL group	-0.057	-0.910	0.213	0.320	-0.165	3.224
	(0.226)	(3.575)	(0.581)	(0.560)	(0.358)	(5.861)
Constant	2.237	38.316	6.797	35.681	5.832	55.312
	(0.163)	(2.539)	(0.411)	(0.447)	(0.240)	(3.698)
Observations	316	316	316	314	316	318
R-squared	0.001	0.001	0.001	0.011	0.005	0.005
F-test: p-value	0.966	0.958	0.969	0.356	0.571	0.669

Robust standard errors in parentheses. The omitted category is the CON group. For two participants, the device was not able to record information on baseline showers, but we could extract valid data on showers in later stages; hence the number of observations is only 316 in most columns. In addition, two participants with initially defective temperature sensors are excluded in column 4.

rate. Shower duration is calculated from dividing water volume by average flow rate. On average, showers in the baseline stage feature 7 minutes of water flow, using in total 37.82 liters of water. On average, water is heated up to a temperature of 36.16 degrees Celsius, resulting in energy use of 2.21 kWh per shower. There is substantial variation across showers, as becomes obvious when looking at standard deviations and different quantiles of the distributions. Water and energy consumption follows a right-skewed distribution, thus the median energy use per shower (1.71 kWh) is substantially lower than the mean. The average flow rate of 5.74 liters per minute is relatively low, likely due to dorm infrastructure not being up to modern standards; flow rates of 10-12 liters per minute are more typical for German households.

4.4. Randomization checks

Our experimental identification strategy assumes that randomization produced treatment groups that are comparable with regard to observable and unobservable subject characteristics. Although it is naturally impossible to test the latter, we can check balance on observable baseline characteristics. Panel A of Table 2 shows results from regressing various measures of subjects' baseline behavior on assigned treatment groups. The differences between groups are very small and treatment assignment is insignificant for predicting any of the behavioral measures, so randomization seems to have worked well. We also check for balance along background characteristics and survey responses (see Appendix 8), and again find that treatment assignment is statistically insignificant. Importantly, environmental attitude and shower comfort are comparable across groups.

4.5. Number of showers

On average, we observe 56.8 showers per individual over the roughly 12 weeks of our study, which corresponds to a frequency of about two showers every three days. However, the net frequency (i.e. adjusting for absences) might be closer to one shower per day, as our study period included a two week Christmas break, and individuals might also leave the city during weekends. Unfortunately, we have no reliable information on absence times. In Panel B of Table 2, we check whether the number of showers per individual differs across experimental conditions. Consistent with Tiefenbeck et al. (2018), we find that treatments have no effect on the number of showers ($p = 0.669$). Hence, our interventions do not seem to induce adjustments along the extensive margin, and we do not need to worry about subjects compensating shorter showers with more showers, or about them compromising on basic hygiene needs. This means that we can make use of the full panel structure of our data and analyze (intensive-margin) water and energy conservation effects on the level of individual showers.

4.6. Presence of information and attention barriers

Before moving on to the analysis of our experimental interventions, we provide some descriptive evidence that imperfect information and limited attention are likely to be significant behavioral barriers to resource conservation in our setting.

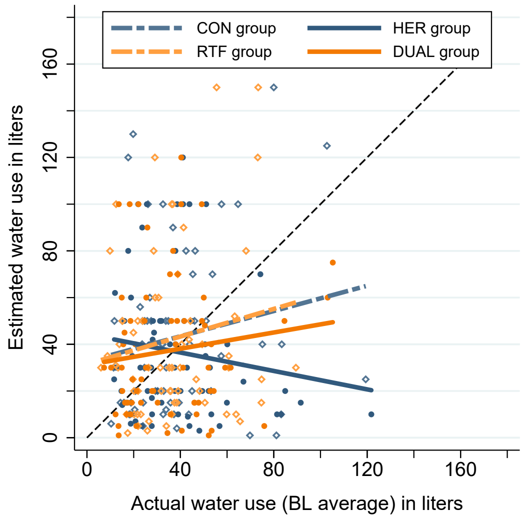
First, we make use of the pre-intervention questionnaire and compare subject's self-assessments of their water use per shower to their actual baseline water use as measured by the smart meter. Figure 5 shows that subjects' estimates are all over the place, and we cannot reject the null hypothesis that estimated and measured water use are in fact uncorrelated (Pearson's $\rho = 0.08$, $p = 0.1825$), which demonstrates that subjects were not well informed about their own behavioral outcomes prior to any intervention.[29] Interestingly, however, the mean estimate across all subjects (39.8 liters) is close to the actual mean water use per shower in the baseline stage (37.8 liters). This is reminiscent of a "wisdom of crowds" phenomenon and suggests that, on average, our interventions should not work through debiasing beliefs about water use.

Furthermore, it is safe to assume that subjects are unaware of how much energy is consumed (and hence CO_2 emitted) in a typical shower, as Attari et al. (2010) show that consumers are highly prone to underestimating the amount of energy required for heating up water (e.g. water boilers, dishwashers). We did not elicit beliefs about energy intensity or carbon emissions in the original experimental sample, to avoid the risk of undermining our home energy report treatments.

Although anecdotally compelling, finding direct evidence for inattention about re-

[29] We excluded 35 subjects who responded to the baseline survey more than 2 weeks after we distributed shower meters, as they have likely reached the intervention stage by then. We also exclude 3 extreme outliers with estimates above 200 liters. The corresponding regression results are presented in Appendix A Table A5.

Figure 5: Pre-intervention awareness about water use per shower



Notes. This figure compares estimated water use from the baseline survey with actual water use in the baseline stage (showers 2 to 10), excluding late survey responders. 3 subjects with extreme estimates (above 200 liters) are excluded. Point clouds consist of individual observations (hollow diamonds for CON and RTF, solid circles for HER and DUAL) and lines represent separate regression fits for each treatment group. The dashed line starting from the origin is the 45 degree line.

source consumption costs in the shower is tricky. The closest we have is a baseline survey item asking subjects how much they agree with the statement "I like to let my mind wander when I shower" on a five-point Likert scale. 59% of our sample states that they agree or strongly agree to the statement (34% agree, 25% strongly agree), whereas only 18% of subjects disagree or strongly disagree (13% disagree, 5% strongly disagree), indicating that lack of focus while showering is prevalent. We further find that response to this item is significantly correlated with baseline energy use in the shower (Pearson's $\rho = 0.17$, $p = 0.003$). In fact, it is the single most predictive item for energy and water use in our entire baseline survey.

5. Estimation approach

We now describe our strategy for estimating the effects of our interventions on resource use in the shower. The empirical results will be presented in the following section.

5.1. Basic estimation strategy

To formally estimate the effects of different intervention regimes, we exploit the staggered introduction of real-time feedback and home energy reports in the experimental design, which gives us a double-layered difference-in-differences setup. The differential changes in consumption behavior across conditions from baseline stage to intervention stage 1 identify the causal effect of real-time feedback (RTF/DUAL versus CON/HER), and the additional changes from intervention stage 1 to stage 2 identify the the causal effect of home energy reports, both in isolation (HER versus CON) and in conjunction with real-time feedback (DUAL versus RTF).

For estimating the effect of real-time feedback in isolation, the most straightforward and easy-to-interpret approach is to simply compare subjects in the RTF and CON groups over the entire experimental period, as these subjects never received home energy reports in any form. We do this by estimating the equation

$$y_{it} = \alpha_i + \beta_0 IN_{it} + \beta_1 IN_{it} \times T_i^R + \varepsilon_{it}\,, \tag{11}$$

where the outcome variable y_{it} is energy use (water use) by individual i for shower number t, α_i is the individual fixed effect, IN_{it} is an indicator that takes the value 1 if observation it falls into the intervention stage (i.e. $t > 10$), and T_i^R is an indicator for being assigned to treatment group RTF. The coefficient of interest is β_1, which gives us a clean estimate for the average treatment effect of real-time feedback over the entire three months of the study. In this specification, we do not have to deal with issues relating to non-compliance and timing of reports, though it comes at the cost of disregarding half of the sample in intervention stage 1.

To make use of the full sample when estimating the effect of real-time feedback, we can compare differential changes in consumption behavior from baseline stage to intervention stage 1 for the pooled RTF/DUAL group versus the pooled CON/HER group, because real-time feedback had started but there were no home energy reports yet. But in intervention stage 2, when home energy reports start flying in, we need to split up the pooled groups again, so the regression equation is

$$\begin{aligned} y_{it} \;=\; \alpha_i \;+\; IN_{it} \times \Big(\beta_0 \;+\; \beta_1 T_i^{R/D}\Big) \\ +\; IN_{it}^{s2} \times \Big(\gamma_0 \;+\; \gamma_1 T_i^{R/D} \;+\; \gamma_2 T_i^H \;+\; \gamma_3 T_i^D\Big) \;+\; \varepsilon_{it}\,. \end{aligned} \tag{12}$$

IN_{it} is again the indicator for the entire intervention stage, and on top of that, IN_{it}^{s2} is an indicator for showers that fall into intervention stage 2 (post-report). As IN_{it} remains switched on for the entire intervention period, all terms that are multiplied by IN_{it}^{s2} need to be interpreted as incremental changes from intervention stage 1 to intervention stage 2, compared to the baseline period. $T_i^{R/D}$, T_i^D and T_i^H are treatment group indicators, where superscript R/D denotes the combined groups RTF and DUAL, superscript D denotes the DUAL group, and superscript H denotes the HER group.

Equation (12) incidentally also includes an estimate for the effect of home energy reports, but one concern here is that the differences between RTF and DUAL or between CON and HER in the first intervention stage are not captured. Although the pooled groups in intervention stage 1 should behave the same before reports are sent out, random differences are likely to exist in finite samples, and these would propagate to the estimates of γ_2 and γ_3. For estimating the effects of home energy reports we therefore prefer the more flexible model in which treatment groups are considered separately from the start:

$$\begin{aligned} y_{it} \;=\; \alpha_i \;+\; IN_{it} \times \Big(\beta_0 \;+\; \beta_1 T_i^{R/D} \;+\; \beta_2 T_i^{H} \;+\; \beta_3 T_i^{D}\Big) \\ +\; IN_{it}^{s2} \times \Big(\gamma_0 \;+\; \gamma_1 T_i^{R/D} \;+\; \gamma_2 T_i^{H} \;+\; \gamma_3 T_i^{D}\Big) \;+\; \varepsilon_{it}\,. \end{aligned} \tag{13}$$

Given the model formulation, we can interpret β_1 as treatment effect of real-time feedback on energy (water) use per shower in the first half of the study, while γ_1 is the change in treatment effect in the second half. γ_2 is the treatment effect of home energy reports in isolation, and γ_3 is the additional effect of adding home energy reports to real-time feedback. The relevant comparisons of interests are between HER and CON on the one hand — for the effect of reports without real-time feedback — and between DUAL and RTF on the other hand — for the marginal effect of adding reports to reinforce the already existing real-time feedback.

5.2. Estimating treatment effects on the treated

A major complication in estimating the effect of home energy reports is that 28% of subjects did not succeed in uploading any data to the Amphiro smartphone app before we sent out the reports, mostly due to technical problems (e.g., Bluetooth connection failure).[30] For these "non-uploaders", we were unable to provide informative home energy reports. As the emails were generated automatically, non-uploaders in HER and DUAL groups received report templates with blanks instead of actual statistical figures about their resource use and environmental impacts. Effectively, this leads to imperfect treatment take-up of home energy reports, although being less the result of deliberate non-compliance than unfortunate circumstances. For participants in the CON and RTF groups it is inconsequential whether they successfully uploaded data, as we only asked them to do so in order to hold constant the general experimental procedure for all participants.

One possible approach to estimate treatment effects under imperfect treatment take-up is to run an intention-to-treat (ITT) analysis, which ignores that some participants did not actually receive informative home energy reports and simply uses treatment assignment to estimate treatment effects. However, this is not very appealing in our context, as failure of information provision due to technical problems is in principle an avoidable problem. The policy-relevant treatment effect is the effect of delivering informative home energy reports. Therefore, our preferred approach is to estimate the treatment effect on the treated (TOT), i.e. on subjects who managed to upload data through the app and thus received proper home energy reports with all the information we wanted to convey.

One way to estimate the TOT is to simply compare resource use of uploaders in HER

[30] Out of the 90 non-uploaders in our estimation sample, 63 have explicitly contacted us for technical problems encountered during their upload attempts.

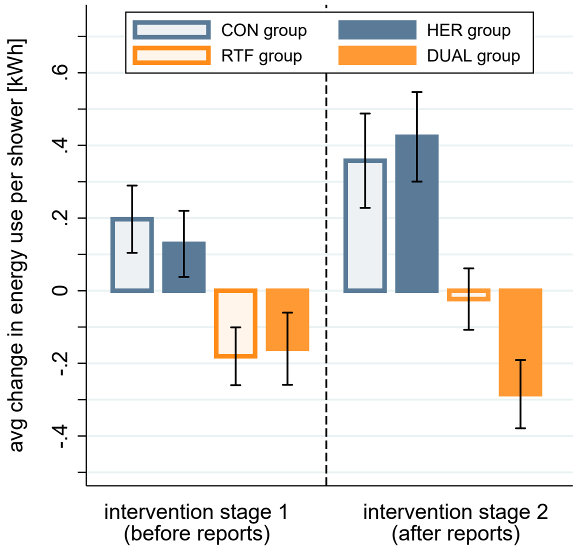
and DUAL groups with resource use in the no-report CON and RTF groups. The usual concern at this point would be that treatment-take up is not random. Fortunately, our setting limits potential endogeneity concerns, for three reasons. Firstly, we include individual fixed effects, so our estimates would still be unbiased if differences between uploaders and non-uploaders do not interact with the treatment. Secondly, subjects only knew that they should use the smartphone app to upload data, but we did not announce that we would use this data to construct home energy reports. Thirdly, the main cause for non-compliance is not the lack of willingness to use the smartphone app, but unexpected technical failure, which is unlikely to be selected on the trend. Hence, the first way in which we assess the effect of home energy reports on the treated is by estimating equation (13), including only uploaders in the HER and DUAL groups. To alleviate the most blatant endogeneity issue, we also exclude non-uploaders in the CON and RTF groups, who did not report any technical problems.

A second way to estimate the TOT is by using random treatment assignment as instrument for actual take-up. This can be shown to identify the so-called local average treatment effect (LATE), i.e. the average treatment effect for the subpopulation of compliers, in our case the uploaders, even under endogenous treatment take-up (Imbens and Angrist, 1994).[31] Compared to the "uploaders-only"-approach, the instrumental variables approach is always consistent, but potentially inefficient. We will report the results from both TOT-approaches, but the estimates are very similar, suggesting that endogeneity is not a large issue in our setting.

To validate that home energy report take-up is not influenced by assignment into treatment group, we can check for differential levels of compliance. The fraction of compliers are 76.6% in CON, 74.4% in HER, 68.4% in RTF, and 67.9% in DUAL. The difference across treatment groups is statistically insignificant ($p = 0.594$). It seems that, at least conditional on (not) receiving real-time feedback, uploading of data is orthogonal to assignment into a group that receives the home energy report. Furthermore, we can compare the subpopulations of compliers and non-compliers along observable characteristics. For one, this is indicative of how relevant the endogeneity issue is. It also gives us a sense of how representative the estimated TOT is for our experimental sample. Table A2 in Appendix A compares shower behavior and subject characteristics. There is a slight tendency for uploaders to take shorter showers at higher flow rate, which could be due to the fact that the shower meters tend to work more reliably at higher flow rates. There is also a slightly lower share of female and international students in the group of uploaders. Importantly, however, energy and water use per shower of uploaders and non-uploaders were not significantly different prior to administering the home energy reports.

[31] This identification result holds under the condition that there are no "defiers", subjects who always do the opposite of what they are prescribed. This monotonicity condition holds by design in our study, because we control the eligibility of home energy report treatment, so any participant in the sample can be classified either as complier or as never taker in the LATE framework.

Figure 6: Descriptive evidence on energy conservation effects



Notes. The bars represent changes in average energy use per shower compared to the baseline period. The error whiskers show standard errors of the mean. Non-uploaders in HER and DUAL as well as non-uploaders without technical problems in CON and RTF are excluded.

6. Empirical results

In this section, we present our empirical results on the effect of our interventions on resource conservation in the shower. We focus on energy and water use per shower as outcome variables and report average treatment effects for the main results. Additionally, we investigate treatment effect dynamics as well as heterogeneous effects along the dimensions of baseline consumption and environmental attitude.

6.1. Main results

Before proceeding to the formal estimates, we present descriptive evidence on the conservation effects of our interventions. Figure 6 shows subjects' average changes in energy consumption per shower in intervention stage 1 (pre-report) and intervention stage 2 (post-report) compared to the baseline period. The difference-in-differences across treatment groups then corresponds to the average treatment effect. In order to show the TOT for home energy reports, we use the uploaders-only approach of excluding non-compliers in HER and DUAL as well as non-compliers without technical problems in CON and RTF. The graph essentially summarizes our main results in eight bars.

The four bars to the left of the dashed vertical line represent the change in energy use per shower in intervention stage 1 compared to the baseline stage. We can see that relative to subjects in the CON and HER groups, subjects in the RTF and DUAL groups reduced their energy consumption drastically, by almost 0.4 kWh per shower. Recall that there were no home energy reports yet at this point. The four bars to the right of the dashed vertical line represent the change in energy use per shower from baseline stage

Table 3: Effect of real-time feedback and ITT estimates

	only RTF & CON		*Intention to treat*	
	(1) Energy [kWh]	(2) Water [liter]	(3) Energy [kWh]	(4) Water [liter]
Intervention	0.283***	4.453***	0.179***	2.915***
	(0.104)	(1.597)	(0.067)	(1.049)
Intervention × RTF/DUAL	-0.397***	-6.346***	-0.309***	-4.628***
	(0.125)	(1.926)	(0.087)	(1.387)
IN stage 2			0.187*	3.157**
			(0.097)	(1.441)
IN stage 2 × RTF/DUAL			-0.071	-1.745
			(0.118)	(1.854)
IN stage 2 × HER			0.038	0.147
			(0.130)	(2.006)
IN stage 2 × DUAL			-0.133	-2.302
			(0.093)	(1.555)
Individual fixed effects	*yes*	*yes*	*yes*	*yes*
Clusters	156	156	318	318
Observations	8446	8446	17942	17942
R^2	0.379	0.375	0.403	0.404

Standard errors in parentheses are clustered at the individual level. Columns (1) and (2) only include individuals in the RTF or CON group.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to intervention stage 2, after home energy reports were sent out. The first observation is that average energy use in the control group further increased, which could be driven by weather effects or by impending exams leaving students stressed and in need for a long, warm shower.[32] The second observation is that the RTF group moved up in parallel with the CON group, hence the effect of real-time feedback in isolation remains nearly constant. The third observation is that providing home energy reports in isolation does not seem to result in effective behavioral change: energy consumption of subjects in the HER group also follows the CON group in close synchronization. In light of this, the fourth and final observation is particularly striking: home energy reports are highly effective when combined with real-time feedback. In fact, subjects in the DUAL group are the only ones to defy the general upward trend and reduce their consumption considerably compared to subjects in the RTF group.

Our formal empirical results come from implementing the estimation strategies outlined in Section 5. In general, they confirm the patterns in Figure 6. We first focus on estimating the effect of real-time feedback in isolation, before turning to the effect of home energy reports, for which we need to account for imperfect compliance.

The simplest and cleanest way to estimate the effect of real-time feedback is to only

[32] While the baseline phase fell mainly into an unusually warm and dry December, the main intervention months of January and February saw much higher precipitation. Exam periods at the universities began in mid-February.

compare subjects in the RTF and CON groups over the entire intervention period, by estimating equation (11). Table 3 columns 1 and 2 show that real-time feedback in isolation reduces resource use by 0.40 kWh of energy and 6.3 liters of water per shower compared to the CON group, which corresponds to about 17-18% of baseline use. This effect size is consistent with previous studies using the same smart shower meters. Tiefenbeck et al. (2019) report a conservation effect of 0.2 kWh (11%) per shower in a sample of Swiss hotel guests without financial incentives, whereas Tiefenbeck et al. (2018) find 0.59 kWh (22%) lower energy use in a sample of Swiss households, which did have financial incentives for resource conservation.[33]

In addition, columns 3 and 4 present the results from estimating equation (12) on the full sample, using treatment assignment as the independent variable. Our estimates show that subjects in RTF and DUAL conserved about 0.31 kWh of energy and 4.6 liters of water per shower in intervention stage 1, compared to subjects in CON and HER. This is slightly lower than the estimates in columns 1 and 2, partly due to the inclusion of the DUAL and HER groups, and because the effect increases in intervention stage 2, albeit statistically indistinguishable from zero. With the advent of home energy reports in intervention stage 2, we split the pairs up into the four separate groups again, which incidentally gives us ITT estimates for the effect of home energy reports; but as discussed earlier, this misses the policy-relevant effect of actually receiving information through home energy reports. That said, the ITT estimates for the effect of home energy reports are neither significant for HER nor DUAL. Still, the point estimates for the DUAL group look quantitatively relevant and already hint at something potentially going on.

Result 1. *Real-time feedback through the smart meter display leads to a reduction in energy (water) consumption by about 0.3-0.4 kWh (4.5-6.3 liters) or 14-18% per shower.*

To estimate the effect of (sending information through) home energy reports on conservation behavior, we move on to the TOT analyses described in Section 5. Table 4 columns 1 and 2 show the estimates obtained by using the uploader-only approach, in which we estimate regression equation (13) on the restricted sample that excludes non-uploaders in HER and DUAL, as well as non-uploaders in RTF and CON without technical issues. Columns 3 and 4 display the LATE estimates, for which we use random treatment assignment to the HER or DUAL group as instruments for actually uploading data and receiving informative home energy reports. While the LATE approach is consistent even under endogenous treatment take-up, the uploaders-only approach is potentially more efficient.

Both approaches produce nearly identical results, suggesting that endogeneity of treat-

[33]The Swiss household sample of Tiefenbeck et al. (2018) also had a higher average baseline energy use of 2.66 kWh per shower, and they estimate that conservation decreases by 0.031 kWh for each 0.1 kWh decrease in baseline use. Projected to our sample (baseline 2.21 kWh), the predicted conservation effect would be 0.45 kWh per shower. Also, the smart meter display in Tiefenbeck et al. (2018, 2019) contained additional information on energy use as well as graphical feedback in the form of a polar bear sitting on a slowly melting ice floe.

Table 4: Treatment on the treated (TOT) estimates

	Uploaders-only		*LATE*	
	(1) Energy [kWh]	(2) Water [liter]	(3) Energy [kWh]	(4) Water [liter]
Intervention	0.179	2.628	0.172*	2.533
	(0.111)	(1.702)	(0.102)	(1.565)
Intervention × RTF/DUAL	-0.388***	-5.753***	-0.365***	-5.481***
	(0.134)	(2.124)	(0.125)	(1.981)
Intervention × HER	0.027	0.837	0.016	0.733
	(0.154)	(2.415)	(0.134)	(2.082)
Intervention × DUAL	0.035	0.576	0.109	2.159
	(0.113)	(1.860)	(0.107)	(1.751)
IN stage 2	0.150	2.770*	0.189*	3.273**
	(0.093)	(1.422)	(0.098)	(1.460)
IN stage 2 × RTF/DUAL	-0.021	-1.142	-0.053	-1.463
	(0.118)	(1.913)	(0.120)	(1.908)
IN stage 2 × HER	0.090	0.714	0.042	-0.084
	(0.137)	(2.168)	(0.162)	(2.510)
IN stage 2 × DUAL	-0.222**	-3.702**	-0.215*	-3.836*
	(0.100)	(1.756)	(0.116)	(2.037)
Individual fixed effects	*yes*	*yes*	*yes*	*yes*
Clusters	261	261	318	318
Observations	14712	14712	17942	17942
R^2	0.413	0.415	0.004	0.004

Standard errors in parentheses are clustered at the individual level. In columns (1) and (2), we exclude all non-uploaders in HER and DUAL as well as all non-uploaders in RTF and CON who did not report a technical problem. In columns (3) and (4), we use treatment assignment to HER and DUAL, respectively, interacted with the IN stage 2 indicator as instrument for receiving informative home energy reports. The reported R^2 in Columns (3) and (4) is the within R^2.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

ment take-up is not a major issue. The conservation effect of real-time feedback in isolation is also similar to the ones reported in Table 3. The results show that home energy reports had no significant effect in the HER group, and the point estimates go into the wrong direction. Even in the less precise LATE specification, we can rule out energy use reductions of more than 7.5% per shower with 90% confidence. Furthermore, we can reject the hypothesis that home energy reports in isolation were as effective as real-time feedback in isolation ($p < 0.003$ in all specifications).

Result 2. *Home energy reports in isolation do not induce any significant reduction in energy and water consumption per shower.*

In stark contrast, subjects in the DUAL group further reduced energy use by around 0.22 kWh (water use by around 3.8 liters) per shower in intervention stage 2, which corresponds to another 10 percentage points reduction from baseline consumption. Put differently, home energy reports boosted the effectiveness of real-time feedback by more than 50%. The difference between energy conservation effects in the DUAL group and the HER group is weakly significant in the uploaders-only specification ($p = 0.067$). This contrast between home energy reports with and without real-time feedback is all the more remarkable given that subjects in DUAL had already cut their energy consumption per shower significantly and thus had less room for further behavioral adjustments. This is exactly the ceiling effect we described earlier in the theoretical framework.

Result 3a. *Combining real-time feedback with home energy reports further reduces energy (water) use by around* 2.2 *kWh (3.8 liters) per shower and thus boosts the conservation effect of real-time feedback in isolation by more than* 50%.

Result 3b. *Home energy reports lead to a larger reduction in resource consumption in the shower for subjects who receive real-time feedback (DUAL group) than for subjects who do not receive real-time feedback (HER group).*

Overall, there seems to be strong complementarity between real-time feedback and home energy reports. This is consistent with our theoretical framework, which shows that in the presence of multiple barriers to resource conservation, behavioral interventions may need to overcome all significant barriers simultaneously in order to unfold their full effect. While home energy reports provide information about resource use and associated environmental impacts, the lack of salience in resource consumption is likely to hinder conservation efforts. Real-time feedback through smart meters could thus turn environmental considerations into action by channelling attention at the moment of decision-making. We will analyze the underlying mechanisms more closely in Section 7.

6.2. Treatment effect dynamics

We now investigate whether the conservation effects of real-time feedback and home energy reports remain stable over the three-month period of our study. The previous subsection already documents that the effect of real-time feedback does not drop from the first to the second intervention stage. Therefore, we now focus on the 5-6 week period of IN stage 2. To estimate dynamic effects, we extend the empirical model for average treatment effects i.e. equation (13) by interacting with a time variable Z_i:

$$\begin{aligned} y_{it} = \alpha_i &+ IN_{it} \times \left(\beta_0 + \beta_1 T_i^{R/D} + \beta_2 T_i^H + \beta_3 T_i^D\right) \\ &+ IN_{it}^{s2} \times \left(\gamma_0 + \gamma_1 T_i^{R/D} + \gamma_2 T_i^H + \gamma_3 T_i^D\right) \\ &+ IN_{it}^{s2} \times Z_i \times \left(\delta_0 + \delta_1 T_i^{R/D} + \delta_2 T_i^H + \delta_3 T_i^D\right) + \varepsilon_{it} \,. \end{aligned} \tag{14}$$

We explore two variants of Z_i. In the first variant, we look additionally at energy use per shower after the second home energy report that was sent about two weeks after the first report. In the second variant, we interact each treatment group indicator with a linear time trend, so the δ coefficients can be interpreted as weekly depreciation (or appreciation) rate of energy conservation effects by intervention regime.

Table 5: Treatment effect dynamics

	$Z_i = \mathbb{I}\{post\ 2nd\ report\}$		$Z_i = \#\ weeks\ after\ 1st\ report$	
	(1) Uploaders	(2) LATE	(3) Uploaders	(4) LATE
IN stage 2	0.139	0.176	0.065	0.109
	(0.103)	(0.110)	(0.124)	(0.127)
IN stage 2 × RTF/DUAL	-0.027	-0.053	0.047	0.019
	(0.128)	(0.134)	(0.156)	(0.159)
IN stage 2 × HER	0.092	0.048	0.198	0.174
	(0.148)	(0.169)	(0.181)	(0.202)
IN stage 2 × DUAL	-0.068	-0.041	0.030	0.075
	(0.123)	(0.135)	(0.166)	(0.177)
IN stage 2 $\times Z_i$	0.019	0.022	0.032	0.029
	(0.093)	(0.090)	(0.027)	(0.026)
IN stage 2 × RTF/DUAL $\times Z_i$	0.012	0.000	-0.026	-0.026
	(0.123)	(0.119)	(0.037)	(0.035)
IN stage 2 × HER $\times Z_i$	-0.002	-0.010	-0.041	-0.051
	(0.126)	(0.136)	(0.042)	(0.047)
IN stage 2 × DUAL $\times Z_i$	-0.279	-0.316	-0.099	-0.114*
	(0.209)	(0.215)	(0.064)	(0.067)
Individual fixed effects	*yes*	*yes*	*yes*	*yes*
Clusters	261	318	261	318
Observations	14712	17942	14712	17942
R^2	0.413	0.005	0.413	0.005

Standard errors in parentheses are clustered at the individual level. The results are obtained by estimating equation (14). The full table with all the coefficients is presented in Appendix A Table A3. In columns (1) and (3), we exclude all non-uploaders in HER and DUAL, as well as all non-uploaders in RTF and CON who did not report a technical problem. In columns (2) and (4), we use treatment assignment to HER and DUAL, respectively, interacted with the IN stage 2 indicator as instrument for receiving informative home energy reports. The reported R^2 in Columns (2) and (4) is the within R^2.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5 shows that the effect of home energy reports in the DUAL group seems to gradually unfold over time. In fact, subjects do not reduce their energy use significantly in the first two weeks of intervention stage 2, but the average conservation effect is driven largely by lower energy consumption in the last 3-4 weeks of the study, after the second reports were sent out. The conservation effect in the DUAL group *increases* by around 0.1 kWh per shower every week. This time trend is not estimated very precisely, and hovers around the 10% significance level. There are several potential explanations for a pattern of increasing behavioral responses over time. For one, subjects may have skimmed through the email reports initially and only looked at it more carefully later, or it may have required some experimentation to discover strategies for further reducing energy use. The additional peer comparison in the second report may also have played a role. Importantly, the results seem inconsistent with pure Hawthorne effects or short-lived attention boosts, as these would predict an "action-and-backsliding" pattern (Allcott and Rogers, 2014; Schwartz and Loewenstein, 2017). Home energy reports in isolation (HER group) do not appear to exhibit strong dynamic patterns; their effect is identical before and after the second report. As in previous Amphiro shower meter studies, the effect of real-time feedback in isolation appears to stay constant over time, showing no signs of weakening within the 3 months of our study.

6.3. Heterogeneous treatment effects

Particular subgroups of individuals may have responded more strongly to our interventions than others. Previous studies often find that households or individuals with high baseline consumption tend to respond more strongly to policy interventions targeted at their conservation behavior (e.g. Allcott 2011*b*; Ferraro and Price 2013; Andor et al. 2017; Tiefenbeck et al. 2018). For example, Allcott (2011*b*) reports that Opower home energy reports achieved virtually no savings for households in the bottom decile of baseline energy use, whereas the treatment effect for top-decile users is 6.3% savings. Tiefenbeck et al. (2018) estimate an additional conservation effect of 0.31 kWh for a 1 kWh increase in baseline energy use per shower. Policy makers concerned about cost-effectiveness can therefore purposefully target high-baseline users. Strong pro-environmental motivation is also frequently associated with higher conservation responses to feedback interventions (e.g. Abrahamse et al. 2005; Costa and Kahn 2013; Tiefenbeck et al. 2018). This is of particular relevance to our setting, in which environmental concerns may be a critical lever to bridge the motivational void left by the absence of monetary incentives.

We focus on these two dimensions here, although it should be noted that our study is not powered to precisely detect heterogeneous treatment effects. To estimate heterogeneity, we extend the basic statistical model in equation (13) with interactions terms:

$$\begin{aligned} y_{it} = \alpha_i &+ IN_{it} \times \left(\beta_0 + \beta_1 T_i^{R/D} + \beta_2 T_i^H + \beta_3 T_i^D\right) \\ &+ IN_{it} \times X_i \times \left(\lambda_0 + \lambda_1 T_i^{R/D} + \lambda_2 T_i^H + \lambda_3 T_i^D\right) \\ &+ IN_{it}^{s2} \times \left(\gamma_0 + \gamma_1 T_i^{R/D} + \gamma_2 T_i^H + \gamma_3 T_i^D\right) \\ &+ IN_{it}^{s2} \times X_i \times \left(\mu_0 + \mu_1 T_i^{R/D} + \mu_2 T_i^H + \mu_3 T_i^D\right) + \varepsilon_{it} \end{aligned} \tag{15}$$

where variable X_i is either the subject's average baseline consumption per shower or her environmental attitude index. As measure of baseline consumption, we use a subject's average energy use in the 9 baseline showers (the first shower is excluded), recentered around the sample mean (2.21 kWh) so that intercept terms can be interpreted as effects at the mean. As proxy for environmental attitude, we use the standardized index con-

Table 6: Treatment effect heterogeneity

	X_i : *baseline energy use*		X_i : *envir. attitude*	
	(1) continuous	(2) $\mathbb{I}\{> median\}$	(3) continuous	(4) $\mathbb{I}\{> median\}$
Intervention × RTF/DUAL	-0.403***	-0.254***	-0.392***	-0.324
	(0.127)	(0.096)	(0.134)	(0.222)
IN stage 2 × RTF/DUAL	-0.014	0.171*	-0.036	-0.032
	(0.117)	(0.102)	(0.115)	(0.196)
IN stage 2 × HER	0.095	0.267**	0.074	0.214
	(0.139)	(0.121)	(0.133)	(0.221)
IN stage 2 × DUAL	-0.239**	-0.156*	-0.225**	-0.313**
	(0.102)	(0.093)	(0.105)	(0.157)
Intervention × RTF/DUAL $\times X_i$	-0.164	-0.247	-0.210	-0.176
	(0.119)	(0.266)	(0.145)	(0.269)
IN stage 2 × RTF/DUAL $\times X_i$	-0.094	-0.385*	0.084	-0.002
	(0.101)	(0.228)	(0.129)	(0.237)
IN stage 2 × HER $\times X_i$	-0.021	-0.368	0.083	-0.363
	(0.124)	(0.268)	(0.144)	(0.260)
IN stage 2 × DUAL $\times X_i$	-0.097	-0.166	0.024	0.146
	(0.092)	(0.203)	(0.083)	(0.207)
Individual fixed effects	*yes*	*yes*	*yes*	*yes*
Clusters	260	260	257	257
Observations	14675	14675	14501	14501
R^2	0.413	0.413	0.414	0.415

Standard errors in parentheses are clustered at the individual level. The coefficients are obtained by estimating equation (15). The full table with all coefficients is presented in Appendix A Table A4. All non-uploaders in HER and DUAL as well as all non-uploaders in RTF and CON who did not report a technical problem are excluded. The environmental attitude index is normalized to mean 0 and standard deviation 1.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

structed from subjects' baseline survey responses to questions on their willingness to engage in pro-environmental behavior. We report a specification with X_i as continuous terms, as well as a specification where X_i is an above-median indicator. Table 6 reports TOT estimates with heterogeneity along baseline energy use (columns 1 and 2) and environmental attitude (columns 3 and 4). Note that we only report the main coefficients of interests here to keep the table visually tractable, but the full set of coefficients can be found in Table A4 in the Appendix.

Consistent with previous literature, we find that the effect of real-time feedback in isolation increases with baseline use. In intervention stage 2, compounding the effects over both periods ($\hat{\lambda}_1 + \hat{\mu}_1$), subjects with 1 kWh higher baseline reduce their energy use per shower by an additional 0.26 kWh ($p = 0.069$) on average. Above-median baseline users (mean 3.30 kWh) save 0.63 kWh ($p = 0.039$) of energy more per shower compared to subjects with below-median baseline use (mean 1.17 kWh). This is consistent with the notion that real-time feedback reduces "slack" in resource use, but does not lead subjects to compromise on basic needs. It also appears that providing information through home energy reports in the DUAL condition induces about double the conservation effect for above-median users ($\hat{\gamma}_3 + \hat{\mu}_3 = -0.322$ kWh, $p = 0.075$), compared to below-median baseline users ($\hat{\gamma}_3 = -0.156$ kWh, $p = 0.096$) in intervention stage 2, although the difference is far from significant ($p = 0.414$). Home energy reports in isolation (HER group), on the other hand, are neither effective for low- nor high-baseline users. In fact, it seems that subjects with below-median baseline use tend to increase their energy use in intervention stage 2 ($p = 0.028$).

For the interactions with environmental attitude, the picture becomes even less clear. While the point estimates suggest that pro-environmental subjects may be more responsive to real-time feedback, the standard errors are simply too large to draw any meaningful conclusion. Ultimately, we are severely underpowered given the noisy nature of our survey proxy for pro-environmental motivation.

7. Underlying mechanisms

Through the lens of our conceptual framework, the empirical findings invite the interpretation that subjects already attached some value to conserving energy and water prior to any intervention, but only paid limited attention when showering, which was corrected by providing them with real-time feedback. This lack of attention could have prevented knowledge gains through home energy reports from taking effect in the absence of real-time feedback as the "enabler". To take a closer look at the mechanisms behind our main results, we now conduct a number of additional analyses to support our proposed channels and to rule out alternative explanations.

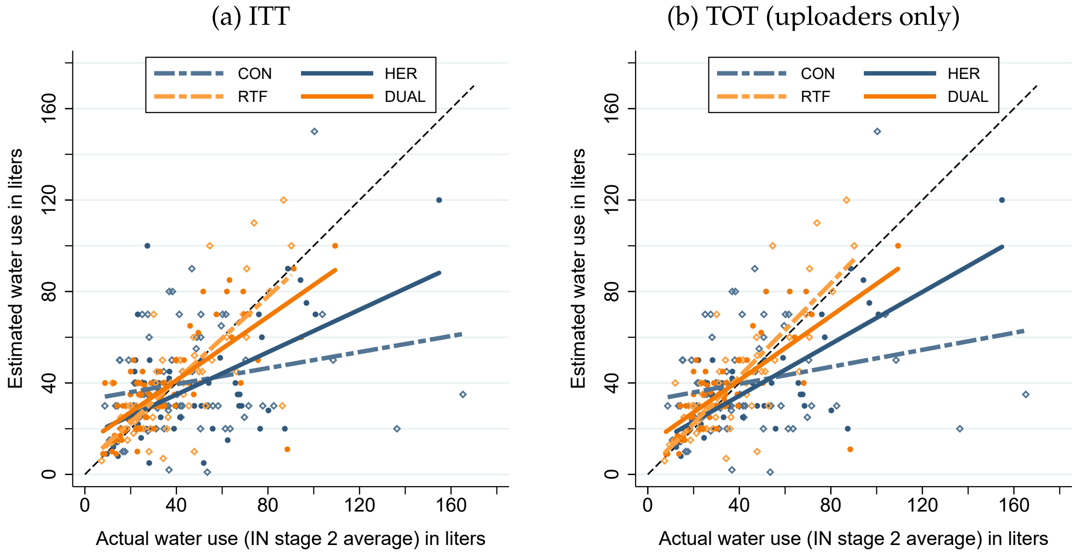
7.1. Awareness about resource intensity and environmental impacts

A crucial element of both interventions in our study is to enable learning about the outcomes of one's behavior. Real-time feedback through the smart meter provides immediate display of water use (and water temperature) for the current shower. Home energy reports also contain information on individuals' entire history of water use per shower since the start of the study, with the difference that it comes in retrospect. Nevertheless, both interventions should increase subjects' awareness about their own water use per shower.

To evaluate the impact of the treatments on resource use awareness, we make use of the post-intervention survey, where we again asked subjects to estimate the amount of water they typically use per shower. Figure 7 plots individuals' estimates as a function of their average water use per shower as measured by the smart meter. We show this separately for each experimental condition, both using the ITT as well as the TOT (uploaders-only) approach. The corresponding regression table A5 is presented in Appendix A.

Before the interventions, subjects' assessments were virtually uncorrelated with their

Figure 7: Post-intervention awareness about water use per shower



Notes. Both graphs compare subject's water use estimates from the endline questionnaire with their actual water use in intervention stage 2. In graph (b), we only use the subsample defined for the uploaders-only approach. Subjects with extreme estimates (above 200 liters) are excluded. Point clouds consist of individual observations (hollow diamonds for CON and RTF, solid circles for HER and DUAL) and lines represent separate regression fits for each treatment group. The dashed line starting at the origin is the 45 degree line.

actual water use, with low-baseline users overestimating and high-baseline users underestimating their water use (see Figure 5). The picture changes completely after the interventions. Whereas subjects in the CON group remain as ignorant as before, subjects who received real-time feedback are now able to estimate their water use almost without bias, so the fitted regression lines are close to the 45 degree line. While the slope looks slightly flatter for the DUAL group compared to the RTF group, the difference is not statistically significant. Importantly, home energy reports in isolation (HER group) already induce strong learning effects about water use, as estimated water use increases visibly in actual water use per shower (TOT slope 0.57), and significantly more strongly than in the CON group ($p = 0.025$). We cannot reject that learning through home energy reports is more effective with real-time feedback than without ($p = 0.497$). While these analyses focus on the bias of subjects' estimates (conditional on actual water use), we obtain similar results when we look at the size of estimation errors across groups (see Table A6 in Appendix A). Subjects in the three intervention treatments are on average about 27-30 percentage points closer to their actual water use than subjects in the CON group, and notably, the effect is virtually the same for HER, RTF, and DUAL groups.

Taken together, these results show that both types of interventions are successful in creating awareness about individual's own behavioral outcomes. However, learning about water use alone neither explains why real-time feedback is much more effective than home energy reports, nor why home energy reports induce an extra conservation effect when combined with real-time feedback. Good information per se does not seem to be sufficient to realize an individual's conservation potential, as the intervention also has to be timely and be able to overcome further barriers like limited attention and self-control problems.

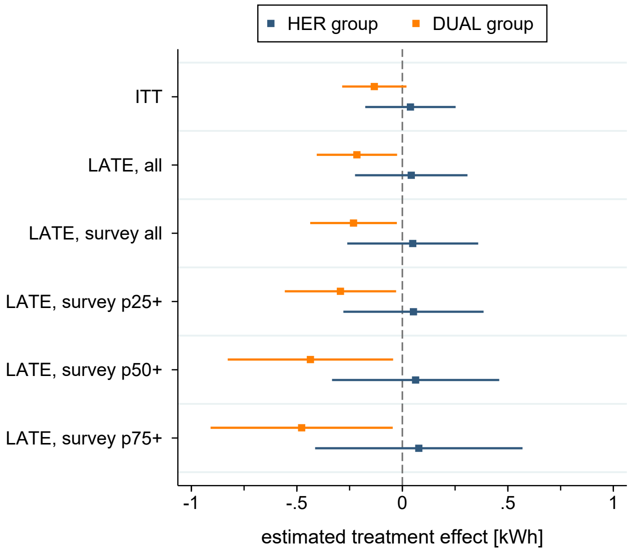
Home energy reports did not only contain information about water use, but also on energy use and environmental impacts in terms of CO_2 emissions. In addition, the second home energy report contained social comparison feedback with a random anonymous peer. These additional informational elements could explain why subjects in the DUAL group reduce their energy consumption even further after receiving the reports. Knowledge gains about the environmental impact of showering are only associated with conservation effects when home energy reports are combined with real-time feedback. One of the key insights of our theoretical framework is that if multiple barriers are significant, different behavioral interventions can become complements, because a single narrowly-targeted intervention is undermined by the presence of other significant barriers. Hence, our empirical results suggest that, in the absence of real-time feedback, barriers like limited attention have prevented knowledge gains through home energy reports from translating into actual behavior.

7.2. Engagement with home energy reports

One potential confounder is differential treatment engagement. By that, we refer to how much attention subjects pay to the treatments per se, e.g. how carefully they read the reports. If previous exposure to real-time feedback induced subjects in the DUAL group to engage much more strongly with the home energy reports than subjects in the HER group, we would then expect stronger conservation effects in the former group. This would not be the source of complementarity we want to highlight in this paper. The previous subsection shows that home energy reports induced similar learning effects about water use per shower in the HER and DUAL groups, which already suggests similar levels of scrutiny. To assess this crucial assumption more directly, we make use of the mini-survey that came with each of the two report emails. As described before, each email included a link to a survey in which we asked subjects to give an estimate of the amount of water they use in a typical shower. The survey link was at the bottom of the email, so subjects had to scroll through all the statistics on resource use and CO_2 emissions before clicking on it. We therefore use survey responses as proxy for the level of engagement with the feedback email.

Table A7 in Appendix A shows response rates by treatment group in the uploaders-only sample. Recall that subjects in the RTF and DUAL groups received Placebo emails containing a link to the same mini-survey. The overall response rates of uploaders was 87% for the first email and 71% for the second email. The share of respondents in the HER group was 8.4% lower than in the DUAL group for the first email ($p = 0.203$), and 9.4% higher for the second mail ($p = 0.308$); both differences are statistically in-

Figure 9: Effects for different levels of engagement with home energy reports



Notes. The points represent estimated regression coefficients for the effects of home energy reports in intervention stage 2, where treatment engagement status is instrumented with treatment assignment. Range lines indicate 90% confidence intervals.

significant. Apart from the extensive margin, Table A7 further shows subjects' relative estimation error by treatment group, defined as percent deviation of estimated water use in the mini-survey from the actual water use per shower.[34] Smaller estimation errors are an indication of subjects paying closer attention while reading the reports. Respondents in the HER group were only 10% off on average, and they actually gave more precise estimates than respondents in the DUAL group ($p = 0.039$), who were 21% off on average. Notwithstanding, both groups still outperform the CON group (49% off on average) by miles. Overall, we find no evidence that uploaders in DUAL group studied reports more carefully than uploaders in HER group.

To further explore whether it is the actual engagement with the information provided by home energy reports that matters, or whether the reports simply serve as cue or reminder for students to pay more attention to their shower behavior, we check whether subjects who studied the reports more closely also engaged more strongly in conservation actions. For this purpose, we again make use of subjects' water use assessments in the mini-surveys and regress energy use per shower on several new home energy report treatment indicators that increase in their level of strictness. Specifically, we define an indicator for whether subjects uploaded data *and* clicked on the mini survey in their report, and additional indicators for whether a subject's estimate precision, defined as distance between estimated and measured water use per shower, was above the 25th, 50th, or 75th percentile of all subjects, respectively. To avoid the endogeneity issue at hand, we use treatment assignment as instrument for level of engagement with reports. Figure 9 plots the coefficients and confidence intervals for the effect of home energy reports in

[34] As measure for actual water use per shower, we take the number that was calculated for each subject when sending out the home energy reports.

Table 7: Change in self-reported attitudes (from baseline to post-intervention survey)

	shower comfort		*environmental attitude*	
	(1)	(2)	(3)	(4)
	ITT	TOT	ITT	TOT
RTF group	0.042	0.047	-0.340***	-0.345***
	(0.117)	(0.119)	(0.117)	(0.119)
HER group	0.085	0.090	-0.277**	-0.253*
	(0.134)	(0.136)	(0.133)	(0.145)
DUAL group	-0.097	-0.011	-0.225*	-0.239*
	(0.138)	(0.150)	(0.129)	(0.144)
Constant	0.026	0.030	0.139	0.143
	(0.086)	(0.088)	(0.094)	(0.095)
F-test: p-value	0.641	0.896	0.034	0.039
Observations	300	255	304	257
R^2	0.007	0.003	0.027	0.031

Robust standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

HER and DUAL group, respectively. Even the most studious subjects in the HER group do not reduce their energy user in response to the reports, which corroborates our finding that home energy reports in isolation are insufficient to induce behavioral change. In contrast, the estimated conservation effect in the DUAL group increases monotonically with strictness of our indicator, reaching almost 0.5 kWh for the strictest definition. Together with the previous finding that the effects of home energy reports in DUAL tend to unfold over time, this speaks strongly in favor of the interpretation that what matters is the actual information content, not potential reminder or Hawthorne effects.

7.3. Environmental attitude and consumption value

Our interventions presented all information in a neutral and factual way, and we specifically refrained from including any normative element. Nevertheless, another confounder could be that the interventions either (differentially) increased subjects' pro-environmental motivation or somehow made showering less pleasurable to them, which would correspond to changes in $V(e)$ or $C(e)$ in the theoretical framework. To check if this could confound our results, we analyse subjects' survey responses before and after the study. The outcome variable of interest is the change in environmental attitude index or shower comfort index, respectively. All indices are normalized by subtracting the pre-intervention mean and dividing by the pre-intervention standard deviation.

The first two columns in Table 7 show difference-in-differences estimates for the experimental conditions, with the dependent variable being change in the indices from baseline to endline survey. Both in the ITT (column 1) and in the TOT (column 2) regressions for subjective shower comfort, we find no significant differences across experimental condition, and all point estimates are virtually zero. Hence, at least based on self-reported measures, our interventions do not seem to have diminished the consumption benefits of showering, which is also relevant for welfare considerations.

The other two columns in Table 7 show the difference-in-differences estimates for impacts on environmental attitude, with ITT estimates in column (3) and TOT estimates in column (4). Surprisingly, we find that subjects in the treated groups become *less* pro-environmental relative to the control group based on their survey responses. The magnitude of this decrease ranges from 22% to 35% of a (pre-study) standard deviation, which is not exactly quantitatively large, but also not negligible. We can only speculate about what is happening here. At face value, it may seem that feedback makes people less motivated to act pro-environmentally. Of course, we only have self-reported measures and cannot be certain about the underlying latent variable that they proxy for. But as we rather proxy *self-perceived* inclination to act pro-environmentally rather than the actual extent of pro-environmental behavior, one possible interpretation could be that feedback provision curbs the capacity for distorted self-image formation, because people become aware of their intention-action gaps. We caution from overinterpreting the result here, and we did not have any ex ante hypothesis along these lines. Still, we can conclude that the conservation effects we observe are unlikely due to higher pro-environmental motivation.

7.4. Other channels

Our leading interpretation of the underlying mechanisms is as follows: home energy reports operate through manipulating knowledge about resource intensity and environmental impacts, whereas real-time feedback works through channeling attention and raising water use awareness. We believe this is the most straightforward interpretation, but a number of alternative channels are conceivable. For example, the social comparison component included in the second home energy report could in principle add another motive for adjusting behavior, although we believe this to be unlikely, as Tiefenbeck et al. (2018) find no effect of including comparisons with the co-resident in a two-person household; hence, social norms are even less likely to arise in our setting, with the peer being assigned randomly and anonymously. Real-time feedback may also work in more ways apart from channelling attention. For example, it might counter self-control problems by offering immediate satisfaction (or disappointment) when the amount of water stayed below (or exceeded) a self-set target. Instant feedback may also facilitate experimentation with effective strategies for reducing resource consumption. Regardless of the channels through which the two interventions affect behavior, the complementarity between them remains an intriguing finding that is consistent with the core of our theoretical framework, i.e. that failure to address all significant barriers may result in failure of an in truth promising policy intervention.

8. Conclusion

In this paper, we have argued that when multiple behavioral barriers, e.g. imperfect information and limited attention, prevent individuals from implementing their values and intentions, then behavioral interventions overcoming different barriers each can be complements, in the sense that the effect of each intervention is boosted by the presence of another intervention. This is because a single intervention cannot always overcome all relevant barriers, so the remaining barriers will attenuate its effectiveness if they are not overcome as well by a complementary intervention. We report evidence from a field experiment on energy conservation in a specific resource-intensive everyday household activity (showering) that is in line with such complementarities in behavioral interventions. Home energy reports appear to be ineffective in isolation, but induce surprisingly large effects when combined with real-time feedback. Real-time feedback targets problems such as limited attention, that could otherwise have prevented knowledge gains through the home energy reports from translating into actual conservation behavior.

Although our interventions were tailored to only one specific resource-intensive activity, the effect sizes are quantitatively meaningful also on the aggregate household level. This is all the more remarkable given that our subjects had no monetary incentives to conserve resource, as they only pay a flat-rate fee for utilities. In our study, real-time feedback in isolation lowered consumption by 0.4 kWh (6.3 liters) per shower; adding home energy reports further lowered consumption by 0.22 kWh (3.8 liters) per shower, although the reports had no effect in isolation. In comparison, total daily energy use for lighting in German households is about 0.33 kWh per person. In his influential evaluation of the Opower home energy reports, which target *aggregate* energy use in U.S. households, Allcott (2011*b*) finds a household-level conservation effect of 0.62 kWh per day. In contrast, other studies find virtually no response to home energy reports in samples that are more similar to ours — German households, whose average baseline consumption level is lower than in the U.S. (Andor et al., 2017), or U.S. college dorm residents without monetary incentives for energy conservation (Myers and Souza, 2019).

One could conclude that home energy reports as policy instruments are relatively ineffective, but our study suggests that such a conclusion for any type of intervention may only hold given the existing policy and choice environment that consumers act in. In fact, our complementarity argument is based exactly on the premise that a well-blended policy mix can alter features of this environment, in a way that brings the best out of each component. More specifically, we highlight that designers of policy interventions not only have to take into account the channels through which interventions affect behavior, but also identify important barriers to behavioral change that still remain, and how these can be overcome by complementary interventions.

Our study shows the existence of complementarities in a very specific setting, but the the notion of barrier multiplicity can be relevant in other contexts, also when involving more standard economic barriers such as time, money, and technology constraints, or lack of financial incentives. Indeed, some empirical findings in the literature are at least suggestive of similar mechanisms at work. For example, Cortes et al. (2019) find that text-message based curricula supporting good parenting practices work less well when parents face high cognitive load than during time periods when the load is lighter. Dupas and Robinson (2013) study financial savings behavior in a developing country and find that simply providing a safe box for storing money is already quite effective for encouraging higher savings, except for the subgroup of individuals with severe present bias, who need additional social commitment. Similarly, prompting deliberation about food choice, to help resist short-run temptations, increases the effectiveness of healthy purchasing subsidies (Brownback, Imas and Kuhn, 2019). We suspect that barrier multiplicity is a pervasive feature in many other domains.

New policies are always introduced to an existing net of policies, institutions, and norms. As social scientists are beginning to pioneer the process from small-scale proof-of-concept studies to large-scale interventions (Banerjee et al., 2017), future research must therefore synchronously advance our knowledge on the interplay of different policy instruments.

References

Abrahamse, Wokje, Linda Steg, Charles Vlek, and Talib Rothengatter. 2005. "A review of intervention studies aimed at household energy conservation." *Journal of Environmental Psychology*, 25(3): 273–291.

Allcott, Hunt. 2011*a*. "Consumers' Perceptions and Misperceptions of Energy Costs." *American Economic Review*, 101(3): 98–104.

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

Allcott, Hunt. 2016. "Paternalism and Energy Efficiency: An Overview." *Annual Review of Economics*, 8(1): 145–176.

Allcott, Hunt, and Sendhil Mullainathan. 2010. "Behavior and energy policy." *Science*, 327(5970): 1204–1205.

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

Altmann, Steffen, Armin Falk, Simon Jäger, and Florian Zimmermann. 2018. "Learning about job search: A field experiment with job seekers in Germany." *Journal of Public Economics*, 164: 33–49.

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

Andor, Mark, Andreas Gerster, Joerg Peters, and Christoph M. Schmidt. 2017. "Social Norms and Energy Conservation Beyond the US." *Ruhr Economic Papers No. 714*.

Ashraf, Nava, B. Kelsey Jack, and Emir Kamenica. 2013. "Information and subsidies: Complements or substitutes?" *Journal of Economic Behavior & Organization*, 88: 133–139.

Attari, Shahzeen Z. 2014. "Perceptions of water use." *Proceedings of the National Academy of Sciences of the United States of America*, 111(14): 5129–5134.

Attari, Shahzeen Z., Michael L. DeKay, Cliff I. Davidson, and Wändi Bruine de Bruin. 2010. "Public perceptions of energy consumption and savings." *Proceedings of the National Academy of Sciences of the United States of America*, 107(37): 16054–16059.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application." *Journal of Economic Perspectives*, 31(4): 73–102.

Bertrand, Marianne, and Adair Morse. 2011. "Information Disclosure, Cognitive Biases, and Payday Borrowing." *Journal of Finance*, 66(6).

Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment." *American Economic Review*, 105(11): 3489–3529.

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. 2018. "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 of the United States of America*.

Brownback, Andy, Alex Imas, and Michael Kuhn. 2019. "Behavioral Food Subsidies." *Working Paper*.

Camilleri, Adrian R., Richard P. Larrick, Shajuti Hossain, and Dalia Patino-Echeverri. 2019. "Consumers underestimate the emissions associated with food but are aided by labels." *Nature Climate Change*, 9(1): 53–58.

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

Cortes, Kalena E., Hans Fricke, Susanna Loeb, David S. Song, and Ben York. 2019. "When Behavioral Barriers Are Too High or Low: How Timing Matters for Parenting Interventions." *IZA Discussion Paper No. 12416*.

Costa, Dora L., and Matthew E. Kahn. 2013. "Energy Conservation "Nudges" And Environmentalist Ideology: Evidence From A Randomized Residential Electricity Field Experiment." *Journal of the European Economic Association*, 11(3): 680–702.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review*, 96(3): 694–719.

Delmas, Magali A., Miriam Fischlein, and Omar I. Asensio. 2013. "Information strategies and energy conservation behavior: A meta-analysis of experimental studies from 1975 to 2012." *Energy Policy*, 61: 729–739.

Dertwinkel-Kalt, Markus, Holger Gerhardt, Gerhard Riener, Frederik Schwerter, and Louis Strang. 2019. "Concentration Bias in Intertemporal Choice." *Working Paper*.

Dietz, Thomas, Gerald T. Gardner, Jonathan Gilligan, Paul C. Stern, and Michael P. Vandenbergh. 2009. "Household actions can provide a behavioral wedge to rapidly reduce US carbon emissions." *Proceedings of the National Academy of Sciences of the United States of America*, 106(44): 18452–18456.

Drexler, Alejandro, Greg Fischer, and Antoinette Schoar. 2014. "Keeping It Simple: Financial Literacy and Rules of Thumb." *American Economic Journal: Applied Economics*, 6(2): 1–31.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. "Education, HIV, and Early Fertility: Experimental Evidence from Kenya." *American Economic Review*, 105(9): 2757–2797.

Dupas, Pascaline, and Jonathan Robinson. 2013. "Why Don't the Poor Save More? Evidence from Health Savings Experiments." *American Economic Review*, 103(4): 1138–1171.

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

Fischer, Corinna. 2008. "Feedback on household electricity consumption: A tool for saving energy?" *Energy Efficiency*, 1(1): 79–104.

Fowlie, Meredith, Catherine Wolfram, C. Anna Spurlock, Annika Todd, Patrick Baylis, and Peter Cappers. 2018. "Default effects and follow-on behavior: Evidence from an electricity pricing program." *Working Paper*.

Frederiks, Elisha R., Karen Stenner, and Elizabeth V. Hobman. 2015. "Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour." *Renewable and Sustainable Energy Reviews*, 41: 1385–1394.

Gabaix, Xavier. 2017. "Behavioral Inattention." *NBER Working Papers 24096*.

Gardner, Gerald T., and Paul C. Stern. 2008. "The Short List: The Most Effective Actions U.S. Households Can Take to Curb Climate Change." *Environment and Behavior*, 50(5): 12–24.

Gifford, Robert. 2011. "The dragons of inaction: Psychological barriers that limit climate change mitigation and adaptation." *American Psychologist*, 66(4): 290–302.

Handel, Benjamin, and Joshua Schwartzstein. 2018. "Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care?" *Journal of Economic Perspectives*, 32(1): 155–178.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning Through Noticing: Theory and Evidence from a Field Experiment." *Quarterly Journal of Economics*, 129(3): 1311–1353.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics*, 123(4): 1373–1414.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(2): 467.

IPCC. 2014. *Climate change 2014: Synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)].* Geneva, Switzerland:IPCC.

Jamison, Julian C., Dean Karlan, and Jonathan Zinman. 2014. "Financial Education and Access to Savings Accounts: Complements or Substitutes? Evidence from Ugandan Youth Clubs." *NBER Working Paper 20135*.

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.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman. 2016. "Getting to the Top of Mind: How Reminders Increase Saving." *Management Science*, 62(12): 3393–3411.

Karlin, Beth, Joanne F. Zinger, and Rebecca Ford. 2015. "The effects of feedback on energy conservation: A meta-analysis." *Psychological Science*, 141(6): 1205–1227.

Keefer, Quinn, and Galib Rustamov. 2018. "Limited attention in residential energy markets: A regression discontinuity approach." *Empirical Economics*, 55(3): 993–1017.

Kollmuss, Anja, and Julian Agyeman. 2002. "Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior?" *Environmental Education Research*, 8(3): 239–260.

Koszegi, B., and A. Szeidl. 2013. "A Model of Focusing in Economic Choice." *Quarterly Journal of Economics*, 128(1): 53–104.

Langenbach, Benedikt P., Sebastian Berger, Thomas Baumgartner, and Daria Knoch. 2019. "Cognitive Resources Moderate the Relationship Between Pro-Environmental Attitudes and Green Behavior." *Environment and Behavior*, 16: 1–17.

Liebman, Jeffrey B., and Erzo F. P. Luttmer. 2015. "Would People Behave Differently If They Better Understood Social Security? Evidence from a Field Experiment." *American Economic Journal: Economic Policy*, 7(1): 275–299.

List, John A., Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer. 2017. "Harnessing Policy Complementarities to Conserve Energy: Evidence from a Natural Field Experiment." *NBER Working Paper 23355*.

Madrian, Brigitte C. 2014. "Applying Insights From Behavioral Economics To Policy Design." *Annual Review of Economics*, 6: 663–688.

Markowitz, Ezra M., and Azim F. Shariff. 2012. "Climate change and moral judgement." *Nature Climate Change*, 2(4): 243–247.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. "Inputs, Incentives, And Complementarities In Education: Experimental Evidence From Tanzania." *Quarterly Journal of Economics*, 134(3): 1627–1673.

Myers, Erica, and Mateus Souza. 2019. "Social Comparison Nudges Without Monetary Incentives: Evidence from Home Energy Reports." *E2e Working Paper 041*.

Newell, Richard G., and Juha Siikamäki. 2014. "Nudging Energy Efficiency Behavior: The Role of Information Labels." *Journal of the Association of Environmental and Resource Economists*, 1(4): 555–598.

Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2008. "Normative social influence is underdetected." *Personality and Social Psychology Bulletin*, 34(7): 913–923.

Pichert, Daniel, and Konstantinos V. Katsikopoulos. 2008. "Green defaults: Information presentation and pro-environmental behaviour." *Journal of Environmental Psychology*, 28(1): 63–73.

Reddy, Sheila M.W., Jensen Montambault, Yuta J. Masuda, Elizabeth Keenan, William Butler, Jonathan R.B. Fisher, Stanley T. Asah, and Ayelet Gneezy. 2017. "Advancing Conservation by Understanding and Influencing Human Behavior." *Conservation Letters*, 10(2): 248–256.

Schwartz, Daniel, and George Loewenstein. 2017. "The Chill of the Moment: Emotions and Proenvironmental Behavior." *Journal of Public Policy & Marketing*, 36(2): 255–268.

Schwartzstein, Joshua. 2014. "Selective Attention and Learning." *Journal of the European Economic Association*, 12(6): 1423–1452.

Taubinsky, Dmitry, and Alex Rees-Jones. 2018. "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." *Review of Economic Studies*, 85(4): 2462–2496.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth and Happiness.* Yale University Press.

Tiefenbeck, Verena, Anselma Woerner, Samuel Schoeb, 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: 35–41.

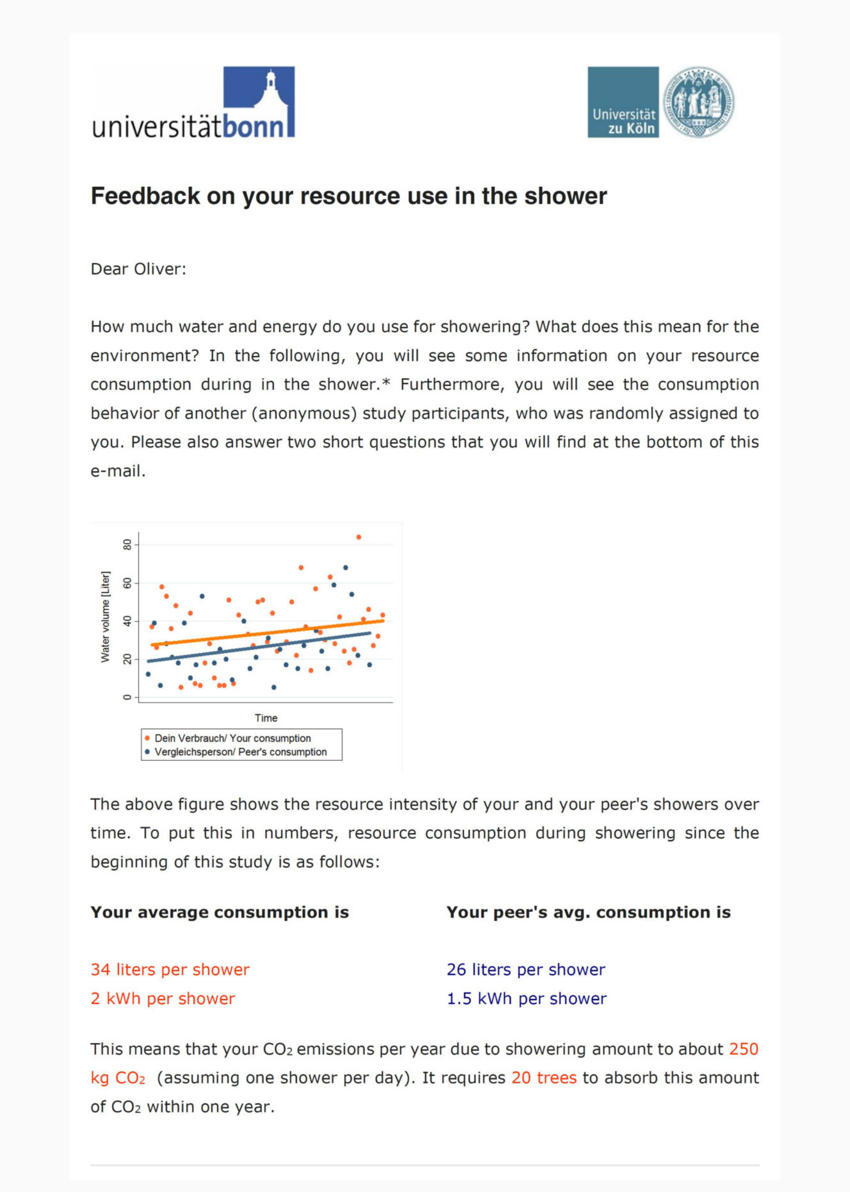
Tiefenbeck, Verena, Lorenz Goette, Kathrin Degen, Vojkan Tasic, Elgar Fleisch, Rafael Lalive, and Thorsten Staake. 2018. "Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation." *Management Science*, 64(3): 1458–1476.

Tonke, Sebastian. 2019. "Imperfect Knowledge, Information Provision and Behavior: Evidence from a Field Experiment to Encourage Resource Conservation." *Working Paper*.

Werthschulte, Madeline, and Andreas Loeschel. 2019. "Cost misperceptions and energy consumption: Experimental evidence for present bias and biased price beliefs." *CAWM Discussion Paper No. 111*.

Appendix A Supplementary figures and tables

Figure A1: Screenshot of a home energy report with peer comparison



universitätbonn

Universität zu Köln

Feedback on your resource use in the shower

Dear Oliver:

How much water and energy do you use for showering? What does this mean for the environment? In the following, you will see some information on your resource consumption during in the shower.* Furthermore, you will see the consumption behavior of another (anonymous) study participants, who was randomly assigned to you. Please also answer two short questions that you will find at the bottom of this e-mail.

The above figure shows the resource intensity of your and your peer's showers over time. To put this in numbers, resource consumption during showering since the beginning of this study is as follows:

Your average consumption is	**Your peer's avg. consumption is**
34 liters per shower	26 liters per shower
2 kWh per shower	1.5 kWh per shower

This means that your CO_2 emissions per year due to showering amount to about 250 kg CO_2 (assuming one shower per day). It requires 20 trees to absorb this amount of CO_2 within one year.

Figure A2: Empirical distribution of report timing

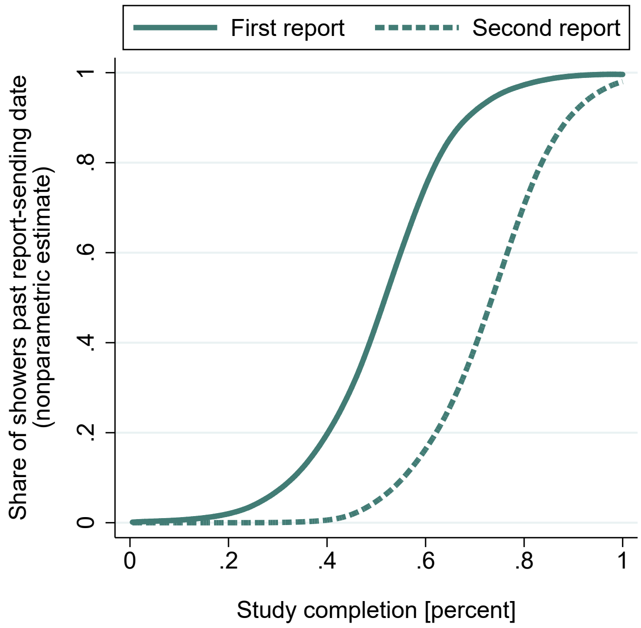


Table A1: Additional randomization checks

	Baseline survey responses				
	environmental attitude	shower comfort	1 if female	age in years	1 if international
HER group	-0.106	0.094	-0.046	0.757	-0.017
	(0.165)	(0.164)	(0.080)	(0.615)	(0.075)
RTF group	0.044	-0.164	-0.015	0.872	0.042
	(0.167)	(0.156)	(0.079)	(0.584)	(0.077)
DUAL group	0.154	0.115	0.117	0.540	0.032
	(0.161)	(0.149)	(0.075)	(0.583)	(0.075)
Constant	-0.041	-0.014	0.597	23.351	0.325
	(0.118)	(0.100)	(0.056)	(0.380)	(0.054)
Observations	307	306	318	307	318
R-squared	0.009	0.012	0.017	0.007	0.003
F-test: *p*-value	0.425	0.327	0.130	0.437	0.847

Robust standard errors in parentheses. The omitted category is the CON group.

Table A2: Comparing uploaders and non-uploaders

	uploaders: mean (sd)	non-uploaders: mean (sd)	diff. in means p-value
Energy [kWh]	2.23	2.20	0.95
	(1.38)	(1.37)	
Water volume [liter]	38.54	37.13	0.87
	(22.36)	(20.73)	
Temperature [Celsius]	35.41	35.94	0.61
	(3.33)	(3.47)	
Flow rate [liter/min]	6.01	5.30	0.11
	(2.34)	(2.19)	
Duration [min]	6.61	7.69	0.10
	(2.98)	(4.54)	
Environmental attitude	-0.04	0.07	0.79
	(1.03)	(0.93)	
Shower comfort	-0.05	0.14	0.55
	(1.05)	(0.87)	
1 if female	0.58	0.70	0.28
	(0.49)	(0.46)	
Age in years	23.93	23.79	0.95
	(3.80)	(3.99)	
1 if international	0.31	0.41	0.42
	(0.46)	(0.49)	
Observations	228	90	

Subject characteristics before sending out home energy reports. p-values adjusted for multiple hypothesis testing (Romano-Wolf procedure using $2,000$ bootstrap repetitions).

Table A3: Treatment effect dynamics

	$Z_i = \mathbb{I}\{post\ 2nd\ report\}$		$Z_i =$ # *weeks after 1st report*	
	(1) Uploaders	(2) LATE	(3) Uploaders	(4) LATE
Intervention	0.179	0.172*	0.178	0.171*
	(0.111)	(0.103)	(0.111)	(0.102)
Intervention × RTF/DUAL	-0.388***	-0.365***	-0.386***	-0.364***
	(0.134)	(0.125)	(0.133)	(0.125)
Intervention × HER	0.027	0.016	0.029	0.019
	(0.154)	(0.134)	(0.154)	(0.134)
Intervention × DUAL	0.046	0.119	0.047	0.120
	(0.113)	(0.108)	(0.112)	(0.108)
IN stage 2	0.139	0.176	0.065	0.109
	(0.103)	(0.110)	(0.124)	(0.127)
IN stage 2 × RTF/DUAL	-0.027	-0.053	0.047	0.019
	(0.128)	(0.134)	(0.156)	(0.159)
IN stage 2 × HER	0.092	0.048	0.198	0.174
	(0.148)	(0.169)	(0.181)	(0.202)
IN stage 2 × DUAL	-0.068	-0.041	0.030	0.075
	(0.123)	(0.135)	(0.166)	(0.177)
IN stage 2 $\times Z_i$	0.019	0.022	0.032	0.029
	(0.093)	(0.090)	(0.027)	(0.026)
IN stage 2 × RTF/DUAL $\times Z_i$	0.012	0.000	-0.026	-0.026
	(0.123)	(0.119)	(0.037)	(0.035)
IN stage 2 × HER $\times Z_i$	-0.002	-0.010	-0.041	-0.051
	(0.126)	(0.136)	(0.042)	(0.047)
IN stage 2 × DUAL $\times Z_i$	-0.279	-0.316	-0.099	-0.114*
	(0.209)	(0.215)	(0.064)	(0.067)
Individual fixed effects	*yes*	*yes*	*yes*	*yes*
Clusters	261	318	261	318
Observations	14712	17942	14712	17942
R^2	0.413	0.005	0.413	0.005

Standard errors in parentheses are clustered at the individual level. In columns (1) and (2), we exclude all non-uploaders in HER and DUAL as well as all non-uploaders in RTF and CON who did not report a technical problem. In columns (3) and (4), we use treatment assignment to HER and DUAL, respectively, interacted with the IN stage 2 indicator as instrument for receiving informative home energy reports. The reported R^2 in Columns (3) and (4) is the within R^2.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Treatment effect heterogeneity

	X_i : *baseline energy use*		X_i : *envir. attitude*	
	(1) linear	(2) median$^+$	(3) linear	(4) median$^+$
Intervention	0.180*	0.272***	0.178	0.243
	(0.105)	(0.072)	(0.112)	(0.203)
Intervention × RTF/DUAL	-0.403***	-0.254***	-0.392***	-0.324
	(0.127)	(0.096)	(0.134)	(0.222)
Intervention × HER	0.012	-0.139	0.003	0.030
	(0.146)	(0.112)	(0.149)	(0.255)
Intervention × DUAL	0.085	0.020	0.049	-0.088
	(0.111)	(0.087)	(0.113)	(0.151)
IN stage 2	0.148	0.001	0.166*	0.140
	(0.091)	(0.065)	(0.089)	(0.172)
IN stage 2 × RTF/DUAL	-0.014	0.171*	-0.036	-0.032
	(0.117)	(0.102)	(0.115)	(0.196)
IN stage 2 × HER	0.095	0.267**	0.074	0.214
	(0.139)	(0.121)	(0.133)	(0.221)
IN stage 2 × DUAL	-0.239**	-0.156*	-0.225**	-0.313**
	(0.102)	(0.093)	(0.105)	(0.157)
Intervention $\times X_i$	-0.016	-0.192	0.031	-0.137
	(0.101)	(0.226)	(0.130)	(0.220)
Intervention × RTF/DUAL $\times X_i$	-0.164	-0.247	-0.210	-0.176
	(0.119)	(0.266)	(0.145)	(0.269)
Intervention × HER $\times X_i$	0.109	0.325	-0.172	-0.039
	(0.140)	(0.301)	(0.166)	(0.296)
Intervention × DUAL $\times X_i$	0.062	0.039	0.103	0.310
	(0.110)	(0.215)	(0.105)	(0.232)
IN stage 2 $\times X_i$	0.056	0.313*	-0.076	0.056
	(0.077)	(0.179)	(0.116)	(0.185)
IN stage 2 × RTF/DUAL $\times X_i$	-0.094	-0.385*	0.084	-0.002
	(0.101)	(0.228)	(0.129)	(0.237)
IN stage 2 × HER $\times X_i$	-0.021	-0.368	0.083	-0.363
	(0.124)	(0.268)	(0.144)	(0.260)
IN stage 2 × DUAL $\times X_i$	-0.097	-0.166	0.024	0.146
	(0.092)	(0.203)	(0.083)	(0.207)
Individual fixed effects	*yes*	*yes*	*yes*	*yes*
Clusters	260	260	257	257
Observations	14675	14675	14501	14501
R^2	0.413	0.413	0.414	0.415

Standard errors in parentheses are clustered at the individual level. The coefficients are obtained using the within estimator. All non-uploaders in HER and DUAL, as well as all non-uploaders in RTF and CON who did not report a technical problem, are excluded.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Estimated vs actual water use per shower

	before study	after study	
		ITT	TOT
Actual volume	0.271	0.175	0.186
	(0.263)	(0.139)	(0.145)
Actual volume × RTF	0.025	0.742***	0.835***
	(0.376)	(0.199)	(0.179)
Actual volume × HER	-0.465	0.289*	0.381**
	(0.292)	(0.174)	(0.169)
Actual volume × DUAL	-0.074	0.520***	0.517**
	(0.299)	(0.182)	(0.230)
RTF group	-0.131	1.694	3.162
	(6.777)	(3.234)	(3.183)
HER group	-7.001	-4.578	-5.200*
	(5.813)	(3.181)	(3.029)
DUAL group	-5.182	1.655	1.588
	(5.851)	(3.136)	(3.826)
Constant	43.436***	39.507***	39.610***
	(4.590)	(2.429)	(2.542)
Observations	267	296	251
R^2	0.030	0.378	0.440

Robust standard errors in parentheses. Actual volume is recentered around 40 liters.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Estimated versus actual water use: relative estimation error

	before study	after study	
		ITT	TOT
RTF group	0.075	-0.283***	-0.296***
	(0.201)	(0.073)	(0.075)
HER group	0.008	-0.172**	-0.281***
	(0.175)	(0.080)	(0.072)
DUAL group	-0.055	-0.214**	-0.270***
	(0.178)	(0.085)	(0.076)
Constant	0.927***	0.577***	0.583***
	(0.136)	(0.061)	(0.064)
Observations	302	296	251
R^2	0.002	0.050	0.101

Robust standard errors in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Response to mini-surveys attached to reports

	Survey response rate			
	(1) first report	(2) second report	(3) any report	(4) estimation error [%p]
RTF group	-1.05	0.53	-2.48	-30.69
	(5.35)	(6.57)	(4.90)	(7.62)
HER group	-7.85	-16.76	-7.18	-38.74
	(6.39)	(7.91)	(5.81)	(7.54)
DUAL group	0.58	-26.17	-0.44	-27.70
	(5.54)	(8.14)	(5.00)	(8.57)
Constant	88.89	80.56	91.67	48.93
	(3.73)	(4.70)	(3.28)	(7.14)
p-value for HER = DUAL	0.203	0.308	0.270	0.039
Observations	261	261	261	231
R-squared	0.009	0.061	0.008	0.139

Robust standard errors in parentheses.

Appendix B Randomization protocol

At the beginning of the study, we randomly assigned subjects into groups that receive or do not receive real-time feedback. Each smart meter was programmed as either treatment or control device. Treatment device started displaying real-time feedback from the eleventh shower onwards, whereas control devices only ever showed the current water temperature. When distributing the smart meters to subjects, we alternated between treatment and control devices after each apartment. Thus, treatment and control devices are by construction balanced within dorms.

We assigned subjects into groups with or without home energy report shortly before we intended to sent out the reports. We used the data that subjects uploaded through the smartphone app to rank them from lowest to highest average water use per shower, split by whether they receive real-time feedback or not. Then, we formed pairs between subjects adjacent to each other in rank and assigned home energy reports to only one member of a pair based on a virtual coin flip. This ensures that the distribution of resource consumption levels remain balanced across experimental conditions. Subjects who had not uploaded any data at that point in time were assigned to a group randomly without prior ranking.

The second home energy report further contained a social comparison component with a random and anonymous peer. This peer was assigned to subjects in the following way: (1) we used uploaded data prior to the second report to rank subjects again by their average water use per shower; (2) we then selected three potential peers for each subject, a subject who was somewhat above him/her in rank, a subject who was somewhat below him/her in rank, and a directly adjacent subject; (3) we then chose one of these three candidates randomly with equal probabilities; (4) subjects who had not uploaded any data received a random peer from the pool of subjects who had uploaded data. This procedure ensured that the direction of peer comparison was orthogonal to subjects' resource use level.

Appendix C Data cleaning procedures

A number of data cleaning steps are performed before running the empirical analyses. In principle, we have access to the smart meter data from two sources: (1) uploads by subjects themselves using the smartphone app, and (2) the data that we read out manually after retrieving the devices. For the large majority of devices, the two sources gave us identical data. In the cases where it differed, we always opted to use the information we read out manually.

We drop the very first data point of each participant, as they usually started with a test run to check if the device was working. Following Tiefenbeck et al. (2018), we further drop any water extraction with volume below 4.5 liters (in total $2,942$ extractions), as these are unlikely to be actual showers but rather minor extractions for other purposes such as cleaning. We further remove 37 extreme outlier points, defined as energy use and water use for that shower being more than 4.5 times the subject-specific interquartile range away from the closest quartile. We are particularly strict in only excluding the most unplausible data points here. Conventionally, 1.5 or 3 times the interquartile range (IQR) are used as criterion for outliers. For a normal distribution, 4.5 times the IQR away from the nearest quartile corresponds to 6.745 standard deviation away from the mean.

We further exclude 1 device with erratic data, as evidenced by huge intra-device variance (the largest for all devices) and some outrageous data points with water volumes of up to above 500 liters for a single shower. In 8 cases, the device's temperature sensor broke at some point, and we impute missing information with the average temperature of showers taken while the sensor was still intact. For some devices, we detected an error through which decimal places of the flow rate are shifted such that the stored number is actually ten times the actual flow rate. We corrected these manually for showers with flow rates that are about ten times the flow rate of other showers stored on the device.

Appendix D Timing of showers

As the smart meter itself has no global time counter and only stores the chronological order of water extractions, we make use of smartphone app information to put a time stamp on each observation. In particular, we need to determine whether a shower took place before or after we sent out the home energy reports, so whether it is in intervention stage 2. The app provides us with information on the date and time of each data upload by subjects. This allows us construct time windows in which a shower observation has plausibly happened. Firstly, a shower must have been taken by the time data was uploaded via the app, so this gives us the upper bound. Secondly, it must have been taken place after the previous data upload, because otherwise it would have been uploaded by then; this gives us the lower bound. To be able to determine the timing relatively reliably around the crucial time period, in which we sent out home energy reports, we sent several upload reminders to all participants. Whenever it was not unambiguously clear, which shower was the first that took place after a home energy report, we assigned the switching point implied by constant shower frequency. For example, if one upload was 1 day before the home energy report and the next upload 1 day after, and there were 2 showers in the window, we assumed that the first shower was before and the second shower after the report.

A complication arising from non-uploaders is that we do not know the timing of showers by these participants, because the shower meter itself only stores the order of showers but not the time and date. We can only infer the earliest and latest possible date of each shower based on when it was uploaded to the smartphone app. Therefore, whenever we want to include non-uploaders in our analyses, we need to impute the timing of showers in one way or another, in particular whether it took place before or after a home energy report.

We use a pragmatic imputation approach based on the assumption that, given the stage of study completion, i.e. which fraction of the number of total recorded showers have been completed, showers by uploaders and non-uploaders have the same probability of having taken place after the first/second home energy report. Formally, we assume that for each stage of study completion π,

$$Pr\left(IN_{it}^{s2}=1|\pi, non\text{-}uploader\right)=Pr\left(IN_{it}^{s2}=1|\pi, uploader\right).$$

To operationalize this approach, we estimate the distribution of uploaders' report timing over study completion non-parametrically, so $\widehat{Pr}\left(IN_{it}^{s2}=1 \mid \pi, uploader\right)$, and, instead of the indicator IN_{π}^{s2} for intervention stage 2, we define

$$\widehat{IN}_{s}^{s2}=\widehat{Pr}\left(IN_{it}^{s2}=1 \mid \pi_{it}^{s}=1, uploader\right)$$

as probabilistic indicator for every shower of non-uploaders in study completion stage π. In other words, the regressor $\widehat{IN}_{\pi}^{s2}$ is the probability that a particular shower by a non-uploader took place after the first home energy report. In all our regressions, we actually use the indicator

$$\widetilde{IN}_{it}^{s2}=\begin{cases} IN_{it}^{s2} & \textit{if uploader} \\ \widehat{Pr}\left(IN_{it}^{s2}=1 \mid \pi, uploader\right) & \textit{if non-uploader}. \end{cases} \tag{16}$$