# Greener Products, Happier Consumers? A Comparative Evaluation of Behavioral Sustainability Interventions

November 6, 2026

**Abstract**

Organizations often aim to promote resource conservation by providing information, but they face uncertainty about which type of information is most effective and most valued by consumers. We provide evidence on these questions by comparing two popular resource conservation interventions that differ in how they provide information: social comparison reports (SC) and real-time feedback (RTF), as well as their combination. Our application targets households' showering behavior. In a field experiment with over 1,000 participants, we compare the two interventions on the same behavioral and welfare outcomes, thereby evaluating both conservation effectiveness and the value consumers place on each intervention. SC reduces water and energy use per shower by 9%, RTF achieves a 29% reduction, and the combination of both achieves a 35% reduction. Incentivized willingness to pay (WTP) elicitations show that participants value RTF and the combined intervention roughly 5–10% more than SC alone. These valuations remain broadly stable before and after participants experience the interventions, suggesting sustained positive consumer valuations.

*Keywords*: Resource Conservation, Welfare, Real-time Feedback, Social Comparison, Behavioral Intervention, Field Experiment

*JEL codes*: D12, C93, Q25

# 1 Introduction

Over the past decade, behavioral economic interventions have become an integral part of governmental policy-making, as illustrated by the establishment of behavioral insights groups in governments around the world (Obama, 2015; OECD, 2017). At the same time, the private sector has increasingly adopted behavioral tools to influence the decisions and actions of customers and employees (e.g. Allcott, 2011; Tiefenbeck et al., 2019; Gosnell et al., 2020; Chen et al., 2021).

Many of these interventions appear highly cost-effective: they typically impose no direct financial costs on consumers, require minimal implementation expenses, and can generate large-scale societal benefits (Allcott and Mullainathan, 2010; Benartzi et al., 2017). However, standard cost-effectiveness analyses often overlook important non-monetary costs and benefits. These include emotional and psychological responses, such as shame, pride, or social pressure (DellaVigna et al., 2012; 2017; Butera et al., 2022), as well as changes in consumer welfare resulting from altered behavior (Allcott and Kessler, 2019), or (dis-)comfort arising from the mere provision of information, independent of its instrumental value (Sharot and Sunstein, 2020). At best, an intervention is perceived as empowering; at worst, it can feel manipulative or intrusive, potentially alienating its intended audience. Consequently, an intervention that produces large conservation effects may nonetheless be socially undesirable if those gains come at the cost of a substantial psychological tax on users. Equivalently, a private firm may be hesitant to implement an intervention that potentially decreases consumer welfare and, correspondingly, customer satisfaction.

How can governments and private firms harness the benefits of behavioral interventions while minimizing potential psychological costs, or even increasing consumer welfare? One approach to address this question is through comparative evaluations of interventions that assess not only their behavioral impact but also their implications for consumer welfare. While most research focuses on reduced-form behavior change, explicit measurement of consumer welfare remains relatively rare. Such evi-

dence is particularly critical in sustainability contexts, where both public agencies and private organizations face mounting pressure to adopt greener practices (Lambin and Thorlakson, 2018), and where interventions that enhance consumer welfare may boost customer loyalty and profitability (Ho and Page, 2025).

This paper is based on an experimental design that allows us to provide a comparative assessment of the behavioral and welfare effects of two popular interventions for resource conservation. This enables a direct comparison of two types of interventions that offer a wide range of applications and differ fundamentally in how they provide information. In addition, our design is portable to other intervention types and other contexts where high-resolution behavioral data over time are available.

Two aspects of our experimental design are crucial for this comparison. First, we compare the behavioral and welfare effects of the interventions within a single field experiment. Conducting this comparison in one population, context, and time period is valuable because behavioral treatment effects are often context-dependent, so differences across separate studies can reflect not only intervention design but also differences in study populations, institutions, or implementation environments (Allcott, 2015; Andor et al., 2020; 2022; Ludwig et al., 2011). Second, we not only test the interventions individually, but also evaluate their combination. The interventions are not mutually exclusive, making it important to understand the effects of potentially employing them jointly. Prior work shows that combining behavioral tools can generate complementarities in some settings, but in others additional intervention elements can strain attention or yield less-than-additive effects (Brandon et al., 2019; Fang et al., 2023; Bonan et al., 2024; Alt et al., 2024).

Our design therefore speaks directly to a practical question faced by firms and policymakers: whether to deploy one intervention, the other, or both, by identifying the incremental behavioral and welfare value of each intervention. To focus on behavior and welfare is important because interventions need not rank similarly in terms of behavioral impact and consumer welfare: a treatment that changes behavior more effectively need not be the one consumers value most, and vice versa. Measuring both

outcomes within a common setting allows us to compare interventions more precisely and to assess whether adding a second intervention meaningfully improves outcomes or instead adds attention, implementation, and psychological costs with limited additional benefit or even adverse effects.

The application in which we evaluate both interventions is household showering, a frequent, resource-intensive activity where small behavioral changes can substantially reduce water and energy use. Because heating water for a single shower consumes more energy than a modern refrigerator does in a day (Michel et al., 2016) or than an average EU household uses for lighting per day (Faberi et al., 2015), showering presents a resource-intensive everyday behavior ideal for testing these interventions.

The first intervention consists of social comparison (SC) reports that provide households with weekly information on their water use per shower and associated $CO_2$ emissions compared to peers. The second intervention provides real-time feedback (RTF) while showering. Specifically, we equipped participants' shower heads to glow in different colors depending on water consumption.

Both interventions convey a normative cue on appropriate amounts of resources to use per shower, enabling (social) learning about optimal energy use and influencing the moral costs of consumption (Allcott, 2011; Allcott and Kessler, 2019). SC delivers this cue retrospectively and periodically, comparing users' consumption to that of their peers and overlaying an injunctive norm emoticon that signals normative judgment of their own consumption relative to others (Schultz et al., 2007). RTF, by contrast, conveys its normative cue through progressive color changes: starting at green, the color signal shifts to blue, purple, red, and finally a flashing red, generally interpreted as a stop signal.[1] More importantly, however, RTF introduces a dimension beyond normative signaling: since it transmits information during the shower itself, it embeds real-time feedback directly into the moment of consumption. This immediacy enables users to adjust their behavior on the spot, better aligning actual use with their pref-

[1]The normative cue in RTF interventions has also taken other forms; for example, Tiefenbeck et al. (2018) used an animation of a polar bear standing on an ice floe that progressively shrank as consumption increased.

erences. Because resource consumption during routine actions is typically a *shrouded* attribute (Gabaix and Laibson, 2006), making the amount consumed salient through actionable, real-time information should facilitate behavioral change (Jessoe and Rapson, 2014). We also test the combination of SC and RTF as a third intervention in our field experiment.

We conducted a field experiment with 1,003 participants across 574 households. Households were randomly assigned to SC, RTF, both (BOTH), or one of two control groups. We tracked shower behavior (water use, temperature, duration) and elicited participants' willingness to pay (WTP) for each intervention, both before and after the intervention experience. The WTP elicitation was incentivized and captures consumers' perceived net welfare effects, including private benefits (lower bills, comfort) and potential psychological costs or gains.

Our findings show that RTF reduces water and energy use by about 29%, whereas SC yields a 9% reduction. The combination yields the largest point estimate, a 35% reduction, although its difference from RTF alone is not statistically significant at conventional levels.

Turning to consumer welfare, our incentivized WTP elicitations show that participants, on average, positively value all three interventions, with average WTP of roughly 9–10 EUR for one month of treatment. Participants attach a roughly 5–10% higher value to RTF and BOTH relative to SC, both in the WTP elicitation preceding and the WTP elicitation following treatment experience. Valuations remain broadly stable between the two elicitations; the only statistically significant change is a small decline in average WTP for SC, which is not attributable to participants who experienced SC. For all three interventions, we find no evidence that experiencing the intervention significantly affects WTP for that intervention, suggesting close alignment between ex ante expectations and ex post assessments. Participants' WTP likely reflects perceived cost savings and non-monetary benefits, net of non-monetary costs such as forgone consumption utility from longer showers. Section 3.5 examines these components. Average WTP exceeds both perceived and actual cost savings (Section 3.5),

suggesting that a sizable proportion of WTP reflects non-monetary benefits. We also examine whether assigning interventions based on positive WTP erodes conservation outcomes. We find no evidence that it does: treatment effects among households with positive WTP are similar to those in the full sample and, for SC, somewhat larger.

Finally, we estimate the marginal value of public funds (MVPF) for these interventions, a benchmark prominently used for the evaluation of government programs (Hahn et al., 2026). For private firms, our findings indicate that real-time feedback and social comparison services can reduce environmental impact while meeting consumer demand, making them relevant both to corporate sustainability initiatives and to commercial product offerings.

We contribute to the literature in three main ways. First, we add to a growing body of research examining the welfare implications of behavioral interventions. We directly elicit consumers' WTP to receive (or avoid) an intervention, following, for example, Allcott and Kessler (2019) and Butera et al. (2022), and contrasting with structural or sufficient-statistics approaches that infer welfare changes solely from observed behavioral changes (e.g. Chetty et al., 2009; DellaVigna et al., 2012; 2017; Rodemeier, 2021; Allcott et al., 2025; Löschel et al., 2023; Rodemeier and Löschel, 2023). This approach allows us to also capture any intangible effects when behavior remains unchanged (e.g., feeling guilty or proud, or experiencing a shift in how one perceives the choice itself). We further measure WTP both before *and* after participants experience the interventions, whereas closely related studies elicit WTP either after prior exposure (Allcott and Kessler, 2019) or contingent on future outcomes before those outcomes are realized (Butera et al., 2022). This within-participant pre- and post-experience elicitation provides first evidence on how ex ante expectations compare with ex post assessments.

Second, our results provide empirical evidence on how differences in the way behavioral interventions deliver information translate into both behavioral responses and consumer welfare. By comparing two widely studied interventions within the same experimental setting, we shed light on how different types of information pro-

vision affect outcomes, holding population, context, and measurement constant. The substantial effects we observe for the RTF and BOTH interventions underscore the importance of behavioral interventions that make shrouded consumption quantities salient at the appliance level and in real time (see also Jessoe and Rapson, 2014; Gerster et al., 2026). Our findings also speak to conceptual frameworks that distinguish interventions by how they influence choice, for example, by motivating behavioral change through messages or by facilitating it through the choice environment (see Congiu and Moscati, 2020; compare also Johnson et al., 2012; Hansen and Jespersen, 2013; Mongin and Cozic, 2018). While both of our interventions likely motivate resource conservation through the provision of normative cues signaling appropriate resource use while showering, the RTF feature adds a facilitating dimension by providing information that consumers can act upon while showering (see also section 3.4 for suggestive evidence on the behavioral patterns driving the results). The results of this comparison are thus consistent with the notion that, in this context, the addition of facilitating features increases both intervention effectiveness and consumer valuation.

Finally, we extend the literature on resource conservation. Prior studies have analyzed social comparison and real-time feedback in various domains (e.g., Allcott, 2011; Ferraro and Price, 2013; Allcott and Rogers, 2014; Brent et al., 2015; Tiefenbeck et al., 2018; 2019; Byrne et al., 2024; Goette et al., 2021; Andor et al., 2020), but rarely compare these interventions head-to-head on the same behavioral and welfare outcomes. Recent studies either compare only the behavioral effects of interventions (Brandon et al., 2019; Fang et al., 2023), or assess the welfare effects of a single intervention relative to no intervention (Allcott and Kessler, 2019). We instead compare social comparison, real-time feedback, and their combination on both dimensions, using the same domain and metrics. For our consumer welfare metric, we draw on incentivized willingness to pay elicitations, conducted both before and after participants experience the interventions. Our design evaluates, within one experiment, conservation effectiveness, consumer valuation, the stability of valuations after experience, and the effects of combining the interventions.

The remainder of the paper proceeds as follows: In Section 2, we detail our experimental design and study sample. Section 3 presents the behavioral impact of the interventions and investigates the underlying behavioral patterns. We then report participants' willingness to pay for receiving (or avoiding) the interventions and assess the interventions' impact on consumer welfare and welfare more generally. Section 4 concludes.

# 2 The field experiment

## 2.1 Technical equipment

We provided all study participants with a smart shower head[2], a WiFi gateway, and installation instructions. Details on the functioning and installation of the shower head are included in Appendix A6. The shower head and WiFi gateway, connected to the participants' WiFi network, formed a data infrastructure that allowed the shower head to continuously send data on the water use in liters, date, time, and temperature of every shower session to the researchers.[3] The participants did not have access to this data during the study.

[2]Each participating household received only one shower head. If the household had multiple showers, it was asked to install it on the shower they use most frequently. In principle, it is conceivable that participants could increase their use of the non-monitored shower in order to appear more resource-efficient. However, we consider such a reaction unlikely, as this behavior would not provide participants with any material advantage and would likely cause them inconvenience – both because of the effort required to change their habits and because of moral concerns, since they would be knowingly cheating. Table A4 in the appendix presents the estimates of the main treatment effects described in Section 3.2, distinguishing between households with only one shower and those with multiple showers; the results show only small and non-significant differences in treatment effects.

[3]We define a shower session as ended if the shower head is turned off for at least 180 seconds, i.e. the shower heads are configured so that they begin measuring a new shower session after a pause of at least 180 seconds. Any interruption of less than 180 seconds is counted as an interruption within a shower session, i.e. we assume that such an interruption likely constitutes a break for soaping rather than one person ending their shower and another person beginning their shower within this time frame. Naturally, any approach to separating different shower events carries the risk that some shower events will be misclassified. For example, a single shower event with a long soaping pause could be misclassified as two separate shower events, or two shower events that occur in quick succession could be misclassified as a single shower event. In Appendix A1.2, we provide some analyses to assess whether this occurs frequently in our data and affects our results. We conclude from these analyses that such misclassifications are unlikely to occur frequently in the data and that this does not have a significant impact on our results.

## 2.2 Sample

Participants were recruited in collaboration with the German survey institute *forsa*, which maintains a sample of panelists that is representative of German-speaking internet users aged 14 and above. In late August 2020, *forsa* sent out an invitation e-mail to 9,376 panelists, living in the Rhine-Ruhr-Area, Germany's largest metropolitan area.[4]

The invitation email conveyed some general information about the environmental impact of showering, information on the compensation for participation,[5] and information about the tasks as a study participant, which included installing the shower head and data infrastructure and completing several surveys throughout the study. The study was presented to the participants under the name "Project Sustainable Showering". Those who were interested in participating were referred in the email to a project website where more detailed information was provided and where they had the opportunity to register for study participation. The invitation email and the project homepage neither explained that a randomized experiment was to be conducted, nor which interventions were planned. Registration took place via an online form on the project homepage in which we queried contact details, consent to data processing within the study, and initial information on household size and technical circumstances to ensure that participants met the requirements to take part in the study. The content of the invitation email is depicted in Appendix A3, and the content of the project website is presented in Appendix A4.

1,100 of the invited persons registered for the study, representing 11.7% of those originally invited. We selected the final study participants among the eligible registrations[6] targeting smaller households with fewer showers in the home because we

[4]This area was chosen because two of the research institutions involved in the research project, [redacted] and [redacted], are located in this region, thus enabling potential technical support at the premises of the participating households.

[5]Participants were compensated for study participation in two ways. First, they were allowed to keep the smart shower head after the study period and could then control it independently using an app. Secondly, the participants received a monetary amount of 35 EUR, which was paid out to them at the end of the study in the form of a voucher of their choice.

[6]We had to delete 25 registrations because they were either obvious duplicates or the participants provided an address outside the Rhine-Ruhr-Area. Further, 203 registered persons did not meet the technical requirements needed for successful installation of the shower head and the data infrastructure.

consider it more likely for these households that the information provided to the participants in the course of the experiment would reach all shower users. Since we targeted a total sample size of around 600 households and expected some attrition during the study, we confirmed the participation of a total of 685 participants and sent them an invitation for the baseline survey in October 2020. 647 participants filled out the baseline survey within the given time frame. Those who did not fill out the survey after a reminder email were not considered for study participation anymore. From the outset, we divided the participants into two study waves. The first study wave began immediately, while the second wave began in January 2021. The division into the two study waves was made to logistically simplify the roll-out of the smart shower heads. This division allowed for regional clustering of the respective waves, which in turn minimized travel time for on-site technical support. Apart from the difference in timing, the research design was identical for both study waves. The final study sample consists of 574 households.[7]

**Table 1:** Socio-economic summary statistics

| Variable | Explanation | Mean | Std. Dev. | N |
|---|---|---|---|---|
| Age | Age of respondent | 53.85 | 13.31 | 567 |
| Female | Dummy: 1 if respondent is a woman | 0.39 | – | 572 |
| Household size | Number of persons living in the household | 1.75 | 0.59 | 574 |
| Children | Dummy: 1 if children live in the household | 0.07 | – | 574 |
| College degree | Dummy: 1 if respondent has a university degree | 0.49 | – | 565 |
| Number of showers | Number of showers installed in the household | 1.25 | 0.52 | 574 |
| Income | Monthly net household income in EUR | 3489.93 | 1445.03 | 541 |

Note: This table shows summary statistics for the final sample of study participants, i.e., participants who filled out the baseline survey and successfully installed the shower head infrastructure. The number of observations per variable varies, because participants had the option of not answering some of the questions. Since our main specification relies on fixed-effects models that do not require information on individual time-constant covariates, the lack of information on socioeconomic characteristics of some participants does not limit the size of our estimation sample. Household income was determined using an interval scale and converted into a continuous income measure for the empirical analysis, in which a participant's income was set to the value of the midpoint of the selected income interval.

Table 1 provides summary statistics on the study sample. As a result of our selection process, the average household size is smaller than that of the German population (1.75 compared to 2.0; German Federal Institute for Population Research, 2023b) as is

[7]The remaining 73 households either dropped out after the baseline survey or were unable to install the smart shower heads or the data infrastructure.

the share of households with children (7% compared to 29%; German Federal Statistical Office, 2022). Correspondingly, the monthly net household income is slightly lower than in the population (3,490 EUR compared to 3,612 EUR; German Federal Statistical Office, 2021). The study sample furthermore consists of slightly older participants (54 years compared to 51 years; own calculation of the average age of the German adult population based on German Federal Institute for Population Research, 2023a), a higher share of participants with a college degree (49% compared to 19%; German Federal Statistical Office, 2022) and a lower share of female participants (39% females compared to 51%; German Federal Statistical Office, 2022).

Given the differences between the study sample and the general population – which are almost inevitable in field trials that require active participant involvement – we advise against relying solely on the results of our experiment to justify a large-scale, mandatory implementation of our interventions. Before doing so, it would be important to conduct pilot studies with individuals who are unlikely to voluntarily participate in studies like ours, to ensure that these groups would not suffer disproportionately negative welfare effects as a result of the interventions. Yet we are confident that our recruitment of participants from a representative population sample makes our results highly informative for attempts to roll out our interventions on a voluntary basis or as a commercial offer.

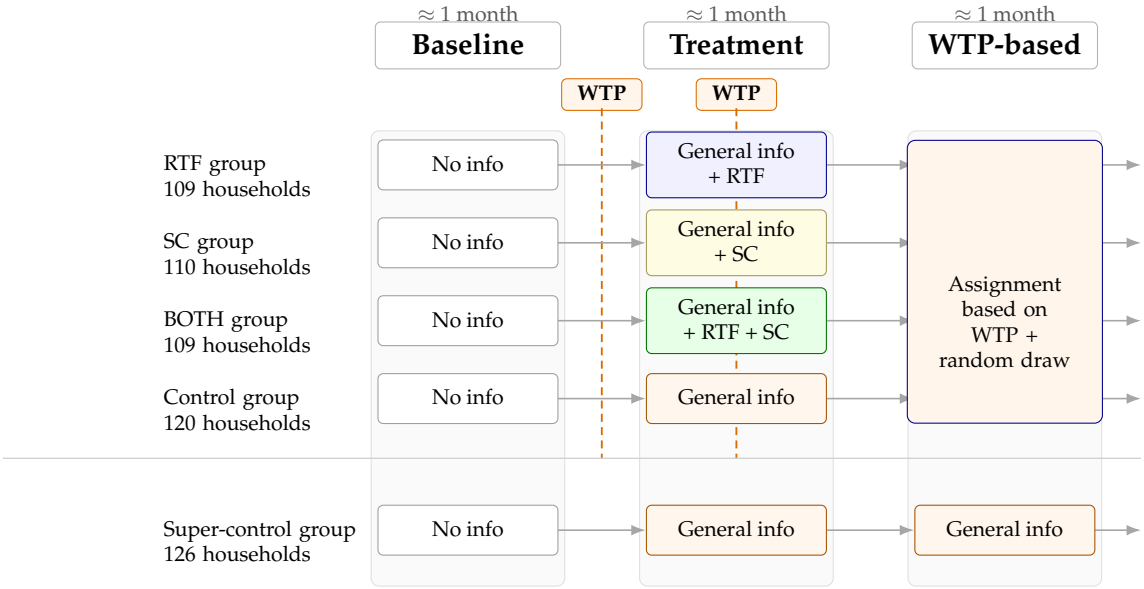
## 2.3 Experimental design

Our experiment consisted of three phases illustrated in Figure 1: A baseline phase, a treatment phase, and a WTP-based phase.[8] In the baseline phase, we tracked showering behavior without introducing any behavioral intervention. The baseline phase was followed by the treatment phase in which households were randomly assigned to receive or not receive the behavioral interventions. After that came the WTP-based phase in which we reassigned participants in the RTF, SC, BOTH, and Control groups

[8]Baseline phase of wave 1 (wave 2): From October 15, 2020 to November 20, 2020 (from January 1, 2021 to March 11, 2021). Treatment phase of wave 1 (wave 2): From November 21, 2020 to December 21, 2020 (from March 12, 2021 to April 14, 2021). WTP-based phase of wave 1 (wave 2): From December 22, 2020 to January 21, 2021 (from April 15, 2021 to May 14, 2021).

to interventions based on the combination of their willingness to pay to receive them and a random draw. Participants in the Super-control group remained untreated by design, because their behavior formed the reference benchmark for the other groups. Correspondingly, we did not conduct incentivized WTP elicitations in the Super-control group.[9] The purpose of the WTP-based phase was to render the WTP elicitation consequential and thus incentive compatible for participants outside the Super-control group in the sense of Carson and Groves (2007).[10] The experimental design was pre-registered with the AEA RCT Registry as trial [redacted] (Anonymous, 2020) and IRB approval was obtained from the GfeW ([redacted]).

**Figure 1:** Overview of the experimental design




Note: This figure shows the experimental design. Baseline duration varied depending on when households successfully connected their shower heads. In the analysis sample, the observed baseline period averaged 39 days, measured from each household's first to last valid baseline shower, inclusive. The analyses use all available valid baseline observations. See footnote 8 for exact dates. Receipt of the interventions during the treatment phase was determined by a random draw. Receipt of the interventions during the WTP-based phase depended on WTP indicated during the two elicitations and a random draw, for participants in the RTF, SC, BOTH, and Control group. "WTP" denotes consequential WTP elicitations for these groups. Participants in the Super-control group filled out hypothetical WTP elicitations that were not incentivized; these are omitted from the figure for clarity.

[9]Super-control group members instead complete hypothetical WTP elicitations, which we do not use in the main analyses.

[10]Details on the assignment mechanism used to implement the WTP elicitation are provided in Appendix A1.3.4.

### 2.3.1 Experimental groups

For the treatment phase, we split the sample into five approximately equal-sized experimental groups. We allocated the experimental groups randomly, with stratification ensuring that they did not significantly differ from each other concerning baseline water use per shower, age, household size, number of showers installed in the household, environmental attitudes, perceived social norms regarding energy use, and education. The five experimental groups are characterized as follows:

**Super-control group:** This group received no behavioral intervention throughout the study and served two purposes: First, it provided data on showering behavior over time in the absence of any intervention and second, this group was used to form reference groups needed for the behavioral interventions: Each participant who was not in the Super-control group was randomly assigned to a reference group consisting of nine members of the Super-control group. Participants who received the social comparison reports were compared to their reference group in the comparison reports. Similarly, the thresholds in the real-time feedback intervention were based on the behavior of the reference group. Because the Super-control group was not eligible for an intervention in either phase of the study, the WTP elicitations filled out by these households were purely hypothetical.[11] Yet, to ensure that these participants stayed at a similar attention level regarding their study participation as the other participants, they received weekly emails ("newsletters") with general information about showering and its energy intensity during the treatment and WTP-based phases; as did all other study participants. The upper panel of the exemplary newsletter in Appendix A5 contains one example of the general information texts.

**Control group:** Like the Super-control group, this group received weekly newsletters but no intervention in the treatment phase. However, just as the SC, RTF and BOTH group, this group was eligible for receiving an intervention in the WTP-based phase, making this group's WTP elicitation binding and thus incentive compatible. We

[11] This concerns only participants in this group - all other participants are incentivized to indicate their true WTP values since these co-determine the interventions these households receive during the WTP-based phase.

therefore pool the Control and Super-control groups only in analyses of baseline and treatment-phase shower behavior that do not rely on incentivized WTP or treatment-specific reference points. In WTP, welfare, opt-in, threshold, and reference-point analyses, we keep the groups distinct and state the respective sample restrictions.

**Social comparison (SC) group:** During the treatment phase, participants assigned to this group received a social comparison section in their weekly newsletters in addition to the general information section. In this social comparison section, the average water use per shower (and corresponding $CO_2$ emissions, given the average water temperature) was reported and compared to the average of the participant's reference group, which consisted of untreated households from the Super-control group. Furthermore, weekly averages of the water use per shower of both the participant household and its reference group were plotted in a time diagram. In addition, participants were shown how their water consumption per shower ranks compared to the nine reference households, i.e., they were assigned a rank between one (lowest use) and ten (highest use). For each rank, an additional emoticon was displayed: A smiling face for comparatively low consumption, a neutral face for medium consumption, and a frowning face for high consumption. To ensure that enough data was available to calculate the social comparison, the information was updated every two weeks, so that the second newsletter repeated the information of the first newsletter, the fourth newsletter repeated the information of the third newsletter, and so on. The frequency with which we send and update the social comparison reports is similar to, or even relatively frequent compared to, other studies and applications from commercial companies (e.g., Allcott, 2011; Allcott and Rogers, 2014; Andor et al., 2020). The middle panel of the exemplary newsletter in Appendix A5 is an example of the social comparison information.

**Real-time feedback (RTF) group:** For participants assigned to this intervention, the real-time feedback function was activated in participants' shower heads. The LEDs built into the shower head signaled the water consumption of the shower in real time. Shower heads were configured to light up green for the first 15 liters of a shower.

Thereafter they turned blue, purple, red and eventually started to flash in the red color. The thresholds at which the colors changed were determined by the behavior of the participant's untreated reference group. In detail, the shower head would start to flash red once the reference group's average water use per shower over the past two weeks was exceeded. The thresholds of the remaining colors were allocated in regular intervals between 15 liters and the average water use per shower of the reference group. For these participants, an RTF section was added to the weekly newsletter. This section contained information on the water consumption corresponding to each of the color thresholds, but participants were not informed about the origin of the color thresholds. For each threshold in liters, the corresponding average $CO_2$ emissions caused by heating the water to the usual water temperature of the household were also detailed. To ensure that enough data was available to reliably construct the thresholds, the thresholds were updated every second week, as they were for the SC group so that the newsletter in between was a repetition of the newsletter of the week before. The lower panel of the exemplary newsletter in Appendix A5 is an example of how the real-time feedback functionality was communicated.

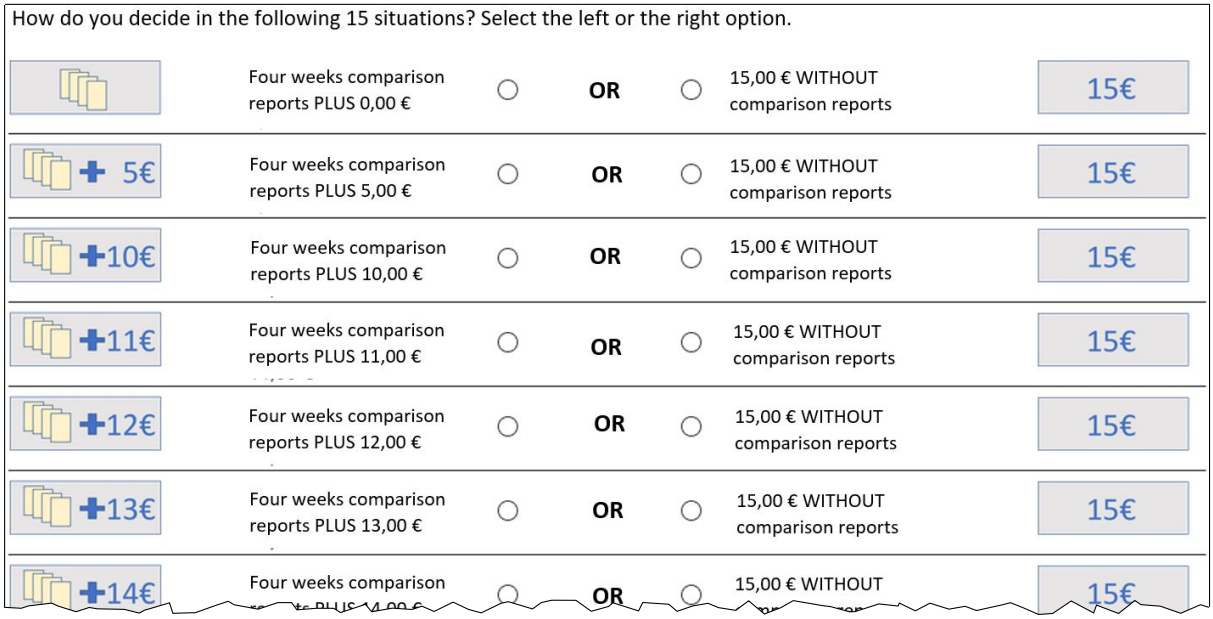
**BOTH group:** For participants assigned to this intervention, the SC section was added to the weekly newsletters, and the RTF function in the smart shower heads was activated. Participants thus received both the interventions of the SC group and the RTF group. The weekly newsletters of this group thus contained an SC section as well as an RTF section, and the full newsletter in Appendix A5 is a representative newsletter of a participant that received the BOTH intervention. Having the information from both interventions, the participants in this group were able to infer that their shower thresholds depended on their reference group's average water use per shower.

### 2.3.2 Elicitation of willingness to pay for the interventions

Towards the end of the baseline phase as well as the end of the treatment phase, participants completed a willingness to pay (WTP) elicitation, in which they indicated how much they were willing to pay to receive each of the interventions (SC, RTF, and BOTH) in the WTP-based phase. As described above, these WTP decisions were con-

sequential for participants in the RTF, SC, BOTH, and Control groups, but hypothetical for participants in the Super-control group.

**Figure 2:** Illustration of the multiple price list for SC




Note: This figure is abridged - left side options continue with 15 €, 16 €, 17 €, 18 €, 19 €, 20 €, 25 €, 30 €. Multiple price lists for RTF and BOTH were identical apart from the left-hand-side illustrations.

The WTP elicitation was conducted in the form of a survey which consisted of several multiple price lists (MPL), with one MPL for each of the three interventions. In each MPL, the participants were asked to decide if they preferred to receive four weeks of the intervention in question plus a varying monetary amount or whether they preferred to forego the intervention and receive an amount of 15 EUR. An example of an MPL eliciting WTP for the social comparison reports is provided in Figure 2. The participants had to fill out an MPL for each treatment twice, once before and once towards the end of the treatment phase, totaling six MPLs. Within each WTP elicitation, the order of the SC and RTF MPLs was randomized, while the MPL for BOTH was always shown last because it combined the two preceding interventions. Except for the Super-control group, these elicitations were binding in that one of the decisions made, i.e., the decision of a randomly drawn row (of 15) in one randomly selected MPL (out of six), was actually implemented in the WTP-based phase. So the participant would, depending on her decision, receive the treatment and the corresponding

monetary amount or the fixed amount of 15 EUR.[12]

Our WTP elicitations thus allow us to elicit, in an incentivized manner, how much consumers are willing to pay to receive any of the interventions for a four-week period. For example, if in a given choice a participant would prefer to receive RTF and 12 EUR to receiving 15 EUR, we interpret this as the consumer having at least a WTP of 3 EUR for receiving RTF for 4 weeks. If this choice was selected for implementation, we would activate the RTF feature of this participant's shower head during the WTP-based phase, without this requiring any effort by the participant. Since participants' decisions are consequential, it is costly for participants to under- or overstate their WTP, and we thus avoid hypothetical response bias. This incentive argument applies to all participants except those in the Super-control group, whose WTP elicitations were explicitly hypothetical and are not used as incentivized WTP measures in our welfare analyses. Also, note that all participants kept the shower head at the end of the study as part of their study compensation (see A3 for the study invitation and information on participant compensation), and it was made clear to participants that this would not be affected by the WTP elicitation. Before eliciting participants' WTP, we explained the structure in detail and provided examples.

In our interpretation of WTP values, we take a conservative approach and set the WTP values to the respective lower limit of the respective interval. Thus, those who always opt for the intervention in an MPL would receive a WTP value of 15 (since they are willing to forego at least 15 EUR for receiving the intervention), those whose choices imply, for example, a WTP between 3 and 4 EUR receive a WTP value of 3 EUR, and so forth. Since the lower end of the MPL is an open interval limit for those who always chose the voucher without an intervention, we set the WTP to -20 in this case (since they are willing to forego at least 15 EUR for not receiving the intervention). In Appendix A1.3.6, we verify the robustness of our conclusions with regard to an alternative, less conservative approach to coding WTP values.

[12]The survey interface enforced monotonicity in the MPL responses. After the first switch between the intervention option and the monetary alternative, the remaining rows were auto-completed consistently. Participants could revise their choices before leaving the page, but inconsistent MPL responses could not be submitted.

# 3 Results

## 3.1 Data

**Table 2:** Baseline summary statistics by treatment-phase group

| Variable | Unit | (Super-) control | SC | RTF | BOTH |
|---|---|---|---|---|---|
| **Panel A: Baseline shower behavior** | | | | | |
| Baseline water use | Liters per shower | 34.07 | 32.16 | 31.89 | 32.31 |
| | | | (-0.73) | (-0.93) | (-0.34) |
| Baseline water flow | Liters per Minute | 9.18 | 9.21 | 9.05 | 8.87 |
| | | | (0.28) | (-0.04) | (-1.19) |
| Baseline Shower frequency | Number of showers per day | 0.96 | 1.12 | 1.08 | 0.99 |
| | | | (1.13) | (1.31) | (0.38) |
| Baseline shower water temperature | °C | 38.12 | 37.17 | 36.88 | 37.81 |
| | | | (-1.74) | (-1.18) | (-0.84) |
| No. of households | | 236 | 110 | 109 | 109 |
| No. of observations | | 8,416 | 4,207 | 4,360 | 4,056 |
| **Panel B: Baseline willingness to pay** | | | | | |
| Baseline WTP for SC | EUR | 9.64 | 8.93 | 9.35 | 10.41 |
| | | | (-0.71) | (-0.29) | (0.81) |
| Baseline WTP for RTF | EUR | 10.80 | 9.62 | 9.63 | 10.37 |
| | | | (-1.23) | (-1.14) | (-0.47) |
| Baseline WTP for BOTH | EUR | 10.58 | 10.06 | 9.41 | 11.13 |
| | | | (-0.56) | (-1.15) | (0.61) |
| No. of households | | 118 | 109 | 105 | 107 |
| No. of observations | | 118 | 109 | 105 | 107 |

Note: Panel A shows baseline shower averages. The test statistics in parentheses are t-values for comparison to the (Super-)control group obtained from regressions with baseline water use, water flow, shower frequency, or water temperature as the outcome variable and group membership as the explanatory variable. The regressions for water use, water flow, and water temperature include calendar-date fixed effects and household-clustered standard errors. The shower-frequency comparison uses household-level OLS with heteroskedasticity-robust standard errors. Panel B shows baseline WTP averages in EUR. In Panel B, the control column includes only the Control group, not the Super-control group, because WTP was hypothetical for Super-control households. Panel B t-values are obtained from robust OLS regressions of baseline WTP on treatment-group indicators. The p-values from joint multinomial assignment-prediction tests are 0.15 for Panel A and 0.49 for Panel B. No. of observations refers to shower sessions in Panel A and baseline WTP responses in Panel B. 10 households in the (Super-)control group were only able to properly connect their shower heads after the baseline phase, so they are excluded from Panel A and all analyses in this section that use baseline shower data.

The four treatment-phase groups are balanced on baseline shower behavior and baseline willingness to pay for receiving each of the interventions (Table 2).[13] For the baseline phase as well as the treatment phase, we pool the Super-control group and the control group to maximize power and refer to this pooled group as the "(Super-)con-

trol group". We see a slightly higher baseline shower temperature in the (Super-)control group compared to the three treatment groups. The joint randomization checks do not reject balance in either panel, and baseline differences between households are controlled for by household fixed effects in the regressions. See Appendix A1.2 for further technical details on how showers are captured.

## 3.2 Impact of the interventions on behavior: Average Treatment Effects

Figure 3 depicts the daily average water use per shower in the different treatment groups in the baseline phase and in the treatment phase.[14] During the baseline phase, average water use per shower is not distinguishable between the (Super-)control group and the treatment groups. The start of the treatment phase, as indicated by the dashed vertical line in Figure 3 leads to a sharp and immediate drop in water use per shower for all groups that received treatment, while the behavior of the (Super-)control group does not change systematically. Moreover, visual inspection suggests that RTF and BOTH lead to stronger reductions in water use per shower than SC.

In a next step, we estimate the treatment effects using a difference in differences regression model of the following form:

$$y_{jit} = \alpha_i + \beta_1 SC_{it} + \beta_2 RTF_{it} + \beta_3 BOTH_{it} + \tau_t + \epsilon_{jit}, \tag{1}$$

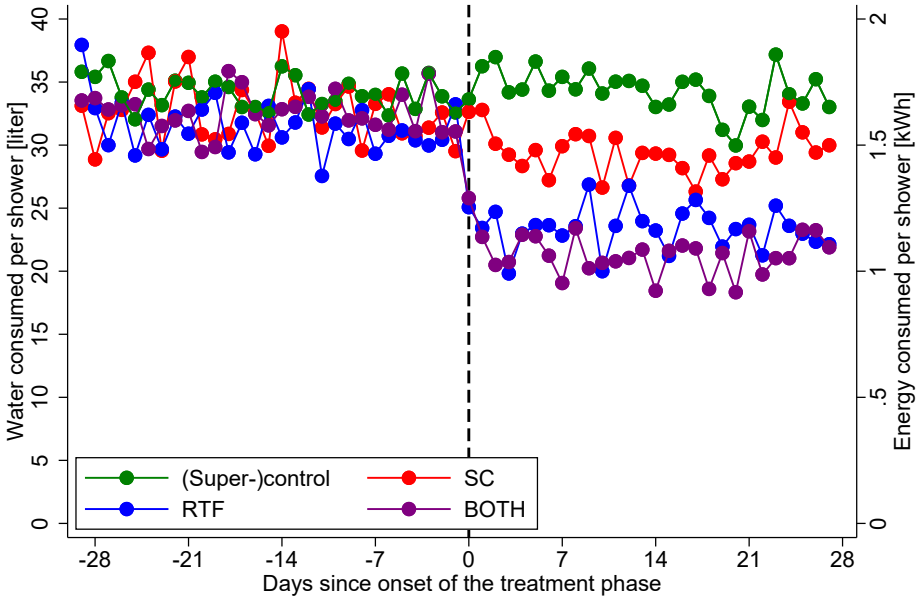
where $y_{jit}$ represents the outcome variable, i.e., water used during shower $j$ by household $i$ on day $t$. $\alpha_i$ represents household-specific fixed-effects. $SC_{it}$, $RTF_{it}$ and $BOTH_{it}$

[13]We cleaned the raw shower data by removing showers with water consumption above 200 liters or below 5 liters. This is necessary so that water withdrawals that are very likely not showers do not add additional noise to the results. A similar approach is taken by Fang et al. (2023). It should also be noted that showers taken while the WiFi gateway was not connected to the Internet were, for technical reasons, recorded with the time stamp of the next successful data transmission from the WiFi gateway and not the actual time stamp of the shower. This is the case for 17.2% of showers taken throughout the entire study period. To correct this, we redefine the time stamps of such offline showers and allocate them evenly over the period between the last and the next successful data transmission.

[14]This graph pools data from the two study waves, such that the onset of the treatment phase was on November 21, 2020 for the first study wave and on March 12, 2021 for the second study wave.

are indicators of household $i$ receiving the respective treatment at time $t$. $\tau_t$ represents day-specific fixed effects and $\epsilon_{jit}$ is the error term. Standard errors are clustered at the household level.

**Figure 3:** Treatment phase - average water use per shower by day and treatment group




Note: This figure compares average water use per shower between treatment groups, shown on the left vertical axis. The right vertical axis translates water consumption into energy consumption required to heat the water, given a water temperature of 38 °C (the average water temperature in our sample). Day 0 marks the start of the treatment phase; negative values indicate days before treatment onset. The vertical dashed line separates baseline and treatment phase.

The estimates of the average treatment effects are shown in Table 3. The SC treatment reduces water use per shower by 3.19 liters, the RTF treatment by 9.79 liters, and the BOTH treatment by 11.91 liters. Relative to the average water use of the (Super-)control group in the treatment phase, which is 34.1 liters, these effects correspond to a 9.4%, 28.7%, and 35.0% reduction in water use per shower. All effects are statistically significant at the 1% level. The effect of SC is significantly smaller than the effects of RTF and BOTH (p-values below 0.01) and the effect of BOTH is somewhat larger than that of RTF (p-value of the difference: 0.13). The robustness checks presented in Appendix A1.3.1 indicate that the estimates of the effects remain largely unchanged if household or time fixed effects are excluded from the regression model. In addition, the interventions have only a negligible effect on shower temperature and shower frequency. We can thus interpret a decrease in water use per shower as an equivalent decrease in energy consumption. Effects are qualitatively similar when restricting the

sample to only one- or multi-person households. [15]

**Table 3:** Average treatment effects on water use per shower

| | (I) Volume (Liters) | (II) Relative to control group |
|---|---|---|
| SC | -3.19*** | -9.4% |
| | (0.66) | |
| RTF | -9.79*** | -28.7% |
| | (0.88) | |
| BOTH | -11.91*** | -35.0% |
| | (1.22) | |
| (super-)control group average | 34.1 | |
| No. of households | 564 | |
| No. of observations | 38,453 | |

Note: OLS estimates with household and calendar-date fixed effects included. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The (Super-)control group average corresponds to the average water use per shower of the control and Super-control group in the treatment phase. Column (II) contains manual calculations obtained by dividing the treatment effects in column (I) by the (Super-)control group average. These percentages are indistinguishable from those obtained when the dependent variable is divided by the average of the (Super-)control group before performing the regression analysis.

The effect of 28.7% that we find for RTF is slightly larger than that found in Tiefenbeck et al. (2018), who estimate that RTF reduces water use per shower by 22%. The 9.4% reduction in water use per shower caused by the SC intervention is substantially larger than that reported by Allcott (2011), who finds that SC reports decrease aggregate household electricity use by 2%. One reason for this difference could be that our SC reports are appliance-specific and thus could make behavioral adjustments easier (cf. also Gerster et al., 2026), and changing showering behavior might generally be easier than changing overall electricity consumption behavior. Further, our study design, and in particular our weekly communications, may have made showering resource use particularly salient to participants. While our design abstracts from any general salience effects due to the Control group also receiving shower-focused communications at the same frequency, increased salience may have moderated treatment

[15] As an additional check on the treatment-phase results, we estimate intervention effects separately in the subsequent WTP-based phase. In that phase, a random draw determined which intervention an eligible participant could receive, while actual receipt also depended on the participant's WTP. Using the random draw as an instrument for actual receipt yields causal estimates that are close to the treatment-phase estimates for RTF and BOTH, while the estimate for SC is larger. See Appendix A1.3.4.

effects.

## 3.3 Impact of the interventions on behavior: Evidence on opt-in designs

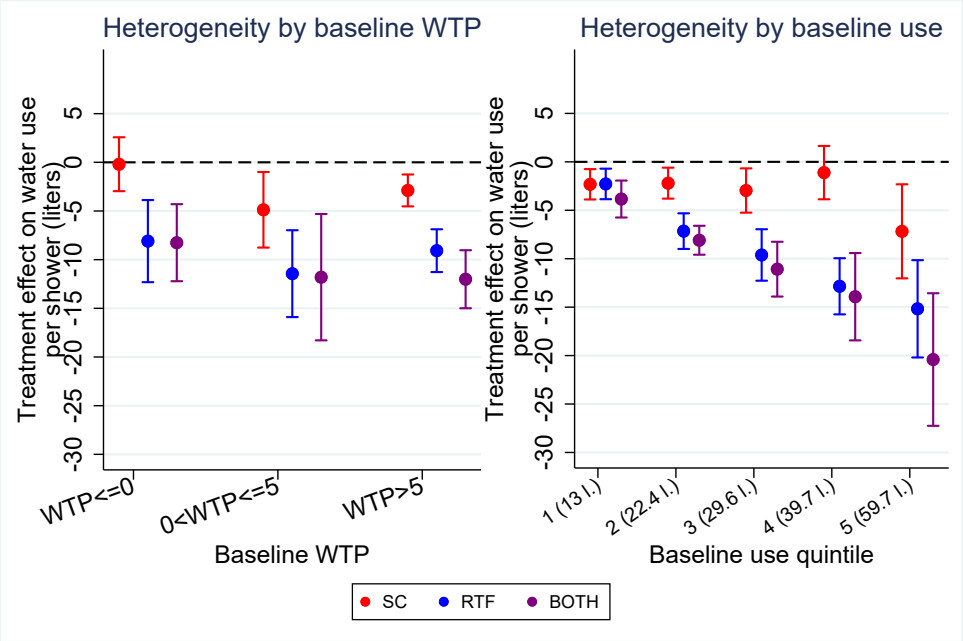
Non-mandatory behavioral interventions are typically adopted on an opt-in basis or allocated through market-based mechanisms when provided by private firms. The aggregate effects of an intervention allocated through market-based mechanisms thus do not depend primarily on average treatment effects, but on the treatment effects observed among those households with a sufficiently high WTP for the respective intervention. However, households with a high WTP for these interventions likely also have a high underlying motivation to conserve. Thus, they may have already adjusted their behavior previously, forming habits that conserve resources; or resource conservation may already be particularly salient to them in the absence of an intervention. Such mechanisms would decrease potential treatment effects. Our experiment design allows us to directly examine heterogeneity in treatment effects by baseline WTP, providing evidence on the effects one may expect with an opt-in design. Results are shown in the left panel of Figure 4.

We find no evidence of attenuated treatment effects at higher WTP for any intervention, indicating that conservation effects are not weaker among households that would self-select into treatment. In contrast, for the SC intervention, treatment effects are instead higher for households with a positive WTP. [16] For BOTH, effects are also somewhat larger for households with positive WTP levels, although the difference is not statistically significant. Treatment effects for RTF do not vary meaningfully with WTP. Appendix A1.3.4 provides further suggestive evidence on delivering interventions on an opt-in basis, estimating the causal effect of the interventions on participant behavior for those with a sufficiently high WTP to receive treatment during the WTP-based phase. We estimate larger effects for SC than in the main analysis that assigns

[16] Lim (2025) uncovers a similar pattern for workers' WTP for peer comparison information and its effectiveness. They find no effect of the information on workers' productivity when workers' WTP for it is zero or negative.

treatment receipt randomly (Table 3), while estimates for RTF and BOTH are similar.

**Figure 4:** Treatment effect heterogeneity




Note: The left panel examines heterogeneity in treatment effects by households' WTP for the allocated intervention. It reports separate household fixed-effects regressions for each intervention and baseline-WTP category. Each specification compares households assigned to the respective intervention with Control households in the same intervention-specific baseline-WTP category, which accounts for potentially differing time trends between baseline-WTP groups. Super-control households are excluded from the left panel because their WTP elicitation was hypothetical. The right panel examines heterogeneity in treatment effects by baseline water use quintile. The specification includes quintile-specific treatment-phase controls and household fixed effects. In both panels, standard errors are clustered at the household level. Markers indicate point estimates, with bars denoting 95% confidence intervals.

## 3.4 Impact of the interventions on behavior: Behavioral patterns

In line with prior evidence on the impact of making shrouded consumption quantities salient at the moment of choice (e.g. Jessoe and Rapson, 2014), we find that RTF reduces resource use substantially. In addition, the effect of RTF strongly exceeds that of SC. Furthermore, the point estimate for the combined intervention exceeds that of RTF alone, although the difference is not statistically significant ($p = 0.13$; Section 3.2). This pattern is consistent with an additional role of the cues provided by SC. In this section, we examine treatment-effect heterogeneity and explore secondary outcomes to assess behavioral patterns consistent with the results.

First, we examine whether experiment participants shower in a more targeted and quantity-oriented manner when experiencing the interventions. If RTF and BOTH in-

fluence consumption behavior by making quantities salient, we would expect to see evidence of participants targeting a certain water consumption. Table 4 shows that RTF and BOTH reduce within-household variation in water use per shower substantially more than SC, consistent with participants targeting a certain self-set quantity threshold. Table 5 points in the same direction: participants receiving RTF and BOTH are much more likely to end showers in lower color ranges and substantially less likely to continue until the flashing-red range than participants receiving SC. One effective strategy to decrease overall resource consumption can be to take short pauses within the shower (e.g. for shampooing).[17] We find that SC does not significantly affect pauses, whereas RTF and BOTH increase the incidence of short pauses (Appendix Table A3). All three patterns suggest that RTF and BOTH allow individuals to exercise greater control over their behavior and act in a more targeted manner.

**Table 4:** Average treatment effect on the standard deviation of water use per shower within households

| | (I) Volume (Liters) | (II) Relative to control group |
|---|---|---|
| SC | -2.23*** | -13.3% |
| | (0.68) | |
| RTF | -4.29*** | -25.6% |
| | (0.71) | |
| BOTH | -5.83*** | -34.8% |
| | (0.74) | |
| (super-)control group average | 17 | |
| No. of households | 564 | |
| No. of observations | 1,127 | |

Note: OLS estimates with household and phase fixed effects included. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The (Super-)control group average corresponds to the standard deviation of water use per shower of the control and Super-control group in the treatment phase. Column (II) contains manual calculations obtained by dividing the treatment effects in column (I) by the (Super-)control group average.

Second, we investigate how the treatment effects relate to baseline water use per shower. Numerous resource-conservation studies have shown that resource-saving interventions are particularly effective in households with high resource consumption

[17] Note that our analysis as well as the smart shower head accommodate pauses up to 180 seconds, i.e. we still count such occurrences as a single shower and the smart shower head also does not reset after a pause. See section 2 for details.

**Table 5:** Average treatment effect on the proportion of showers ended within a certain (hypothetical) color threshold

| | (I) | (II) | (III) | (IV) | (V) |
|---|---|---|---|---|---|
| | Effect on the proportion of showers ended while the shower head was / would have been... | | | | |
| | ...green | ...blue | ...purple | ...red | ...flashing red |
| SC | 0.04** | 0.02 | -0.00 | 0.01 | -0.07*** |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.02) |
| RTF | 0.11*** | 0.07*** | 0.03 | -0.00 | -0.21*** |
| | (0.02) | (0.02) | (0.02) | (0.02) | (0.03) |
| BOTH | 0.24*** | 0.01 | -0.02 | -0.01 | -0.22*** |
| | (0.03) | (0.03) | (0.02) | (0.02) | (0.03) |
| Control group average | 0.24 | 0.12 | 0.16 | 0.12 | 0.36 |
| No. of households | | | 446 | | |
| No. of observations | | | 19,190 | | |

Note: OLS estimates with household and phase fixed effects included. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. Each shower session is retained as a separate observation. The control-group averages are calculated using Control households in the treatment phase; Super-control households are excluded from both the regressions and these averages. The dependent variable is a dummy variable indicating whether a shower ended within a given color threshold. Color classifications use the reference value applicable at the time of each shower. Because reference values were updated every two weeks, a given water volume can correspond to different color categories on different dates. For the baseline phase and for the treatment groups that did not receive real-time feedback, i.e., SC and the control group, the color thresholds represent only hypothetical values, that is, the thresholds indicate the colors that a given shower head would have signaled if real-time feedback had been enabled. We calculated these hypothetical color thresholds as we did for the groups receiving real-time feedback, i.e., we based them on the average water consumption per shower of a participant's reference group, which consisted of participants from the Super-control group. Therefore, participants in the Super-control group are not included in these analyses because there are no valid hypothetical color thresholds available for them. See Table A11 in the Appendix for an overview of the individual transitions between color thresholds.

(see, for example, Allcott, 2011; Andor et al., 2020). The right panel of Figure 4 shows that baseline consumption plays a role in all treatments, but there is a clear difference between the SC treatment and the other two interventions. With RTF and BOTH, the conservation effect increases almost linearly with rising baseline consumption. For the social comparison treatment, treatment effects are highest in the group with the highest consumption; for the two groups with lower consumption, estimated treatment effects are either non-significant or relatively small. This suggests that the normative cue from social comparison information seems to mainly affect high consumers. This pattern is consistent with the results of the literature on social comparisons, which show that such comparisons lead to energy savings, particularly among high-consumption

households (Allcott, 2011; Andor et al., 2020; Byrne et al., 2018; Schultz et al., 2007). It also appears to carry over, to some extent, to the BOTH group, where the treatment effect in the quintile with the highest consumption appears to exceed that of the RTF treatment by about one-third, while the effects in the other quintiles are closer together; however, this difference is not statistically significant at conventional significance levels.

Third, we examine differences in the stringency of the normative cue communicated to treatment households. The idea is to shed further light on the patterns identified regarding baseline use per shower: One reason high baseline users may be displaying larger treatment effects is that they are more likely to experience a more stringent normative cue. In the SC treatment, the normative cue is conveyed through a comparison to nine Super-control group households that the treatment household is randomly matched to. In the RTF treatment, the normative cue is conveyed through the shower head changing colors once a certain threshold is reached. These thresholds are also based on the behavior of nine Super-control group households, although this connection is not communicated to households unless they receive BOTH (see Section 2.3.1 and A1.3.2 for details). Since households were randomly matched to different reference groups, this process created exogenous variation in treatment stringency. For SC, being matched to a reference group with frugal showering behavior means a higher likelihood of being ranked less favorably in the SC reports.[18] For RTF, being matched to a reference group with frugal showering behavior means that the color phases have shorter durations, with the shower head blinking red sooner.

Although the variation in the reference points is not very pronounced, Appendix Table A10 shows suggestive evidence of stronger effects among treatment households matched to more frugal Super-control households, supporting the idea that normative cues play a role in driving treatment effects across all treatments. The importance of this channel seems to be stronger (in relative terms) for SC than for RTF and BOTH.

[18]Note that directly examining heterogeneity by reported ranking would be confounded by several factors, such as showering behavior at baseline, while the reference group is randomly assigned and thus exogenously influences the favorability of the received ranking. We thus focus on heterogeneity by reference group behavior in our analysis. See section A1.3.2 for details.

## 3.5 Impact of the interventions on consumer welfare

As a further step, we analyze the WTP data, which we expect to subsume the utility loss and the perceived monetary savings due to shorter showers as well as any psychological costs and benefits caused by the interventions. This information is central to assessing the interventions' consumer welfare effects and – if consumers exhibit a positive willingness to pay – can be useful for companies in developing business models for offering such interventions. To provide a clean and incentivizable measure of WTP, we elicit participants' WTP to experience each of the interventions for one month, not for the shower head itself.[19] Our study is unique in that we elicit participants' WTP both before and after experiencing the interventions. By eliciting WTP before experiencing the interventions, we gather information on the anticipated utility of the interventions for the participants, or in other words, how much they are willing to pay in the hopes of gaining from the interventions. By comparing these values to WTP elicited after the intervention, we can ascertain if the participants' perception of the interventions has shifted.

Table 6 depicts the average WTP values and the shares of participants with negative WTP for the full sample of participants with incentivized WTP elicitations, i.e., excluding the Super-control group. The first column depicts WTP elicited in the baseline phase and the second column depicts WTP elicited at the end of the treatment phase. From panel I, we find that WTP for SC is significantly lower than that for RTF (10%-level) and BOTH (1%-level), but all average WTP figures are at around 10 EUR. In addition, WTP for SC is significantly lower in the treatment phase than in the baseline phase (5%-level). However, as indicated by the small and non-significant regression results in Table 7, this decrease does not appear to be driven by those participants who experienced SC.

[19] One advantage of this approach is that WTP values become more comparable between individuals: WTP for the shower head itself will naturally be influenced by participants' expectation of the durability of the shower head and future consumption utility, which will likely differ across participants. As the participants were informed that they could keep the shower heads after the end of the study, we do not believe that the participants included their evaluation of the shower head itself in their WTP. See section 2.3.2 for details.

**Table 6:** Average willingness to pay for the interventions

| | | WTP elicited in the baseline phase | | WTP elicited at the end of treatment phase | | |
|---|---|---|---|---|---|---|
| | | WTP amounts in EUR | | | | |
| | | Mean | Std. dev. | Mean | Std. dev. | Obs. |
| I | WTP for SC | 9.58 | 7.29 | 8.92 | 7.40 | 439 |
| | WTP for RTF | 10.12 | 7.49 | 9.53 | 7.59 | 439 |
| | WTP for BOTH | 10.31 | 7.19 | 9.76 | 7.39 | 439 |
| | | Differences in WTP between interventions | | | | |
| | | Difference | p-value | Difference | p-value | |
| | RTF-SC | 0.54 | 0.07 | 0.62 | 0.01 | |
| | BOTH-SC | 0.72 | 0.01 | 0.84 | 0.00 | |
| | BOTH-RTF | 0.18 | 0.52 | 0.23 | 0.20 | |
| | | Participants with negative WTP | | | | |
| | | Share | Std. dev. | Share | Std. dev. | Obs. |
| II | WTP for SC | 6.8% | 25.3% | 5.9% | 23.6% | 439 |
| | WTP for RTF | 5.9% | 23.6% | 5.7% | 23.2% | 439 |
| | WTP for BOTH | 6.2% | 24.1% | 5.5% | 22.8% | 439 |
| | | Differences in negative WTP between interventions | | | | |
| | | Difference | p-value | Difference | p-value | |
| | RTF-SC | -0.9 | 0.41 | -0.2 | 0.78 | |
| | BOTH-SC | -0.7 | 0.55 | -0.5 | 0.59 | |
| | BOTH-RTF | 0.2 | 0.85 | -0.2 | 0.76 | |

Note: The sample consists of all participants who completed both WTP elicitations and who were not part of the Super-control group, as the WTP elicitations for the Super-control group were purely hypothetical. Six participants did not complete either WTP elicitation, and three of the remaining participants did not complete the second WTP elicitation. The reported differences and p-values are based on within-participant paired tests.

Focusing on the proportion of participants who indicated a negative WTP for the interventions (panel II of Table 6), i.e., participants who were willing to forgo money to avoid receiving the interventions, we find that these shares vary between 5 and 7 percent, with hardly any systematic differences between the interventions.

Participants' WTP for an intervention may depend on both monetary and non-monetary factors. Some factors may reduce WTP for an intervention, such as an expected loss of consumption utility from shorter showers or anticipated direct psychological costs from the intervention experience. Other factors may increase it: Conserving water and energy yields monetary savings and may bring emotional rewards. Participants may also simply want to learn more about their showering behavior.

A comparison of WTP with expected monetary savings helps distinguish the role of monetary from non-monetary considerations in driving WTP. Consider first the

**Table 7:** Average treatment effects on willingness to pay

| | Effect of treatment experience on WTP for | | | Effect of treatment experience on share with negative WTP for | | |
|---|---|---|---|---|---|---|
| Treatment group | SC | RTF | BOTH | SC | RTF | BOTH |
| SC | -0.14 | 0.02 | 0.36 | 0.06 | -0.03 | -0.00 |
| | (0.98) | (0.99) | (0.90) | (0.05) | (0.04) | (0.04) |
| RTF | -0.83 | 0.05 | 0.22 | 0.01 | -0.03 | -0.03 |
| | (0.82) | (0.93) | (0.76) | (0.04) | (0.04) | (0.03) |
| BOTH | -0.94 | -0.26 | -1.02 | 0.03 | 0.01 | -0.00 |
| | (0.83) | (0.90) | (0.76) | (0.04) | (0.04) | (0.04) |
| Control group average | 9.44 | 10.25 | 10.14 | 0.04 | 0.04 | 0.06 |
| No. of households | | | 439 | | | |
| No. of observations | | | 878 | | | |

Note: OLS estimates with household and elicitation-period fixed effects included. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The omitted group is the control group who also completed incentivized WTP elicitations but did not experience any treatment during the treatment phase. The sample consists of all participants who completed both WTP elicitations and who were not part of the Super-control group, as the WTP elicitations for the Super-control group were purely hypothetical.

simplifying assumption that households expect no non-monetary costs from the intervention. This assumption is unrealistic, since households plausibly derive enjoyment from long showers even when they are willing to shorten them for budgetary or environmental reasons. Nonetheless, it provides a useful benchmark. Under this assumption, a positive gap between WTP and expected monetary savings, that is, WTP exceeding expected savings, identifies expected non-monetary benefits. When non-monetary costs are allowed, the same positive gap no longer identifies non-monetary benefits exactly, but provides a conservative lower bound: expected non-monetary benefits must be at least as large as the positive gap.[20]

To assess perceived savings, we leverage a survey question about the perceived monetary savings that we elicited at the end of the treatment phase.[21] On average,

[20]Formally, WTP = expected monetary savings + expected non-monetary benefits - expected non-monetary costs, which implies WTP $-$ expected monetary savings $\leq$ expected non-monetary benefits. As an example, suppose a household would pay 10 EUR to receive the intervention and expects monetary savings of 6 EUR, so that its WTP exceeds expected savings by a positive gap of 4 EUR. If the household expects no non-monetary costs, its expected non-monetary benefits equal 4 EUR. If instead it expects a non-monetary cost of 1 EUR from shorter showers, its expected non-monetary benefits equal 5 EUR, since 10 EUR = 6 EUR + (5 EUR - 1 EUR). In this case the positive gap of 4 EUR is a conservative estimate of the 5 EUR in benefits.

[21]The question read: "Do you think INTERVENTION has helped you reduce your electric and water bills?" and if yes "Please estimate: "By how many euros did your energy and water costs decrease

participants in the SC group reported perceived savings of 4.12 EUR in the treatment month. Average perceived savings in the RTF group were 5.48 EUR and perceived savings in the BOTH group were 6.23 EUR. Using the average number of 31 showers per household in the treatment phase, the treatment effects in liters per shower from Table 3, and an estimated cost of about 0.01 EUR per liter[22] of warm shower water, we estimate the actual monetary savings for one month of treatment to be around 1 EUR for SC, 3 EUR for RTF, and 3.7 EUR for BOTH. This indicates that participants tended to overestimate the monetary savings from the interventions.

**Table 8:** Comparison of willingness to pay with perceived savings for those experiencing treatment

| Treat. (N) | Post-treatment WTP | | Perceived savings | | WTP minus perceived savings | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| SC (89) | 8.46 | (6.93) | 4.12 | (6.62) | 4.34 | (9.95) |
| RTF (92) | 9.07 | (7.88) | 5.48 | (7.61) | 3.59 | (10.82) |
| BOTH (93) | 9.61 | (7.03) | 6.23 | (6.69) | 3.38 | (8.67) |

Note: *N* denotes the number of participants. In contrast to Table 6, only participants who were part of the respective treatment groups and who answered the survey question eliciting the perceived savings are included in this table. We elicit perceived savings after intervention experience. Respondents who reported no savings are included with perceived savings equal to zero.

In Table 8, we contrast the perceived savings with WTP elicited in the second WTP elicitation from those who experience the respective treatment. As described above, subtracting the perceived savings from WTP provides a lower bound of the non-monetary benefits of the interventions (third column of Table 8). This back-of-the-envelope calculation suggests that a substantial non-monetary benefit is created by the interventions, which consumers may experience from adjusting their behavior (e.g., warm glow from consuming fewer resources, time savings, or internalized environmental and social benefits from consuming less energy),[23] or from psycholog-

last month thanks to the INTERVENTION?"; where INTERVENTION was the specific intervention experienced by the respondent during the treatment phase.

[22]See Appendix A7 for the cost assumption and supporting sources.

[23]In our welfare calculations in Table 9, we assume that consumers on average do not fully or correctly internalize the social cost of carbon, and thus add it as an additional benefit in the MVPF calculation. Taking the opposite assumption, i.e. that consumers fully and correctly internalize the social cost of carbon, leads only to small changes. For instance, for the conservative scenario, MVPF for SC decreases from 1.88 to 1.82, for RTF from 2.12 to 1.94, and BOTH from 2.27 to 2.05 (see Appendix Table

ical benefits that are independent of the interventions' impact on showering behavior (e.g., a feeling of reward for those who already use resources efficiently, the non-instrumental value of having more information available, even if it does not directly change behavior (Sharot and Sunstein, 2020), and other types of benefits experienced independent of the behavioral change).

For participants with a positive WTP for the interventions, monetary and non-monetary benefits seem to outweigh total costs, including potential psychological costs experienced by adjusting behavior (e.g., loss of comfort due to a shorter shower), as well as psychological costs to consumers that are independent of the interventions' impact on showering behavior (e.g., continuing to take long showers but feeling bad about it, being bothered by the light, disliking receiving emails).

The WTP values seem relatively high compared to, for example, Allcott and Kessler (2019) who estimate an average WTP of 2.81 USD, or 2.42 EUR, for receiving four home energy reports on residential natural gas consumption. This figure is 72% lower than the WTP for the SC intervention in our study. Yet, our interventions are appliance-specific while Allcott and Kessler (2019) target aggregate household natural gas use. Participants may be more attracted to appliance-specific interventions because they – correctly (Gerster et al., 2026) – anticipate that appliance-specific feedback helps them to better conserve energy than aggregate feedback. Moreover, Tomberg (2024) finds an average WTP of 60 to 70 EUR for an RTF shower head in a large German population sample, corroborating that consumers are willing to pay considerable amounts to experience RTF and curb their resource use while showering. To address remaining uncertainty about external validity of our WTP values, we analyze a range of sensitivity scenarios in the welfare analysis below.

## 3.6 Impact of the interventions on social welfare

To estimate the effect of the interventions on welfare, we need to consider not only the change in consumer surplus but also the changes in externalities and the implementation costs of the interventions. In doing so, we lean on the approach by Hendren

A20).

and Sprung-Keyser (2020) and Hahn et al. (2026) and calculate the marginal value of public funds (MVPF) for our interventions. The MVPF provides us with an estimate of the social benefit per unit of public funds invested in the implementation of the respective intervention, which has the advantage that it allows a unified analysis of the welfare effects of different policies and makes them directly comparable across domains. We argue that this value is also relevant for private companies seeking a way to maximize the impact of their corporate social responsibility spending.[24]

We calculate the MVPF as:

$$MVPF = \frac{\Delta V_j - \phi_R \times \Delta R_j}{c_j}, \tag{2}$$

where $\Delta V_j$ is the experienced change in consumer surplus, which we measure as WTP elicited after treatment experience of intervention $j$.[25] $\Delta R_j$ is negative and describes the change in resource use while showering, i.e., the treatment effect of intervention $j$. $\phi_R$ is the external effect resulting from the carbon emissions associated with heating one liter of shower water and $c_j$ is the cost associated with the intervention, i.e., the cost of the smart shower head.

We use three scenarios to parameterize the MVPF. The first is the "base scenario", which relates directly to the results of our experiment. The second scenario is a "conservative" scenario, in which we make pessimistic assumptions to obtain a lower bound estimate of the welfare effects. The third scenario is an "opt-in" or continuation-demand scenario, in which we assume that only households with positive post-treatment WTP for an intervention receive that intervention, reflecting voluntary continuation after treatment experience rather than representative initial uptake

[24]We are not the first to take this view: Among the environmental policies whose MVPFs are quantified by Hahn et al. (2026) are also interventions that are implemented by private companies.

[25]Consumer surplus is defined as the difference between maximum WTP and the price paid. We assume in our welfare analysis that the costs of the interventions are borne by the government, i.e., considered in $c_j$. Therefore, consumers do not bear a direct cost, which is why WTP is an appropriate measure of consumer surplus. It could be debated whether we should adjust the WTP values to account for participants' overestimation of the financial savings achieved through the measures; however, WTP need not be a one-for-one function of perceived savings: it may also reflect non-monetary benefits, information value, warm glow, internalized environmental benefits, or psychological costs. Nevertheless, Appendix Table A21 presents a robustness test of the MVPF results, obtained by subtracting the overestimation of financial savings from the WTP values.

in the full population.

For the base scenario, the values for $\Delta V_j$ are based on WTP elicited in the treatment phase and are reported in the upper panel of Table 9. With regard to the external effects, we focus on the effects of carbon emissions associated with the energy use required to heat water for showering. Based on the average shower temperature of 38 °C, averages for the energy sources used to heat water in Germany and (marginal) emission factors for the respective energy sources, we assume that one liter of water consumption is, on average, associated with 15 grams of $CO_2$ (see Appendix A7, Table A22, for details). To obtain a monetary value, we multiply the 15 grams of $CO_2$ by the social cost of carbon estimated in Rennert et al. (2022) of 160 EUR per ton of $CO_2$, resulting in external costs of 0.0024 EUR/liter.[26] $\Delta R_j$ is the effect of the respective intervention in liters per shower taken from Table 3 and multiplied by 31, which is the average number of showers taken in the treatment month. The costs associated with the intervention are not easily quantified. To approximate these costs, we assume the retail costs of the smart shower heads of 80 EUR and a lifetime per shower head of 5 years. Thus, we assume costs per month of $80/(5 \times 12) = 1.33$ EUR.[27] Operational costs are abstracted from, as they are expected to be low on a large scale since the interventions can be fully automated. We assume the same device costs for RTF, BOTH, and SC, since smart shower heads are also required to provide SC, as households' individual showering behavior needs to be tracked.

The results of the welfare change due to the interventions in the base scenario are depicted in the upper panel of Table 9 and indicate that the MVPF of SC is 6.64 EUR per euro invested in the intervention, the MVPF of RTF is 7.41 and the MVPF of BOTH is largest at 7.94. These values indicate that all three interventions lead to substantial welfare gains in our sample; as discussed in Section 2.2, these magnitudes are most

[26]Specifically, Rennert et al. (2022) estimate the social cost of carbon to be 185 USD. We transform this value to EUR assuming the long-term average exchange rate of USD to EUR over the past 10 years, which is 0.8627 (European Central Bank, 2023).

[27]This approach conservatively assumes that the new shower head itself does not provide a welfare benefit independent of the interventions. Another approach would be to assume that the shower head provides additional welfare benefits comparable to those of a regular shower head, which we assume to be around 20 EUR. Amortized over the same five-year lifetime, this adds 0.33 EUR per month to the benefit numerator, increasing all welfare effects accordingly (see Appendix Table A19).

informative for households similar to our participants. MVPF values that exceed 5 are typically considered large (Hendren and Sprung-Keyser, 2020; Hahn et al., 2026) and are, for example, in the order of magnitude of several policies targeting children's health and education outcomes analyzed by Hendren and Sprung-Keyser (2020). In terms of environmental policies, MVPFs of more than 5 are found, for example, for nudges to reduce energy consumption in regions of the US with a dirty electricity grid (Hahn et al., 2026). In contrast, many other widely used environmental policies, such as subsidies to promote the uptake of clean technologies or efficiency improvements, often result in MVPFs of around 1 or 2 (Hahn et al., 2026).

**Table 9:** Monthly welfare inputs in EUR and resulting MVPF

| Base scenario | | | | |
|---|---|---|---|---|
| | $\Delta V$ | $\phi_R \times \Delta R$ | $c$ | $MVPF$ |
| SC | 8.59 | $-0.24$ | 1.33 | 6.64 |
| RTF | 9.13 | $-0.73$ | 1.33 | 7.41 |
| BOTH | 9.67 | $-0.89$ | 1.33 | 7.94 |
| Conservative scenario | | | | |
| SC | 2.42 | $-0.08$ | 1.33 | 1.88 |
| RTF | 2.57 | $-0.24$ | 1.33 | 2.12 |
| BOTH | 2.73 | $-0.29$ | 1.33 | 2.27 |
| Opt-in scenario | | | | |
| SC | 11.98 | $-0.26$ | 1.33 | 9.20 |
| RTF | 12.21 | $-0.74$ | 1.33 | 9.74 |
| BOTH | 11.89 | $-0.86$ | 1.33 | 9.58 |

Note: The welfare calculations combine inputs estimated using the relevant samples. In the base scenario, $\Delta V$ is mean post-treatment WTP among households that experienced the respective intervention and completed both incentivized WTP elicitations (109 SC households, 105 RTF households, and 107 BOTH households). The treatment effects used for $\Delta R$ are estimated using the full sample of 564 households. The conservative scenario rescales these base-scenario inputs as described in the text. In the opt-in scenario, $\Delta V$ is mean post-treatment WTP among treatment households with positive post-treatment WTP for the corresponding intervention (81 SC households, 85 RTF households, and 89 BOTH households). The corresponding treatment effects are estimated separately by comparing treatment and Control households with positive post-treatment WTP for that intervention; these estimation samples contain 179 households for SC, 184 for RTF, and 188 for BOTH (see Table A15). MVPFs are calculated as follows: $MVPF = \frac{\Delta V - \phi_R \times \Delta R}{c}$.

For the conservative scenario, we scale our WTP values and treatment effects down, recognizing that our sample of volunteers may be particularly motivated to conserve resources and to receive the interventions. To this end, we compare our WTP estimate for the SC intervention with a broadly comparable estimate of WTP for

energy-use social-comparison reports from Allcott and Kessler (2019). While we find an average WTP of 8.59 EUR for four shower-related reports, they find an average WTP for receiving four home energy reports on residential natural gas consumption of 2.81 USD, or 2.42 EUR,[28] which is 72% lower than the WTP for the SC intervention in our study. Since there are no comparable WTP estimates available for the RTF and BOTH interventions, we reduce all of our WTP values in the conservative scenario by 72%. To scale down the treatment effects, we draw on the results by Tiefenbeck et al. (2019), who conducted a study on the effectiveness of RTF while showering among hotel guests, i.e., in a population without volunteer selection bias and no monetary incentive to conserve water. They find a conservation effect of RTF of 11.4%. Given the control group's water use of 34.1 liters during the treatment period, an 11.4% reduction would correspond to water savings of 3.89 liters per shower, which is 60.3% less than the effect we found in the treatment phase (Table 3). In our conservative scenario, we thus assume 3.89 liters as the treatment effect of RTF and also reduce the effects of the other interventions by 60.3%, resulting in an SC effect of 1.27 liters and a BOTH effect of 4.73 liters per shower. Furthermore, we reduce our emission factor for one liter of shower water by 16%, which is the share of energy used to heat water in private households that comes from renewable energy or district heating and that thus is associated with low or zero carbon emissions (see Appendix A7 for details). The calculation in the conservative scenario results in external costs of 0.002 EUR/liter.

The results of the conservative scenario, which we interpret as a lower bound on the average welfare effects of the interventions, are shown in the middle panel of Table 9 and indicate that all three interventions still yield modest welfare gains with MVPFs ranging from 1.88 (SC) to 2.27 (BOTH) euros of social value per euro spent, even under conservative assumptions.

For the opt-in scenario, we use the same parameters as in the base scenario. However, we prune the sample so that only individuals with a positive WTP after treat-

[28] Assuming the long-term average exchange rate of USD to EUR over the past 10 years, which is 0.8627 (European Central Bank, 2023).

ment experience remain in the sample.[29] We therefore interpret this scenario as a continuation-demand sensitivity exercise: it asks what welfare would look like among households that still value the intervention after having experienced it, not among the full population or among all households that would initially opt in before experience. The results of the continuation-demand opt-in scenario are shown in the bottom panel of Table 9 and can be interpreted as upper-bound MVPF estimates rather than representative-population estimates. As expected, these are much larger than in the base scenario, ranging from 9.20 (SC) to 9.74 (RTF) euros of social value per euro spent. However, these welfare effects are not average effects for the entire sample, but only for those with positive WTP and thus for a smaller overall group.

As a final check, note that Table 9 compares one-month WTP and one-month treatment effects with the monthly amortized cost of the device, obtained by amortizing the 80 EUR shower head over a five-year lifetime. This implicitly assumes that treatment effects and WTP stay constant over the five-year period. While we find no evidence of attenuation in our data or in related research,[30] we additionally provide a break-even calculation to shed light on the implications of this assumption. We find that, in the base scenario, monthly welfare gains would need to persist for only about 10 months for SC, 9 months for RTF, and 8 months for BOTH for cumulative benefits to equal the upfront device cost.

# 4 Conclusion

This study provides a comparative evaluation of two widely used behavioral sustainability interventions, social comparison (SC) and real-time feedback (RTF), in a high-impact domain: household water and energy use during showering. We find

[29] Table A15 (Columns IV to VI) in the Appendix reports the treatment effects for this subsample, which we use to calculate $\Delta R$ in the opt-in scenario. Estimated treatment effects are similar when restricting the sample based on households' WTP before, rather than after, treatment experience (Columns I to III).

[30] In our data, treatment effects do not decline over the study period, and WTP does not significantly decline after one month of treatment experience. Related evidence also points to persistent effects of Home Energy Reports over two years (Allcott and Rogers, 2014) and to no decline in the effects of real-time feedback on showering behavior over four months (Byrne et al., 2024), although evidence on longer-run WTP remains limited.

that while both interventions reduce consumption, real-time feedback is significantly more effective than social comparison, and the combination yields the largest estimated reduction, although its difference from RTF alone is not statistically significant at conventional levels. Participants not only conserve more with RTF and the combined intervention but also value these tools more highly, even when psychological costs are taken into account.

From a company perspective, these findings carry several implications. First, our suggestive evidence is consistent with both SC and RTF motivating sustainable behavior through normative cues, while RTF additionally facilitates behavioral adjustment by embedding actionable information in the consumption context. With RTF producing greater behavioral impact and being more positively valued by consumers than SC, this suggests that some consumer-facing products might become more attractive through the addition of RTF-like features aimed at facilitating sustainable consumption, particularly in markets where environmental performance and user engagement can serve as differentiators, and in consumption contexts where fine-grained data is accessible at low cost. As real-time measurement technology becomes more widely available, interventions of this kind become feasible at larger scale for policymakers and firms alike.

Second, the combined intervention performs at least as well as RTF alone on both the consumption and the valuation dimension, suggesting that hybrid designs are a viable option for organizations pursuing both environmental and commercial objectives. Third, average willingness to pay exceeds perceived and actual cost savings, implying that consumers derive non-monetary benefits from these interventions; this suggests that sustainability features of this kind can be offered without reducing customer satisfaction, at least among consumers similar to our participants. Because the study sample selected into a sustainable-showering experiment, the WTP and MVPF magnitudes should be interpreted as most directly informative about an opt-in segment rather than the full population; the across-intervention comparisons are less sensitive to this selection. Finally, the behavior change we observe under market-based alloca-

tion supports the feasibility of demand-driven business models that rely on voluntary uptake rather than mandates or subsidies.

Our results align with earlier work showing the importance of making shrouded attributes (e.g., the quantity of energy or water consumed) salient to consumers (e.g. Gabaix and Laibson, 2006; Chetty et al., 2009; DellaVigna and Pollet, 2009; Grubb, 2009; Brown et al., 2010; Taubinsky and Rees-Jones, 2018). In contexts where consumption remains hidden or only roughly observable, informing people in real time lowers barriers to behavior change and improves responsiveness to prices (Jessoe and Rapson, 2014) and environmental impact (Fang et al., 2023).

While our specific focus is showering, the pattern of results may extend to other behaviors characterized by shrouded and high-cost energy use, such as room heating or cooling, where real-time information could similarly boost salience. Conversely, in settings where consumption is already salient (e.g., airfare), the difference between interventions that provide immediate and actionable information and interventions that target consumers' perception of the moral costs of overconsumption may be less pronounced. Future work can explore these variations more systematically.

Finally, our study adds to the small but growing literature directly measuring welfare effects of behavioral interventions through willingness-to-pay (Allcott and Kessler, 2019; Butera et al., 2022). We document a setting in which reductions in externalities coincide with positive consumer valuations, and we find little evidence that personal experience with the interventions alters WTP.

More broadly, our study illustrates the importance of evaluating the effects of interventions on behavior and households' willingness to pay. These two dimensions do not necessarily align: an intervention that is highly valued by households need not produce the largest reduction in negative externalities, and an intervention that offers larger social benefits need not be the one most strongly demanded in the market. Jointly analyzing behavioral effects and willingness to pay helps to identify when market demand is likely to be sufficient for adoption, and when policy support may be warranted to encourage households to adopt interventions with larger social returns.

In this sense, our approach provides a useful framework for evaluating sustainability interventions more generally, whenever private valuations and social benefits diverge.

# References


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

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3):1117–1165.

Allcott, H., Cohen, D., Morrison, W., and Taubinsky, D. (2025). When do "nudges" increase welfare? *American Economic Review*, 115(5):1555–1596.

Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1):236–76.

Allcott, H. and Mullainathan, S. (2010). Behavior and energy policy. *Science*, 327(5970):1204–1205.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–37.

Alt, M., Bruns, H., DellaValle, N., and Murauskaite-Bull, I. (2024). Synergies of interventions to promote pro-environmental behaviors – a meta-analysis of experimental studies. *Global Environmental Change*, 84:102776.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2006). Elicitation using multiple price lists. *Experimental Economics*, 9(4):383–405.

Andor, M. A., Gerster, A., and Peters, J. (2022). Information campaigns for residential energy conservation. *European Economic Review*, 144:104094.

Andor, M. A., Gerster, A., Peters, J., and Schmidt, C. M. (2020). Social norms and energy conservation beyond the US. *Journal of Environmental Economics and Management*, 103:102351.

Anonymous (2020). Pre-analysis plan and trial registration for the study reported in this manuscript. AEA RCT Registry. Identifying information omitted for double-blind review. An anonymized version is provided in the supplementary review materials.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232.

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., and Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, 28(8):1041–1055.

Bonan, J., Cattaneo, C., d'Adda, G., Galliera, A., and Tavoni, M. (2024). Widening the scope: The direct and spillover effects of nudging water efficiency in the presence of other behavioral interventions. *Journal of Environmental Economics and Management*, 127:103037.

Brandon, A., List, J. A., Metcalfe, R. D., Price, M. K., and Rundhammer, F. (2019). Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity. *Proceedings of the National Academy of Sciences*, 116(12):5293–5298.

Brent, D. A., Cook, J. H., and Olsen, S. (2015). Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. *Journal of the Association of Environmental and Resource Economists*, 2(4):597–627.

Brown, J., Hossain, T., and Morgan, J. (2010). Shrouded attributes and information suppression: Evidence from the field. *Quarterly Journal of Economics*, 125(2):859–876.

Butera, L., Metcalfe, R., Morrison, W., and Taubinsky, D. (2022). Measuring the welfare effects of shame and pride. *American Economic Review*, 112(1):122–68.

Byrne, D. P., Goette, L., Martin, L. A., Delahey, L., Jones, A., Miles, A., Schob, S., Staake, T., Tiefenbeck, V., et al. (2024). How nudges create habits: Theory and evidence from a field experiment. Working paper.

Byrne, D. P., La Nauze, A., and Martin, L. A. (2018). Tell me something I don't already know: Informedness and the impact of information programs. *Review of Economics and Statistics*, 100(3):510–527.

Carson, R. T. and Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, 37(1):181–210.

Chen, Y., Cramton, P., List, J. A., and Ockenfels, A. (2021). Market design, human behavior, and management. *Management Science*, 67(9):5317–5348.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–77.

Congiu, L. and Moscati, I. (2020). Message and environment: A framework for nudges and choice architecture. *Behavioural Public Policy*, 4(1):71–87.

Consumer Protection Agency NRW (2019). Warmwasser. Brochure, p. 5. `https://www.verbraucherzentrale.nrw/sites/default/files/2019-04/201904_Warmwasser_Energie_Wasser-sparen_Brosch%C3%BCre_VZ-NRW.pdf`, accessed on October 26, 2026.

Consumer Protection Agency RLP (2020). Warmwasser: Komfortables Sparen – So geht's! As of: February 2020. Verbraucherzentrale Rheinland-Pfalz. `https://www.verbraucherzentrale-rlp.de/sites/default/files/2020-03/VZ_WarmWasser_2020_12_web.pdf`, accessed on September 22, 2020.

DellaVigna, S., List, J. A., and Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *Quarterly Journal of Economics*, 127(1):1–56.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2017). Voting to tell others. *Review of Economic Studies*, 84(1):143–181.

DellaVigna, S. and Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *Journal of Finance*, 64(2):709–749.

European Central Bank (2023). Euro foreign exchange reference rates: US dollar (USD). Change from 24 January 2013 to 24 January 2023. European Central Bank (ECB). https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/eurofxref-graph-usd.en.html, accessed on January 24, 2023.

Faberi, S., Paolucci, L., Lapillonne, B., and Pollier, K. (2015). Trends and policies for energy savings and emissions in transport. Report, ODYSSEE-MURE project. https://www.odyssee-mure.eu/publications/archives/energy-efficiency-trends-policies-transport.pdf, accessed on October 12, 2023.

Fang, X., Goette, L., Rockenbach, B., Sutter, M., Tiefenbeck, V., Schoeb, S., and Staake, T. (2023). Complementarities in behavioral interventions: Evidence from a field experiment on resource conservation. *Journal of Public Economics*, 228:105028.

Ferraro, P. J. and Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1):64–73.

Gabaix, X. and Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, 121(2):505–540.

German Environment Agency (2019). Emissionsbilanz erneuerbarer Energieträger: Bestimmung der vermiedenen Emissionen im Jahr 2018. Table 62: Emission factors of heat supply from fossil fuels in private households, the tertiary sector and industry. Umweltbundesamt. https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-11-07_

cc-37-2019_emissionsbilanz-erneuerbarer-energien_2018.pdf, accessed on September 22, 2020.

German Federal Institute for Population Research (2023a). Altersstruktur der Bevölkerung. Demografieportal. Bundesinstitut für Bevölkerungsforschung. https://www.demografie-portal.de/DE/Fakten/bevoelkerung-altersstruktur.html, accessed on November 11, 2023.

German Federal Institute for Population Research (2023b). Zahl der Privathaushalte und durchschnittliche Haushaltsgröße in Deutschland (1871-2021). Bundesinstitut für Bevölkerungsforschung. https://www.bib.bund.de/DE/Fakten/Fakt/L49-Privathaushalte-Haushaltsgroesse-ab-1871.html, accessed on February 2, 2023.

German Federal Ministry for Economic Affairs and Energy (2020). Energiedaten. As of: May 2020. Table 7b: Endenergieverbrauch nach Anwendungsbereichen II. Bundesministerium für Wirtschaft und Energie. https://www.bmwk.de/Redaktion/DE/Artikel/Energie/energiedaten-gesamtausgabe.html, originally retrieved on September 22, 2020; the URL was last accessed on January 26, 2023 and contains a newer version of the document.

German Federal Statistical Office (2021). Laufende Wirtschaftsrechnungen: Einkommen, Einnahmen und Ausgaben privater Haushalte - 2020 - Fachserie 15 Reihe 1. Published 15. Dezember 2021. Article number: 2150100207004. Statistisches Bundesamt (Destatis).

German Federal Statistical Office (2022). Bevölkerung und Erwerbstätigkeit - 2020 (Endergebnisse) - Fachserie 1 Reihe 3. Published 31. Januar 2022. Article number: 2010300207004. Statistisches Bundesamt (Destatis).

Gerster, A., Andor, M. A., and Goette, L. (2026). Disaggregate consumption feedback. *The Economic Journal*, 136(675):1066–1086.

Goette, L., Han, H.-J., Lim, Z. H., et al. (2021). The dynamics of goal setting: Evidence from a field experiment on resource conservation. Working Paper 283, CRC TR 224 Discussion Paper Series University of Bonn and University of Mannheim.

Gosnell, G. K., List, J. A., and Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. *Journal of Political Economy*, 128(4):1195–1233.

Grubb, M. D. (2009). Selling to overconfident consumers. *American Economic Review*, 99(5):1770–1807.

Hahn, R. W., Hendren, N., Metcalfe, R. D., and Sprung-Keyser, B. (2026). A welfare analysis of policies impacting climate change. *American Economic Review*, 116(7):2368–2421.

Hansen, P. G. and Jespersen, A. M. (2013). Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, 4(1):3–28.

Hendren, N. and Sprung-Keyser, B. (2020). A unified welfare analysis of government policies. *Quarterly Journal of Economics*, 135(3):1209–1318.

Ho, L. and Page, L. (2025). Got beef with beef? evidence from a large-scale carbon labeling experiment. *Working Paper*.

Holland, S. P., Kotchen, M. J., Mansur, E. T., and Yates, A. J. (2022). Why marginal CO2 emissions are not decreasing for US electricity: Estimates and implications for climate policy. *Proceedings of the National Academy of Sciences*, 119(8):e2116632119.

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655.

Huber, J., Lohmann, K., Schmidt, M., and Weinhardt, C. (2021). Carbon efficient smart charging using forecasts of marginal emission factors. *Journal of Cleaner Production*, 284:124766.

Jessoe, K. and Rapson, D. (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review*, 104(4):1417–1438.

Johnson, E. J., Shu, S. B., Dellaert, B. G., Fox, C., Goldstein, D. G., Häubl, G., Larrick, R. P., Payne, J. W., Peters, E., Schkade, D., et al. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters*, 23:487–504.

Lambin, E. F. and Thorlakson, T. (2018). Sustainability standards: Interactions between private actors, civil society, and governments. *Annual Review of Environment and Resources*, 43(1):369–393.

Lim, Z. H. (2025). To each their own: Heterogeneity in worker preferences for peer information. *arXiv preprint arXiv:2508.06162*.

List, J. A. (2025). *Experimental economics: Theory and practice*. University of Chicago Press.

Löschel, A., Rodemeier, M., and Werthschulte, M. (2023). Can self-set goals encourage resource conservation? Field experimental evidence from a smartphone app. *European Economic Review*, 160:104612.

Ludwig, J., Kling, J. R., and Mullainathan, S. (2011). Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, 25(3):17–38.

Michel, A., Attali, S., and Bush, E. (2016). Energy efficiency of white goods in Europe: monitoring the market with sales data. Final report. Study realised on behalf of ADEME by SOWATT and Bush Energie. `https://storage.topten.eu/source/files/Market-Monitoring-2016-EN-Topten.eu.pdf`, accessed on October 12, 2023.

Mongin, P. and Cozic, M. (2018). Rethinking nudge: Not one but three concepts. *Behavioural Public Policy*, 2(1):107–124.

Obama, B. (2015). Executive order – Using Behavioral Science insights to better serve

the American people. Media Release, White House, Washington, DC, September 15, 2015.

OECD (2017). *Behavioural insights and public policy: Lessons from around the world*. OECD Publishing, Paris.

Rennert, K., Errickson, F., Prest, B. C., Rennels, L., Newell, R. G., Pizer, W., Kingdon, C., Wingenroth, J., Cooke, R., Parthum, B., Cromar, K., Diaz, D., Moore, F. C., Müller, U. K., Plevin, R. J., Raftery, A. E., Sevcikova, H., Sheets, H., Stock, J. H., Tan, T., Watson, M., Wong, T. E., and Anthoff, D. (2022). Comprehensive evidence implies a higher social cost of CO2. *Nature*, 610(7933):687–692.

Rodemeier, M. (2021). Buy baits and consumer sophistication: Theory and field evidence from large-scale rebate promotions. Working Paper 124, CAWM Discussion Paper Series.

Rodemeier, M. and Löschel, A. (2023). Information nudges, subsidies, and crowding out of attention: Field evidence from energy efficiency investments. Working Paper 16141, IZA Discussion Paper Series.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5):429–434.

Sharot, T. and Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behaviour*, 4(1):14–19.

Taubinsky, D. and Rees-Jones, A. (2018). Attention variation and welfare: Theory and evidence from a tax salience experiment. *Review of Economic Studies*, 85(4):2462–2496.

Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., and Staake, T. (2018). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Science*, 64(3):1458–1476.

Tiefenbeck, V., Wörner, A., Schöb, S., Fleisch, E., and Staake, T. (2019). Real-time feedback promotes energy conservation in the absence of volunteer selection bias and monetary incentives. *Nature Energy*, 4(1):35–41.

Tomberg, L. (2024). Resource conservation through improved efficiency, behavioral change, or both: Willingness to pay for (smart) efficient shower heads. *Resources, Conservation and Recycling*, 203:107387.

# Appendix

## A1 Supplementary analyses

### A1.1 Analyses to assess the sensitivity of our results with regard to excluding shower sessions with very low or very high water use

For our main analyses, we exclude shower sessions with water use below 5 or above 200 liters. In Table A1, we assess the sensitivity of our results with regard to this approach. For reference, column I of Table A1 reiterates the results when we exclude shower sessions with water use below 5 or above 200 liters. Column II presents results, when we use stricter cutoffs of 10 and 100 liters. We find that the estimated treatment effects of RTF and BOTH remain virtually unchanged compared to column I, while the effect of SC reduces. The average baseline water use per shower increases slightly by around 1.7 liters.

In column III, we do not condition on water use per shower but on the more intuitive measure of shower length. Limiting the data set to showers between 3 and 15 minutes, the effects of SC and RTF remain relatively stable, while the effect of BOTH increases slightly to 14.1 liters per shower. The baseline average water use per shower increases substantially to 48 liters, which is in the ballpark of the baseline average of 44.8 liters found in Tiefenbeck et al. (2018). The final column of Table A1 presents the results if no exceptionally short or long recorded shower sessions are excluded. While the baseline average water use per shower reduces substantially to 26.9 liters, the treatment effects remain remarkably stable.

Concluding, we find that the size of the treatment effects is to a limited extent sensitive to the approach of excluding very short or long shower sessions. Yet, the overall treatment patterns remain stable so that we do not expect this sensitivity to jeopardize the validity of our takeaways.

**Table A1:** Treatment effects robust to exclusion of very small or very large water extractions / very short or very long showers

| | Exclude shower session if water use in liters $< 5 \vee > 200$ | Exclude shower session if water use in liters $< 10 \vee > 100$ | Exclude shower session if shower duration in seconds $< 180 \vee > 900$ | No exclusion |
|---|---|---|---|---|
| SC | -3.19*** | -1.73*** | -3.12*** | -2.76*** |
| | (0.66) | (0.53) | (0.98) | (0.62) |
| RTF | -9.79*** | -9.80*** | -10.60*** | -8.80*** |
| | (0.88) | (0.80) | (0.89) | (0.80) |
| BOTH | -11.91*** | -11.38*** | -14.09*** | -11.17*** |
| | (1.22) | (1.22) | (1.78) | (1.17) |
| Baseline average | 32.9 | 34.6 | 48.4 | 26.9 |
| No. of households | 564 | 563 | 522 | 564 |
| No. of observations | 38,453 | 32,284 | 18,262 | 47,060 |

Note: OLS estimates with household and calendar-date fixed effects included. All columns start from the same 564 households used in the main analysis, with retained counts reported. The no-exclusion specification retains all recorded shower sessions for these households. Standard errors are clustered at the household level and are reported in parentheses. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

## A1.2 Analyses to assess the validity of our approach to separating different shower sessions

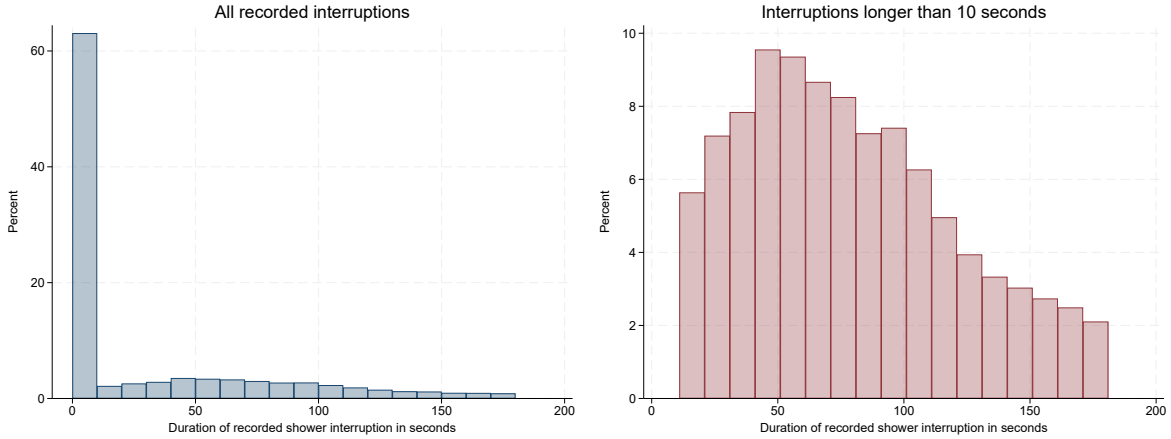
In order to distinguish between individual shower sessions, the shower heads are configured so that they begin measuring a new shower session after a pause of 180 seconds. Any interruption of less than 180 seconds is counted as an interruption within a shower session.

Under this rule, the average pre-intervention water volume per identified shower session in our data is 33 liters. This magnitude is in line with prior shower-level field studies. Fang et al. (2023) report an average baseline shower volume of 37.8 liters, Goette et al. (2021) report baseline shower water use of around 32 liters, and Tiefenbeck et al. (2018) report 44.8 liters per baseline shower. Our estimate therefore lies well within the range observed in the literature, which supports the plausibility of the session-separation rule used in our main analysis.

Figure A1 depicts histograms of the recorded shower interruptions. For around 55% of shower sessions there is no recorded interruption. In addition, around 4,500

shower events exhibit a shower interruption of less than 10 seconds. Since the electronic components for measuring showering activity are powered by a generator that uses the water flowing through the shower head, a certain water pressure was required for them to become active. The considerable number of very short shower breaks could thus be caused by short-term reductions in water pressure, e.g., at the beginning or end of the shower. For this reason, we condition the right-hand histogram to longer shower breaks of more than 10 seconds, as we assume that especially these breaks represent significant interruptions of the shower session, e.g., for soaping up.

**Figure A1:** Histogram of durations of recorded shower interruptions




Note: The data shown here covers the shower sessions of all three experimental phases. Like the other shower data, this data was also recorded automatically by the shower heads.

From the right-hand histogram in Figure A1, we see that the distribution of shower breaks seems to be to some extent truncated due to the automated recording of a new shower event after an interruption of more than 180 seconds. Thus, it may be the case that some shower sessions are recorded as two single sessions due to longer shower breaks, yet considering the left-hand histogram it becomes clear that this is likely to be the case for only a very small fraction of all recorded shower events and it is thus highly unlikely that a possible misclassification of these showers distorts our results.

To provide some evidence to underline that our results are robust to the recorded shower interruptions, Table A2 presents our central estimation results from Table 3

for subsamples of shower sessions with no or shorter interruptions, assuming that these shower sessions are unlikely to, in truth, be two subsequent shower sessions by different persons. There are no conspicuous changes in the effect sizes in the columns of Table A2, which suggests that possible misclassification due to rapid changeovers between shower users, and thus the counting of two separate showering sessions as one, is unlikely to distort our results.

**Table A2:** Treatment effects robust to exclusion of longer interruptions of shower sessions (pauses)

| | No pause | Pauses $<$ 30 seconds | Pauses $<$ 60 seconds | Pauses $<$ 90 seconds |
|---|---|---|---|---|
| SC | -2.51*** | -2.77*** | -2.89*** | -3.07*** |
| | (0.77) | (0.75) | (0.71) | (0.69) |
| RTF | -9.79*** | -9.57*** | -9.53*** | -9.48*** |
| | (1.03) | (0.98) | (0.92) | (0.89) |
| BOTH | -11.66*** | -12.07*** | -11.85*** | -11.58*** |
| | (1.36) | (1.36) | (1.16) | (1.06) |
| Baseline average | 35.6 | 34.9 | 33.8 | 33.3 |
| No. of households | 554 | 563 | 564 | 564 |
| No. of observations | 21,264 | 26,147 | 29,900 | 33,247 |

Note: OLS estimates with household and calendar-date fixed effects included. Standard errors are clustered at the household level and are reported in parentheses. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

In Table A3 we examine how our interventions affect shower pause occurrence and duration. SC shows no significant effect on pauses, whereas both RTF and BOTH markedly reduce the likelihood of a shower occurring without any pause. This suggests that participants receiving RTF or BOTH are more likely to insert pauses to cut resource use during their shower sessions. The induced pauses are largely between 1 and 74 seconds. Importantly, none of the interventions significantly increases the probability of longer pauses, indicating that they do not provoke extended breaks that could lead to misclassification of shower sessions.

**Table A3:** Treatment effects on shower interruptions (pauses)

| | Pause duration (seconds) | no pause | Probability (y/n) of positive-pause interval (seconds) (0,14] | (14,46] | (46,74] | (74,108] | (108,180] |
|---|---|---|---|---|---|---|---|
| SC | -0.57 | 0.00 | 0.00 | -0.00 | 0.00 | -0.00 | -0.00 |
| | (1.24) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| RTF | -0.27 | -0.04** | 0.03*** | 0.01 | 0.01 | 0.01 | -0.01* |
| | (1.46) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| BOTH | 1.16 | -0.08*** | 0.04*** | 0.02* | 0.02** | -0.01 | -0.00 |
| | (2.00) | (0.02) | (0.01) | (0.01) | (0.01) | (0.01) | (0.01) |
| (super-)control group average | 27.31 | 0.60 | 0.07 | 0.07 | 0.08 | 0.08 | 0.09 |
| No. of households | | | | 564 | | | |
| No. of observations | | | | 38,432 | | | |

Note: OLS estimates with household and calendar-date fixed effects included. The first column reports treatment effects on recorded pause duration in seconds. The remaining columns report linear probability estimates for no recorded pause and for the five positive-pause duration brackets. These brackets are based on quintiles of positive recorded pauses up to the 180-second session-separation threshold. Standard errors are clustered at the household level and are reported in parentheses. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

## A1.3 Further analyses on the impact of the interventions on behavior

### A1.3.1 Robustness checks

**Table A5:** Treatment effects robust to different specifications

| | Volume (Liters) | | |
|---|---|---|---|
| SC | -3.79** | -4.66** | -3.19*** |
| | (1.55) | (1.98) | (0.66) |
| RTF | -9.91*** | -10.79*** | -9.79*** |
| | (1.26) | (1.68) | (0.88) |
| BOTH | -12.01*** | -12.88*** | -11.91*** |
| | (1.33) | (1.74) | (1.22) |
| Time FE | No | Yes | Yes |
| Household FE | No | No | Yes |
| No. of households | 564 | 564 | 564 |
| No. of observations | 38,453 | 38,453 | 38,453 |

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

**Table A4:** Average treatment effects on water use per shower – differentiated by number of showers in household

| | (I) Full sample | (II) Only one shower in household | (III) Multiple showers in household |
|---|---|---|---|
| SC | -3.19*** | -2.85*** | -4.24*** |
| | (0.66) | (0.73) | (1.40) |
| RTF | -9.79*** | -10.25*** | -8.41*** |
| | (0.88) | (1.00) | (1.73) |
| BOTH | -11.91*** | -12.13*** | -10.87*** |
| | (1.22) | (1.43) | (1.87) |
| (super-)control group average | 34 | 35 | 32 |
| No. of households | 564 | 442 | 121 |
| No. of observations | 38,453 | 30,561 | 7,822 |

Note: OLS estimates with household and calendar-date fixed effects included. One household whose survey responses indicate zero showers and bathtubs is included in the full-sample specification but cannot be classified into either subgroup. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The (Super-)control group average corresponds to the average water use per shower of the control and Super-control group in the treatment phase.

**Table A6:** Negligible effect on shower temperature and frequency

| | Shower temperature | Showers per day |
|---|---|---|
| SC | -0.17 | 0.05 |
| | (0.15) | (0.04) |
| RTF | -0.19 | -0.03 |
| | (0.18) | (0.03) |
| BOTH | -0.29* | -0.02 |
| | (0.15) | (0.03) |
| No. of households | 564 | 564 |
| No. of observations | 38,446 | 32,148 |

Note: OLS estimates with household and calendar-date fixed effects included. For Spec. (2), we construct a data set including the number of showers per household per day during the baseline phase and treatment phase of the study. Standard errors are clustered at the household level and are reported in parentheses. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

**Table A7:** Average treatment effects on water use per shower in the treatment phase split by household size

| | (I) Full sample | (II) 1-person households | (III) 2- and 3-person households |
|---|---|---|---|
| SC | -3.19*** | -3.39*** | -3.14*** |
| | (0.66) | (1.09) | (0.79) |
| RTF | -9.79*** | -8.36*** | -10.14*** |
| | (0.88) | (1.60) | (1.03) |
| BOTH | -11.91*** | -15.34*** | -10.93*** |
| | (1.22) | (3.50) | (1.08) |
| (super-)control group average | 34 | 31 | 35 |
| No. of households | 564 | 188 | 376 |
| No. of observations | 38,453 | 9,081 | 29,372 |

Note: Dependent variable is shower volume in Liters. OLS estimates with household and calendar-date fixed effects included. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The (Super-)control group average corresponds to the average water use per shower of the control and Super-control group in the treatment phase, restricting to the respective households. Column (II) restricts the sample only to one-person households, and column (III) restricts the sample to only 2- or 3-person households. Of these, 329 are two-person and 47 are three-person.

**Table A8:** Average treatment effects on water use per shower in the treatment phase split by gender (1-person households)

| | (I) All 1-person households | (II) Male 1-person households | (III) Female 1-person households |
|---|---|---|---|
| SC | -3.39*** | -4.18*** | -1.88 |
| | (1.09) | (1.52) | (1.42) |
| RTF | -8.36*** | -7.19*** | -8.85*** |
| | (1.60) | (2.39) | (2.01) |
| BOTH | -15.34*** | -16.78*** | -12.75** |
| | (3.50) | (4.93) | (4.93) |
| (super-)control group average | 31 | 32 | 29 |
| No. of households | 188 | 108 | 79 |
| No. of observations | 9,081 | 5,598 | 3,428 |

Note: Dependent variable is shower volume in Liters. OLS estimates with household and calendar-date fixed effects included. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The (Super-)control group average corresponds to the average water use per shower of the control and Super-control group in the treatment phase, restricting to the respective households. Column (I) includes all one-person households, (II) restricts the sample only to male one-person households, and column (III) restricts the sample to only female 1-person households. Respondent gender is missing for one one-person household; this household is included in column (I) but excluded from columns (II) and (III).

**Table A9:** Baseline willingness to pay regressed on treatment-group indicators

| | Regressions of baseline WTP outcomes on treatment-group indicators | | | | | |
|---|---|---|---|---|---|---|
| | WTP for | | | share with negative WTP for | | |
| Treatment group | SC | RTF | BOTH | SC | RTF | BOTH |
| SC | -0.72 | -1.17 | -0.53 | -0.01 | 0.05 | -0.00 |
| | (1.01) | (0.95) | (0.94) | (0.03) | (0.03) | (0.03) |
| RTF | -0.29 | -1.17 | -1.18 | 0.01 | 0.05 | 0.04 |
| | (1.00) | (1.03) | (1.03) | (0.04) | (0.03) | (0.04) |
| BOTH | 0.77 | -0.42 | 0.55 | -0.03 | 0.00 | -0.02 |
| | (0.95) | (0.91) | (0.90) | (0.03) | (0.02) | (0.03) |
| Control group average | 9.64 | 10.80 | 10.58 | 0.08 | 0.03 | 0.06 |
| No. of households | 439 | | | | | |

Note: OLS estimates with robust standard errors. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The omitted group is the control group who also completed incentivized WTP elicitations but did not experience any treatment during the treatment phase. The sample consists of all participants who completed both WTP elicitations and who were not part of the Super-control group, as the WTP elicitations for the Super-control group were purely hypothetical.

### A1.3.2 Heterogeneity by exogenous variation in the communicated reference point

As shown in the right panel of Figure 4, households with higher baseline water consumption exhibit larger treatment effects in absolute terms. Within the SC treatment, households receiving less favorable rank information also reduce consumption more. However, this relationship should not be interpreted causally, because the rank information communicated in SC is mechanically linked to baseline consumption: households with higher baseline consumption tend to rank worse relative to their comparison group and also have greater scope to reduce consumption.

A more relevant question for the interpretation of our treatment differences is therefore whether, holding baseline consumption fixed, the specific reference point communicated to a household affects treatment intensity. In the SC treatment, this would mean that two otherwise similar households respond differently because one is shown a less favorable relative position than the other. In the RTF treatment, the analogous question is whether two otherwise similar households respond differently because the shower head begins flashing red at a lower water-consumption threshold for one household than for the other. If such differences mattered quantitatively, then part of the gap between SC and RTF could reflect differences in the stringency of the communicated norm.

To examine this possibility, we exploit exogenous variation in the reference groups that underlie both the SC reports and the RTF thresholds. As described in Section 2.3.1, we first partitioned the Super-control households into 14 reference groups with 9 households each. Treatment households were then assigned to one of these reference groups. Because assignment to reference groups is exogenous, households with similar consumption behavior could nevertheless face different communicated benchmarks. This design therefore generates variation in the information conveyed by the intervention that is orthogonal to households' own prior water use.

The role of the reference group differs somewhat across treatments. For households in SC, the assigned reference group determines both the average peer consumption used for comparison and the household's rank relative to the 9 comparison house-

holds. For households in RTF, the average consumption of the assigned reference group determines the water-use threshold at which the shower head changes color, although this benchmark is not itself stated explicitly. For households in BOTH, both features are present, and participants were able to infer that the real-time threshold corresponds to average peer consumption.

Table A10 uses this design-based variation to examine whether treatment effects differ systematically with the reference point faced by the household. Specifically, we split households into three roughly equal-sized groups based on the threshold implied by their assigned comparison group. This keeps the heterogeneity analysis focused on exogenous differences in the stringency of the communicated norm, rather than confounding it with differences in households' own baseline behavior.

**Table A10:** Heterogeneity by assigned threshold

| | Low threshold | Middle threshold | High threshold |
|---|---|---|---|
| SC | -3.44** | -3.47** | -2.08** |
| | (1.42) | (1.53) | (0.86) |
| RTF | -10.45*** | -9.14*** | -8.76*** |
| | (1.64) | (1.60) | (1.57) |
| BOTH | -12.14*** | -11.83*** | -11.13*** |
| | (2.41) | (2.37) | (1.35) |
| Average assigned threshold | 30.5 | 34.7 | 41.2 |
| Assigned threshold range | 27.5–33.5 | 33.5–37.0 | 37.0–42.5 |
| No. of households | 145 | 145 | 146 |
| No. of observations | 8,233 | 11,247 | 11,436 |

Note: OLS estimates with household fixed effects and threshold-group-specific treatment-phase indicators included. Standard errors are clustered at the household level and reported in parentheses. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The dependent variable is liters per shower. Households are sorted into three roughly equal-sized groups by their assigned threshold/reference point, calculated as the average of the two comparison-group thresholds used during the treatment phase. The sample includes households with both treatment-phase reference values and usable shower observations in both the baseline and treatment phases. This household-level classification does not require a separate color classification for each shower. Table 5, by contrast, requires usable shower-level color classifications based on the reference value applicable at the time of each shower. The household and observation counts therefore need not coincide across the two analyses. The low, middle, and high threshold groups therefore differ only in the randomly assigned comparison-group benchmark, not in households' own baseline behavior. Because assigned thresholds take a discrete set of values, tied threshold values can appear in adjacent groups at the group boundaries. We exclude the Super-control group from this analysis, since they form the basis of the threshold calculation. The median baseline-phase shower consumption across baseline showers in the full analysis sample is 27.0 liters per shower.

### A1.3.3 Analysis of the transition between color thresholds due to the interventions

**Table A11:** Transition matrix: Changes in the color threshold in which a participant's showers ended most often

| | | | to | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | green | blue | purple | red | flashing red | N |
| Cont | **from** | green | 95% | | | | 5% | 22 |
| | | blue | 38% | 50% | 12% | | | 8 |
| | | purple | 8% | 8% | 46% | 8% | 31% | 13 |
| | | red | 14% | | 21% | 21% | 43% | 14 |
| | | flashing red | 4% | 2% | 4% | 2% | 88% | 51 |
| SC | **from** | green | 95% | | | 3% | 3% | 38 |
| | | blue | 86% | 14% | | | | 7 |
| | | purple | 15% | 46% | 23% | | 15% | 13 |
| | | red | | | 100% | | | 1 |
| | | flashing red | 8% | 12% | 4% | 2% | 75% | 51 |
| RTF | **from** | green | 97% | | | | 3% | 31 |
| | | blue | 89% | 11% | | | | 9 |
| | | purple | 20% | 60% | | 10% | 10% | 10 |
| | | red | 30% | 40% | 20% | 10% | | 10 |
| | | flashing red | 29% | 4% | 18% | 6% | 43% | 49 |
| BOTH | **from** | green | 100% | | | | | 25 |
| | | blue | 87% | 13% | | | | 15 |
| | | purple | 54% | 38% | 8% | | | 13 |
| | | red | 100% | | | | | 1 |
| | | flashing red | 29% | 18% | 9% | 5% | 38% | 55 |

Note: We present the changes between the baseline phase and the treatment phase. For each shower, color categories are defined using the reference value applicable at that time, as in Table 5. For the baseline phase and households without active real-time feedback, the categories indicate the colors that would have been displayed if feedback had been enabled. We then identify each household's most frequent category separately in the baseline and treatment phases. In contrast to Table 5, this table additionally requires at least one classifiable shower in each phase. The ten households contributing only baseline observations to Table 5 are therefore excluded, leaving 436 households. Each household contributes one transition between its most frequent baseline and treatment-phase color categories. The sum of the individual rows is 100% (if not, this is due to rounding error). As an example, we can interpret the last row of the table as follows: Of the 55 participants in the BOTH group who finished most of their showers in the baseline phase while the shower head would have flashed red, 29% (18%, 9%, 5%, 38%) finished most showers in the treatment phase while the shower head glowed green (blue, purple, red, flashing red). We calculated the hypothetical color thresholds in the baseline phase as we did in the treatment phases, that is, we based them on the average baseline water consumption per shower of a participant's reference group, which consisted of participants from the Super-control group. Therefore, participants in the Super-control group are not included in this table because there are no valid hypothetical color thresholds available for them.

### A1.3.4 Behavioral responses in the WTP-based phase and the IV approach

The primary purpose of the WTP-based phase was to make the willingness to pay (WTP) decisions of those households assigned to RTF, SC, BOTH or Control in the treatment phase consequential and thus incentive compatible. These households experienced interventions in the WTP-based phase based on their prior WTP decisions. Super-control households completed the same elicitations hypothetically and remained ineligible for an intervention.

Participants completed six multiple price lists (MPLs): one for each intervention (SC, RTF, BOTH), elicited once before and once during the treatment phase. Each MPL elicited WTP for receiving the respective intervention for one month. For households who were in the RTF, SC, BOTH or Control group during the treatment phase, one of the decisions made in one of these six MPLs was randomly chosen and implemented. This random draw worked as follows: First, one of the two elicitation blocks (i.e. completed before or during the treatment phase) was randomly determined, then one of the three intervention-specific MPLs completed in this block was randomly chosen, and, finally, one of its 15 price rows was randomly selected.[31] The decision made in this row was implemented: The household received the selected intervention if it had chosen it over the monetary alternative in that row; otherwise, it received no intervention. Further, the monetary amount corresponding to the household's choice was added or deducted from participants' payment.

Because all of the decisions participants made could theoretically be the decision selected for implementation, this procedure made the WTP choices consequential for eligible households. It follows standard practice in experimental economics to ensure incentive compatibility when multiple decisions are elicited (see e.g. Becker et al., 1964; Holt and Laury, 2002; Andersen et al., 2006).

Intervention receipt in the WTP-based phase therefore depended on both participants' decisions and the random draws. While this is intentional and in line with

[31] We used this three-step procedure for the random draw to make it easier to explain the procedure to participants.

the WTP-based phase's primary purpose of making WTP elicitations consequential, it makes an analysis of households' responses to intervention receipt in the WTP-based phase less straightforward than under exogenous assignment. We thus regard the below analysis of treatment effects in the WTP-based phase as additional evidence that complements our main analysis in section 3.

We address selection into intervention receipt by reporting assignment intention-to-treat (ITT) estimates and instrumental-variable (IV) estimates, following standard approaches in experimental economics with endogenous treatment uptake or non-compliance (see, e.g., List, 2025, chapter 13), as detailed below. In our main specifications, we restrict the analysis to households who were in the Control or Super-control group during the treatment phase, i.e. households who did not receive any intervention during the treatment phase. While the Control group could receive an intervention in the WTP-based phase depending on their WTP and the outcome of the random draw, the Super-control group remained ineligible to receive an intervention during the WTP-based phase.

We restrict the analysis to these two groups because behavior of those households who experienced RTF, SC or BOTH during the treatment phase may, in the WTP-based phase, still be confounded by possible lasting effects of the interventions experienced earlier (see e.g. Byrne et al., 2024, for evidence on post-intervention effects). In our ITT analysis, we define a household as intended to receive a certain treatment (SC, RTF, or BOTH), if the MPL corresponding to this intervention is selected in the random draw, irrespective of actual receipt. Super-control households form the no-offer comparison group. These assignment comparisons draw on the random allocation to Control or Super-control in the treatment phase and the subsequent random MPL selection among Control households.

The ITT models in Table A12 correspondingly use indicators for selection of the SC, RTF, or BOTH MPL for implementation as the explanatory variables, with Super-control as the omitted category. The identifying assumption is that the Control and Super-control group are comparable, i.e. the mere fact that one group fills out

incentive-compatible MPLs and the other group fills out hypothetical MPLs does not influence behavior apart from the resulting differences in treatment assignment and experience.

The IV model instruments the three indicators of actual intervention receipt with the ITT assignment indicators using conventional two-stage least squares. Under the standard IV assumptions, the IV estimates identify average effects among households induced to receive the respective intervention due to the respective MPL assignment. With one-sided uptake,[32] these estimates correspond to the average treatment effect on the treated (ATT). Because both the instrument (intervention assignment) and the endogenous variable (actual intervention receipt) are time-invariant at the household level, a panel two-stage least squares estimator with household fixed effects has no within-household variation in the first stage and cannot be estimated in its standard form. We therefore collapse the data to one observation per household, averaging water use per shower within each household and phase and subtracting the baseline average from the WTP-based average. This yields a single change score per household, on which we estimate a conventional cross-sectional two-stage least squares model. To gauge the influence of this procedure on estimated effects, Col. 2 first reports the intention-to-treat (ITT) estimates from this cross-sectional data set, which can be compared directly with the panel-based ITT estimates in Col. 1. The two sets of estimates differ only slightly. The existing differences are likely due to the fact that after collapsing, each household receives equal weight in the regression analysis. Col. 3 then reports the IV analysis identifying ATT effects.

The ITT estimates in Col. (1) are perhaps most straightforward to compare with our treatment phase results in 3.2, since they also use a panel sample and include households without any prior intervention experience. Results are similar to our treatment phase results, with the ranking between the interventions unchanged. However, due to the ITT nature of assignment and the smaller sample, estimates are more imprecise. Standard errors further increase in Col. (2) and (3), yielding similar but more imprecise

[32]Specifically, participants in the Super-control group had zero access to the treatments, while participants in the ITT groups could choose whether or not to take up the respective treatment.

estimates.

For additional, suggestive evidence, Table A13 also reports specifications that include households who received RTF, SC, or BOTH during the treatment phase. Unadjusted comparisons with Super-control in these specifications capture a mixture of prior and posterior effects: the WTP-based outcomes reflect both the lasting influence of earlier interventions and any effects arising in the WTP-based phase itself. In an attempt to separate the two components, Columns (IV), (V) and (VI) add indicators for earlier treatment experience to the ITT regressions and to both stages of the 2SLS IV regression. These indicators add or subtract a fixed amount to or from outcomes for each prior-treatment group, so the model assumes that a household's earlier treatment changes its outcome by the same amount regardless of the treatment it receives in the WTP-based phase. This assumption is restrictive. It rules out interactions between prior and current treatment, so the incremental effect of receiving SC in the WTP-based phase cannot vary with whether a household previously received RTF or BOTH, although their outcome levels may differ. The controls are also imperfect: properly isolating the influence of treatment history would require additional randomly assigned no-offer groups among households who experienced SC, RTF, or BOTH during the treatment phase, rather than households who endogenously declined an intervention in the WTP-based phase. We therefore treat both full-sample specifications as suggestive evidence and base our principal robustness check of the main treatment phase results in section 3.2 on the Control–Super-control comparison from Table A12.

**Table A12:** Treatment effects on water use per shower in the WTP-based phase, including former Control and Super-control households

| | (I) Panel ITT | (II) Cross-section ITT | (III) Cross-section ATT |
|---|---|---|---|
| SC | -4.20*** | -5.13*** | -5.67*** |
| | (1.30) | (1.68) | (1.94) |
| RTF | -10.09*** | -11.20*** | -11.56*** |
| | (2.26) | (2.38) | (2.40) |
| BOTH | -11.42*** | -10.51*** | -11.43*** |
| | (1.68) | (1.68) | (1.77) |
| IV Approach | No | No | Yes |
| No. of households | 228 | 228 | 228 |
| Observations | 14,520 | 228 | 228 |

Note: Robust standard errors are reported in parentheses. The regressions in this table use observations from the baseline phase and the WTP-based phase and include only households who were part of the Control or Super-control group in the treatment phase. The dependent variable is liters per shower. Column (I) includes household and phase fixed effects and estimates ITT effects for the WTP-based phase, based on the intervention selected for potential implementation. Columns (II) and (III) use the change in each household's average water use per shower between the baseline and WTP-based phases as the outcome. Because both the instrument (intervention assignment) and the endogenous variable (actual intervention receipt) are time-invariant at the household level, a panel two-stage least squares estimator with household fixed effects has no within-household variation in the first stage and cannot be estimated in its standard form. We therefore collapse the data to one observation per household, averaging water use per shower within each household and phase and subtracting the baseline average from the WTP-based average. This yields a single change score per household, on which we estimate a conventional cross-sectional two-stage least squares model (Column III). Column (II) is included to check that the cross-sectional transformation yields ITT estimates similar to the panel-based results in Column (I). The first stage estimates of the two-stage least squares estimator are presented in Table A14. All columns use the same household sample, requiring a valid assignment and observed water use in both phases: Households in the Super-control group are included as assigned to no intervention. Households who were in the Control group in the treatment phase are included based on their observed WTP-based assignment. Four households were excluded because they received a different intervention during the WTP-based phase than the one to which they had been assigned. Specifically, these households received RTF instead of BOTH, because at the start of the WTP-based phase, there was insufficient data available for these households to generate credible comparison reports, which were required for the BOTH intervention. In Column (I), standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

**Table A13:** Treatment effects on water use per shower in the WTP-based phase, suggestive check including all households

| | (I) Panel ITT | (II) Cross-section ITT | (III) Cross-section ATT | (IV) Panel ITT | (V) Cross-section ITT | (VI) Cross-section ATT |
|---|---|---|---|---|---|---|
| SC | -5.84*** | -5.87*** | -7.33*** | -5.17*** | -5.31*** | -5.84*** |
| | (1.15) | (1.24) | (1.57) | (1.25) | (1.46) | (1.70) |
| RTF | -9.07*** | -9.75*** | -11.42*** | -8.32*** | -9.11*** | -9.85*** |
| | (1.19) | (1.30) | (1.52) | (1.48) | (1.63) | (1.75) |
| BOTH | -12.23*** | -12.79*** | -14.34*** | -11.45*** | -12.11*** | -12.81*** |
| | (1.14) | (1.35) | (1.55) | (1.36) | (1.51) | (1.57) |
| Controlling for prior treatment experience | No | No | No | Yes | Yes | Yes |
| IV Approach | No | No | Yes | No | No | Yes |
| No. of households | 550 | 550 | 550 | 550 | 550 | 550 |
| Observations | 36,376 | 550 | 550 | 36,376 | 550 | 550 |

Note: Robust standard errors are reported in parentheses. The regressions in this table use observations from the baseline phase and the WTP-based phase. The dependent variable is liters per shower. Column (I) includes household and phase fixed effects and estimates assignment ITT effects based on the intervention selected for potential implementation in the WTP-based phase. Columns (II) and (III), and (V) and (VI) use the change in each household's average water use per shower between the baseline and WTP-based phases as the outcome. Because both the instrument (intervention assignment) and the endogenous variable (actual intervention receipt) are time-invariant at the household level, a panel two-stage least squares estimator with household fixed effects has no within-household variation in the first stage and cannot be estimated in its standard form. We therefore collapse the data to one observation per household, averaging water use per shower within each household and phase and subtracting the baseline average from the WTP-based average. This yields a single change score per household, on which we estimate a conventional cross-sectional two-stage least squares model (Columns III and VI). Columns (II) and (V) provide the corresponding household-level ITT estimates for comparison with the panel estimates in Columns (I) and (IV). Columns (IV)–(VI) control for prior treatment assignment; the panel model interacts these controls with the WTP-based phase, and the cross-sectional IV model includes them in both stages. The first stage estimates of the two-stage least squares estimator are presented in Table A14. All columns use the same household sample, requiring a valid assignment and observed water use in both phases: Households who were in the Super-control group in the treatment phase are included as assigned to no intervention. Households who were in Control, SC, RTF or BOTH are included based on their observed WTP-based assignment. Five households were excluded because they received a different intervention during the WTP-based phase than the one to which they had been assigned. Specifically, four of these households received RTF instead of BOTH, and one household received RTF instead of SC, because at the start of the WTP-based phase, there was insufficient data available for these households to generate credible comparison reports, which were required for the SC and BOTH interventions. In Columns (I) and (IV), standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

**Table A14:** First-stage estimates for the IV analysis of treatment effects in the WTP-based phase

| | Treatment received in WTP-based phase | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | (I) | (II) | (III) | (IV) | (V) | (VI) | (VII) | (VIII) | (IX) |
| MPL selected for implementation | SC | RTF | BOTH | SC | RTF | BOTH | SC | RTF | BOTH |
| SC | 0.90*** | – | – | 0.80*** | – | – | 0.84*** | 0.03*** | 0.01 |
| | (0.05) | | | (0.03) | | | (0.03) | (0.01) | (0.01) |
| RTF | – | 0.97*** | – | – | 0.85*** | – | 0.04** | 0.89*** | 0.01 |
| | | (0.03) | | | (0.03) | | (0.02) | (0.03) | (0.01) |
| BOTH | – | – | 0.92*** | – | – | 0.89*** | 0.04** | 0.03*** | 0.90*** |
| | | | (0.05) | | | (0.03) | (0.02) | (0.01) | (0.03) |
| Controlling for prior treatment experience | Control/Super-control | | | No | No | No | Yes | Yes | Yes |
| F-test of excluded instruments | 392.00 | 974.60 | 411.98 | 564.90 | 837.29 | 1212.12 | 231.53 | 431.25 | 442.52 |
| No. of households | 228 | 228 | 228 | 550 | 550 | 550 | 550 | 550 | 550 |

Note: Each column reports a separate first-stage cross-sectional regression. Columns I to III are the first stage regressions underlying Column III in Table A12, Columns IV to VI are the first stage regressions underlying Column III in Table A13, and Columns VII to IX are the first stage regressions underlying Column VI in Table A13. The dependent variable indicates whether the household received the respective intervention, and the explanatory variables indicate which MPL was randomly selected for implementation. For Columns I to VI, the reported diagonal coefficients are identical to the household-level compliance rates and because the selected MPL completely determined which intervention a household was eligible to receive, the off-diagonal cells are structural zeros and are shown as dashes. Columns VII to IX additionally control for prior treatment assignment, so their coefficients are conditional linear-projection coefficients, not unconditional receipt probabilities. The F-statistic tests the excluded MPL-assignment indicators. In Columns I to VI, the off-diagonal coefficients are structural zeros, so the F-statistic tests whether the coefficient on the corresponding MPL-assignment indicator is zero. Robust standard errors are reported in parentheses. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

### A1.3.5 Treatment effect for the opt-in scenario (subsample of those with WTP>0)

**Table A15:** Average treatment effects on water use per shower in the opt-in scenario (subsample of those with WTP>0)

| | Baseline WTP | | | Post-treatment WTP | | |
|---|---|---|---|---|---|---|
| | (I) SC | (II) RTF | (III) BOTH | (IV) SC | (V) RTF | (VI) BOTH |
| Treatment effect | -3.50*** | -9.88*** | -11.97*** | -3.44*** | -10.01*** | -11.53*** |
| | (0.79) | (1.02) | (1.40) | (0.86) | (1.06) | (1.44) |
| Control group average | 34 | 33 | 33 | 33 | 32 | 32 |
| No. of households | 182 | 191 | 200 | 179 | 184 | 188 |
| No. of observations | 13,290 | 13,963 | 13,987 | 12,664 | 13,286 | 13,165 |

Note: OLS estimates with household and calendar-date fixed effects included. The results show treatment effects for households who would choose to take up an intervention in an opt-in scenario, i.e., the estimation sample consists of those with a positive WTP before (Column I to III) or after (Columns IV to VI) treatment experience. We estimate separate regressions for each intervention, comparing intervention recipients with positive WTP for that intervention against control group members who also have positive WTP for that intervention. Consequently, the control group composition varies across regressions. Super-control households are excluded because their WTP elicitation was hypothetical. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively.

### A1.3.6 Sensitivity with regard to assumptions about WTP

For our analysis of the WTP for the interventions in Section 3.5, we took a conservative approach and set the WTP values to the respective lower limit of the respective interval. Thus, those who always opt for the intervention in an MPL received a WTP value of 15 (since they are willing to forego at least 15 EUR for receiving the intervention), those whose choices imply, for example, a WTP between 3 and 4 EUR receive a WTP value of 3, and so forth. Since the lower end of the MPL is an open interval limit for those who always chose the voucher without an intervention, we set the WTP to -20 in this case (since they are willing to forego at least 15 EUR for not receiving the intervention).

To examine the extent to which these assumptions influence our conclusions, we conduct a robustness analysis using a less conservative assumption. In the following tables, we present results on WTP, using the upper bound of the respective intervals as WTP values instead of the lower bound. Since the upper bound of the highest interval is open, we set the WTP in this case to +20, analogous to our treatment of the open lower bound of the interval in our main analysis.

Comparing Tables A16, A17, and A18 to Tables 6, 7, and 8, respectively, we find that all WTP values increase by approximately 4 EUR, with neither the quantitative order of the WTP values nor the significance of the treatment effect on WTP changing. The proportion of participants with negative WTP decreases by 2 to 3 percentage points (Table A16), underscoring that very few participants would actually be willing to pay money to avoid receiving one of the interventions.

**Table A16:** Average willingness to pay for the interventions – robustness check using the upper bound of the respective intervals as WTP values

| | | WTP elicited in the baseline phase | | WTP elicited at the end of treatment phase | | |
|---|---|---|---|---|---|---|
| | | WTP amounts in EUR | | | | |
| | | Mean | Std. dev. | Mean | Std. dev. | Obs. |
| I | WTP for SC | 13.51 | 8.66 | 12.78 | 8.77 | 439 |
| | WTP for RTF | 14.36 | 8.53 | 13.57 | 8.81 | 439 |
| | WTP for BOTH | 14.54 | 8.30 | 13.87 | 8.57 | 439 |
| | | Differences in WTP between interventions | | | | |
| | | Difference | p-value | Difference | p-value | |
| | RTF-SC | 0.85 | 0.01 | 0.79 | 0.00 | |
| | BOTH-SC | 1.03 | 0.00 | 1.10 | 0.00 | |
| | BOTH-RTF | 0.18 | 0.56 | 0.31 | 0.14 | |
| | | Participants with negative WTP | | | | |
| | | Share | Std. dev. | Share | Std. dev. | Obs. |
| II | WTP for SC | 3.4% | 18.2% | 3.2% | 17.6% | 439 |
| | WTP for RTF | 3.4% | 18.2% | 3.6% | 18.8% | 439 |
| | WTP for BOTH | 3.9% | 19.3% | 3.4% | 18.2% | 439 |
| | | Differences in negative WTP between interventions | | | | |
| | | Difference | p-value | Difference | p-value | |
| | RTF-SC | 0.0 | 1.00 | 0.5 | 0.41 | |
| | BOTH-SC | 0.5 | 0.59 | 0.2 | 0.66 | |
| | BOTH-RTF | 0.5 | 0.62 | -0.2 | 0.71 | |

Note: The sample consists of all participants who completed both WTP elicitations and who were not part of the Super-control group, as the WTP elicitations for the Super-control group were purely hypothetical. Six participants did not complete either WTP elicitation, and three of the remaining participants did not complete the second WTP elicitation.

**Table A17:** Average treatment effects on willingness to pay – robustness check using the upper bound of the respective intervals as WTP values

| | Effect of treatment experience on WTP for | | | Effect of treatment experience on share with negative WTP for | | |
|---|---|---|---|---|---|---|
| Treatment group | SC | RTF | BOTH | SC | RTF | BOTH |
| SC | -0.53 | -0.10 | 0.56 | 0.02 | -0.01 | 0.04 |
| | (1.13) | (1.12) | (1.02) | (0.03) | (0.03) | (0.03) |
| RTF | -0.78 | -0.08 | 0.58 | 0.02 | -0.04 | -0.01 |
| | (0.95) | (1.03) | (0.83) | (0.03) | (0.03) | (0.03) |
| BOTH | -1.26 | -0.42 | -1.23 | -0.01 | -0.02 | -0.01 |
| | (0.99) | (1.03) | (0.91) | (0.03) | (0.03) | (0.02) |
| Control group average | 13.32 | 14.34 | 14.26 | 0.02 | 0.03 | 0.03 |
| No. of households | | | 439 | | | |
| No. of observations | | | 878 | | | |

Note: OLS estimates with household and elicitation-period fixed effects included. Standard errors are clustered at the household level. ***, ** and * denote statistical significance at the 1%, 5% and 10% level, respectively. The omitted group is the control group who also completed incentivized WTP elicitations but did not experience any treatment during the treatment phase. The sample consists of all participants who completed both WTP elicitations and who were not part of the Super-control group, as the WTP elicitations for the Super-control group were purely hypothetical.

**Table A18:** Comparison of willingness to pay with perceived savings for those experiencing treatment – robustness check using the upper bound of the respective intervals as WTP values

| Treat. (N) | Post-treatment WTP | | Perceived savings | | WTP minus perceived savings | |
|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD |
| SC (89) | 12.07 | (8.65) | 4.12 | (6.62) | 7.94 | (11.30) |
| RTF (92) | 13.02 | (9.07) | 5.48 | (7.61) | 7.55 | (11.68) |
| BOTH (93) | 13.54 | (8.45) | 6.23 | (6.69) | 7.31 | (9.66) |

Note: In contrast to Table A16, only participants who were part of the respective treatment groups and who answered the survey question eliciting the perceived savings are included in this table.

### A1.3.7 Further welfare scenarios

**Table A19:** Monthly welfare inputs in EUR and resulting MVPF – Including the value of the shower head itself

| | $\Delta V$ | $v_H$ | $\phi_R \times \Delta R$ | $c$ | $MVPF$ |
|---|---|---|---|---|---|
| Base scenario | | | | | |
| SC | 8.59 | 0.33 | −0.24 | 1.33 | 6.89 |
| RTF | 9.13 | 0.33 | −0.73 | 1.33 | 7.67 |
| BOTH | 9.67 | 0.33 | −0.89 | 1.33 | 8.19 |
| Conservative scenario | | | | | |
| SC | 2.42 | 0.33 | −0.08 | 1.33 | 2.13 |
| RTF | 2.57 | 0.33 | −0.24 | 1.33 | 2.37 |
| BOTH | 2.73 | 0.33 | −0.29 | 1.33 | 2.52 |
| Opt-in scenario | | | | | |
| SC | 11.98 | 0.33 | −0.26 | 1.33 | 9.45 |
| RTF | 12.21 | 0.33 | −0.74 | 1.33 | 9.99 |
| BOTH | 11.89 | 0.33 | −0.86 | 1.33 | 9.83 |

Note: MVPFs are calculated as follows: $MVPF = \frac{\Delta V + v_H - \phi_R \times \Delta R}{c}$. In Table 9, we conservatively assumed that the new shower head itself does not provide a welfare benefit independent of the interventions. Here we assume that the shower head provides additional welfare benefits comparable to those of a new regular shower head (20 EUR). Amortized over the five-year lifetime, this adds 0.33 EUR per month to the benefit numerator.

**Table A20:** Monthly welfare inputs in EUR and resulting MVPF – Assuming that externalities are already fully internalized

| Base scenario | | | | |
|---|---|---|---|---|
| | $\Delta V$ | $\phi_R \times \Delta R$ | $c$ | $MVPF$ |
| SC | 8.59 | 0.00 | 1.33 | 6.46 |
| RTF | 9.13 | 0.00 | 1.33 | 6.87 |
| BOTH | 9.67 | 0.00 | 1.33 | 7.27 |
| Conservative scenario | | | | |
| SC | 2.42 | 0.00 | 1.33 | 1.82 |
| RTF | 2.57 | 0.00 | 1.33 | 1.94 |
| BOTH | 2.73 | 0.00 | 1.33 | 2.05 |
| Opt-in scenario | | | | |
| SC | 11.98 | 0.00 | 1.33 | 9.00 |
| RTF | 12.21 | 0.00 | 1.33 | 9.18 |
| BOTH | 11.89 | 0.00 | 1.33 | 8.94 |

Note: MVPFs are calculated as follows: $MVPF = \frac{\Delta V - \phi_R \times \Delta R}{c}$. Compared to Table 9, the monetized externality term $\phi_R \times \Delta R$ is set to 0, assuming that the emissions caused by showering are already internalized by the participants.

**Table A21:** Monthly welfare inputs in EUR and resulting MVPF – Subtracting overestimated cost savings from $\Delta V$

| Base scenario | | | | |
|---|---|---|---|---|
| | $\Delta V$ | $\phi_R \times \Delta R$ | $c$ | $MVPF$ |
| SC | 5.45 | $-0.24$ | 1.33 | 4.28 |
| RTF | 6.69 | $-0.73$ | 1.33 | 5.58 |
| BOTH | 7.14 | $-0.89$ | 1.33 | 6.03 |
| Conservative scenario | | | | |
| SC | 1.54 | $-0.08$ | 1.33 | 1.21 |
| RTF | 1.89 | $-0.24$ | 1.33 | 1.60 |
| BOTH | 2.01 | $-0.29$ | 1.33 | 1.73 |
| Opt-in scenario | | | | |
| SC | 8.84 | $-0.26$ | 1.33 | 6.84 |
| RTF | 9.77 | $-0.74$ | 1.33 | 7.91 |
| BOTH | 9.35 | $-0.86$ | 1.33 | 7.68 |

Note: MVPFs are calculated as follows: $MVPF = \frac{\Delta V - \phi_R \times \Delta R}{c}$. Perceived monthly monetary savings are depicted in Table 8 and amount to 4.12 EUR (SC), 5.48 EUR (RTF), and 6.23 EUR (BOTH), while we estimate the actual monthly savings to be 1 EUR (SC), 3 EUR (RTF), and 3.7 EUR (BOTH). In this table, we subtracted the difference between perceived and actual monetary savings from $\Delta V$ for each intervention. For the conservative scenario, we first adjust the base-scenario WTP for overestimated savings and then apply the same proportional reduction as in Table 9.

## A2 Relation to the Pre-Analysis Plan

The experimental design and planned analyses were registered before the randomized treatment phase began (Anonymous, 2020). This section describes how the analyses reported in the paper relate to the pre-analysis plan (PAP). The central elements of the PAP are retained. The paper compares social comparison reports (SC), real-time feedback (RTF), and their combination; uses water consumption per shower as the primary behavioral outcome; estimates treatment effects with household fixed effects; and examines willingness to pay (WTP) before and after treatment experience. The combination treatment is called DUAL in the PAP and BOTH in the paper. The remaining differences concern the scope of the reported experimental phases, the realized sample, construction of the shower data, the welfare analysis, and a set of additional analyses developed after fieldwork.

### A2.1 Scope of the reported experiment and realized sample

The PAP described four experimental phases. The present paper covers the baseline phase, the randomized treatment phase, and the WTP-based phase. It focuses on the three core informational interventions, SC, RTF, and BOTH. The supplementary fourth phase set out in the PAP comprised carbon offsetting and changes in showering costs. The corresponding RTF+CO, RTF+CC, and CO comparisons are outside the scope of the current paper.

The WTP-based phase corresponds to the phase in the PAP in which intervention receipt depends on a randomly selected WTP decision, making the WTP elicitations consequential for all groups except the Super-control group. The paper also uses the random implementation draws as instruments when studying behavior in this phase. This instrumental-variables analysis was developed after registration and is presented as a robustness exercise in the Appendix. The main behavioral estimates continue to come from the randomized treatment phase.

The recruitment and sample figures in the manuscript reflect the completed fieldwork and differ from the preliminary flow recorded in the PAP. The PAP reported

6,300 invitations, 946 interested households, 753 eligible households, and 640 accepted households, with a target of about 600. The final manuscript reports 9,376 invited panelists, 1,100 registrations, 685 confirmed registrants, 647 completed baseline surveys, and 574 households with a successful installation. The main behavioral specification uses 564 households because ten households established a valid data connection only after the baseline phase.

The final randomization allocated approximately equal numbers to the five experimental groups, which produced a larger Super-control group than the group of 63 described in the PAP. Treatment assignment remained randomized and stratified on the covariates listed in the PAP. Inference in the paper is based on the realized sample and household-clustered standard errors. The ex ante minimum detectable effect calculations remain in the PAP and are not repeated in the manuscript.

## A2.2 Primary behavioral analysis

The primary behavioral outcome and identifying comparison follow the PAP. The outcome is water use in liters per shower, and the treatment effects compare SC, RTF, and BOTH with untreated households during the randomized treatment phase. The main specification is a difference-in-differences model with household and day fixed effects, with standard errors clustered at the household level. The equation is estimated at the shower-session level, so each recorded session enters as one observation. For analyses that do not rely on incentivized WTP or treatment-specific reference points, the Control and Super-control groups are pooled. The PAP specified a fixed-effects approach without fixing the precise time controls, error structure, or pooling rule; the manuscript provides these implementation details. The predicted ordering in the PAP, with the largest absolute effect for BOTH, followed by RTF and SC, is evaluated directly, including pairwise comparisons across treatment coefficients.

The PAP left several rules for translating the device records into an analysis data set open. A new shower session begins after the shower head has been inactive for at least 180 seconds. The main sample excludes recorded sessions below 5 liters or above 200 liters, since these observations are especially likely to represent water withdrawals

other than showers or technical recording problems. When a gateway was offline, the shower head stored the records and transmitted them later; the corresponding timestamps are distributed evenly over the interval between the last and next successful transmissions.

The appendix examines the sensitivity of the main estimates to these choices. It reports results under alternative volume cutoffs, shower-duration restrictions, no volume exclusions, and samples defined by the length of interruptions within a recorded session. These checks leave the ranking and the broad magnitude of the treatment effects intact. Shower temperature and daily shower frequency are also analyzed to assess whether reductions in water use per shower are offset along other margins.

## A2.3 Willingness-to-pay analyses

The WTP analyses retain the two objectives set out in the PAP: comparing valuations of SC, RTF, and BOTH before treatment experience, and estimating how experience with each intervention changes WTP for all three interventions. The Super-control group completed hypothetical WTP elicitations and is excluded from analyses that interpret WTP as an incentivized measure. The main WTP sample contains participants who completed both elicitations. Experience effects are estimated in a two-period household fixed-effects specification using the consequential Control group as the untreated comparison. The PAP left the regression equation for this outcome open.

The coding of interval-valued WTP responses was determined during analysis. The main results use the lower bound of each interval. Participants who always selected the intervention are assigned a WTP of 15 EUR, and participants who always selected the monetary alternative are assigned a WTP of -20 EUR. An appendix analysis uses the upper interval bounds and obtains the same ordering across interventions and the same conclusions about experience effects.

The manuscript reports several WTP-based outcomes beyond those listed in the PAP. These include the share of participants with negative WTP, perceived monetary savings, the difference between post-treatment WTP and perceived savings as a lower-bound measure of non-monetary benefits, heterogeneity in behavioral treatment ef-

fects by WTP, and a continuation-demand scenario restricted to households with positive post-treatment WTP. These analyses provide additional information about consumer welfare and market-based allocation.

## A2.4 Welfare analysis

The PAP proposed a structural model that would separately recover the removal of salience bias, consumption-dependent psychological costs or benefits around a reference point, and fixed psychological costs or benefits. The structural decomposition was not carried forward to the final analysis. Post-treatment WTP directly captures the net value participants assign to each intervention and supports the paper's comparative welfare analysis. Separating this net value into the three latent channels would require additional functional-form and identifying assumptions. The welfare analysis therefore leaves the channels aggregated in WTP.

The paper evaluates welfare through a transparent reduced-form accounting and reports the marginal value of public funds (MVPF). Post-treatment WTP measures the value households place on receiving another month of the intervention after treatment experience. The treatment-induced reduction in hot-water use is converted into avoided carbon externalities, and the resulting social benefit is related to the monthly cost of providing the intervention. This approach preserves the PAP's objective of combining behavioral effects, consumer valuation, and externalities while avoiding a parametric decomposition of the psychological channels.

The base, conservative, and continuation-demand scenarios were developed after registration. The same applies to the assumptions on the social cost of carbon, the lifetime and cost of the smart shower head, and the sensitivity analyses that vary device costs, assume full internalization of the carbon externality, or adjust WTP for overestimated monetary savings. These scenarios make the role of the welfare assumptions explicit and show how the conclusions change under alternative parameter values.

### A2.5 Secondary and additional analyses

The PAP listed shower frequency, possible switching to another shower, perceived effectiveness, and participants' self-assessment of resource use as secondary outcomes. The paper reports effects on daily shower frequency. Since each household received one monitored shower head, switching to an unmonitored shower cannot be observed directly; the appendix instead compares treatment effects between households with one shower and households with multiple showers. The paper also uses the survey measure of perceived monetary savings. The broader set of subjective effectiveness and self-assessment measures listed in the PAP remains outside the scope of the current manuscript.

Further analyses examine within-household variation in water use, the color range in which showers end, pauses during a shower, heterogeneity by baseline consumption, heterogeneity by the randomly assigned reference point, heterogeneity by WTP, and differences by household size, number of showers, and gender in one-person households. These analyses extend the PAP. They are used as exploratory or robustness evidence and support the interpretation of the main randomized treatment effects.

Overall, the main randomized treatment and WTP analyses align closely with the PAP. The differences concern the realized sample, the scope of the reported phases, the rules used to construct shower sessions, the reduced-form welfare accounting, and the additional robustness and exploratory analyses described above.

## A3 Invitation for study participation

Dear XYZ!

With this email we would like to invite you to participate in a new study by our long-time scientific partner, [redacted]. The study is financed by the Ministry of Culture and Science of the State of North Rhine-Westphalia and examines ways to reduce resource consumption. In a first step, only our panelists

have the opportunity to participate.

WHAT EXACTLY IS THE STUDY ABOUT?

As a core element of the study, participants will be provided with a high-quality shower head that measures water consumption per shower. During the four-month study phase, this information will be automatically shared with [redacted]. Additionally, you will be provided with various information about your showering behavior during the course of the study. At the end of the study, you will be able to keep the shower head, which then ceases to share information. This will allow you to keep track of your water consumption yourself on a permanent basis.

HOW CAN I SIGN UP?

You can find all further information and registration form on the project website: [URL]

YOUR BENEFITS

As an **expense allowance**, you will receive vouchers or additional resource saving tools worth a total of €35 in addition to the shower head before the end of the study. You will receive these vouchers as part of your participation in short, monthly surveys conducted by [redacted] to accompany the study. At the end of the study, we will offer you a selection of popular suppliers where you can redeem your vouchers. Furthermore, you will of course be informed about the results of the study in the form of a short report. So again in a nutshell:

- Modern shower head with innovative functions
- Vouchers as an expense allowance - choose your preferred provider from our list of providers
- Report with the results of the study
- Participation in a study, the findings of which will make the world a little bit more sustainable

WHY SHOWERS?

What's not commonly known is that showering involves a relatively high consumption of resources, as not only water but also significant amounts of energy are used to heat the water. For example, with the energy consumed by an ordinary shower, you could light an average household for two and a half days, or watch TV for 26 hours.

HOW IS MY DATA PROTECTED?
The study is subject to strict data protection regulations. More detailed information can be found on the project website: [LINK].

DON'T FORGET TO REGISTER!
You can register on the project homepage [LINK].

If you have any questions about the study, please contact the project team at [E-MAIL ADDRESS].

Your forsa.omninet Team

*

## A4 Website to register for the study

The project website aimed at participants included the following subpages:

- **Home page:** Welcome message and general introduction to the study
- **Project goals:** Short overview of the experiment's objectives
- **Your contribution:** Information on environmental potential of the study
- **FAQ:** Answers to typical questions
- **Who we are:** Short overview of the research team
- **Sign up:** Overview of the general terms & conditions and signup form

The following subsection features a translated version of selected pages.

### A4.1 Website text content

**Home Page**

Welcome to the "sustainable showering" project homepage!

We are pleased that you are interested in our research project! On this website, we would like to introduce the project to you and hope to win you as a study participant.

This project is a scientific research project of [redacted]
in research partnership with [redacted].

The project is funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia under the title "Digitization of Sustainable Behavior". [Click here to go to the project page on the [redacted] homepage. Please note that you can only register on the current page].

As a core element of your study participation, you will be provided with a high-quality shower head that measures water consumption per shower. During the four-month study phase, this information will be automatically shared with [redacted]. Additionally,

you will be provided with various information about your showering behavior during the course of the study. At the end of the study, you will be able to keep the shower head, which then ceases to share information. This will allow you to keep track of your water consumption yourself on a permanent basis.

As an expense allowance, in addition to the shower head, you will receive optional vouchers and/or additional resource saving tools worth a total of €35 until the end of the study. You will receive the vouchers as part of your participation in short, monthly surveys that will be conducted to accompany the study.

Feel free to take a further look at the project page and sign up!

**Project goals**

"Sustainable showering":
Of course, showering is an indispensable part of personal hygiene. Simultaneously however, showering is a very resource-intensive activity and hardly perceived as such. Many people underestimate the associated energy consumption, caused in particular by heating water. For example, an average shower consumes about 2.6 kilowatt hours. The same amount of energy could be used to light an average household for two and a half days, or to watch television for 26 hours. So the question is whether it is possible to shower more "sustainably."

The goals of the project:

- Gather information on the exact amount of energy and water used in showering.
- Development of helpful measures to reduce resource consumption.

By saving energy consumption while showering, an individual's CO2 emissions can be significantly reduced. An average shower taken daily causes annual emissions of 325 kg of CO2. It takes 26 trees to reabsorb this amount of CO2 within one year. Or put another way: Just 14 showers emit as much CO2 as a tree can absorb within a year.

**Your contribution**

Your contribution to climate protection:
By participating in this study, you are making a concrete contribution to climate protection. This happens in two ways:

- You gain exclusive access to the savings measures investigated in this study, which can be used to reduce your personal resource consumption and the associated energy costs and CO2 emissions.
- The knowledge generated in the study can be used to formulate concrete recommendations for resource conservation in the context showering and other energy-intensive activities. In this way, policy measures can be developed to effectively save CO2 emissions.

Every reduction in CO2 emissions makes a valuable contribution to combating climate change. Further information on the causes and consequences of climate change can be found, for example, on the website of the Federal Agency for Civic Education: https://www.bpb.de/gesellschaft/umwelt/klimawandel/.

Our project is primarily focused on saving energy, i.e., measures are being developed to save hot water. However, water is also a valuable resource in itself and is scarce in many parts of the world, as well as regionally in Germany, at least at times. See, for example, here: https://www.umweltbundesamt.de/presse/pressemitteilungen/wassersparen-sinnvoll-ausgereizt-uebertrieben. Thus, with the study, we will identify scientific findings that are also of great interest to other parts of the world.

**FAQ**

Do I receive an expense allowance for my participation?
Yes. You will be compensated in two ways for the time and effort you have incurred. First, you get to keep the high-quality shower head used in the study after completion. With this shower head, you can continue to keep track of your water consumption by using a smartphone app. Second, you will receive optional vouchers and/or additional resource saving tools worth €35 until the end of the study. You will receive the vouchers as part of your participation in short, monthly surveys that accompany the study. At the end of the study, we will offer you a selection of popular suppliers where you can redeem your vouchers.

How much effort will it take for me to participate?
At the beginning of the study, we will send you a shower head and a WiFi gateway, which you can easily install yourself with the help of the enclosed instructions. If you encounter any difficulties, we will gladly assist you with the installation. During the course of the study, we will send you monthly online surveys by e-mail, each of which can be completed in about 15 minutes. In addition, you will receive newsletters from us at regular intervals, which you only need to read. The study involves no further effort on your part and is completed after four months.

How is my data collected?
The study is subject to strict data protection regulations. For details, please read the project's privacy policy.

I am going on vacation within the study phase. Is this a problem?
No. However, it would be nice if you let us know when you register. You have the option to do this on the registration form under "Notify the Project Team" or by emailing [E-MAIL ADDRESS].

There is more than one shower in my household. Can I still participate?
Of course. In this case, we ask you to install the new shower head at the shower that you personally use most often.

My shower is not compatible with conventional shower heads.

If you for example have a Rain Dance shower that requires a special shower head, participation is unfortunately not possible. If you are unsure about your shower head model, please feel free to contact us. You can find the contact information here.

## A4.2 Signup form content

1. We kindly ask you to provide information about the number of persons living in your household permanently:

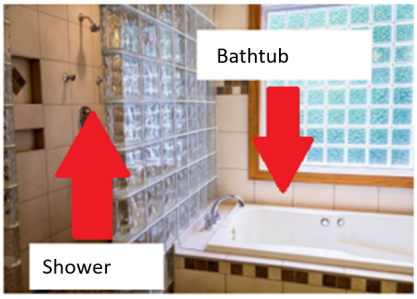
Household size

- Total number of persons (adults and children)
- Number of children (age below 18)

2. We kindly ask you to provide information about your shower setup:




Picture 1: Shower and bathtub combined




Picture 2: Shower and bathtub separate

How many separate or combined showers and bath tubs are there in your home?

- Number of bath tubs with shower (combination, see picture 1)
- Number of separate showers (not in bath tub, as seen in picture 2)
- Number of bath tubs (without shower or not used as shower, as seen in picture 2)

3a. This question is about the shower you use most frequently:

How many household members use this shower at least once a week?

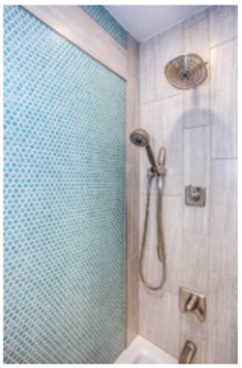
- Number of shower users



Picture 1: The shower head is screwed onto a shower hose.



Picture 2: The shower head protrudes from the wall.



Picture 3: There are two shower heads. One is screwed onto a shower hose, and the other protrudes from the wall.

Which of these descriptions fits the shower in question?

- No selection made
- The shower head is screwed onto a shower hose (as seen in picture 1).
- The shower head protrudes from the wall (as seen in picture 2).
- There are two shower heads: One is screwed onto a shower hose and one protrudes from the wall (as seen in picture 3).

Visually, the shower head we are going to provide you with resembles an ordinary hand shower (see picture below). It can be screwed onto ordinary 1/2 inch shower hoses, which can be found in most households.



3b. We kindly ask for your assessment about the shower you personally use the most

- No selection made
- I think my shower is compatible with this shower head
- I am unsure whether my shower is compatible with this shower head or not
- I do not think my shower is compatible with this shower head

4. Technical requirements

To set up data transmission, you need to connect the shower head to your WiFi network using a smartphone app. This step can be completed in a minute and is necessary for study participation. To perform this step, you will need your WiFi password and a smartphone. The WiFi password will not be shared with us.

- No selection made
- I own a smartphone and have access to my WiFi password
- I own a smartphone and am unsure about my WiFi password
- I do not own a smartphone and/or do not have access to my WiFi password

In addition to the smart shower head, we will provide you with a WiFi gateway, which is required to connect the shower head to your network. This device is very small (dimensions: 4.5 x 4.5 x 1cm, see picture below). It should be located within a radius of about 5 meters around your shower (ideally in the same room) and should be continuously powered.



- I have access to an electrical outlet within a radius of 5 meters around my shower

## A5 Exemplary newsletter

### Part of the newsletter sent to all study participants



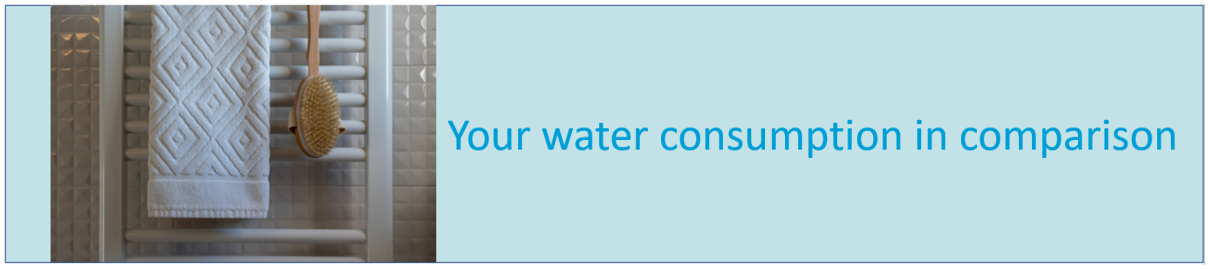

Dear Mr/Ms XYZ,

do you remember how much energy is consumed in a five-minute shower at a water temperature of 38 degrees Celsius? That's right, it's 2.2 kWh!
With these 2.2 kWh you could also

- toast 132 slices of bread
- boil 16.5 litres of water at 100 degrees Celsius with an electric kettle

The generation of 2.2 kWh of energy emits 650 g of CO2. It takes 19 beech trees to reabsorb this amount of CO2 within one day.

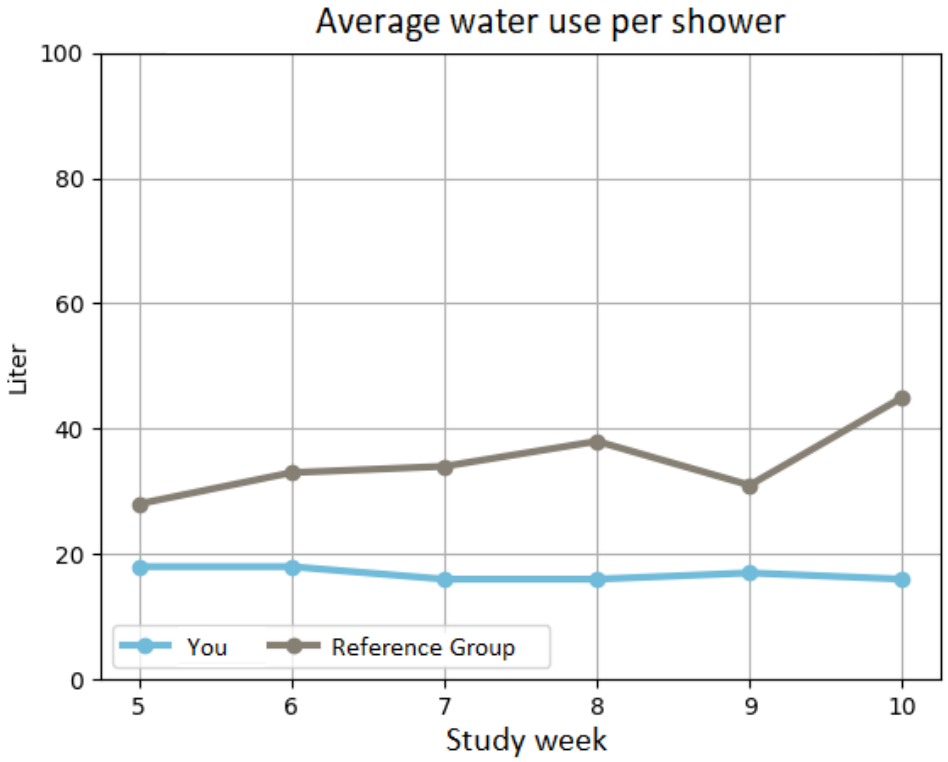
### Part of the newsletter sent only to SC and BOTH



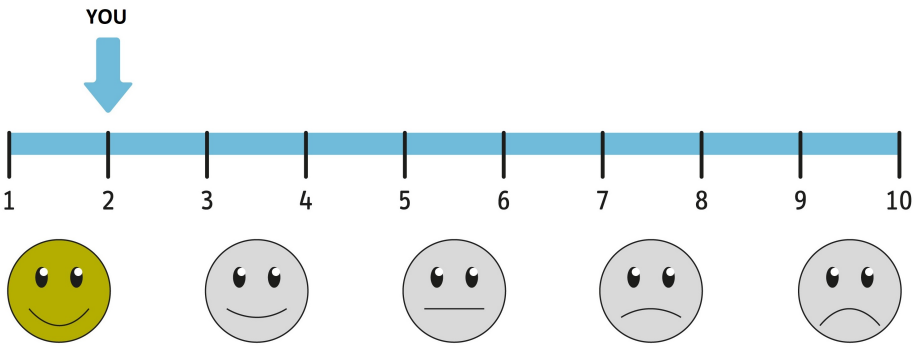

In the last two weeks, you have used an average of **17 litres of water** per shower. At your average water temperature and the type of water heating used by an average German citizen, this corresponds to emissions of **244.5 g of CO2 per shower**. It takes 7.1 beeches to reabsorb this amount of CO2 within one day.

The average consumption of 9 other randomly selected participating households was **36 litres**. The average CO2 emission per shower was **537.0 g CO2 per shower**. It takes 15.7 beeches to reabsorb this amount of CO2 within one day.

In the following figure you can see how your average consumption and the average consumption of your comparison group developed in the last weeks. The last two weeks were weeks 9 and 10 of the study.

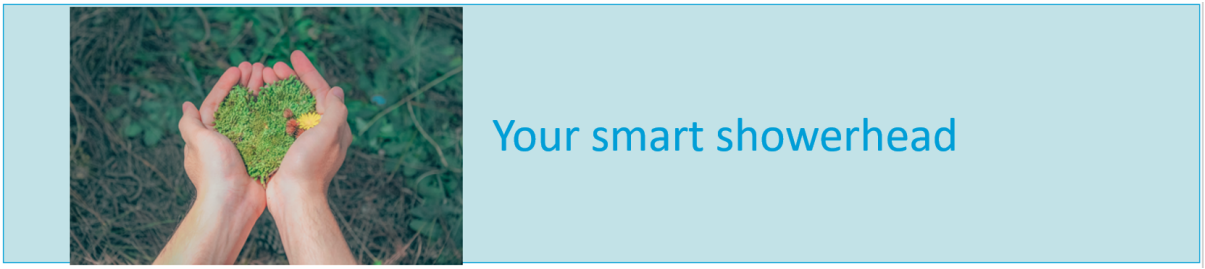



**Compared to 9 other participating households, your rank is 2nd place:**

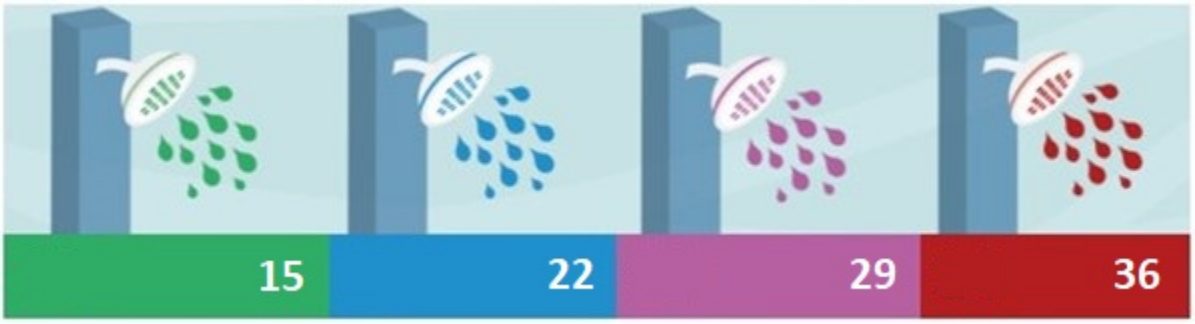



One of these households had lower average water use per shower than you. Eight of these households had a higher average water consumption.

**Part of the newsletter sent only to RTF and BOTH**




Your shower head will help you keep track of your water consumption while showering. It will light up in color and as your water consumption increases, the color will change. Finally, once your water consumption exceeds 36 litres, it will start flashing red. 36 litres of water consumption corresponds to 537 g of CO2 emissions based on your average water temperature and the energy consumption of an average German citizen. It takes 15.7 beech trees to reabsorb this amount of CO2 within one day. The exact sequence of colours is as follows:




During the first 15 litres of your shower, the shower head will glow green.
After **15 litres** (225 g $CO_2$ emissions), your shower head will begin to glow blue.
After **22 litres** (330 g $CO_2$ emissions), your shower head will begin to glow purple.
After **29 litres** (435 g $CO_2$ emissions), your shower head will start to glow red.
Finally, after **36 litres** (537 g $CO_2$ emissions), your shower head will begin to flash red.

Sincerely,
Your "sustainable showering" team

Please do not reply to this mail. For questions and comments, please contact us at [E-MAIL ADDRESS].

## A6 Technical details and installation of the smart shower heads

Participants received a package with the smart shower head and their WiFi gateway. In parallel, participants received the installation instructions in digital form via email. Upon request, a printed version of the instructions was enclosed in the package.
The installation of the smart shower heads and the infrastructure for data transmission took place in the following steps.

1. The existing shower head had to be unscrewed from the shower hose and the smart shower head had to be screwed on.
2. The proper functioning of the shower head could be confirmed by brief light signals, which the shower head emits during start-up. Apart from this short signal, the shower head functioned like any regular shower head during the Baseline Phase.
3. The WiFi gateway had to be plugged into a power outlet near the shower head. Participants then had to download an app to connect the WiFi gateway to their home network. This was done by selecting the correct WiFi network and entering the network key. A successful connection of the gateway was indicated by light signals.
4. If problems arose, participants could reach the project team by e-mail. The problems were then solved either by e-mail, phone call or personal visit. Only for a minority of the study participants were the problems not solvable. In most cases, the issues resulted from WiFi configuration errors or too much distance between the gateway and the smart shower head, which could be easily solved.

After successful installation, the shower head transmitted information for every shower taken (time stamp, amount of water used, average water temperature, water flow, length of shower breaks) via Bluetooth to the WiFi gateway, which then transmitted the information to the research team via the Internet. Participants had no way

to access this information during the study.

If the WiFi gateway was not plugged in or had no internet connection during a shower, the data from that shower was stored in the shower head (up to 200 showers can be stored) and transmitted with the next successful connection. Interruptions of showers of less than 3 minutes were interpreted by the smart shower head as shower breaks, while an interruption of more than 3 minutes signaled the start of a new shower.

## A7 Details on the calculation of shower costs and carbon intensity

In the following, we describe our approach and assumptions to calculate (1) the costs and (2) the carbon intensity per liter of shower water. For the monetary-savings calculations, we assume an approximate cost of 0.01 EUR per liter, including water, wastewater, and heating energy. For comparison, Consumer Protection Agency NRW (2019, p. 5) report total costs of 0.60 EUR for an 80-liter shower at 38°C with gas- or oil-fired condensing heating, and 1.12 EUR with an electric instantaneous water heater. These examples imply costs of 0.0075 and 0.014 EUR per liter, respectively. Our assumption is a simplified benchmark rather than a household-specific cost estimate. For the separate calculation of carbon intensity, we use the heating-energy costs per shower and energy prices reported by Consumer Protection Agency RLP (2020) to infer energy consumption. The assumed temperature in their calculations corresponds to the average baseline temperature in our control group of 38°C (Table 2). For each energy source (electricity, gas, and oil), we divide the heating-energy cost per shower by the price per kilowatt-hour and the shower volume to obtain energy demand per liter of shower water. We then multiply the energy demand per liter of shower water with the $CO_2$ intensity of one kilowatt-hour of heat supply with the respective energy source. For gas- and oil-fired water heating, we use the estimates of the $CO_2$ intensity provided by the German Federal Environment Agency (German Environment Agency, 2019; Table 62 "Emission factors for the provision of heat from fossil fuels in private households, the tertiary sector and industry"). For electric water heating, we use an estimate for the marginal emission factor of electricity generation in Germany provided by Huber et al. (2021). The use of a marginal emission factor compared to average emissions per kWh is appropriate because a reduction in electricity demand leads to a reduction in electricity generation at the marginal power plant, which is typically fossil fuel-fired and therefore more carbon-intensive than the average electricity mix (see, e.g., Holland et al., 2022). We then use data from the German Federal Ministry

for Economic Affairs and Energy (2020) on the distribution of energy consumption for water heating in private households to calculate a representative $CO_2$-factor per liter of shower water. In the base scenario, we make the assumption that shower water in Germany is heated only by electricity, gas or oil. This is a simplification, but it includes a large 84% of the final energy consumption for water heating in German households in 2018. The remaining 16% of final energy consumption for water heating was supplied by district heating or renewable energy sources. As a representative $CO_2$-factor, this calculation yields a value of 0.015 kg $CO_2$ per liter of shower water.

In the conservative scenario, we reduce our emission factor for one liter of hot water consumption by 16%, assuming that the share of energy used to heat water in private households that comes from renewable energy or district heating is associated with zero carbon emissions. While district heating does in principle entail carbon emissions, there may be cases where it is fed entirely by industrial waste heat, so that the use of this waste heat does not result in further emissions, which is what we assume in the conservative scenario. The calculation in the conservative scenario results in external costs of 0.002 EUR/liter.

**Table A22:** Estimation of $CO_2$ per liter

| | (I) Energy cost per shower (40 L at 38°C) | (II) Cost per kWh | (III) kWh per L at 38°C | (IV) kg $CO_2$ per kWh | (V) kg $CO_2$ per L | (VI) PJ used for water heating in 2018 | (VII) Share | (VIII) kg $CO_2$ per L (weighted avg.) |
|---|---|---|---|---|---|---|---|---|
| Elec. | 0.40 EUR | 0.28 EUR | 0.036 | 0.550 | 0.020 | 55.307 | 17.9% | |
| Gas | 0.10 EUR | 0.05 EUR | 0.050 | 0.246 | 0.012 | 190.384 | 61.6% | 0.015 |
| Oil | 0.15 EUR | 0.07 EUR | 0.054 | 0.318 | 0.017 | 63.307 | 20.5% | |
| | | | $\frac{(I)/40}{(II)}$ | | $(III) \times (IV)$ | | $\frac{(VI)}{\sum(VI)}$ | $\sum[(V) \times (VII)]$ |

Sources: First column: Consumer Protection Agency RLP (2020), Second column: Consumer Protection Agency RLP (2020), Fourth column: Huber et al. (2021) for Elec. & German Environment Agency (2019) for Gas/Oil, Sixth column: German Federal Ministry for Economic Affairs and Energy (2020). In this table, $CO_2$ is used as shorthand for $CO_2$-equivalent emissions. The gas and oil emission factors include non-$CO_2$ greenhouse gases. For the welfare calculation, we apply the social cost per tonne of $CO_2$ to these emissions as an approximation: equivalence in global-warming potential does not necessarily imply identical monetary damages across greenhouse gases.