Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Umwelt, Verkehr, Energie und Kommunikation UVEK

Bundesamt für Energie BFE

Final Report July 06, 2014

ewz-Amphiro Study

On the Effectiveness of Real-Time Feedback:

The Influence of Demographics, Attitudes, and Personality Traits

Auftraggeber:

Bundesamt für Energie BFE

Forschungsprogramm EWG

CH-3003 Bern

www.bfe.admin.ch

Auftragnehmer:

Bits to Energy Lab

Information Management

Department of Management, Technology, and Economics, ETH Zurich

Weinbergstrasse 56/58

CH-8092 Zürich

www.bitstoenergy.ch

Département d'économétrie et économie politique

Université de Lausanne

Internef

CH-1015 Lausanne

http://www.hec.unil.ch/deep/home

Autoren:

Verena Tiefenbeck, ETH Zürich

Lorenz Götte, Université de Lausanne

Kathrin Degen, Université de Lausanne

Vojkan Tasic, ETH Zürich

Thorsten Staake, Universität Bamberg

BFE-Bereichsleiterin Anne-Kathrin Faust / **BFE-Programmleiterin** Anne-Kathrin Faust

BFE-Vertrags- und Projektnummer: 8100087/500791-01

Contact:
Verena Tiefenbeck
ETH Zurich
Weinbergstrasse 56/58, WEV G 218
CH-8092 Zurich
Phone: +41 44 632 3953
Email: vtiefenbeck@ethz.ch

Please cite this manuscript as follows: Tiefenbeck, Verena; Degen, Kathrin; Tasic, Vojkan; Goette, Lorenz; Staake, Thorsten, "On the Effectiveness of Real-Time Feedback: The Influence of Demographics, Attitudes, and Personality Traits". Final report to the Swiss Federal Office of Energy, Bern, 2014.

Contents

List of Tables

List of Figures

Acknowledgements We wish to thank the Swiss Federal Office of Energy for financing the research activities of this study and for providing valuable advice throughout all stages of our work. In particular, we want to express our gratitude to Dr. Nicole Mathys for her ideas and input as well as for initiating the collaboration between the two research teams of ETH Zurich and the University of Lausanne. We are also grateful for the valuable reviews by Dr. Boris Krey and Dr. Ann-Kathrin Faust that helped to improve the final report considerably. We would also like to express our gratitude to the utility company ewz for financing the study devices and for their confidence, trust, cooperation and continuous support. Our particular gratitude goes to Dr. Lukas Küng and Dr. Romeo Deplazes (ewz) for their early, outstanding, and continuous support of the project - without them, the endeavor would not have been possible. Special thanks also goes to Mirjam Avdyli, Fabian Frei, Gian Andrea Jäger, and Valentin Peter for all their ideas, suggestions, support, engagement, and flexibility throughout the study implementation. Without them and their outstanding customer insight and experience regarding all ewz-processes the quality of all operations-related activities would have been impossible to maintain. We would also like to thank all members of the ewz Kundenzentrum for their support with questions of study participants and our colleagues at ETH Zurich and the University St. Gallen: First and foremost to Samuel Schöb for numerous programming and read-out night-shifts; thanks as well to Flavius Kehr, Felix Lossin, and Dr. Claire Loock for their critical review of the surveys. Furthermore, we also wish to thank Prof. Elgar Fleisch and Prof. Rafael Lalive for their valuable guidance, support, and encouragement. Last but not least, we also want to acknowledge Dr. Thomas Stiefmeier, Liliane Ableitner and Michael Benjert (Amphiro AG), for their collaboration and support.

Authors Verena Tiefenbeck and Vojkan Tasic are both Ph.D. candidates at the Department of Management, Technology, and Economics (D-MTEC) of ETH Zurich. Kathrin Degen is a Ph.D. candidate at the Faculty of Business and Economics of the University of Lausanne (HEC). Lorenz Götte is professor for microeconomics at the Faculty of Business and Economics of the University of Lausanne and member of the research networks CEPR, IZA, and of the institute of macroeconomics at the University of Lausanne. Thorsten Staake is professor for Energy Efficient Systems at the University of Bamberg, co-founder of Amphiro AG and director of the Bits to Energy Lab at ETH Zurich.

Abstract

Feedback interventions that indicate personal energy consumption have received much attention among scholars and practitioners alike. Due to their cost-effectiveness, political feasibility, and scalability, such programs have been rolled out to millions of households. Recent studies that lasted between 6 months and 5 years have documented reductions in the range of 1-6%. While it has been shown that feedback is more effective when provided on a specific behavior and right at the point and time of use, a demonstration of the impact and cost-effectiveness of feedback in such favorable conditions on a larger scale is still missing.

This study investigates the impact of behavior-specific real-time feedback (here: on hot water consumption in the shower) and evaluates factors and mechanisms facilitating its effectiveness in a randomized controlled trial with 697 households. Overall, participants who received real-time feedback reduced both their energy and their water consumption by 23% compared to the control group. The effects are sustained throughout the study period of two months. Projected to one year and assuming persistence of the effect, this results in yearly savings of 443 kWh, 8,500 liters of drinking water, 94 kg of CO_2, and CHF 110, making the device cost-effective within 6-9 months. Individuals with high baseline consumption show a stronger behavioral response to the intervention, as do participants with a general tendency to monitor progress towards goals. While environmental attitudes drive the effect behind the scenes, they do not significantly affect the overall net treatment effect, as they are strongly negatively correlated with baseline consumption. The intervention also substantially increases knowledge about resource consumption. Conversely, the results do not support any evidence that negative psychological pressure might drive the treatment effect. The baseline data also indicate that the amount of energy and water used is negatively correlated with age, with 20-29 year-olds using 72% more resources per shower than participants over 65. Yet given their higher baseline consumption, young people respond stronger to the feedback, making them a valuable target for feedback campaigns.

The study shows that behavior-specific feedback can prompt substantial behavior change. Moreover, profiling, e.g., targeting households with an above-average baseline consumption, can even further raise program cost-effectiveness. The findings also suggest that positive mechanisms drive the conservation effect, not negative psychological pressure. Altogether, the results indicate that behavior-specific real-time feedback is highly cost-effective and scalable.

Zusammenfassung

Feedback-Interventionen, welche den persönlichen Energieverbrauch als Folge des eigenen Handelns aufzeigen, haben in Literatur und Praxis erhebliche Beachtung gefunden. Aufgrund ihrer Wirtschaftlichkeit, ihrer politischen Durchführbarkeit und ihrer Skalierbarkeit wurden solche Programme bereits mit Millionen von Haushalten durchgeführt. Neuere Feldstudien mit Laufzeiten zwischen 6 Monaten und 5 Jahren erzielten Energieeinsparungen zwischen 1% und 6%. Obwohl gezeigt wurde, dass die Wirksamkeit solcher Programme höher ist, wenn die Verbrauchsinformationen für eine einzelne Handlung am Ort und zum Zeitpunkt der Durchführung zur Verfügung gestellt werden, wurde bislang noch kein Nachweis für die Wirtschaftlichkeit und Skalierbarkeit für diese Art von Feedback in grösserem Massstab erbracht.

Diese Studie untersucht die Wirkung von verhaltensspezifischem Echtzeitfeedback (hier: zum Warmwasserverbrauch beim Duschen) und evaluiert Faktoren und Mechanismen, die dessen Wirkung beeinflussen. Hierzu wurde eine randomisierte kontrollierte Feldstudie mit 697 Haushalten durchgeführt. Teilnehmer, die Echtzeit-Feedback zu ihrem Duschverbrauch erhielten, reduzierten ihren Energie- und ihren Wasserverbrauch gegenüber der Kontrollgruppe um durchschnittlich 23%. Der Effekt ist stabil über die zweimonative Studiendauer. Unter der Annahme, dass der Effekt auch längerfristig anhält, betragen die Einsparungen auf ein Jahr hochgerechnet durchschnittlich 443 kWh, 8,500 Liter Trink- und Abwasser, 94 kg CO_2 und 110 CHF, wodurch sich die Feedback-Anzeige innerhalb von 6-9 Monaten amortisiert. Teilnehmende mit einem hohen Grundverbrauch zeigen eine stärkere Reaktion auf die Intervention als Teilnehmende mit einem niedrigen Grundverbrauch. Die Wirkung ist auch stärker auf Teilnehmende, die generell dazu neigen, ihre persönlichen Leistungen anhand von Zielen zu messen. Während Umwelteinstellung der Teilnehmer zwar im Hintergrund den Spareffekt zu treiben scheint, beeinflusst diese nicht die Nettowirkung der Massnahme, da Umwelteinstellung stark negativ mit dem Grundverbrauch korreliert. Die Massnahme steigert zudem in erheblichem Masse das Wissen zum eigenen Ressourcenverbrauch. Hingegen finden sich keinerlei Belege dafür, dass die Einsparung durch negative psychologische Effekte bedingt werden könnte. Bemerkenswert ist zudem, dass der Grundverbrauch stark negativ mit dem Alter korreliert: 20-29-Jährige verbrauchten 72% mehr Energie und Wasser pro Dusche als Teilnehmende über 64. Dem höheren Grundverbrauch entsprechend reagieren junge Teilnehmende stärker auf Feedback als ältere und bilden somit eine gute Zielgruppe für solche Massnahmen.

Die Studie zeigt, dass verhaltensspezifisches Echtzeitfeedback beträchtliche Verhaltensänderungen bewirken kann. Die Wirtschaftlichkeit solcher Programme kann zudem durch Segmentierung noch beträchtlich gesteigert werden, z.B. indem gezielt Haushalte mit überdurchschnittlichem Grundverbrauch anvisiert werden. Die Ergebnisse weisen auch darauf hin, dass die Spareffekte durch positive Mechanismen bedingt werden, nicht durch negativen psychologischen Druck. Insgesamt zeigen die Ergebnisse auf, dass verhaltensspezifisches Echtzeitfeedback höchst wirtschaftlich und skalierbar ist.

Résumé

Le domaine du feedback comportemental pour des individus au niveau de leur consommation énergétique a reçu une attention considérable venant de la part des chercheurs ainsi que des praticiens. Grâce à leur rentabilité, faisabilité politique et leur évolutivité, des programmes touchant des millions de ménages ont été mis en place. Des études récentes avec une durée entre 6 mois à 5 ans documentent une réduction énergétique allant de 1% à 6%. Bien qu'il a déjà été montré que le feedback est plus efficace lorsqu'il est donné sur un comportement spécifique au moment précis de son apparition, une démonstration de l'impact ainsi que de la rentabilité du feedback sous ces conditions favorables n'a pas encore été réalisée à grande échelle.

Ce rapport porte sur l'impact d'un appareil qui fournit du feedback en temps réel sur un comportement précis (dans ce cas-ci : consommation d'eau chaude dans la douche). Ce rapport évalue également les facteurs influençant l'efficacité lors d'un essai randomisé contrôlé comportant 697 ménages. Les participants ayant reçus de l'information sur leur consommation en temps réel ont réduit leur consommation d'eau et d'énergie de 23% chacune par rapport au groupe de référence. Les effets de l'intervention ont gardé la même intensité pendant toute la période de l'étude de deux mois. Projetés sur un an, l'appareil permettrait au foyer moyen d'économiser 443 kWh d'énergie, 8,500 litres d'eau potable, 94 kg de CO_2 et 110 CHF. L'appareil est ainsi rentabilisé au bout de 6-9 mois. Les individus avec une haute consommation de base montrent une réponse plus importante au feedback, tout comme les individus possédant une tendance générale à surveiller leur progression vers des objectifs. Quant aux attitudes écologiques, elles contribuent à l'effet en arrière plan; cependant, ces dernières n'ont pas d'incidence significative sur l'impact net de l'intervention puisqu'elles ont une forte corrèlation nègative avec la consommation de base. En outre, les résultats indiquent que l'intervention augmente substantiellement le savoir de l'individu au niveau de sa consommation de ressources. Par contre, les résultats ne supportent pas qu'une pression psychologique négative puisse entrainer les effets de l'intervention. Il convient de noter aussi que les données avant l'intervention indiquent une forte corrélation négative entre l'âge et la quantité d'énergie / eau utilisée à la douche. En effet les participants âgés entre 20 à 29 ans utilisent 72% de ressources en plus par douche que les participants âgés de 65 ans ou plus. Cependant, même si les jeunes personnes ont une consommation de base plus importante, elles montrent aussi une réaction plus importante au feedback, les rendant ainsi une cible précieuse pour ce genre de programmes.

L'étude montre que le feedback sur un comportement spécifique peut inciter un changement de comportement considérable. De plus, le profilage, comme par exemple viser des ménages avec une très haute consommation de base, peut d'avantage augmenter la rentabilité du programme. Les résultats suggèrent que ce sont des mécanismes positifs qui entrainent l'effet conservateur et non pas des pressions psychologiques négatives. Dans son ensemble l'étude montre que le feedback sur un comportement spécifique en temps réel est très rentable et évolutif.

Management Summary

Keywords: **Residential energy consumption, energy efficiency, water conservation, water heating, smart shower meter, real-time feedback, behavior change, randomized controlled trial, transformation of conventions.**

The promotion of energy efficiency is one of the central pillars of Switzerland's *Energy Strategy 2050*. Over the past years, behavioral feedback interventions have received much attention, with many pilot studies around the globe assessing their impact and cost effectiveness. Recent large-scale studies with in-home displays to monitor electricity consumption have yielded energy reductions in the range of 1-6%. While many reports state that feedback is particularly effective when it is provided in real-time, on specific behaviors or appliances, and right at the point of use, the large-scale implementation of more timely and detailed feedback in the utility domain has so far been dismissed with the argument of high costs.

This report evaluates the impact of individual real-time feedback on a specific behavior at the point of use: on hot water consumption in the shower. The study also assesses factors and mechanisms that might facilitate the effectiveness of such feedback interventions. In a randomized controlled trial with 697 Swiss households, participants received feedback on their resource consumption in the shower using a display in proximity to the shower head. Three dedicated study versions of the smart shower meter *amphiro a1* were used to assess the impact of the real-time feedback display against baseline use and against a control group. Altogether, nearly 47,000 showers were recorded over a 2-month study period and supplemented with survey data before and after the intervention to assess demographics, personality, attitudes, and users' experience with the device.

Participants who received real-time feedback on their consumption in the shower reduced both their energy and their water consumption by 23% compared to the control group. The effects were sustained throughout the study period. Although an extrapolation of the savings to longer periods is associated with uncertainty, recent studies back the persistence of saving effects over longer periods of time for feedback devices with data push. Under this assumption, projected to one year, this results in yearly energy savings of 443 kWh, 8,500 liters of drinking water, 94 kg of CO_2 and CHF 110 for the average household, making the device cost-effective within 6-9 months and at a cost of 0.041 CHF per kWh saved.

The study also seeks to understand which observable household characteristics and psychological mechanisms lead to the observed behavioral change. Individuals with a high baseline consumption show a stronger response to the intervention, so do participants with a general tendency to monitor their progress towards goals. Conversely, environmental attitudes do not significantly affect the overall treatment effect: Increased efforts by participants with stronger pro-environmental attitudes seem to be compensated by the fact that those individuals tend to start out from a lower baseline, making it harder for them to further reduce their consumption. The results also suggest that the treatment effect is not driven by peer pressure or appeals to adherence of social norms, but by the individual's tendency to quantify goals and monitor progress by making water and energy consumption salient during the act of showering. Notably, the baseline data indicate that the amount of energy and water used per shower is negatively correlated with age: The younger a age cohort, the higher its baseline consumption, with 20-29 year-olds using 72% more energy and water per shower than participants over 65. Yet given their higher baseline consumption, young people also respond stronger to the feedback.

These results have several major implications, also beyond the application domain of showering. The study demonstrates that real-time feedback on a single and very specific behavior can by far exceed the impact of feedback on a broader domain like household electricity usage: Projected to one year, the thermal energy saved by the shower meter outperforms the impact of electricity smart meter data displayed on in-home displays by a factor of 3 to 5.5[1]. The results indicate that the effects were driven by positive mechanisms: Many individuals simply seem to enjoy tracking metrics about their own life, in line with numerous recent reports on the rise of the Quantified Self movement. By contrast, the findings do not support any evidence for mechanisms that operate through negative psychological pressure. While environmental attitudes drive the effect behind the scenes, they do not significantly affect the overall net treatment effect, as they are strongly negatively correlated with baseline consumption. This means that the savings are not driven by a small subset of individuals with a particularly green mindset. Moreover, in the case of shower feedback, age can be used as a good proxy for the effectiveness of the device. That way, households with a particularly high savings potential can easily be identified with good accuracy based on the most basic demographics. The results further indicate that feedback is effective even for individuals who stated upfront that they had no or very limited intention to reduce their consumption with the device. Finally, on a broader perspective, the finding that the energy and water consumption *per shower* of younger people exceeds by far the consumption of elder persons adds to the literature that describes an ongoing transformation of socio-technical regimes and collective conventions. These behavioral shifts have been reported for several related domains like shower frequency, laundry quantities, the use of space heating and air conditioning, and with dramatic increases

[1]Compared to recent large-scale studies on electricity smart metering, which yielded annual electricity savings of 86 kWh Degen et al. (2013) resp. 154 kWh Schleich et al. (2013)

to the 3- or 5-fold within a few decades. While demand forecasts typically take into account different scenarios for technological progress and substitution, behavior is typically assumed to be stable, or at best determined by financial determinants (rising incomes and rebound effects). It might be the case, though, that progressively changing norms and conventions undermine the gains of improved energy efficiency. At the same time, given that young people (or, to be precise, high-consumers) seem to respond more strongly to the intervention, this study also shows that feedback technologies can help to close this intergenerational gap.

Chapter 1

Motivation

With the decision of the Swiss Federal Council and Parliament to phase out the use of nuclear energy, Switzerland's energy system has entered an era of massive transformations, requiring successive restructuring until 2050 (SFOE (2013b)). One of the central pillars of the long-term policy *"Energy Strategy 2050"* is the promotion of energy efficiency. The residential sector, accounting for 28% of Swiss energy end use (SFOE (2013b)), will clearly need to contribute to this transformation. Households have already been identified as a *"huge reservoir of potential for reducing carbon emissions and mitigating climate change that can be tapped much more quickly and directly"* than savings from carbon emissions trading, fuel economy standards, or changes on the energy supply side (Gardner and Stern (2008)). Aside from legal and financial measures (e.g., more stringent standards and target agreements), this also incorporates the need for behavioral change (Fetz (2013)). Over the past years, behavioral interventions have received growing attention for their quick scalability, political feasibility, and cost effectiveness. Yet as Allcott and Mullainathan (2010) stated in their Science magazine article, *"What has been missing is a concerted effort by researchers, policymakers, and businesses to do the 'engineering' work of translating behavioral science insights into scaled interventions, moving continuously from the laboratory to the field to practice. It appears that such an effort would have high economic returns."*.

One of the end-uses that is clearly underrepresented in the promotion of energy conservation measures is water heating. Water heating is the second-largest energy end use in households, accounting for 12-18% of residential energy consumption (Umweltbundesamt Deutschland (2013); eia (2013); Prognos AG (2013); BAFU (2013); BDEW (2010)). This equals the combined consumption of lighting (3%), refrigeration (3%), wet cleaning & drying (2%), cooking (4%), and entertainment / communication / IT (2%) altogether. In 2011, Swiss households used 32 PJ of energy for water heating (Prognos AG (2013)), or 2 500 kWh per home (figure 1.1).

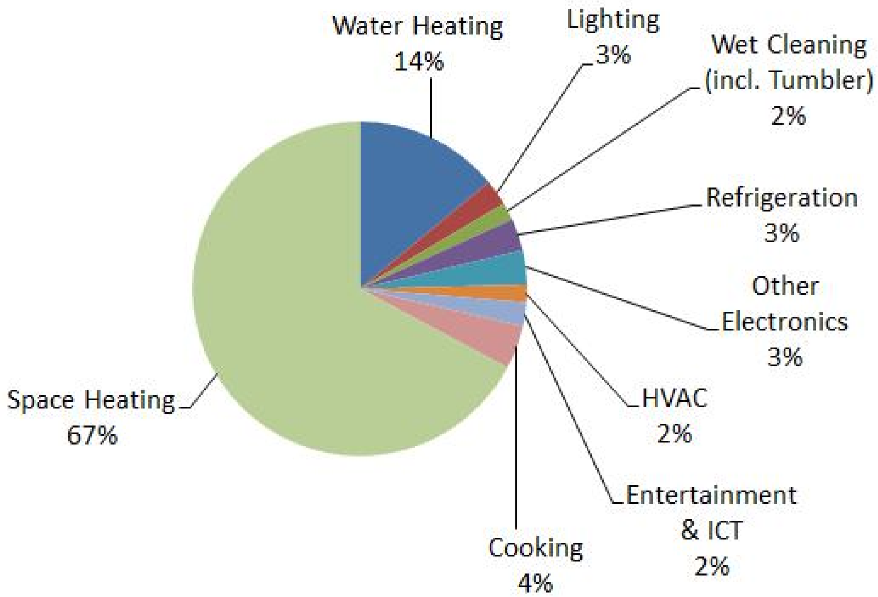


Figure 1.1: Energy end uses in Swiss households in 2011. Source: SFOE data, BAFU (2013)

With the progress of building technology and more and more stringent building codes and standards, water heating increasingly accounts for an ever larger share of energy end uses in residential buildings, amounting to 45% of the energy consumption in a typical Passive House (figure 1.2).

As the breakdown by energy sources in figure 1.3 shows, the majority of Swiss households rely on fossil fuels for water heating: 40% use oil, 25% electricity, and 21% natural gas. Renewable energy sources such as wood, solar thermal, or ambient air only account for a relatively small fraction (Prognos AG (2013)). This implies that the carbon footprint of water heating in Switzerland (per kWh of thermal energy) is nearly twice as high as for one kWh of electricity: Electricity in Switzerland is mainly produced with fossil-free resources - 56% hydro power, (SFOE (2013c)) and 39% nuclear power, (SFOE (2013a)) - resulting in a carbon intensity of 122 g/kWh at the plug level (Frischknecht et al. (2012)). By comparison, the more fossil-based energy mix for water heating results in a carbon intensity of 212 g/kWh (BAFU (2011)).

The diversity of fuels that is used for water heating might be one of the reasons why so far, water heating has been rather neglected for behavioral interventions compared to programs that target electricity consumption alone. It could also be the fact that as of today hot water consumption is barely quantified. One of the main barriers to this are quality and power supply problems in the deployment of electric metering devices in wet or humid environments: While batteries require periodic replacements, plumbers might refuse the installation of line-powered devices in close proximity to water or simply lack the required certification to do so. Water heating may also not have a more prominent position on the energy conservation

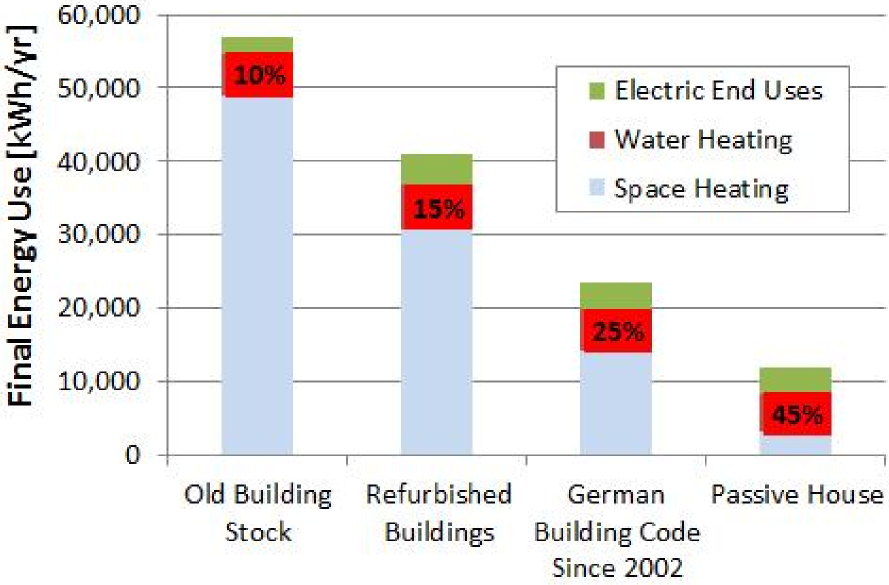


Figure 1.2: Residential end energy use by building type. Source: ASEW (2008)

agenda due to lobby efforts of water utility companies in Central Europe. Their infrastructure is designed for a higher demand (e.g., Fink (2012); Schorsch (2012)) and they argue that water conservation in Central Europe leads to congestion issues of the sewage system and higher water tariffs. Although this may be true for cold water, the conservation of hot water clearly makes sense due to the large amount of incorporated energy.

The majority of hot water is consumed in the shower, yet the general public is not aware of the energy dimension of showering. In comparison with in-home energy displays that visualize electricity consumption, showering particularly qualifies as an ideal domain for real-time feedback interventions for several reasons:

- **Control**: Users can easily influence their energy and water consumption in the shower
- **Immediacy**: Feedback can be provided in real-time and right at the point of consumption
- **Specificity**: Showering is a concrete and delimited behavior
- **Personal**: Showering is typically carried out by a single individual
- **Visibility and tangibility** of water, especially compared to electricity
- **Simplicity**: No abstract concepts like standby loads and metrics that are easy to understand (liters)
- **Fewer distractions**: No cell phones etc. in the shower

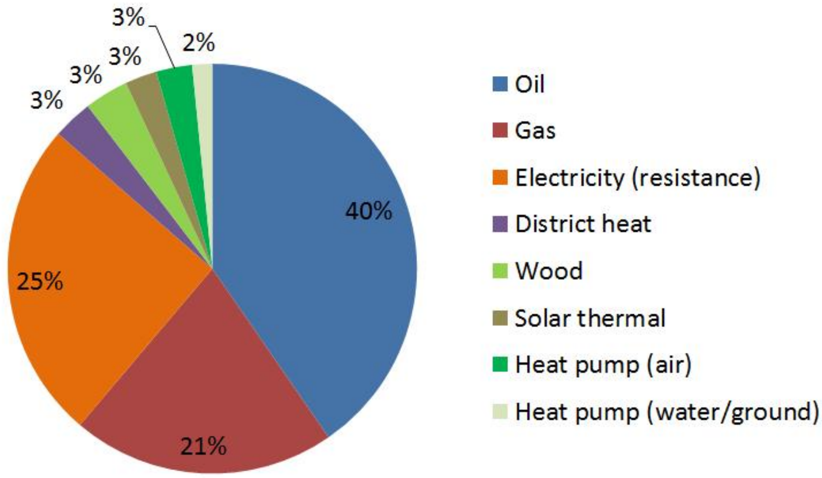


Figure 1.3: Hot water generation in Swiss households by fuel type. Source: Prognos AG (2013)

The combination of favorable conditions for feedback and the large amount of energy that water heating consumes has motivated this research project. In this context, the study investigates how visual feedback on energy and water consumption affects user behavior. Furthermore, the study seeks to identify factors that affect the effectiveness of such feedback technology.

The remaining report is structured as follows. Chapter 2 gives an overview of the related work and the research questions guiding this study. Chapter 3 explains the methodology and implementation of the study, including, e.g., the recruitment of participants, timeline, and research design. Chapter 5 contains the results of the study. Chapter 6 discusses the practical impact and relevance of these findings in a broader context. Finally, chapter 7 concludes with an outlook and a critical discussion of the limitations of this study.

Chapter 2

Related work and research questions

A vast body of literature has investigated the effectiveness of feedback on human behavior in a variety of domains. In the context of residential resource consumption, the main focus of large scale feedback trials has been on electricity consumption. In this domain, a solid number of systematic feedback studies have been carried out to determine the effectiveness of feedback interventions. Section 2.1 gives a brief overview on this closely related body of literature. While several studies have evaluated which factors generally affect residential water consumption (section 2.2), the use of feedback interventions and smart metering technology for water consumption is still in its infancy. Only recently, smart meter data and visual display technologies for consumption feedback have become available for this domain; yet most studies conducted so far still lack a clean research design and / or a meaningful sample size (sections 2.3 - 2.5).

2.1 Feedback trials on residential electricity consumption

Over the past decade, many studies have evaluated the effectiveness of behavioral feedback interventions on household electricity consumption. A growing body of literature indicates that feedback is more relevant when the link between resource consumption and specific appliances and activities is clear (Stewart et al. (2013)). Several studies reported that feedback is particularly effective when it is provided more frequently, in real-time, at a less aggregate level, ideally at the level of individual appliances (Houde et al. (2013); Ehrhardt Martinez et al. (2010)). Yet most reports state that these measures also tend to be more costly (Schleich et al. (2013); EPRI (2008); Fischer (2008)). In the residential context, so far, there are hardly any low-cost and cost-effective devices available that provide users with individual and immediate feedback on a particular action right at the point of consumption.

EPRI (2009) developed a classification of six feedback types, which was also adopted by the meta-analysis of 57 feedback studies by Ehrhardt Martinez et al. (2010). Figure 2.1 shows the classification scheme developed by EPRI. On the left side of this spectrum, they distinguish four types of indirect feedback (ordered by their typical level of information content and implementation cost): standard billing, enhanced billing (e.g., *Opower* or *Ben Energy* home energy reports), estimated feedback (e.g., web-based appliance disaggregation) and daily / weekly feedback (e.g., self-meter reading). In the upper range in terms of information content and cost, they distinguish two categories of direct feedback: real-time feedback (e.g., in-home displays) and "real-time plus" feedback (e.g., appliance disaggregation and / or control). Ehrhardt Martinez et al. (2010) report median electricity savings of 3.8% for enhanced billing, of 6.8% for estimated feedback, of 8.4% for daily / weekly feedback, of 9.2% for real-time feedback, and of 12% for "real-time plus" feedback. These numbers, however, do not take into account the sample size of the 57 studies under review, nor their recruitment method or the existence of a control group. Thus, many of these studies lack a proper research design (see below).

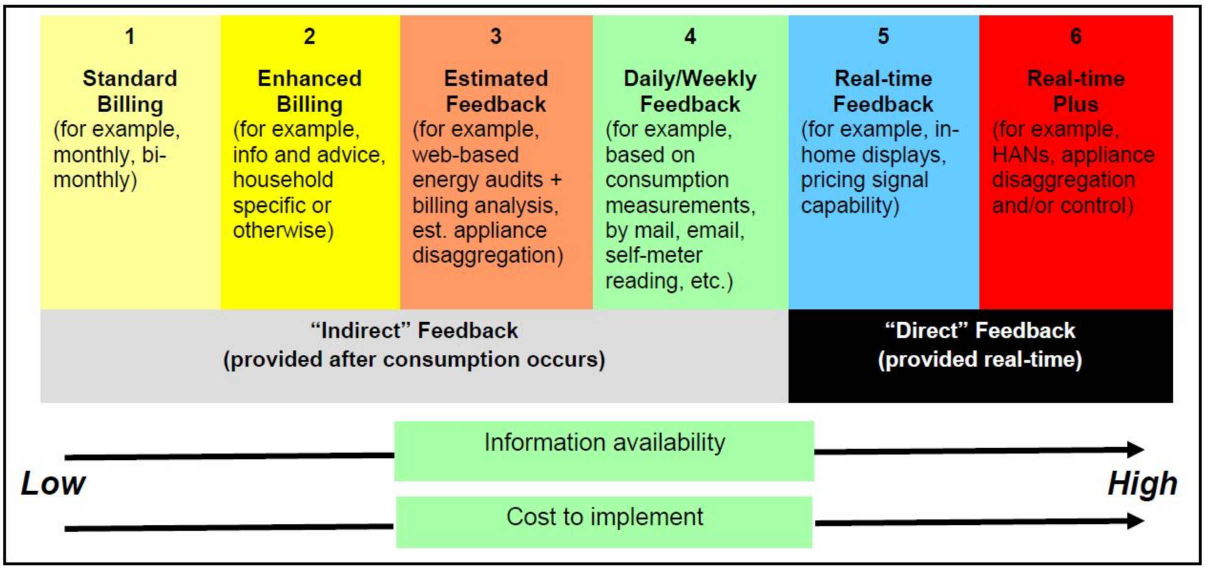


Figure 2.1: Classification of feedback types developed by EPRI. Source: EPRI (2009)

Smart metering pilots typically use in-home displays or web-portals to visualize aggregated data on gas or electricity consumption of the entire household. Faruqui et al. (2010) presents a meta-study on in-home displays and McKerracher and Torriti (2013) gives an overview on 27 completed real-time feedback trials. Overall, existing feedback studies show a large variety with respect to data collected and presented to users, frequency of feedback, content, level of aggregation / breakdown, medium of presentation, inclusion of comparisons, and combination with additional information and other instruments (Fischer (2008)). While earlier reviews reported savings in the range from 5 to 15% (Darby (2006); Ehrhardt Martinez et al. (2010)), more recent large-scale trials measuring savings against control groups find more modest effects for in-home displays in the order of 1 to 6% (Darby (2012); Schleich et al. (2013); Degen et al. (2013); Carabias-Huetter (2013)). A recent meta-study by McKerracher and Torriti (2013) analyzes 27 completed and 7 upcoming in-home displays trials with respect to their treatment effect, sample size, recruitment method, year, peer-reviewed publication and type of in-home displays used. They find that earlier studies were characterized by smaller sample sizes, unrepresentative samples, a higher involvement by the study administrators, more prone than more recent studies to the Hawthorne effect (i.e., the feeling of being observed or merely participating in an experiment may affect individuals' behavior; see Schwartz et al. (2013) for a recent study on that topic in the electricity sector). While earlier metastudies had indicated an expected a higher range of conservation effects, McKerracher and Torriti (2013) conclude that *"3-5% is a more accurate expected conservation figure for a large-scale roll-out of in-home displays"*. To evaluate the effect of recruitment method and sample size, McKerracher and Torriti (2013) classify existing trials into four categories and weigh treatment effects by sample size. While the weighted mean conservation effect for studies with a representative sample ("class A") is 2.6%, studies with opt-in design, 100 or more participants, and a low degree of involvement by the administrators ("class B") typically yield 4.5%; for studies with opt-in design, less than 100 households, and a high degree of involvement by the administrators, the weighted mean treatment effect is 8.2% ("class C"). The forth category contains studies lacking information on sample selection or recruitment methodology.

One of the key questions in the context of the savings achieved is their persistence. Ehrhardt Martinez et al. (2010) report that the vast majority of the savings can be attributed to behavior change, not to the adoption of new, energy-efficient technologies. This implies that the persistence of the reduction depends on the persistence of the change in everyday practices. In their study on water end use feedback, Fielding et al. (2013) found that once the intervention had ended, the effect eventually dissipated and households returned to pre-intervention consumption levels. By contrast, Ayres et al. (2009) reports sustained savings throughout the 7 resp. 12 month study duration. Raw and Ross (2011) also report persistent effects for electricity smart meters to the end of the trial. In their review of various pilots in the U.S., U.K. and Ireland, Foster and Mazur Stommen (2012) report that all but one study that tested for effect persistence showed savings to persist over the course of the pilot. The most comprehensive analysis on long-term effects available so far was undertaken by Allcott and Rogers (2014). They analyze the electricity consumption data of 234,000 households over four to five years and find that the savings are much more persistent than previously generally assumed. They observe an immediate response to the first home energy report, followed by a relatively quick decay. However, they identify cyclical but diminishing patterns of action and backsliding as response to subsequent reports. Overall, the effects become gradually more persistent as the intervention continues. The study also finds that if the intervention ends after two years, the effects are still relatively persistent, with a decay rate of 10-20% per year. The authors conclude that the cost-effectiveness of these programs has been dramatically underestimated in the past. In this context, Boyd (2014) debates the topic of data "push versus pull", which had also been brought up by other authors (Foster and Mazur Stommen (2012); Froehlich et al. (2010)). They argue that currently, most online portals and energy monitors require an additional layer of user interaction to access the feedback information (data pull) and conjecture that the future of real-time feedback lies in systems with data push.

2.2 Factors influencing (shower) water consumption

In the domain of water consumption, several studies have investigated to what extent socio-demographic factors, efficiency of household water stock, and psycho-social profiles affect water use in the household or, more specifically, in the shower. Yet the vast majority of these reports were based on simple correlations, without carrying out any intervention. Many of these studies solely rely on self-reported behaviors and estimates. Jorgensen et al. (2009) propose a water consumption behavioral model that includes issues of inter-personal trust (social norms) and institutional trust with respect to the water utility company. Gilg and Barr (2006) examine social, attitudinal, and behavioral aspects and links between water saving, energy conservation, green consumerism, and waste management with survey data from 1.265 UK households, but do not assess the data against measured consumption information. Randolph and Troy (2008) also use survey data alone, in that case from 2,179 Australian households, to examine attitudes towards conservation and water consumption. Corral-Verdugo et al. (2003) compare survey data with water meter readings. They found that utilitarian water beliefs promoted water consumption, while ecological water beliefs inhibited that behavior. De Oliver (1999) contrast attitudes reported in surveys with municipal water consumption data in Texas. The results *"reveal substantial disparities between survey responses and manifested actions. They also indicated that well-accepted patterns of conservation response ascribed to various demographic segments in the survey format need to be more precisely qualified before equating them to performance."* Beal et al. (2013) compare water usage data with self-reported water consumption and attitudes towards water conservation of a total of 252 households with high-resolution end use data. They find that self-identified high users consistently use *less* water than self-nominated medium and low water users and analyze differences in the socio-demographic and psycho-social profiles of these groups.

Carragher et al. (2012) investigate the degree of influence of household water stock efficiency (e.g., shower heads) in 191 households. They report that households with higher household water stock efficiency consume 25% less water during peak hour periods. Given the lack of randomization, however, these results may be confounded with other factors like higher problem awareness or a stronger engagement of participants who have installed more efficient equipment. Willis et al. (2013) assess the impact of socio-demographic factors and efficient devices on end use water consumption in 151 Australian households. They find that households in lower socio-economics groups tend to use more water for showering than households in higher socio-economic groups and that households with high efficiency shower heads use 48% less water per person (no intervention, no randomization). Mayer et al. (1999) analyze water end uses in 1,188 single-family homes in twelve North American locations. Among other findings, they report that children and teenagers used incrementally more water for showers than adults. Makki et al. (2013) analyze determinants of shower water consumption and reporte that households with teenagers are consuming considerably more water; yet they regress each and every factor (e.g., number of males, number of females, number of children, number of teenagers) separately, without trying to disentangle the actual contribution of the individual factors in a holistic approach. This is just a small and rather arbitrary extract from a wide collection of papers to illustrate that the existing body of literature provides very mixed results on the influence of demographics and other contextual factors on residential resource consumption and households' response to resource conservation programs. Most of these factors are highly correlated, yet in many cases, the factors are analyzed one by one, with the result that the findings are not very conclusive.

Sociologists, on the other hand, try to understand the rise of shower frequency and resource consumption over the past decades in a broader perspective. Several papers embed the topic into a larger context, but with a more qualitative than a quantitative lens. Shove (2003) describes the socio-technical transformation of inconspicuous behaviors in the domains of comfort, cleanliness, and convenience ("the three C's"). She investigates cultural and generational shifts of expectations and practice, in the domains that encompass *"the environmental hot spots of consumption."* Her book describes and analyzes the ongoing *"creep of convention and escalation of standardization of conditions and circumstances that people take for granted"* which *"challenge the established theories of consumption and technology"*. Southerton et al. (2004) analyzed the rising frequency and flow rate of showering in the UK, forecasting a five-fold increase of consumption between 1991 and 2021. Herrington (1996) investigates water demand trends, followed by long-term domestic demand forecasts. The book by Butler and Memon (2006) include technical, social, and legal aspects in order to draw a comprehensive picture of demand management topics, including rising standards of living and changes in lifestyle.

2.3 Interventions other than feedback to reduce water consumption

Consumption feedback is only one option under the umbrella of water demand-side management stategies that seek to influence individuals' water use activities through a range of social marketing, economic, and other conservation programs (Carragher et al. (2012)). Measures include both non-financial (e.g., water-efficient technology) and financial (e.g., incentives, pricing) approaches, as well as mandatory (e.g., regulations) and optional (i.e., market system) instruments. In their review on water demand-side management programs, Inman and Jeffrey (2006) conclude that such programs can be expected to reduce water consumption by 10 to 20% over a 10 to 20 year period; however, they find a higher elasticity for outdoor water use than indoor use. In their literature review on environmental behaviors, Kurz et al. (2005) only identify five studies related to water conservation (out of 87 reviewed); in another literature and research review on water demand-side management, Hurlimann et al. (2009) state the need for research on interventions that positively influence water-related behavior. A more recent review by Dolnicar et al. (2012) only identifies five studies with actual measures of water use. Based on survey data from 430 utilities, Nieswiadomy (1992) analyzes the effects of price structure, conservation, and education. The report states that regions that have previously experienced water shortages feature higher price elasticities; in contrast to service territories that have not been affected by water shortages, households in these regions are even responsive to interventions that are centered around public education. Lee et al. (2011) analyze the impact of water conservation incentives, mainly rebates and unit exchange programs for shower heads and other equipment. They find a 6 - 14% reduction in the first and second years; yet these figures are based on a comparison against the previous year, not against an actual control group. Based on monthly data from 19,000 households, Campbell et al. (2004) investigate different policy instruments for water conservation. They find that pricing and appropriate regulation can be effective, but warn that offsetting behavior can negate engineering solutions to policy problems. On the other hand, their results indicate that adding communication to engineering solutions can overcome such offsetting issues. Fielding et al. (2013) list several drawbacks for pricing mechanisms and mandatory approaches like equity issues involved with their implementation, limits to price elasticity, evidence that they do not necessarily result in long-term change, the political will required for their implementation, and resistance by the public.

2.4 Technology-based feedback on shower behavior and water consumption

More recently, a number of papers with a more practical-technical orientation describe applied feedback devices aiming to influence actual user behavior in the shower or at the tap with self-developed prototypes of feedback devices: Arroyo et al. (2005) present "Waterbot", a system to inform and transform water consumption behavior at the sink. Kappel and Grechenig (2009) introduce an ambient shower display named "show-me"; Kuznetsov and Paulos (2010) a system named "UpStream", a pervasive display for showers and sinks. Laschke et al. (2011)

introduce "Shower Calendar", another pervasive concept study to motivate reduced resource consumption in the shower. Froes Lima and Portillo Navas (2012) describe a system for remote metering of water and electricity consumption, yet only report some preliminary results of two prototype deployment sites, without stating overall savings of the participating households. Despite a plethora of innovative concepts to visualize feedback on shower behavior, all these studies share the limitation of a very small number of participants and a lack of verifiable research hypotheses. A more recent Australian study (Willis et al. (2010); Stewart et al. (2013)) includes a larger number of households (N=151) to quantify baseline water consumption and to evaluate the effect of a shower feedback device (N=44). Households who participate in the second part of the study with the alarming visual display device are reported to reduce their consumption by 27%. However, the subset of households that underwent the treatment with the feedback devices was self-selected and not chosen by random assignment. From a research design perspective, this selection bias violates the internal validity of the second part of the study that evaluates the effectiveness of the intervention.

2.5 Randomized controlled trials on water use feedback

Very recently, two larger studies investigated the impact of feedback on water consumption in the home or, more specifically, in the shower. Both Ferraro and Price (2013) and Fielding et al. (2013) provided mail-based feedback to households in a randomized controlled trial. Ferraro and Price (2013) report the outcome of a large-scale mail-based residential customer conservation education program. They conducted a randomized experimental design with 100,000 households to assess the effectiveness of social comparison messages against simple prosocial messages or technical information alone. They find that the effect of the strong social norm message was a) much higher than technical advice, b) that it was twice as high as in similar programs on electricity conservation (Ayres et al. (2009); Allcott (2011)), and c) that it was much more effective on high users. Fielding et al. (2013) deployed water smart meters in 221 Australian households, which they randomly assigned into one of four conditions: control group, information only group, descriptive norm group, and water end use feedback. During the mail-based intervention (four group-specific postcards), the three treatment groups reduced their consumption between 7-13% relative to the control group. Once the intervention had ended, the effect eventually dissipated and households returned to pre-intervention consumption levels.

As a bottom line, while many studies have investigated water consumption behaviors, the majority of studies that evaluated how different factors affect water consumption did not collect any measurement data, but solely relied on self-reported data from surveys. On the other hand, studies that did incorporate measured consumption data either did not investigate the effectiveness of a specific intervention to change user behavior, or did not use a clean research design with a decent sample size that would allow to attribute savings to a specific intervention or factor. To our knowledge, the present study is the first one that determines the impact of real-time feedback on water consumption in a randomized control trial; even regardless of the study design, it is also by far the largest study that assesses the impact of real-time feedback at least the context of water, if not in general for the environmental impact of a specific household routine.

2.6 Research questions

The goal of this empirical study was to evaluate the effects of disclosing information on individual resource consumption on subsequent resource usage. The following research questions framed this project:

1. How does feedback information affect self-assessment of consumption (learning) and energy / water conservation?
2. To what extent can comparisons and (implicit) competitions between household members increase saving effects?
3. To what extent does initially stated willingness to conserve energy explain the overall saving effects induced by a consumption feedback device?
4. Which demographic factors and personality traits do significantly influence baseline consumption and to what extent?
5. Which demographic factors and personality traits do significantly influence participants' reaction to the feedback and to what extent?

Question 4 and 5 were not part of the initial proposal; instead, we had intended to investigate whether results of the *ewz Studie Smart Metering* might affect or at least be correlated with the findings of the *ewz-Amphiro study*: a) Could participants' responsiveness to consumption feedback on electricity usage help to predict his or her responsiveness to consumption feedback in the shower? b) Would the ability to estimate one's energy demand in the course of the smart metering pilot for electricity translate into a more accurate self-evaluation with respect to personal hot water consumption? And c) Would improvements / positive feedback on electricity consumption have a positive or negative effect on subsequent hot water usage?

However, electricity smart meters were only deployed in one fifth of the households that had participated in the *ewz Studie Smart Metering*, and only a fraction of those households participated in the *ewz-Amphiro study*. Therefore, the sub-sample of households from which both shower data and electricity smart meter data were available is very small. As a consequence, the dataset to answer these sub-questions would not have been sufficient for meaningful conclusions. Instead, we have identified two other and in our opinion more relevant research questions (questions # 4 and 5). By supplementing the data collected for the *ewz-Amphiro study* with the previously collected data of the participating households from the *ewz Studie Smart Metering* (with the consent of the participants), this study combines a large and rare set of measured field data - to our knowledge the world's largest dataset on shower behavior / consumption - with a vast array of data on personality, attitudes and demographics. As a result, the study is not only the first one that evaluates the impact and cost-effectiveness of behavior-specific real-time feedback in the field; moreover, the combined dataset grants a unique opportunity to answer questions 4 and 5, which in our eyes are at least as relevant for researchers and policymakers.

Chapter 3

Methodology and implementation

This chapter gives a brief overview on the study and presents information on the parties involved, timeline, participants, study device, research design, and implementation.

3.1 Overview of the study

This chapter describes the implementation and data collection procedure of the study. Altogether, 697 participating households were recruited among a larger sample of 5,000 ewz-customers that had previously completed the *ewz Studie Smart Metering*. After an initial survey, households were randomly assigned into three different experimental conditions, each of which received a different version of the smart shower meter *amphiro a1*. After a short baseline period, the three device versions showed different feedback information while showering. The smart shower meters stored data of every shower taken throughout the two-month study period. At the end of the study, participants were asked to ship their device back for the data readout and to fill out the final survey.
Figure 3.1 shows a snapshot of the study device display and the device in its position between the shower head and the shower hose; a more detailed device description follows in section 3.5.

3.2 Parties involved

The *ewz-Amphiro-study* was carried out under the lead of researchers of ETH Zurich (Department of Management, Technology, and Economics (D-MTEC)) in close collaboration with *ewz*, researchers from the University of Lausanne (Faculty of Business and Economics (HEC)) and the ETH Zurich spin-off company *Amphiro AG*. The Swiss Federal Office of Energy supported the research activities of this study, while *ewz* funded the study devices.

Figure 3.1: On the left: *amphiro a1* display and body; on the right: the device installed between showerhead and showerhose

3.3 Timeline of the study

The field deployment phase of this study lasted from early December 2012 to early February 2013 (two months). In the preceding months, the researcher team members at ETH Zurich adapted the user manual and the website for each feedback condition (see section 3.6) and reconfigured the study devices. In July and September 2013, two staff training sessions were organized for the employees of the ewz customer support center, who supported the research team from ETH Zurich by addressing study participants' questions and issues.
The recruitment of participants took place as the *ewz Studie Smart Metering* (on electricity consumption) phased out, in staggered (bi-)weekly batches between September and November 2012 (see section 3.4). All devices were shipped on November 29/30. The packages contained a return envelope with prepaid postage and shipping address for the readout at the end of the study.

After a two-month field deployment phase, all study participants received an email asking them to return their shower meter and to fill out a final survey (approx. 20 minutes). Participants who had not shipped back their device or not completed the survey received one or two additional reminders in the course of the following weeks. Data readout, device reconfiguration, and reshipping procedure were completed in April 2013. Thereafter, the individual datasets were merged, anonymized, and analyzed.

3.4 Participants

Participants of this study were recruited among the 5,000 participants of the *ewz Studie Smart Metering* (see report to the Swiss Federal Office of Energy Degen et al. (2013)). They all received the smart shower meter *amphiro a1* as a thank-you gift and were informed about the possibility of voluntarily participating in another study with that device. In order to opt into the study, they had to fill out a short survey.

Based on that survey, 697 households were selected. Households with more than two members could not be admitted: As the study devices could only store the data of 202 showers, not all shower data might have been recorded in larger households. Ideally, an equal number of single- and two-person households was pursued. However, due to an under-representation of single-person households among the pool of participants (and, as a consequence, survey respondents) compared to the number of two-person households, the number of participating single-person households ended up being slightly smaller (324 single- vs. 373 two-person households).

Apart from household size, the criteria for admission were:

- No anticipated relocation during the study period
- Handheld shower head: As the device was designed for handheld shower heads, it could not be installed in wall-mounted showers or body sprays
- Approval of the conditions of the data privacy protection statement
- Stated willingness to ship the device back after two months (for the data readout)

Eligible households were randomly assigned into one of the three feedback conditions (separate assignment process for single- and two-person-households). Households who did not qualify for the study received an invitation to instead participate in a (separate) long-term study.

Regarding recruitment strategy, the approach adopted follows an opt-in approach, which by its nature could potentially be subject to self-selection bias and external validity issues. As outlined in 2.1, the meta-analysis by McKerracher and Torriti (2013) distinguishes between four types of studies, based on their sample size, recruitment method, level of interaction between study administration and participants, year, peer-reviewed publication and type of

IHD used. Based on these criteria, the present study falls into "class B" design (more than 100 households, opt-in, low degree of interaction between participants and study administration); the weighted mean treatment effect for class B studies is 4.5%. We will address the question of potential self-selection bias more in detail in chapter 7.

3.5 Study device

The study was carried out with the smart shower meter *amphiro a1*. The device measures and stores time series data on shower behavior and provides real-time in-situ feedback on resource consumption in the shower. Users can easily install the device between the shower hose and the handheld showerhead in less than a minute and without any tools. The device is energy-autarkic: A built-in micro-generator harvests energy from the water flow, supplying the device with the power required for its processing unit and display. This self-powering concept eliminates the need for a battery and allows tracking user behavior in response to the feedback provided over extended periods of time. Showers with short interruptions up to three minutes (e.g., for lathering up) are stored as one coherent shower; water extractions below five liters are not considered as showers and consequently not stored. The underlying assumption is that most of these occurrences serve other purposes like flower watering or bathtub cleaning. During each shower, the device continuously measures water temperature and generator speed. Based on these data, water consumption, energy consumption, and energy efficiency class of the current shower are permanently calculated. The standard device displays real-time feedback on the current water and energy consumption, water temperature and the current energy efficiency class; the latter is visualized by a letter ranging from A to G and a polar bear animation. Standard devices can store the data of up to 507 showers.

The memory allocation and the display content of the device were modified for the purpose of this study. Section 3.6 describes the memory allocation more in detail. Devices were reconfigured into three different study feedback condition modes, which are explained in the following section. A detailed description of the technical aspects of the reconfiguration process is available in Tiefenbeck et al. (2013). Compared to the standard device, the information displayed by the study devices was modified in several ways:

- Baseline phase: For the first ten showers, all study devices only displayed water temperature, see section 3.6
- Online code: While standard devices display an online code which enables users to access additional information on their shower behavior at the *Amphiro* user portal, this feature was disabled on the study devices to avoid an information bias through the portal.

- Shower data: The display content was modified depending on the feedback condition, as described in section 3.6.

3.6 Research design

In order to answer the research questions outlined in section 2.6, a 2 (household size) x 3 (display content) randomized controlled trial was carried out over two months. Households were randomly assigned to one of the three feedback content conditions, each condition being subdivided into single-person households and two-person households. To determine households' consumption without providing feedback information ("baseline phase"), all devices only displayed water temperature for the first ten showers. In the user manual of the treatment groups, this was described as "initial acclimatization phase of the device", without specifying how much time or how many showers needed to be taken in order to complete this phase. This was based on a lesson learned from the pilot study. First of all, if users were aware of the purpose and specifics of the baseline phase, they might unnaturally alter their behavior; second, curiosity might induce them to skip / short-circuit the baseline phase with a couple of manual water extractions.

Upon completion of the baseline period, the three feedback conditions were exposed to different display contents: The control group continued to see only the current water temperature throughout the study. Control group participants were informed that different display contents were evaluated, allowing participants to test different display versions during and after the study. The devices of the *real-time shower information* condition displayed the same information on the water and energy consumption of the current shower as the standard devices. Households in the *real-time plus previous shower information* condition were additionally exposed to the water volume consumed in the previous shower.

Each of the three feedback conditions was subdivided into single- and two-person households. The goal was a) to break down the anticipated effects into components that are self-internalized by the individual alone and b) into effects that can be attributed to social dynamics between household members, e.g., effects of competition and social norms.

3.7 Logistics and implementation

In order to be able to match the different datasets of each participating household (see section 4.2), survey respondents had to provide their *ewz* customer ID at the beginning of the entry survey. This number was sent to *ewz* together with the randomly assigned feedback condition to create shipping lists for each of the three study groups. The non-profit organization *Drahtzug* packaged and shipped the devices to the participants, each parcel containing a prepaid reply envelope.

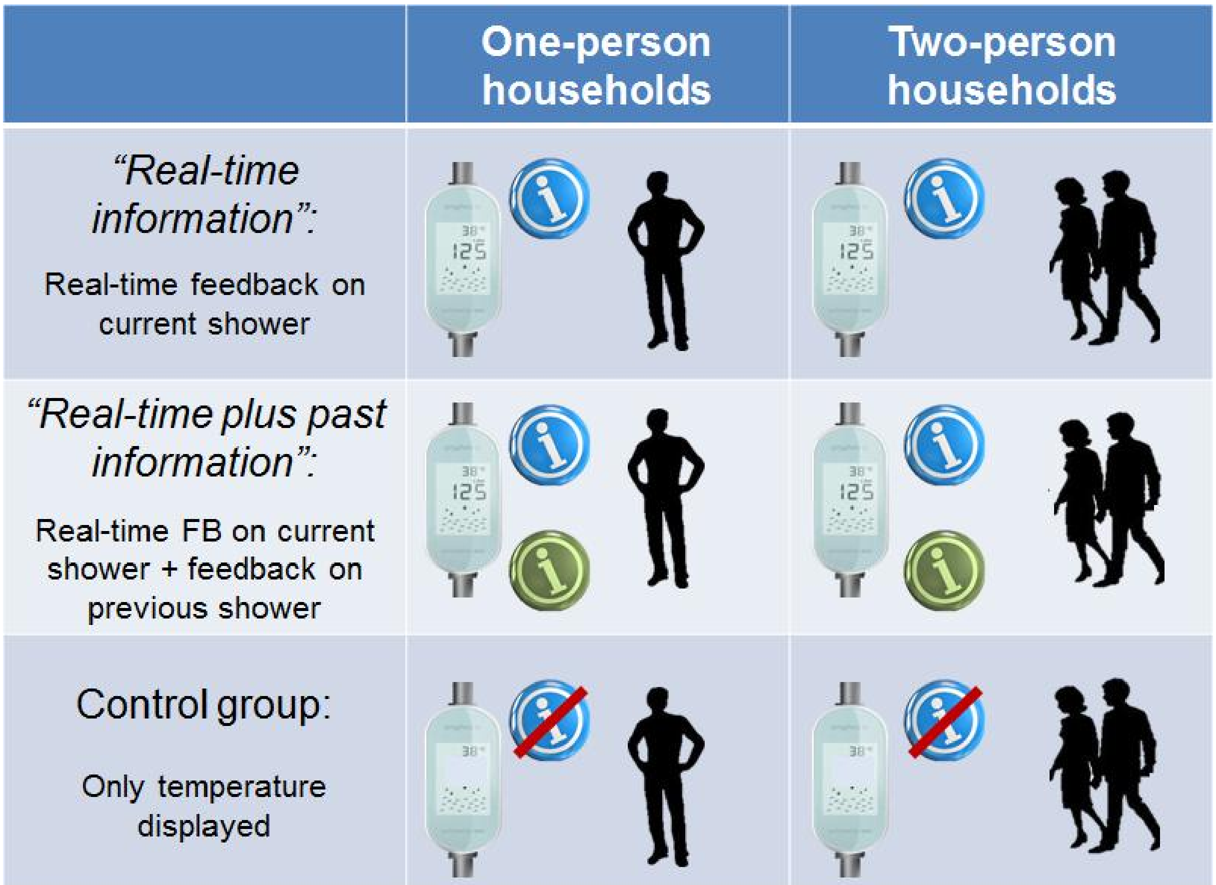


Figure 3.2: 3x2 research design

After the field deployment phase, participants sent back their devices for the data readout. Each device was read out individually (for details see Tiefenbeck et al. (2013)), which took approximately five minutes per device. During that process, the serial number of the device was scanned and linked to the corresponding household's study ID (see section 4.3). The read-out process involved the following tasks:

- Visual data read-out (roughly five minutes per device)
- Data sanity and consistency check
- Linking the shower dataset with the corresponding survey ID
- Verifying whether device was fully functional (based on survey and read-out data)
- Verifying whether the participant wished to receive the device back
- Resetting the memory and configuring the device to standard operation mode
- Cleaning, checking for completeness of the set and adding missing small parts (o-ring seals, sieve)
- Repackaging for reshipping (new envelope with the correct address label)

Overall, great attention was paid to the research design and its implementation to ensure the internal validity of the study, participants' privacy protection, and correct data matching.

Chapter 4

Data collection and analysis

This chapter describes the dataset: Granular measurement data were collected on individual showers and combined with survey information (including survey data from the *ewz Studie Smart Metering*). The chapter further outlines how participants' privacy was protected, describes participation and response rate, data filtering, weather during the study period, and software used.

4.1 Shower data

In addition to the data which the standard devices record by default of every shower taken (water consumption and average temperature), study devices additionally stored the duration of each shower as well as the duration and number of interruptions during each the shower. This reduces the maximum number of storable showers from 507 to 202. More technical details on measurement, data storage, and data read-out can be found in the publication Tiefenbeck et al. (2013). Energy consumption is calculated under the assumption that no energy losses occur[2].

One modification that was made to the dataset was the exclusion of shower #1 from each household: While all subsequent showers had similar frequency distributions for temperature, volume, and flow rate, shower #1 deviated considerably from the typical distribution patterns: An unusually high number of households had only extracted between 5 and 10 liters and at lower temperatures than usually. We assume that a large fraction of participants simply turned on the water once they had completed the installation of the device to check its functionality and display content, without taking an actual shower. Therefore, baseline data are calculated based on the data of shower #2 through shower #10 for all households.

[2]Section 6.1 contains a calculation of the energy consumption per shower, including an assessment of energy savings which takes energy losses into account.

4.2 Survey data

4.2.1 Surveys conducted for the *ewz-Amphiro Study*

Both surveys (before and after the study) were edited and carried out with the online software tool *surveygizmo*. The initial survey contained basic questions on demographics and several questions to check the participation criteria outlined in section 3.4. Two questions assessed whether the household pays for water and heat energy based on its consumption, or whether they pay a fixed rate or rent that is independent of their water and heat energy consumption. The survey also included several questions on attitudes towards water and energy consumption in the household in general and in the shower in particular. Furthermore, participants were asked to estimate some figures on their personal shower behavior (water volume per shower, water temperature, duration), also relative to other study participants.

The final survey covered extended periods of absence by household members and the use of the shower by guests (to check whether the classification of a household as single-person- or two-person-user-household was valid). Otherwise, the survey mainly focused on participants' experience with the smart shower meter; readability and comprehensibility of information content of the display elements; another self-estimate of their shower consumption and behavior; questions on discussions and comparisons within the household (for two-person households); usability; goal-setting and perceived behavior change; their intent to continue using the device, as well as the likelihood of them recommending the device.

It should be noted that only one person per household filled out the survey. For simple demographics (e.g., number of household members), this is irrelevant. As far as attitudes, environmental attitudes, etc., are concerned, the data are based on the survey respondent's answers. However, 92% of the participating two-person households are couples. A large body of literature shows that partners typically show a very high concordance in the realm of social and political attitudes (Alford et al. (2011)). Thus one partner's attitudes can serve as a decent proxy for both partner's attitudes.

4.2.2 Survey data from the *ewz Studie Smart Metering*

In the entry survey, respondents were asked whether they approved that their data from the *ewz Studie Smart Metering* (on electricity consumption, conducted by the University of Lausanne) were also provided to the research team at ETH Zurich, so that these data could also be included in the evaluation of the *ewz-Amphiro-study*. Most relevant to this study, that survey contains extensive data on participants' personality traits and attitudes.

4.3 Privacy protection

The survey at the beginning of the study contained a link to the privacy protection statement; its main points were additionally summarized next to check boxes at the begin of the survey. In order to participate in the study, survey respondents had to check these boxes and to accept the privacy protection statements. The document contained the scope of the data collection, information data storage, data processing, and data deletion. The survey also explained that the data of individual households would not be published and that the data analysis would be carried out with pseudonomized study IDs. In contrast to the *ewz* customer ID, this number cannot be traced back to the individual household.

4.4 Participation and response rate

Overall, 5,919 households were invited to participate in the study. 1,348 filled out the entry survey (a response rate of 23%), 697 of whom were selected for the study. The main exclusion criterion was household size - only 31% of the survey respondents lived in single-person households (the initial goal was to have 50% single-person households in the study).
Altogether, 685 households shipped their device back at the end of the study (98.3%), 636 of them were read out successfully. The entire raw shower dataset contained 46,835 showers. Nearly as many participants (666 households, 95.5%) filled out the final survey. Overall, a complete dataset (entry survey, final survey, shower data, and data from the *ewz Studie Smart Metering*) of 626 households (90%) is available.

4.5 Filtering

Several measures were carried out to ensure data quality and to verify whether the results were not driven by measurement errors or extreme outliers. First of all, 22 devices were discarded as they had experienced water damage; their memories could either not be read out or contained obviously flawed data (e.g., flow rates of 7,000 liters per minute). For those devices that could still be read out, the incidence of water damage was relatively easy to detect: The datasets in question contained perfectly reasonable measurements up to a certain point, then all of a sudden switched to completely unrealistic data. Thanks to this binary state (working properly vs. damaged), readings from defective devices could be discarded easily.
Second, several survey responses indicated inconsistencies in the number of shower users, for instance frequently visiting guests, move-in or move-out of a household member, or one household member being away over extended periods of time. In several two-person households, the shower where the *amphiro a1* was installed was used only by one of the household members (separate bathrooms). Conversely, many one-person households in particular re-

ported frequent visits by partners, friends, or family. As a consequence, shower users changed over time and the assumption of a single person not interacting with other household members was violated. Altogether, 102 households with such inconsistencies were flagged. While we included them in the overall assessment of the treatment effect, we excluded them from the following analyses of the psychological mechanisms due to their unstable composition.

Third, we analyzed the influence of outliers on the results. For that purpose, we calculated the average and the standard deviation both of water temperature and of water volume used per shower for every household. All data entries that deviated from a household's average value by more than two standard deviations were flagged as potential outliers. Analyses were carried out with and without these flagged entries, but the results were hardly affected by this filter and the results were robust to the removal of such outliers. Most outliers can probably be explained by the fact that water is also extracted through the shower head for other purposes than "normal" showering, e.g., to water flowers, to rinse the bathtub, to clean the bathroom, for exceptionally cold showers after exercising, for bathing (if the bathtub is filled through the shower hose), etc.

4.6 Weather

As seasonal fluctuations might also affect shower behavior (e.g., cold outdoor temperature might increase shower duration), we collected weather data from Zurich for the study period from the publicly available website www.freemeteo.com. As figure 4.1 shows, outside temperatures remained relatively stable throughout the study period; there was no particular trend upwards or downwards that might explain a drift towards higher or lower water consumption or temperature over time.

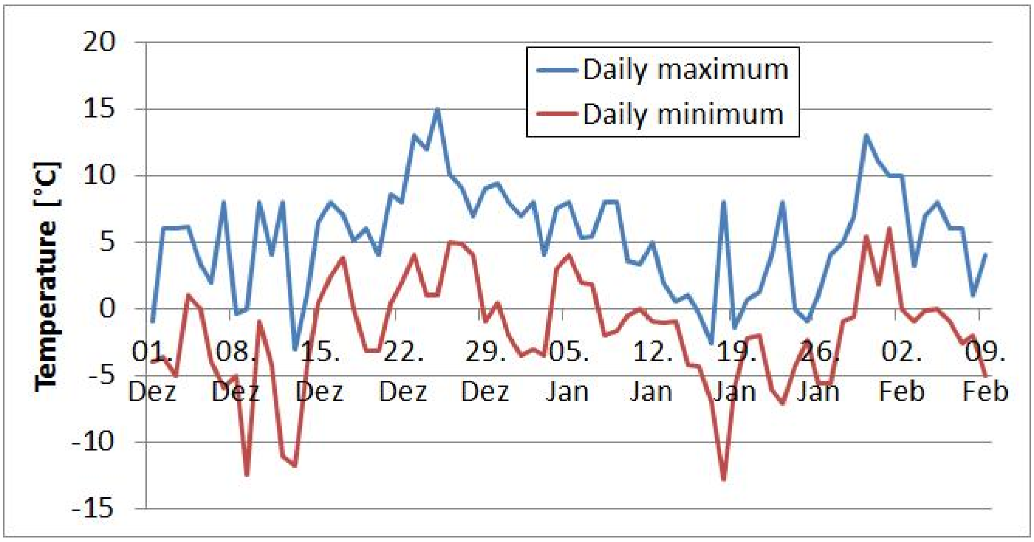


Figure 4.1: Outdoor temperature in Zurich during the study period

4.7 Software

The readout software generated a csv-file for every household's smart shower meter, while the survey data of all households were exported into a single xlsx-file. The data analysis was mainly carried out with Stata 11.0 / Stata 13.0.

Chapter 5

Results

The main goals of this section are to quantify the effect of the intervention on shower behavior, what parameters they adjusted to realize the savings, to evaluate how individuals perceived the device, how they assessed their resource consumption in the shower, and to understand the underlying psychological mechanisms. For these purposes, the set of shower panel data was combined with an extensive set of survey questions. The structure of this chapter is the following: The first section presents participants' evaluation of the device; the second section describes participant's shower behavior before the onset of the treatment; the third section compares participants' self-estimated water use with their actual behavior; the forth section presents descriptive statistics and verifies whether the randomization process has successfully produced balance on observable key characteristics. The fifth section quantifies the main treatment effect before section 5.6 analyzes the underlying psychological mechanisms analyzes underlying psychological mechanisms.

5.1 Device evaluation by the participants

Overall, the device was rated very favorably by the study participants, as figure 5.1 shows.

82% of the treatment group agreed or strongly agreed with the statement *"I'm overall happy with the shower meter"* and 79% of all participants (including control group participants) stated that they intended to continue using the device after the study. Moreover, among those who had indicated that they neither agreed nor disagreed with the later statement or slightly disagreed (14% altogether), the majority still wished to get the device back after the data readout.

Given the sample recruitment (see section 3.4), two aspects should be taken into account with respect to these figures: On the one hand, study participants had probably at least some general interest in energy conservation or technology topics, otherwise they would not have signed up for the antecedent *ewz Studie Smart Metering* in the first place. On the other hand, none of the participants had ever stated any interest in having a feedback device for

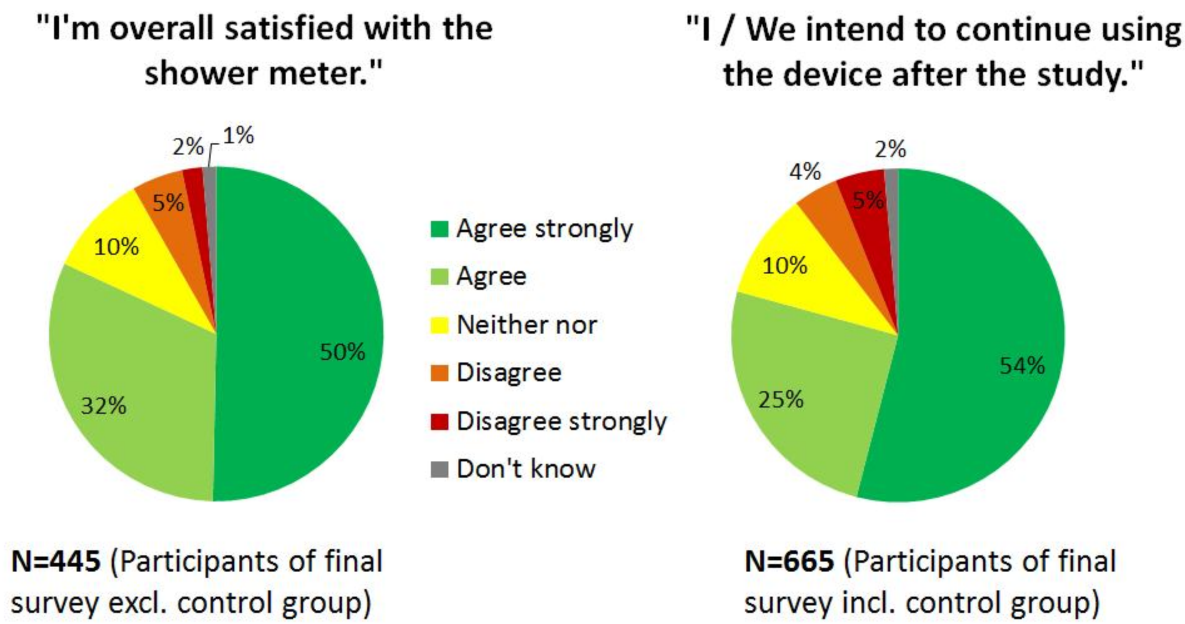


Figure 5.1: Overall user satisfaction and intention to continue using the device after the study

the shower: They simply received it as a thank you gift, independent of their participation in the *ewz-Amphiro study*. It remains an open question to what extent an evaluation by the general public on the one hand, or on the other hand by actual customers who bought the device out of their own interest, would look like. However, it can be concluded that there were no major privacy concerns, and general positive attitude towards the device among the study participants.

5.2 Baseline shower data

The histograms in figure 5.2 show the distribution of liters of water per shower, temperatures, and the implied use of energy (in kWh) for the baseline period (i.e., up to shower #10) over all treatments. The histograms show that water and energy consumption vary widely between showers. As panel (a) visualizes, while some individuals shower with as little as 10 liters of water, others use over 100 liters per shower. Ranked by their per-shower consumption, users in the 90th percentile use over seven times as much energy and water as the ones in the 10th percentile. Shower temperature, depicted in panel (b), varies much less and is between 35 and 40 degrees in most cases. Panel (c) displays the implied energy consumption per shower. The mean is 1.6 kWh, and is thus considerable for an activity that lasts on average four minutes. Panel d) shows the distribution of flow rates, with an average flow rate of 11.0 liters/minute and a standard deviation of 2.3 liters/minute between households.

With respect to policy relevance for Switzerland, the most interesting implications are in terms of the reduction in energy consumption, and this will be the focus of our analysis. We

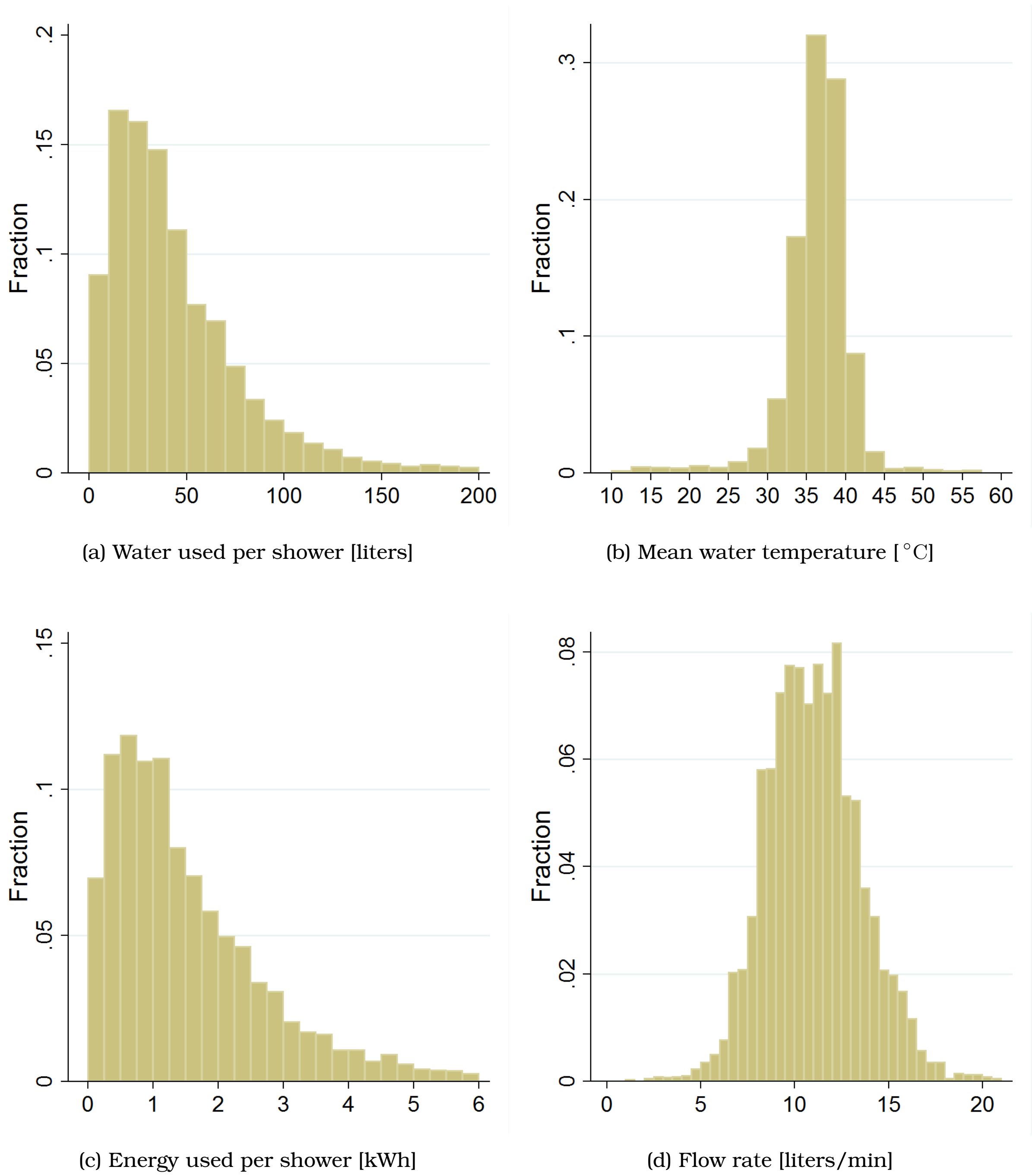


Figure 5.2: Frequency distribution of four key shower characteristics (baseline period)

will also present some of the results for the volume of water. Any of the conclusions will be valid for both outcomes, as they are highly correlated ($\rho = 0.993$).

Table 5.1 presents correlations between baseline energy consumption (per shower) and a range of demographic and personality variables. Most importantly, the table shows that age is very highly correlated with baseline water consumption. The simple correlation is equal to 0.3, and highly statistically significant. Age differences are associated with very large differences in water and energy consumption during a shower. Using the underlying regression from the reported correlation, it implies that raising the age by 10 years reduces predicted energy consumption by 0.25 kWh per shower (and reduces the water consumption by 6.3 liters of water). This is a quantitatively important effect, but is difficult to interpret: It could either be a cohort effect, implying that the younger generation uses more water throughout its life, or it could be a life-cycle effect: Perhaps in every generation, individuals use more water when they are young. Our data offer us no way to distinguish between the two, but we will return to this point in section 6.3.

The table also shows that household structure (one- or two-person household) and the fraction of female inhabitants does not affect behavior during a shower very much. The correlations are small. On the other hand, individuals do seem to realize that they could save water and energy when their baseline consumption is high: There is a strong positive correlation between the degree of agreement with the statement that the household could reduce their water and energy consumption in the shower. Strikingly, an individual's self-reported tendency to protect the environment (environmental attitudes) is strongly correlated with behavior during a shower: The higher the agreement with the statement (see line (5) in table 5.1), the lower the water and energy consumption during a shower. The correlation is equal to 0.2 and statistically highly significant. Again, the effect is also economically large: A one-point increase on the 1-to-5 Likert scale for that question is associated with 0.29 kWh less energy consumption per shower, and 7.7 liters lower water consumption. However, a look at the cross-correlations of this variable with others shows that it is also highly correlated with age and with gender. Thus, it is not clear how these correlations with baseline consumption should be interpreted. These correlations also raise the issue of multicollinearity in regressions where all these variables are included. However, one should keep in mind two things: First of all, the correlations we observe here are not very large. Second, one should recall that multicollinearity is not a source of bias in the estimation of coefficients in a multiple regression. It affects the precision with which the contribution of a variable within a group of collinear variables can be pinned down, and vanishes as the number of observations grows large. In our set of variables, as we will see below, multicollinearity is clearly present, but does not present a problem, as we are still able to assess the contribution of most variables with sufficient accuracy.

Table 5.1: Empirical correlations between energy usage and socio-demographic variables

	Baseline usage (in kWh)	(1)	(2)	(3)	(4)	(5)	(6)
Age (1)	-0.318						
	(0.000)						
Number of household members (2)	-0.018	-0.043					
	(0.659)	(0.295)					
Fraction of women in household (3)	-0.079	0.111	0.028				
	(0.049)	(0.007)	(0.479)				
Self-assessed savings potential (4)	0.183	-0.068	0.059	-0.014			
	(0.000)	(0.097)	(0.144)	(0.732)			
I will protect the environment, even at high personal cost (5)	-0.207	0.236	0.087	0.169	0.046		
	(0.000)	(0.000)	(0.031)	(0.000)	(0.251)		
Tendency to measure goals (6)	0.062	-0.297	-0.058	0.016	0.052	-0.049	
	(0.132)	(0.000)	(0.158)	(0.696)	(0.213)	(0.233)	
Tendency to compare (7)	0.113	-0.329	-0.010	-0.081	0.007	-0.093	0.546
	(0.006)	(0.000)	(0.814)	(0.050)	(0.857)	(0.024)	(0.000)

Notes: Table reports correlation coefficient and in parentheses p-value of the (two-sided) hypothesis that the correlation is zero. The numbering in the first column of each row is used as an abbreviation for the column labels.

5.3 Estimated use vs. actual use

Feedback can work effectively without reflective decision-making (Hansen and Jespersen (2013)) or without users increasing their knowledge about their consumption. Mitchell et al. (2013) for instance describe that while the home water reports used in their program achieved a 5% reduction in water consumption, they did not increase households' ability to provide accurate estimates of their average daily water use. To evaluate whether the information was actively being processed and whether users actually learned about their consumption in the shower in this study, participants' self-estimated water consumption per shower in the pre- (resp. post-) intervention survey was compared with their baseline (resp. intervention) period consumption mean. As figure 5.3 illustrates, the majority of participants underestimate their water consumption per shower (points below the dashed black line). After the intervention, the situation changes quite a bit: Participants who had received the feedback information now report values that are close to their household's actual consumption per shower (solid red and dashed orange line), as figure 5.4 shows. Interestingly, the estimate of the control group also slightly improved over the initial self-estimate (solid blue line). This indicates that the installation of the device has raised their awareness, pushing their estimate towards more realistic values. This also implies that the actual effect of the device might be even larger, as the consumption of the treatment group should ideally be compared with a group whose awareness and consumption pattern is not affected by the participation in the study.

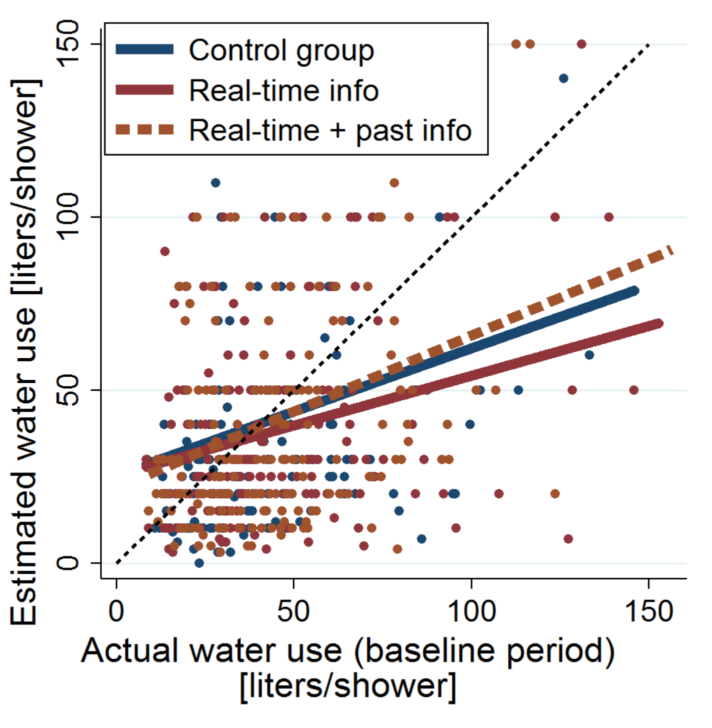


Figure 5.3: Actual consumption during baseline period vs. upfront estimate (initial survey) of water consumption per shower

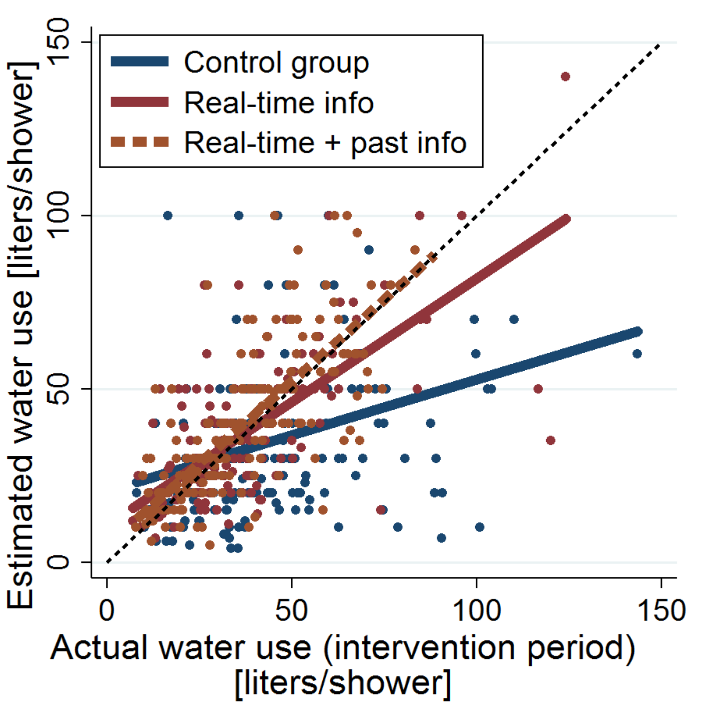


Figure 5.4: Actual consumption during intervention period vs. ex-post estimate (final survey) of water consumption per shower

Two aspects should be pointed out concerning figure 5.3 and figure 5.4: First, the estimated values in both figures are based on the survey respondent's estimate for her personal consumption, while actual usage data are composed of two persons in two-person households. Second, 5.3 compares consumption estimates from the initial survey with baseline use data, while 5.4 compares estimates from the final survey with consumption data after the beginning of the intervention. The reason for that is that upfront estimates should be compared with measurement data before the intervention influenced the shower behavior; on the other hand, ex-post estimates are affected by the information that participants were exposed to on the shower displays; therefore, ex-post estimates should be compared with the study period where (treatment group) participants could monitor their consumption.

Regarding relative consumption, participants seem pretty unaware what "high" or "low" consumption is: In the initial survey, participants were asked to rank their per-shower consumption relative to other participating households of equal size. As figure 5.5 illustrates, participants' perception of their relative position (i.e., consumption compared to other participating housholds) is poor: While low users tend to overestimate their usage relative to others, high users tend to underestimate it. This means that participants have a very limited understanding of what an average shower is: There is no clear social norm that prescribes what would be considered a as long or as a short shower. This is in line with the findings of Beal et al. (2013) who argue that *"self-nominated high users may be setting themselves a higher* *benchmark on what is low or personally acceptable consumption and believe there is always something more they could do to reduce their household's consumption."*

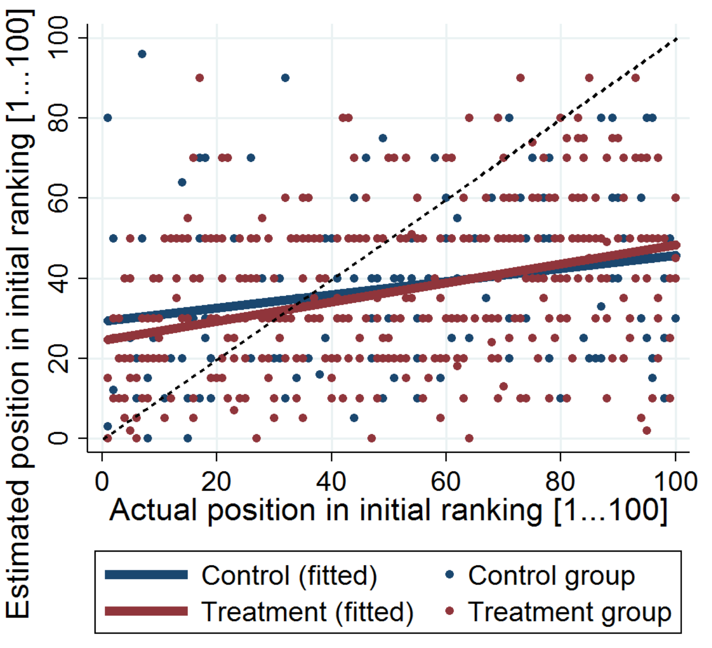


Figure 5.5: Ranking of participating households by water consumption in the shower: estimated position from the initial survey vs. actual rank during the baseline period

Overall, the results show that a) most households are unaware whether their per-shower consumption is high or low and b) that households who were exposed to the real-time information in the shower acquired an improved knowledge on how much water they typically use. This implies that the feedback information is actively being processed and remembered by the users.

5.4 Descriptive statistics and randomization checks

Key to our subsequent analysis is the random assignment to the experimental conditions. In order to be able to estimate the causal effect of the feedback device on shower behavior, one needs to be certain that the assignment to the treatment and control groups was random. We had implemented a random mechanism in the assignment, but we can also verify that observable characteristics before the onset of the treatment (at shower 11) were identical across groups.

In order to do this, we estimate the equation

$$y_{it} = \beta_0 + \beta_1 T_{1i} + \beta_2 T_{2i} + d_t + \epsilon_{it} \tag{5.1}$$

where T_1 and T_2 are indicators for the *real-time information* and *real-time plus past information* treatments, respectively. We use all the variables from table 5.1 as dependent variables in order to examine random assignment: The first dependent variable y_{it} is mean energy use during baseline phase, i.e., the first ten showers. We then examine various individual-level traits that are potentially important for our analysis. These regressions also serve a second purpose in that they give a sense of the means of these variables in the control group by examining the constant term in the regression.

Table 5.2 displays the results. They show that the random assignment to the groups has had its desired effect: None of the regressions finds any significant difference for any of the variables examined across treatments. Most importantly, there are no differences with respect to energy use during showers during the baseline phase. As can be seen in the table, there are only minimal differences between the means of the three treatments. Thus, before the onset of the intervention, all groups had the same shower behavior. The same holds for all the other traits. The F-tests for the regression model also show that we cannot reject $\beta_1 = \beta_2 = 0$ for any of the specifications.

Table 5.2: Randomization checks

	Baseline consumption	Age	Fraction female	Household structure	Environmental attitudes	Measure goals	compare with others
Current shower info (=1)	0.027	-0.011	-0.013	-0.024	0.007	0.010	0.159
	(0.095)	(0.129)	(0.037)	(0.045)	(0.074)	(0.081)	(0.108)
Current plus previous shower info (=1)	0.082	-0.070	0.024	-0.015	0.079	0.048	0.121
	(0.105)	(0.128)	(0.036)	(0.045)	(0.070)	(0.083)	(0.111)
Constant	1.595***	4.036***	0.459***	0.524***	3.492***	3.285***	2.619***
	(0.069)	(0.091)	(0.026)	(0.032)	(0.052)	(0.057)	(0.076)
R^2	0.001	0.001	0.002	0.000	0.002	0.001	0.004
Obs	626	600	626	626	615	588	588

* p<0.10, ** p<0.05, *** p<0.01

5.5 Analysis of the main treatment effects

We proceed in several steps in the presentation of the main results. First, we present evidence from simple shower-to-shower means by treatments in order to simply gauge the rough overall effect. We also examine the distribution of shower outcomes across treatments, in order to get a sense if the treatments affected different aspects of the outcome differently.

The descriptive graph in figure 5.6 illustrates the main treatment effect: It plots the means for each of the three treatment groups by shower. Up to the 10th shower, the device displayed only water temperature in all groups, thus providing little information on energy and water use. At shower 11, the display was activated in the two treatment conditions (*real-time information*, and *real-time plus past information*, respectively). The figure makes it obvious that there is a dramatic impact of the information displayed on shower behavior. Virtually on impact, energy consumption drops by roughly 0.3 kWh per shower. The effect is remarkably stable over time, thus suggesting a highly significant and persistent treatment effect.

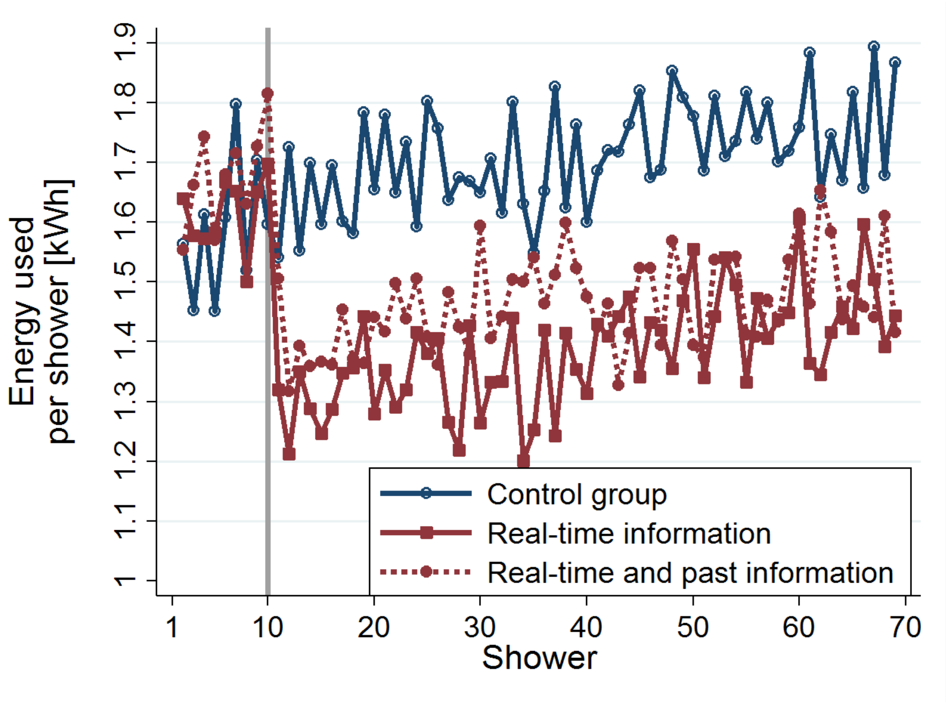


Figure 5.6: Mean energy consumption by shower across all treatments: means over time

A second way to visualize the effect is to compare the cumulative density functions (cdf) of the shower outcomes before and after the treatment. Figure 5.7 shows the results. Panel a) shows the cdf during the baseline period, i.e., before the display of the information in the two treatment groups. The cdfs are identical across the three groups, as could be expected from figure 5.6. Panel a) shows that in all three groups, for instance, for 60 percent of the showers, 60 liters of water or less are consumed and 90 percent of the showers are taken with 90 liters of water or less. Panel b) shows the cdfs over the treatment period. The cdfs of the two treatment groups are clearly shifted to the left of that of the control group line: This implies that there is a clear shift towards using less water during showers, which starts building up from 20 liters onwards, and remains relatively stable across the distribution. Furthermore, the cdfs of the two treatment groups are very similar, suggesting that the two treatment groups have very similar effects on behavior.

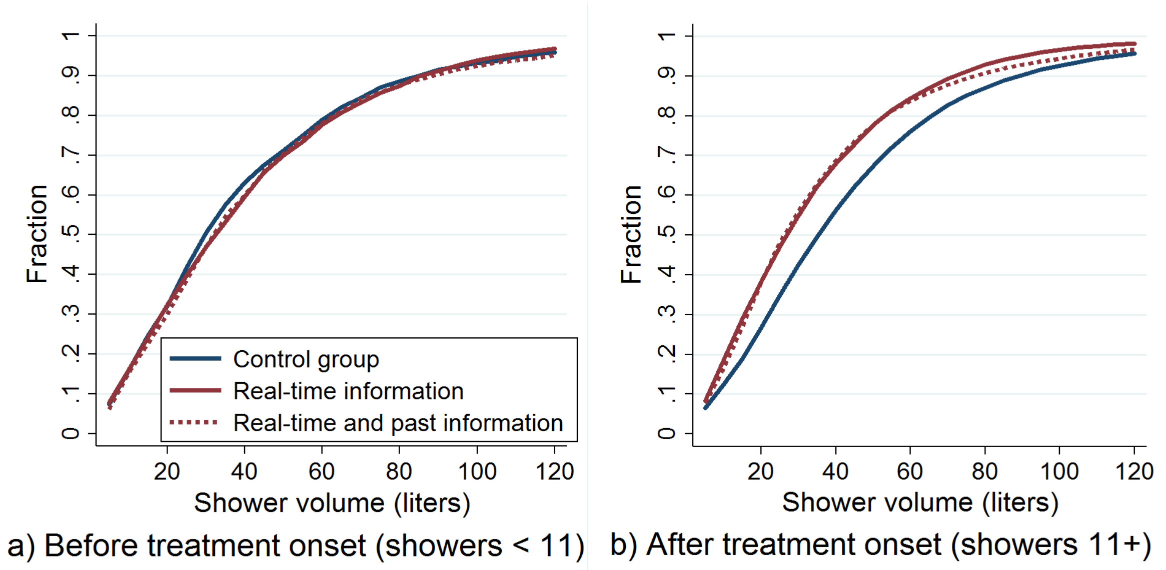


Figure 5.7: Mean energy consumption by shower across all treatments: empirical cumulative density functions

Panel a) and Panel b) also indicate that the cdf of the control group has shifted slightly over the course of this period. This suggests that a better approach to assess the true effect of the feedback device is a difference-in-differences strategy: Comparing the change of the means in the treatment period relative to the baseline period of the treatment group to that of the control group. This is implemented in our main equation that we estimate in (5.3). Before estimating equation (5.3), we also provide a graphical impression of the difference-in-differences estimates in figure 5.8. To this end, we calculate

$$\Delta y_i = \overline{y_{i1}} - \overline{y_{i0}} \tag{5.2}$$

for each household i, where $\overline{y_{i1}}$ is the average energy used from shower 11 until the last observed value, and $\overline{y_{i0}}$ is the average over baseline showers. Thus, the comparison provides us with a valid estimate of the treatment effect, and the standard error of these variables is unaffected by between-subject heterogeneity, as this is differenced out.

The results in figure 5.8 confirm the qualitative finding from figure 5.6. They show that there is a modest increase in energy consumption of the control group over the study period, and this increase is mostly insignificant. However, there is a strong decrease in the treatment groups of each type of household. Thus, it is reasonable to use the difference-in-differences strategy in order to obtain the causal effect unconfounded with time trends that affect the treatment and control groups alike. The true treatment effect is the difference of the two differences, and the standard error bars around the means already indicate that the reduction due to the treatment is highly significant.

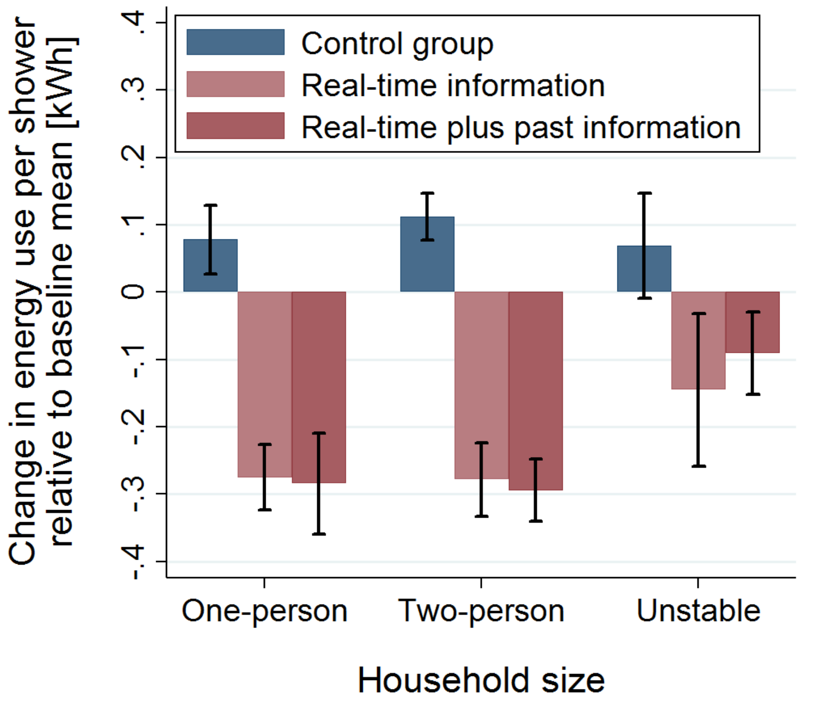


Figure 5.8: Difference-in-differences estimates of the main treatment effect

The treatments work approximately equally well for the single and two-person households, and equally well for the *real-time information* and the *real-time plus past information* condition. Furthermore, the graph already indicates that the treatment appears not to have worked so well in the category of households with an unstable composition (see section 4.5).

5.5.1 Empirical strategy

In this subsection, we set up a statistical model that allows us to fully take advantage of the experimental setup and to test formally the indications of the treatment strengths that we obtained in figure 5.6. We use a fixed effects model, which controls for time-invariant characteristics that are unique to the individual household (Torres-Reyna (2007)). The participating households have, for instance, different kinds of shower heads, which equally affect shower flow rate (and thus water and energy consumption) throughout the study. These characteristics are unobserved, time-invariant, and unique for every household. We define the model as follows

$$y_{it} = \alpha_i + \beta_1 T_{1it} + \beta_2 T_{2it} + d_t + \epsilon_{it} \tag{5.3}$$

where y_{it} is our dependent variable (either energy consumption in kWh or liters of water consumed per shower). We include an individual fixed effect α_i for each household in order to eliminate all variance stemming from fixed differences in shower outcomes between households. The indicators T_{1it} and T_{2it} are all zero for the first 10 showers and then take on the value of 1 if household i is assigned to the *real-time information* and *real-time plus past information* treatment, respectively. We also include a shower fixed effect d_t to capture time trends in the best possible way.

We estimate equation 5.3 by ordinary least squares (OLS) and allow the residuals to be correlated within a household in arbitrary ways. We correct for this by reporting standard errors clustered at the household level.[3] Note that since we use a fully-randomized design, model specification is not an issue here. The treatment is randomized and thus uncorrelated with any other variable, observable or not. The only variables we need to include in the specification for valid inference are the treatment dummy variables. Of course, inclusion of other variables will make the estimation potentially more precise, but they are not necessary to estimate the causal effect of the treatments.

5.5.2 Main treatment effects

We present the results in the following way. In a first step, we report the treatment effects on the two key outcomes: energy used per shower and water used per shower. In a second step, we examine whether the behavioral adjustments come from adjusting shower time, flow rate, or water temperature. We estimate these treatment effects separately for each category of households: single-person households, two-person households, and those unstable households that could not be characterized clearly.

The results are presented in table 5.3. The first three columns present the results for the volume of water used, the next three columns also take into account water temperature and calculate the amount of energy used. For each group, the estimated coefficient is reported, followed by the standard error in parenthesis. Asterisks indicate significant interactions, the corresponding significance levels are indicated below the table. Turning to the results for water volume, the estimates show that the activation of the display had a clear and significant impact on the amount of water used per shower, as was already to be suspected from the graphs presented earlier. However, in this case, the standard errors are calculated in a reliable way and are valid for inference. The results in the first column (single households)

[3] Ex-post analyses show that there is no very strong correlation in residuals. The estimated autocorrelations, while significant, are all in the order of magnitude of -0.1 to +0.1. However, it is still prudent to correct the standard errors, as we have a large enough number of clusters for the asymptotic approximations to work (Angrist and Pischke (2009)).

Table 5.3: Difference-in-differences estimates for water and energy consumption per shower by household type

	Volume [liters]			Energy [kWh]		
	1-person	Unstable	2-person	1-person	Unstable	2-person
Real time info	-9.407***	-5.590*	-10.477***	-0.369***	-0.219*	-0.389***
	(1.789)	(3.018)	(1.685)	(0.069)	(0.115)	(0.066)
RT + past info	-10.474***	-3.644	-10.997***	-0.397***	-0.146	-0.407***
	(2.441)	(2.667)	(1.513)	(0.094)	(0.101)	(0.058)
Constant	44.355***	42.814***	44.508***	1.634***	1.539***	1.611***
	(1.674)	(2.521)	(1.607)	(0.066)	(0.097)	(0.061)
R^2	0.575	0.359	0.359	0.577	0.371	0.370
Obs	13298	6711	25027	13298	6711	25027
Clusters	255	102	269	255	102	269

* p<0.10, ** p<0.05, *** p<0.01

show that information on the current shower alone achieves the same reduction in water consumption as information on the current shower augmented with information on the previous shower. The estimated reduction in water use is 10.5 liters in the *real-time plus past information* condition, compared to a reduction of 9.5 liters in the *real-time information* condition. Statistically, the two effects are indistinguishable, and qualitatively they are also very close. It is worth reiterating the considerable quantitative importance of these effects. As can be seen from the constant, the average water use per shower in the control group was 44.4 liters. We achieve a reduction in water use of 10 liters, i.e., reduce water consumption by almost 25 percent of the mean. This is a very large reduction.

The third column in the table shows the results for the two-person households and highlights that for this category, too, there is a significant reduction in water consumption of virtually the same magnitude. Again, information on the current shower leads to just the same reduction in water use as information on the current plus past shower. This is notable, as in a two-person household, the past information could be more interesting, as it conveys some information about the behavior of the other household member.

The second column shows the outcome for the third group of households which we could not clearly classify as a one-person or two-person household. The treatment appears less effective in these households, but are nevertheless jointly significant ($p = 0.07$, not reported in the table). It is difficult to interpret these weaker results, as household composition in most of these households was unstable, which was the reason why they were classified into this group (section 4.5). In many of the initially two-person households, one of two household members was reported to have been absent over extended periods of time. If the two household members have different shower patterns, this will strongly affect the outcome of this period and of the overall intervention: While some periods of the dataset (e.g., the baseline period) only reflect the shower behavior of one household member, later (or earlier) sections of the dataset incorporate the shower patterns of both household members. The same is true for visitors: Depending on their shower patterns relative to the permanent household member(s) and depending on the timing of their visit, their shower behavior affects the outcome in irregular ways. In all of these cases, the occurrence of visitors/absent household members is erratic. As these erratic events take place both in the treatment and control group as well as during both the baseline and the intervention period, they add "noise" to the dataset and reduce the treatment effect in these households. It might also be the case that more "marginal" members of the household do not really care about the appliance or exhibit some other characteristic behavior that makes them less responsive to the treatment. We favor the first explanation, as we found no other evidence that water consumption differs between household types (see table 5.1).

Columns 4 to 6 of table 5.3 display the impact of the experiment on energy use per shower. Again, the results show a strong and significant impact of the treatment for one-person and two-person households, and a somewhat weaker response among mixed households (but again, the reduction is jointly significant, $p = 0.06$). Energy consumption is simply calculated based on total volume of water and its average temperature, not taking into account energy losses in water heating, distribution, and storage. Thus is it a lower bound on the actual energy savings. However, even these numbers show that the shower feedback device can lead to behavioral changes that have large effects on energy consumption of a household. Energy savings between 0.3 and 0.4 kWh per shower already amount to roughly 5 percent of a household's daily energy consumption.

In table 5.4, we decompose the overall outcomes into different behavioral patterns. Again, the regression coefficients are reported along with the corresponding standard errors. The first three columns of panel A show the regression results with the duration of the shower as the dependent variable. The constant term shows that in the control group, the average duration of a shower is 245 seconds. The activation of the display leads to a sharp reduction in shower duration by roughly 45 to 55 seconds, depending on the treatment and the exact population (again, the results are weakest for the household types with unclear composition). The next three columns in panel A display the results for the flow rate during a shower, measured in liter per minutes. Aside from reducing shower duration, one possibility to reduce water and energy consumption during a shower is to reduce water flow rate, while possibly maintaining the same duration of the shower. The constant term, displaying the mean of the control group, is equal to roughly 11 l/min, in all three samples. The results for the experimental treatments show a modest, but statistically significant, reduction in the flow rate of about 0.2 to 0.3 liters

Panel A	**Duration of shower** (seconds)			**Flow rate of water** (l/min)		
	1-person	Unstable	2-person	1-person	Unstable	2-person
Real-time info (=1)	-48.676***	-35.655**	-59.594***	-0.296***	-0.104	-0.032
	(9.670)	(16.048)	(10.063)	(0.108)	(0.192)	(0.109)
Real-time + past info (=1)	-55.157***	-18.000	-56.348***	-0.294***	-0.160	-0.079
	(12.478)	(14.067)	(8.239)	(0.108)	(0.134)	(0.110)
Constant	244.827***	232.996***	248.240***	10.813***	11.370***	11.003***
	(8.976)	(15.088)	(9.219)	(0.062)	(0.111)	(0.079)
R^2	0.534	0.381	0.361	0.859	0.776	0.737
Obs	13298	6711	25027	13298	6711	25027

Panel B	**Average temperature** (°C)			**Breaks during shower** (seconds)		
Real-time info (=1)	-0.716***	-0.528*	-0.037	2.056	5.219	9.463***
	(0.260)	(0.295)	(0.238)	(2.933)	(3.823)	(2.781)
Real-time + past info (=1)	-0.421**	-0.139	-0.177	1.878	-0.605	4.399
	(0.201)	(0.351)	(0.220)	(3.406)	(5.132)	(3.098)
Constant	36.297***	36.044***	36.257***	34.521***	30.294***	35.501***
	(0.243)	(0.282)	(0.197)	(2.966)	(4.713)	(3.058)
R^2	0.380	0.401	0.298	0.396	0.269	0.305
Obs	13298	6711	25027	13298	6711	25027

* p<0.10, ** p<0.05, *** p<0.01

Table 5.4: DiD estimate of baseline effects by household types on other outcomes

per minute for one-person households. While the point estimates are also negative on the treatments for two-person households, we cannot reject that the flow rate is constant in this sample. Thus, we cannot reject that the responses are different across groups, leaving us in somewhat unclear territory. However, in any case, the relative change induced in the flow rate (roughly two percent) is much smaller than the change in the duration of the shower (nearly one quarter). Again, we find the weakest results for the households with unclear composition, a fact which we will henceforth ignore.

The first three columns in panel B of table 5.4 examine how the intervention affects shower water temperature. If individuals want to save energy, another possibility would be to lower the temperature of the water used in the shower. Our dataset allows us to explore the extent to which individuals use this margin. The constant term shows that the average water temperature in the control group was 36 degrees. The experimental intervention for single households reduces the water temperature slightly, by 0.4 to 0.7 degrees, depending on the treatment. However, again, the results are more muddled for the two-person households: Even though the point estimate of all the treatment effects are negative, they fail to be significant. Again, as with the flow rate, the temperature doesn't appear to be an important margin of adjustment. Columns 4 to 6 of panel B show the results for possible stops of water flow during a shower. Another possibility to conserve water and energy is to stop the water flow while applying shampoo or soap. The mean time during which no water flows during a shower was 30 to 35 seconds, depending on the type of household. Again, the results show that the display tend to push individuals towards conserving energy, but the results are again quite disparate across treatment groups. This time, the largest effect appears in the group of two-person households, while one-person households display a somewhat smaller treatment effect. But again, the margin of adjustment does not appear to be particularly important.

In summary, the activation of the display leads to a dramatic change in behavior by the participants. In all experimental groups, we observe a large and significant drop in water and energy consumption per shower. Water and energy drop by approximately 23 percent and thus amount to a quantitatively significant reduction in water consumption and energy consumption at the same time. This change in behavior is primarily driven by a change in the duration of the shower by over 20 percent. Other margins of adjustment, such as flow rate, water temperature, or stops of water flow during a shower, all point in the direction of more conservation, but fail to show a clear an important contribution to the overall outcome that we observed. Overall, adjustment of shower time appears to be the major source of water and energy reduction.

5.6 Identification of psychological mechanisms

The previous subsection has identified the causal treatment effect on behavior. While this is already a useful piece of information for policy analysis, more can be obtained from the setup of our study. The setup allows us to provide a more detailed picture on the most cost-effective use of the shower meter, the psychological mechanisms that lead to the large observed treatment effect, and also provide us with the possibility to assess the external validity of the results.

5.6.0.1 a) Responsiveness of different subgroups

Policymakers face budget constraints and often need to decide how to allocate a limited budget. In our case, this may be the decision to whom to allocate a *amphiro a1* device. To maximize impact, the device should be allocated to individuals who will create the largest savings with it. Our setup allows us to examine for which groups of individuals the intervention was most effective by combining the behavioral data recorded by the shower monitor with the information collected in the surveys. For example, it may simply be easier to save energy for individuals who start from a very high level of water consumption per shower. Alternatively, it is possible that young users find the device more appealing, as they may be more familiar with this type of technology in their daily lifes. If the results reveal the relevance of an easily observable characteristic, which is accessible to policymakers, such as gender, age, and baseline water (or energy) consumption, this can help to identify the most responsive group and consequently boost the cost-effectiveness of a related campaign.

5.6.0.2 b) Understanding psychological mechanisms

It is important for economic theory and, ultimately, policy analysis to understand the psychological mechanisms that lead to the observed behavioral change. Since the intervention leads to a large change in behavior, it is important to understand the channel through which this occurs. Thi contributes to an understanding of how prosocial preferences, such as energy conservation, are translated into behavior. Furthermore, this is also important for designing new interventions that could then specifically target these mechanisms.

Since all the participants in this study had previously participated in the *ewz Studie Smart Metering*, we have access to a rich set of personality traits and and other individuals characteristics, which allows us to address these questions. This dataset, combined with the large number of households, the accurate behavioral data, and in particular with the large effect size allows us to examine in detail which psychological mechanisms play an important role.

The key impact of the *amphiro a1* is to make water and energy consumption salient to the individuals while they are taking a shower. This focus on energy savings may help users to curtail energy consumption for several reasons. Our first hypothesis is that anxiety (susceptibility to pressure) or a feeling of being observed creates the impetus to saving. Anxious individuals may strongly react to this information because they may think reducing water consumption is expected from them and they want to fulfill that expectation. Thus, it is possible that the mechanism, while behaviorally effective, relies on creating negative sensations in terms of creating a heightened disutility from using water while showering. Thus, while behaviorally effective, the device may not be desirable because it relies on generating negative affect that is then partially offset by changing behavior. Alternatively, and this is our second hypothesis, it could also be that the information supplied helps individuals to better fulfill their goal of conserving energy, thus creating a utility gain that accompanies the behavior change. The device enables individuals to measure how much energy they use in the shower, thus making it particularly useful for individuals who rely on strategies of quantifying and monitoring progress towards goals.

Our data allows us to test these two hypothesis: We operationalize anxiety with the trait "emotionality" from the HEXACO personality scale available from the *ewz Studie Smart Metering*. If the treatment causes individuals feel observed and anxiety causes the behavioral change, then we should observe that individuals with a higher score in emotionality respond more strongly to the treatment. We also have various indicators that measure individuals' tendency to quantify goals and monitor progress using specific metrics. If the second hypothesis is correct, then we should observe that individuals with a higher tendency to monitor progress respond more strongly to the treatment.

5.6.0.3 c) Assessing the external validity

The results presented in this study so far satisfy the criterion of internal validity: The experiment was properly randomized so that the treatment effect could be estimated *on this population*. One may argue that our sample may be particularly interested in protecting the environment and that it is due to these attitudes that the treatment is particularly effective, while using the device on the general population may generate much smaller benefits. Thus, while our experiment may be internally valid, it may not satisfy the criterion of external validity in the sense that a wide rollout of the feedbacl device may not trigger the same behavioral response in the general population.

Again, the information obtained in the surveys accompanying the study and the information from the *ewz* Studie Smart Metering allow us to address this issue. From that study, we have several measurements, all prior to the present study, that assess individuals' willingness to protect the environment even in situations that may be costly to her/him. If the effectiveness of the feedback device is driven by a selection bias with respect to interest in protecting the environment, then the effect should be larger for individuals with stronger environmental attitudes.

5.6.1 The empirical strategy

We proceed in three steps in this analysis. In a first step, in order to make the analysis more intuitive, we provide a graphical assessment of the interaction effects with respect to the most important variables of interest. We split the sample by the median of each of the interacting variables and display the $\overline{y_i}$ for the treatment and control group in that respective category. In this step, we collapse one-person and two-person households and do not distinguish between *real-time information* and *real-time plus past information*, in order to be able to focus on whether one can detect an interaction effect.

In a second step, in order to estimate the interaction effects and to perform formal tests, we augment the previous statistical model to

$$y_{it} = \beta_0 + \beta_1 T_{1it} + \gamma_1 T_{1it} \cdot z_i + \beta_2 T_{2it} + \gamma_2 T_{2it} \cdot z_i + d_t + \epsilon_{it} \tag{5.4}$$

where the coefficients γ_1 and γ_2 measure how variable z_i interacts with the treatment. In all empirical specifications, we center the interacting variables at their sample mean, such that β_1 and β_2 still have the interpretation of indicating the treatment effect for an individual with average characteristics including z. As can be seen in equation (5.4), here we allow for separate interaction effects with the two treatments, and we also estimate the equation separately for single-person and two-person households.

Two specifications are being estimated for interaction effects: First of all, we estimate equations with interaction terms as specified in (5.4). We then augment the specification to allow for the time trend in the control group to depend on the characteristic z_i as well, i.e. we estimate

$$y_{it} = \beta_0 + \beta_1 T_{1it} + \gamma_1 T_{1it} \cdot z_i + \beta_2 T_{2it} + \gamma_2 T_{2it} \cdot z_i + d_t + +\delta t \cdot z_i + \epsilon_{it} \tag{5.5}$$

.

This could be particularly important in light of the possibility of Hawthorne Effects in the control group: It is possible that the same personality trait that affects responsiveness to the treatment also affects the behavior in the control group: Anxious individuals may respond more strongly to the treatment. They may also, in general, react differently to being studied, thus inducing different time trends in the control group depending on the personality trait z_i. The addition of the $\delta t \cdot z_i$ term allows time trends to differ by personality and thus resolves this issue.

While we perform these regressions for each trait z_i separately (in line with the majority of previous studies which only had access to a subset of information), it is also important to assess the interaction effects in one large regression that includes all the candidates for interaction effects:

$$y_{it} = \beta_0 + \beta_1 T_{it} + \gamma_1{}' T_{it} \cdot \mathbf{z_i} + d_t + \epsilon_{it} \tag{5.6}$$

.

This allows us to distinguish, for example, whether an interaction of the treatment is mediated by age or by baseline consumption: These two variables are highly correlated, thus in order to assess the source of a potential interaction effect, one needs to include both in the estimation. In this final estimation, we do not distinguish between *real-time information* and *real-time plus past information* treatments in order to maximize the statistical power on teasing out the interaction effects.

5.6.2 Treatment interaction effects

In the analysis below, we consider three groups of interacting variables of interest.

- **Responsiveness in different subgroups:** We examine whether there are age-related or gender-related differences in responsiveness to the treatment. We also examine whether savings effects are higher among individuals with high baseline consumption.
- **Understanding psychological mechanisms:** We examine whether there are differences in response to the treatment related to the HEXACO personality traits. Of particular interest is the interaction with emotionality, as discussed previously. We also examine whether an individual's tendency to monitor and quantify progress towards goals affects how he or she reacts to the treatment.
- **External validity:** We examine whether the treatment interacts with individuals' self-reported willingness to protect the environment.

Figure 5.9 illustrates the first set of interactions.[4] Panel (a) shows the interaction for the baseline consumption. As can be seen in panel (a), there is a strong increase in the absolute reduction in energy used per shower due to the feedback device in high-usage households. The treatment effect for above-median households in terms of baseline water consumption is around 0.45 kWh – a very large effect. By contrast, the treatment effect for households with a below-median consumption, while significant, is much smaller and only about 0.20 kWh.[5] Thus, it appears that high-utilization households react much stronger to the shower

[4]In the subsequent analysis, we exclude households with an unstable composition.

[5]We also find that households with below-median consumption in the control group tend to show an increase in their average consumption during the experimental period, whereas households with above-median consumption in the baseline period tend to have roughly constant consumption. This may simply reflect mean reversion and highlights the importance of also estimating an equation that allows the control group to have differing time trends depending on the interacting variable z_i.

feedback device than low-utilization households. Likewise, in panel (c), households with a higher subjective savings potential seem to react more strongly to the treatment compared to below-median households. Similarly, we see strong and significant interactions for age in panel (b). Younger households show a much larger treatment effect. This could be for two reasons: First, young individuals may genuinely respond more strongly to the device, because it appeals to them more than to older persons. On the other hand, age is also highly correlated with baseline consumption. Thus (second), the graph might simply be reflecting this feature. We will be able to address this issue in the final step of the analysis.

In figure 5.10, we provide a first look at the psychological mechanisms behind the treatment effect. We examine whether the impact of the treatment depends on individuals' tendency to monitor their progress towards goals. Indeed, the evidence in panel (a) suggests that this is the case. The treatment effect appears about 0.15 kWh stronger for individuals with an above-median tendency to keep track of progress towards goals than for those with a below-median tendency. We find similar results for the tendency to compare oneself to others (panel (b)): Individuals with a higher self-reported tendency to compare themselves to others display a much larger reduction in water and energy consumption in response to the feedback device. In panel (d), we also examine whether the treatment effect is stronger for individuals with a higher propensity to protect the environment. As the figure shows, this does not appear to be the case. However, one needs to recall that these individuals also tend to exhibit a lower baseline consumption. Therefore, it is possible that the potential for them to save was lower from the beginning. We will return to this point in section 6.3.3.

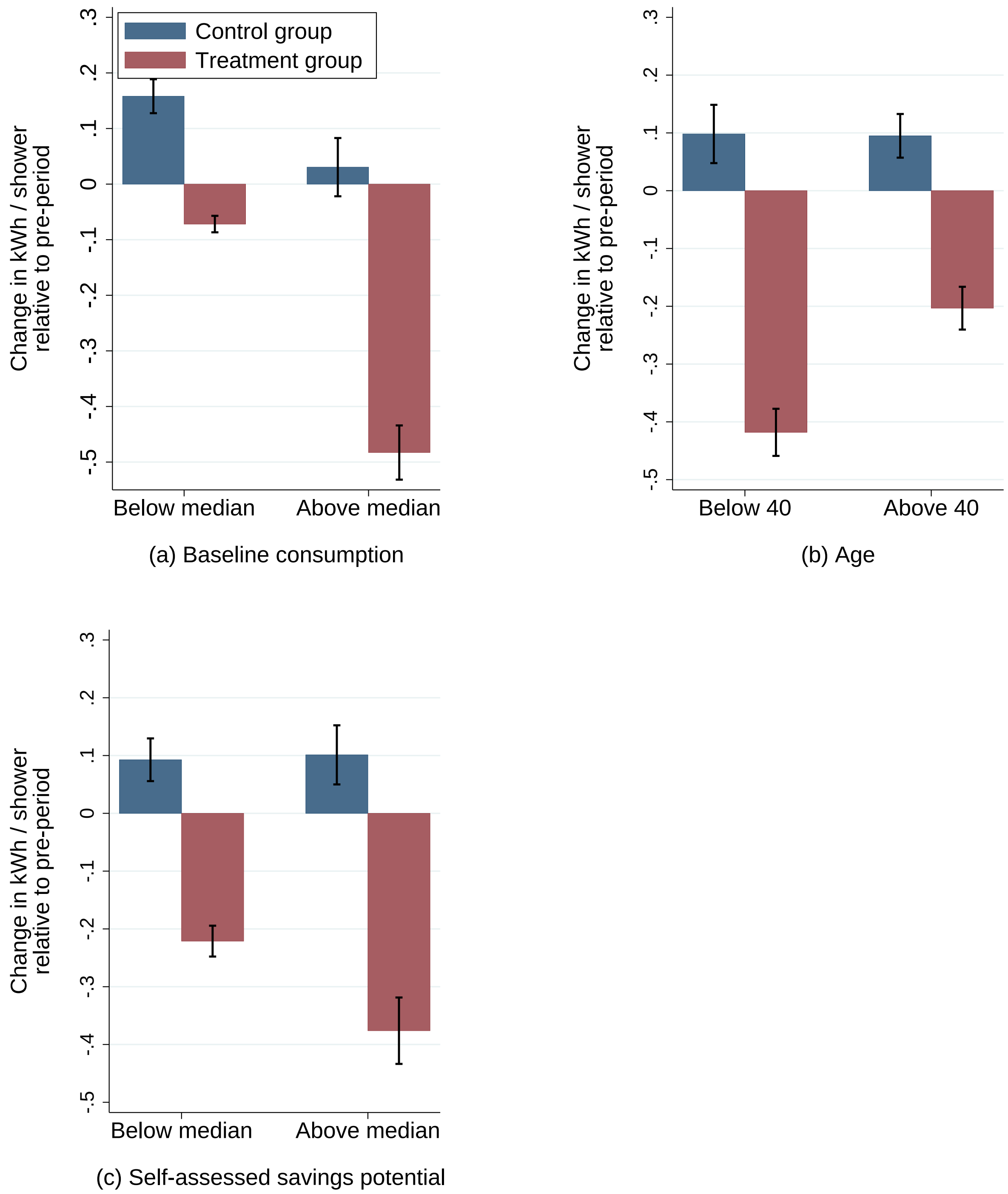
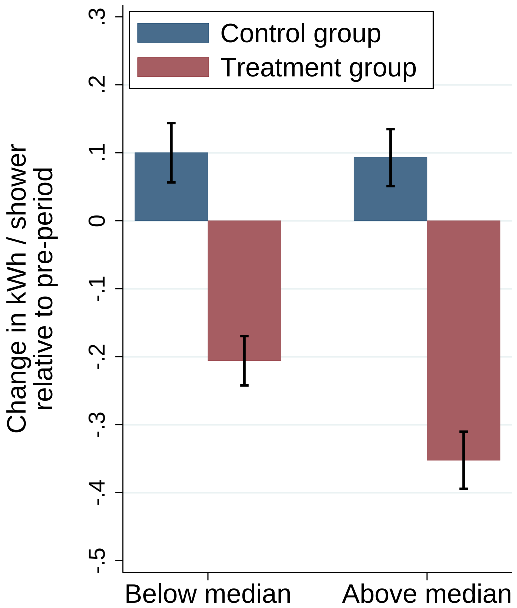
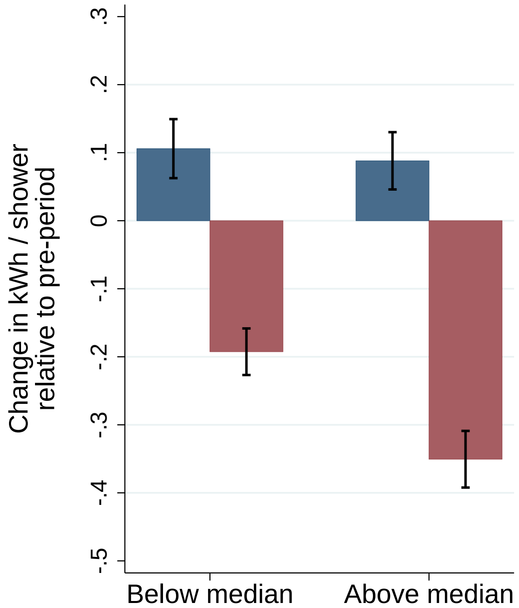


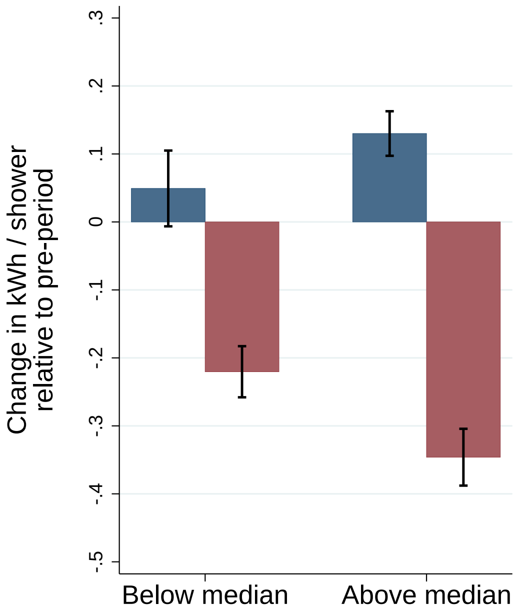
Figure 5.9: Difference-in-differences estimates of the interaction of treatment with baseline consumption, age, and self-assessed conservation potential



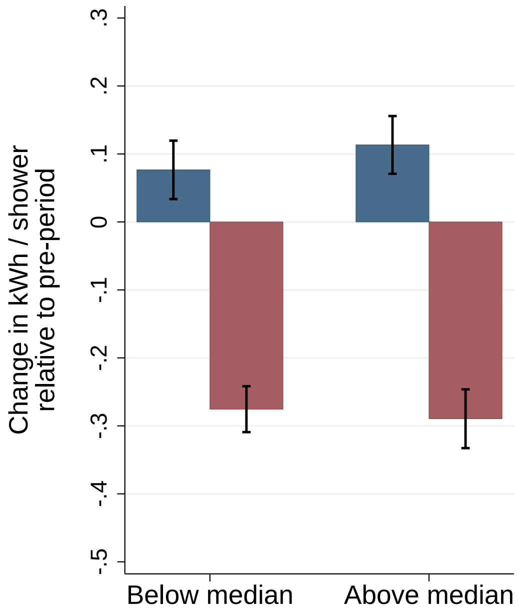
(a) Agreement: measuring and comparing goals



(b) Agreement: I am often comparing myself to others



(c) Satisfaction with quality of life



(d) Agreement: protect the environment even at considerable cost to me

Figure 5.10: Difference-in-differences estimates of the interaction of treatment with personality characteristics

In a second step, we estimate the regression models (5.4) and (5.5) in order to see a more complete picture of the interaction effects for each of the different variables. The results are displayed in table A.1 to A.12 in the appendix. The tables show univariate interaction effects for different variables which are typically associated with heterogeneity in the treatment effect. The first and second item ("RT info (=1)" resp. "RT & past info (=1)") display the regression coefficient β_1 and β_2 from equation 5.4 (column 1, 3 and 5) and 5.5 (column 2, 4 and 6) respectively, along with their standard errors. The following items contain the (more relevant) regression coefficients γ_1 and γ_2 of the interaction terms from equation 5.5 along with their standard errors. The results confirm the findings from the graphs: When the graphs indicate a significant interaction, it is typically found in single-person and two-person households, and mostly significant in each of the treatments.

It is important to point out a null result that is reassuring: We do not find evidence that individuals who are more susceptible to psychological pressure respond more to the treatment. In particular, we do not find any interaction of the treatment effect with the HEXACO personality trait emotionality. By contrast, we observe a rather counter-intuitive outcome for the personality trait conscientiousness: We find that *less* conscientious individuals exhibit a stronger reaction to the treatment (table A.12). Yet if the effect was driven by psychological pressure, we would expect a stronger response from individuals who score high on emotionality and high on conscientiousness. Both the result that the emotionality trait does not interact with the treatment and the fact that there is a negative interaction between the conscientiousness trait and the treatment are important pieces of information to understand the mechanism behind the result: They show that the conservation effect is likely not caused by the generation of psychological pressure to shorten a shower. We also do not find any unspecific interaction effects with personality traits such as honesty or openness, with respect to which we had no hypothesis.

In Table 5.5, we estimate equation (5.6). We select the set of variables to include all interaction terms that turned out to be individually significant in the appendix tables A.1 through A.12. Since many of these variables are correlated, the estimation of equation (5.6) will show us which of the variables genuinely mediate the effect of the treatment. The results are quite surprising. The estimation in column (1) of table 5.5 shows that three variables interact with the treatment: baseline consumption, tendency to measure progress towards goals, and environmental attitudes; the latter was not significant individually. This may be because of its strong correlation with baseline consumption, and the opposite effect on the treatment effect. The significance of these three variables is unaltered when we add group-specific time trends in column (2). In columns (3) and (4), we repeat the estimation without resp. with group-specific time trends, but now exclude all the interaction terms (and their respective time trend in column 4) that were not significant in column (1). None of the results changes.

	(1)	(2)	(3)	(4)	(5)	(6)
Display is on (=1)	-0.445***	-0.424***	-0.426***	-0.411***	-0.385***	-0.385***
	(0.051)	(0.053)	(0.049)	(0.052)	(0.039)	(0.039)
Display × pre-consumption	-0.312***	-0.294***	-0.309***	-0.301***	-0.349***	-0.340***
	(0.056)	(0.058)	(0.048)	(0.051)	(0.048)	(0.050)
Display × monitoring goal progress	-0.072*	-0.077**	-0.061*	-0.067**	-0.064**	-0.069**
	(0.037)	(0.038)	(0.031)	(0.033)	(0.031)	(0.032)
Display × environmental attitude	-0.080**	-0.091**	-0.081**	-0.083*	-0.059*	-0.063*
	(0.039)	(0.043)	(0.039)	(0.043)	(0.030)	(0.034)
Display × fraction female	0.110	0.070	0.099	0.067		
	(0.069)	(0.076)	(0.065)	(0.073)		
Display × age	0.002	0.011				
	(0.028)	(0.029)				
Display × comparing	0.010	-0.001				
	(0.024)	(0.025)				
Display × exchange	-0.012	0.004				
	(0.034)	(0.033)				
Display × conservation potential	0.005	-0.006				
	(0.013)	(0.015)				
Display × satisfaction	-0.002	0.007				
	(0.032)	(0.036)				
Display × monitor × pre-cons					-0.127***	-0.127***
					(0.036)	(0.036)
Display × environment × pre-cons					-0.128***	-0.129***
					(0.048)	(0.048)
Constant	1.632***	1.631***	1.634***	1.633***	1.633***	1.633***
	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
R^2	0.453	0.453	0.452	0.452	0.452	0.452
Obs	35288	35288	36167	36167	36167	36167

* p<0.10, ** p<0.05, *** p<0.01

Table 5.5: Full analysis of interaction effects

Because of the strong interaction with baseline consumption, we explore whether, in general, households with a higher baseline consumption are more responsive to the treatment than households with lower baseline consumption. Figure 5.11 shows that this appears to be the case: It shows the double interaction between high and low baseline consumption *and* the tendency to measure goals. The lightly shaded bars in the left half of the chart display the treatment effect for high- and low-measurement individuals whose baseline consumption was below the median. It shows a small treatment effect, but it hardly appears to change with the tendency to measure progress towards goals. The solid bars in the right half of the chart display the treatment effects for high- and low-measurement individuals whose baseline consumption was above the median: One can easily see that, for the same tendency to measure progress, the treatment effect is higher when the baseline consumption is higher. However, more importantly, the graph also shows that the entire interaction effect of the tendency to measure progress stems from households with a high baseline consumption. Thus, we included a triple interaction between tendency to measure progress and baseline consumption, and also for environmental attitudes and baseline consumption. The results are displayed in columns (5) and (6) of table 5.5. Indeed they show a strong triple interaction: Individuals' tendency to measure progress towards goals leads to a stronger reaction to the treatment when baseline consumption is high. Similarly, environmental attitudes lead to a stronger reaction to the treatment when the baseline consumption level is high. The triple interaction is highly significant in both cases, as the table shows.

In summary, the combination of survey information with behavioral data from the randomized intervention allows us to better understand the psychological mechanisms behind the observed treatment effect. We find no evidence that the generation of peer pressure or other forms of pressure to adhere to social norms would drive the treatment effect, as personality traits measuring susceptibility to pressure do not mediate the treatment effect. Our results are rather suggestive of a mechanism that operates through making energy conservation easier and its measurement more salient. This interpretation is corroborated by the fact that the treatment effect is stronger on individuals who (six months before the intervention) stated a high tendency to monitor progress towards goals. Our best interpretation of the behavioral response is in line with Woodside (2011) and Taubinsky (2013): Many individuals have a strong desire to conserve energy (and water), but they lack the necessary information or the behavioral cues to bring these preferences to their attention in the rush of their daily routines. When provided through a feedback system like the *amphiro a1*, individuals change their behavior quite radically.

This interpretation is also corroborated by evidence on participants' subjective evaluation of the feedback device. As a final piece of corroboration, table 5.6 displays the results from

Likert ratings of various statements. Panel (a) of the table shows that, in general, individuals in the treatment group evaluate the feedback device more positively. They are more likely to recommend it to others than the control group, find the device more helpful, and paid more attention to it in the final two weeks of the study period. Panel (b) estimates these regressions with the tendency to quantify also included as a regressor and as an interaction effect with the treatment. One result stands out in particular: Individuals with a high tendency to quantify tend to find the treatment device much more useful in order to achieve their conservation goals. In other specifications, the interaction is not significant though. We also included baseline consumption as regressor and as an interaction with the treatment effect, but this did not provide relevant new results.

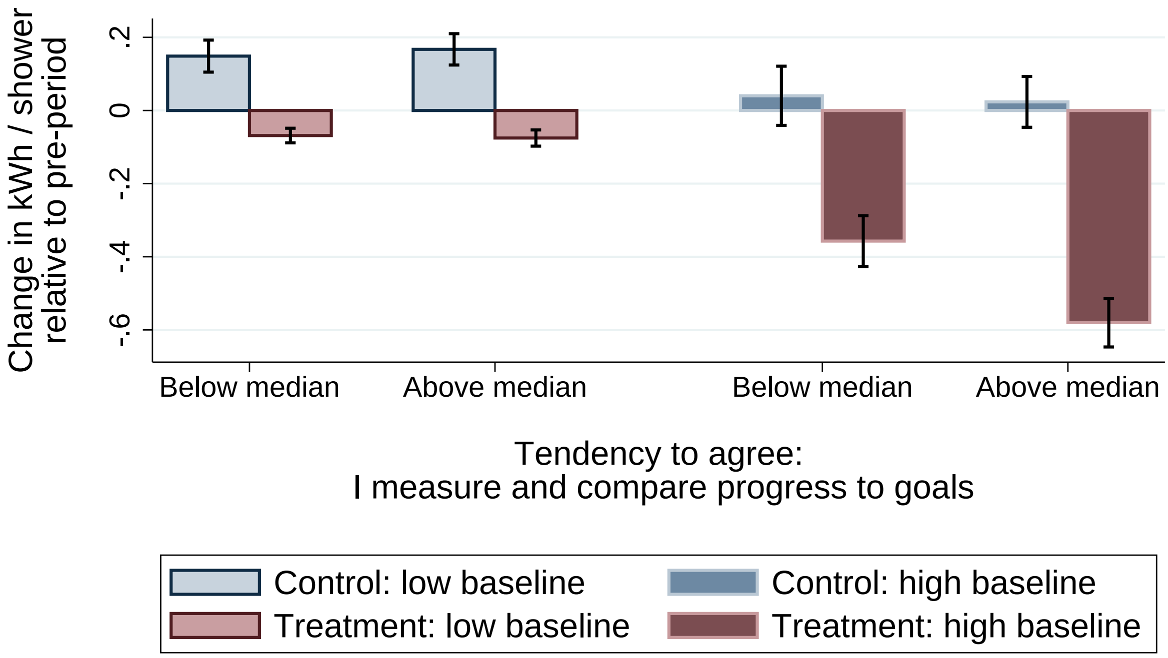


Figure 5.11: Difference-in-differences estimates of the interaction of treatment with baseline consumption and tendency to monitor progress towards goals

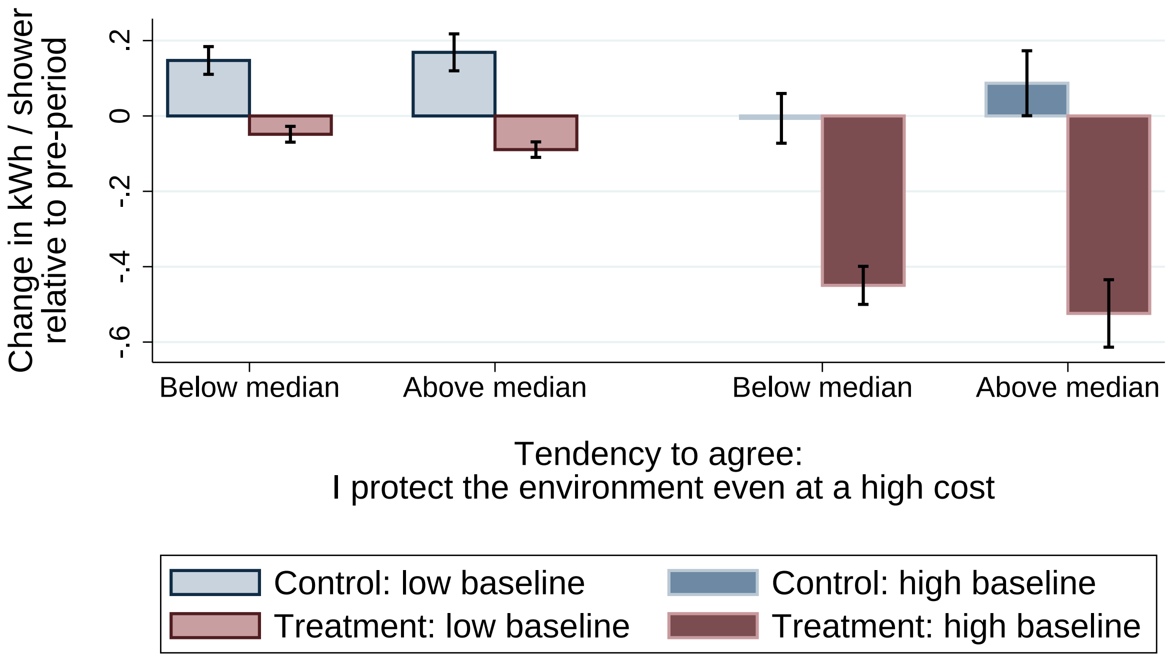


Figure 5.12: Difference-in-differences estimates of the interaction of treatment with baseline consumption and environmental attitude

Table 5.6: Participants' subjective evaluation of *amphiro a1*

Panel (a) **Baseline results**	Device is annoying	Overall I'm happy with the device	I would recommend it	Helpful for saving energy	In final two weeks no attention	We talked about it
Treatment (=1)	-0.041	0.236***	0.170**	1.239***	-0.428***	0.523***
	(0.110)	(0.088)	(0.079)	(0.100)	(0.130)	(0.119)
Constant	2.015***	4.048***	3.571***	2.274***	3.102***	1.696***
	(0.092)	(0.075)	(0.064)	(0.082)	(0.107)	(0.095)
R^2	0.000	0.013	0.007	0.195	0.017	0.058
Obs	620	626	620	612	623	307
Panel (b) **Interaction effects**						
Treatment (=1)	-0.135	0.406	-0.543	0.026	-1.058*	-0.227
	(0.519)	(0.404)	(0.350)	(0.486)	(0.598)	(0.529)
messen_mean	-0.118	0.013	-0.075	-0.328***	-0.056	-0.022
	(0.117)	(0.094)	(0.077)	(0.102)	(0.140)	(0.116)
treat_messen	-0.019	-0.015	0.222**	0.391***	0.044	0.214
	(0.140)	(0.111)	(0.097)	(0.129)	(0.169)	(0.145)
pre_kwh	-0.196**	0.067	0.002	-0.006	-0.195	0.075
	(0.097)	(0.072)	(0.066)	(0.091)	(0.124)	(0.108)
treat_base	0.068	-0.049	-0.030	-0.055	0.293**	0.036
	(0.109)	(0.083)	(0.078)	(0.105)	(0.142)	(0.138)
Constant	2.774***	3.857***	3.847***	3.368***	3.618***	1.651***
	(0.435)	(0.351)	(0.275)	(0.391)	(0.498)	(0.420)
R^2	0.024	0.019	0.017	0.203	0.027	0.085
Obs	583	588	582	575	585	288

* p<0.10, ** p<0.05, *** p<0.01

Chapter 6

Impact and relevance

This chapter discusses the implications of the findings of this study. In a first step, average savings per household and year are calculated, followed by a quantification of the expected savings for a large-scale implementation and a quantitative comparison against existing electricity smart metering programs. Then six findings of particular interest for research and policy are discussed.

6.1 Quantification of direct savings per household

As outlined in section 5.5, the real-time feedback provided on the shower device yields a reduction of 23% or 0.38 kWh of thermal energy consumption per shower on average. This number, however assumes 100% boiler efficiency and zero losses in generation, distribution, and storage. Actual boiler efficiency depends upon boiler size, fuel type, and age; it averages 65% efficiency in Swiss households (Prognos AG (2013)). Again depending on the infrastructure and heating system, average distribution losses amount to 24-36% (Tschui and Stadelmann (2006)). Taking a rather conservative estimate for distribution losses of 20%, a cold water temperature of 12 degrees Celsius (Geberit (2011)), and the Swiss average for boiler efficiency, the actual mean reduction is 0.56 kWh per shower. Assuming one shower per person per day and the Swiss average household size of 2.2 persons (SFSO (2012)), and extrapolating the persistence of the effect from the observed two months to a full year, this results in energy savings of 443 kWh per household per year. We will address the assumption of effect persistence in chapter 7. In terms of carbon abatement, given the carbon intensity of water heating in Switzerland (see section 1), the device thus abates 94 kg of CO_2 per household and year.

In addition to heat energy, reduced water consumption per shower results in a yearly per-household reduction of 8,500 liters (again, under the same assumptions on shower frequency, household size, and persistence of the effect). Water conservation may less of a critical issue in a water-rich country like Switzerland than on an international level: According to UN estimates, half the world's population will be living in areas of high water stress by 2030 (United

Nations (2013)). In their article in *Nature*, Voeroesmarty et al. (2010) even find that nearly 80% of the world's population is exposed to high levels of threat to water security.

Based on the breakdown of fuel types outlined in section 1.3 and current utility prices, the reduced energy and water consumption results in average savings of CHF 110 per household and year[6]. As a consequence, the average payback period ranges between six months (bulk purchase) and nine months (individual purchase).

6.2 Savings at scale

6.2.1 Quantification of the impact of a large-scale deployment

Based on the figures calculated in 6.1, the cost per kWh saved is CHF 0.041 (dividing the initial cost by the kilowatt-hours saved over the course of a 3-year lifetime period and assuming persistent saving effects). Although this number does not reflect the additional benefits through water conservation, it compares favorably to the marginal generation costs of most energy sources: It is nearly half of the marginal cost of current electricity production of 0.074 CHF/kWh (Kost et al. (2012)). For electric water heaters, the resulting carbon abatement costs are thus negative, generating net savings of 0.033 CHF/kWh or 159 CHF/t of CO_2 abated. Probably as one of the first studies worldwide, these findings show that individual and immediate feedback on a particular action at the point of consumption is feasible at scale and at low (or even negative) costs.

Given the standardized threads for shower fittings, the device can technically be deployed in 97% of Swiss (and European) showers. Among the 697 study participants, a single household reported having issues with the tool-free DIY installation of the device. Another crucial aspect for scaling up the study to a large-scale program roll-out is external validity of the findings: To what extent the treatment effect observed can be generalized to the general population. This aspect will be addressed in section 6.3.3. Assuming external validity of the findings, a large-scale roll-out with deployment of the device in 10% of Swiss households would yield a reduction of 170 GWh of on-site thermal energy (25% of which are generated with electricity, see previous section). For comparison, total production of all Swiss wind power plants in 2012 was 85 GWh of electricity.

6.2.2 Comparison with electricity smart metering pilots

As outlined in section 2.1, randomized controlled trials that evaluate the effectiveness of feedback on residential energy consumption typically report reductions between 1-6%. In relative

[6]Underlying assumptions: fuel mix for water heating shown in figure 1.3 with the efficiency factors given in Prognos AG (2012), table 4.9; 20% distribution losses; and current utility prices as follows: water 1.45 CHF/m3, waste water 2.35 CHF/m3 (Statistisches Amt des Kantons Basel Landschaft (2013); Zanzi (2011)); oil 0.105 CHF/kWh (Hauseigentuemerverband Schweiz (2012, 2013)); natural gas: 0.10 CHF/kWh (Fischer (2013); Stadtwerk Winterthur (2014)); electric resistance heating 0.20 CHF/kWh (Elektrizitätskommission (2013)); wood/pellets: 0.07 CHF/kWh (Holzenergie Emmental (2014)), district heating 0.087 CHF/kWh (iwb (2012)).

terms, this is a much smaller change than the 23% reduction observed in this study. However, also in absolute terms, the device compares favorably to the kWh-reduction of other smart metering studies that provide users with electricity consumption feedback on the household level (using e.g., in-home displays, emails, web portals, smart phone applications): The 3.2%-reduction reported by Degen et al. (2013) translates into a reduced electricity consumption of 86 kWh, and the 4.5% reduction reported by Schleich et al. (2013) into a 154 kWh-reduction per year. For the sake of brevity, we will not go into the details of converting one form of energy into another, but highlight that at least for the 25% of Swiss households that have electric water heating, these figures can be directly compared with the 3- to 5.5-fold reduction of 443 kWh calculated above for the shower feedback device used in this study. Moreover, as outlined in chapter 1, carbon intensity of water heating in Switzerland is 212 g/kWh, compared to 122 g/kWh for household-level electricity.

As a bottom line, the study shows a) that a simple device applied to a very specific but relevant domain can achieve energy savings that are quantitatively important, even for a household's total energy consumption and b) that the impact of such a specific intervention can by far exceed the impact of interventions that aim at broader domains such as overall household electricity usage. Further research is needed to find out whether the magnitude of the savings are proprietary to the very specific shower context, and to what extent the findings can also be applied to other domains such as smart metering for electricity or consumption feedback in electric vehicles.

6.3 Research and policy implications

In addition to evaluating the main treatment effect of the intervention, this project also investigates how household-level and individual-level factors affect the effectiveness of the treatment. First of all, it allows to segment households into different group and to target specific population group, to further increase the effectiveness of such interventions. Second, this is relevant to evaluated the external validity of the findings. Third, an assessment of these factors also helps to identify the underlying mechanisms: This way, we can identify whether the intervention appealed to intrinsic motivations of the participants, or whether its effectiveness was due to the creation of negative feelings and peer pressure. We consider six findings of particular interest for policymakers and researchers alike: The stronger treatment effect on high users and younger people, the influence of the tendency to measure progress towards goals, the role of environmental attitudes, the different pieces of evidence that the device operates through a positive mechanism, a clear user preference for time reduction over flow reduction, and implications of the findings on profiling. All six of these findings will be briefly discussed in the following paragraphs.

6.3.1 Stronger treatment effect on high users and younger people

In line with other studies (Davis (2011); Allcott (2011)), the absolute reduction in energy used per shower is much higher for ex-ante high-usage households. However, the baseline data of this study suggest that younger consumers use much more energy and water in the shower, with 20-29 year-old participants using 2.3 times as much energy as participants over 64. To the best of our knowledge, this is the first study that provides data from a larger number of households on the higher resource consumption per shower of younger people. While we cannot determine with certainty whether this is an age effect or a cohort effect, a growing body of literature strongly supports the cohort effect explanation. Our data confirm what sociologists have described in the last years as a substantial change of norms and conventions for perceptions of comfort, cleanliness and convenience, leading to increasingly resource-intensive consumption patterns (Shove (2003)). However, while most of the existing literature focuses on increasing shower *frequency* over the past decades, our data suggest that there is an even bigger shift of the resource consumption *per shower* that even multiplies with shower frequency. Also, the existing literature that describes changes of similar patterns mainly focuses on qualitative aspects to understand this paradigm shift in the society, whereas this study analyzes field data to quantify the magnitude of this trend.

Aside from showering, this behavioral shift has also been reported for the use of space heating, air conditioning, and laundry quantities. This poses significant challenges to policy, as it could imply that energy efficiency gains achieved through technological progress and through policy interventions are not only thwarted by the different forms of rebound effects that are currently widely discussed: It may be the case that even more profound changes of lifestyle, norms, and paradigms additionally undermine these energy efficiency gains. In contrast to different rebound mechanisms, these changes do not appear to be related to financial factors. Neither in the existing literature on the change of shower behavior, nor in our dataset, shower behavior is correlated with income (neither baseline consumption, nor treatment effect). There is also no difference between households who pay a fixed price for water / water heating, and households utility bills are variable, based on their consumption.

This kind of transformation of sociotechnical regimes and collective conventions could have profound repercussions on future energy demand: Our dataset indicates that resource consumption in the shower has more than doubled within one generation; along similar lines, U.S. studies observe an increase in the amount of laundry per person to the threefold within 50 years (Biermeyer (2001)). Along similar lines, Shove (2003) reports a five-fold increase in the frequency of bathing, showering, and washing clothes over the last century. Given the magnitude of these changes within a few decades, these trends should be taken into account in projections of future energy demand. Instead of limiting the focus solely on the technological changes, it might be wise to also "pay attention to the transformation of these habits and the conventions associated with them" (Shove (2003)). Today, a growing body of literature has established that financially driven rebound effects can erode or even negate the technical potential of emissions reductions (e.g., Jenkins et al. (2011)). It might be the case that these are not the only mechanisms that can undermine the gains of improved energy efficiency. It is not clear to what extent changing lifestyle conventions, norms and paradigms also apply to other fields such as mobility, for instance.

The results are also interesting from another point of view: There is a widespread belief that young people are particularly concerned about the environment (Irvine (2012)): In attitudinal polls, they generally show a high level of concern for environmental issues (Partridge (2008)); in elections, young and especially first-time voters tend to be core supporters of green parties (Schlieben (2009)); they value sustainable products and companies that engage in these issues (Hewlett et al. (2009)). "Generation Green" enrolls in environmental studies in soaring numbers (Galbraith (2009)), is reported to *"plan to be more engaged than did youth 20 years ago"* (Salmond et al. (2009)) and surveys find that *"young people are leading the way in their attitudes to the environment"*. As a consequence, today's young generation is often considered as *"pivotal in leading the environmental movement forward"* (McKay (2010)). Yet the study results suggest that despite their good intentions and a higher degree of awareness, younger people may not live up to their ideals, using by far more resources for daily actions than older generations, as in the case of showering.

Regarding response to the intervention, however, there is also a more encouraging finding of the study: The results indicate that the device seems to have a bigger impact on young people. Yet this is actually confounded with their higher baseline use: The apparent stronger reaction of younger people is actually driven by their high baseline usage. Similar to the findings of the large-scale intervention on water usage carried out by Ferraro and Price (2013), the treatment effect is stronger on users with a high baseline usage. Nevertheless, the discrepancy in the reaction of different age groups is relevant from a practitioner's point of view. We will return to this point in section 6.3.6.

6.3.2 Impact of tendency to monitor progress towards goals

As section 5.6 shows, individuals with a higher tendency to measure their performance relative to self-set goals respond stronger to the treatment. In our sample, 74% of the participants indicated upfront that they frequently compare their performance against self-set goals and 28% reported that they frequently compare themselves against their peers' performance. Although we do not know to what extent these numbers are representative of the overall population, there seems to be a considerable fraction of individuals who in general tend to compare their performance with their own goals or with their peers' performance. For reference purposes, U.S. data from a representative survey show that 69% of Americans track at least one health indicator like weight, diet, exercise routine, or a symptom (Fox and Duggan (2013)).

This is also in line with numerous recent newspaper reports that observe a *"mainstreaming of the Quantified Self Movement"*, the tendency to track metrics about one's own life using technology (Bradley (2013); Snyder (2013); Hay (2013)). As Woodside (2011) stated in a Nature Climate Change feature article, *"people do want to use less energy, but forget or put it aside in the rush of the routine, or they don't know, because in industrialized countries society's systems aren't set up that way."* The shower feedback device thus addresses an intrinsic need of many individuals by providing them with the necessary information to reduce their consumption in the shower.

6.3.3 Impact of environmental attitudes and external validity

The results show that environmental attitudes (measured before the intervention) strongly correlate with lower baseline consumption. Also, after controlling for baseline consumption, environmental attitudes interact with the treatment effect. And yet, environmental attitudes do *not* influence the overall treatment effect.

These findings are relevant for several reasons: First of all, they show that the treatment effect is not driven by individuals with an extremely strong motivation to conserve energy. Participants of this study reduced their consumption independent of their environmental attitudes. This is fundamental, as studies like this one often face criticism with regards to a self-selection bias towards green consumers: It is argued that participants of such studies tend to have a more pro-environmental mindset than the general population. However, the sample recruited in fact scores even slightly lower on pro-environmental attitudes ($M_A = 3.48, N = 643, SD = 0.9$) than the nationally representative Swiss sample presented in Diekmann et al. (2008) ($M_D = 3.80, N = 3352, SD = 1.0$). Second, the results suggest that environmental attitudes do not affect the net treatment effect. Consequently, self-selection bias - at least with respect to the main source of concern for external validity, the green mindset dimension - is not an issue. More research is needed to determine whether this finding is only valid up to a certain threshold of minimum pro-environmental attitudes, whether it holds true also for other countries and for other user characteristics.

Nevertheless, after controlling for baseline consumption, the treatment effect is positively correlated with pro-environmental attitudes. An explanation might be that individuals with stronger pro-environmental attitudes might actually have cared more about the feedback and actually have paid more attention to it. In the final survey of this study, users with stronger pro-environmental attitudes reported paying more attention to the device and discussing their consumption more frequently within the household. However, they probably had already paid more attention to their energy and water consumption in the shower upfront, which is in line with their lower baseline usage. As users with less strong pro-environmental attitudes start out from a higher baseline use, they have more possibilities (and a higher margin) to reduce their consumption. Overall, the higher conservation potential of the ex-ante high users may simply have balanced out the stronger efforts of those individuals who (before and during the study) were more driven by their stronger pro-environmental attitudes.

6.3.4 Positive mechanism

The results do not support any evidence that the conservation effect might be driven by negative psychological pressure: First of all, two-person households and participants who had access to the additional information on the previous shower did not show a stronger reaction. This is a first indicator that peer pressure is not the main mechanism driving the effect. Second, individuals who are more subjective to psychological pressure in general (measured by the personality traits emotionality and conscientiousness) did not respond more to the treatment. In the case of conscientiousness, we even found the contrary: A stronger reaction from less conscientious users, which suggests that the device is particularly useful for individuals with a stronger need for behavioral cues to follow through on their intentions. This is in line with the inattentive choice model by Taubinsky (2013) and with previous research that states that while many people are motivated to reduced their environmental impact, they fail to do so in their daily lives (Woodside (2011)). Furthermore, we found that individuals with a stronger tendency to monitor their progress towards goals respond more to the feedback device. Moreover, the results show that after controlling for baseline consumption, environmental attitudes drive the treatment effect. All of these findings indicate that the device serves as a behavioral cue that makes information salient that is relevant to users' long-term preferences. That way, users can incorporate the information into the decision-making context at the moment when the behavior takes place.

Overall, none of the results supports evidence for negative psychological pressure. The findings on individuals' preference to monitor progress towards goals, on environmental attitudes, and on conscientiousness all indicate that the feedback device operates through positive mechanisms, helping individuals act in line with their preferences.

6.3.5 Clear user preference for time reduction over flow reduction

The study device does not prescribe in any way how users can or should reduce their energy consumption in the shower. In theory, they could reduce water temperature, shorten the shower duration, reduce the flow rate, or temporary turn off the water, e.g., while they are soaping. Yet the data show a clear user preference for simply shortening shower duration, whereas flow rate does not change significantly. This is interesting from a policy point of view: Many states increasingly limit the maximum flow rate of shower heads, hoping to reduce water and energy consumption that way. Yet this has stirred quite some controversy around the government dictating conservation (Power (2011)); the phenomenon of direct rebound is also discussed, as individuals might respond by extending their shower time. The study device lets users decide to what extent and how they want to change their shower behavior.

And given that choice, it appears that users prefer by far taking shorter showers over lower flow rates. While technical solutions and policy measures often privilege automation and standards, reducing the need (but also freedom) for active user decisions, this study is an example that behavioral interventions that actively involve the user can be very cost-effective and, at the same time, yield a high impact.

6.3.6 Implications for profiling

Chapter 5 identified several variables that significantly contribute to the variance in the treatment effect. This knowledge can be applied to further increase the treatment effect by profiling: Based on these variables, it is possible to define target groups for which the treatment effect can be expected to be considerably higher than for the average participant. Instead of administering a program to an arbitrary set of households, one could identify strategies to mainly target individuals or households with these characteristics, for instance by promoting the device or the program in media channels whose audience tend to exhibit these traits.

The results of this study show that the best predictor of the treatment effect is baseline consumption. Exclusively administering the intervention to to the 50% of the households with a baseline consumption above the median would raise the treatment effect by 74% (using the same cost calculation as in section 6.2.1). Restricting the program even further to those households with an above-average baseline consumption (39% of the households in this dataset) would literally double the treatment effect (+99%). This would reduce the cost per kWh saved by 41% in the case of the median split and by 48% in the case of the mean split. These numbers are in line with Allcott (2011) who found that restricting the *Opower* program to half the eligible population - namely to those with a baseline consumption above the median - would increase the treatment effect by 74% and the cost-effectiveness by 43% .

From a practical point of view, however, in the case of showering, it may not be as easy to identify households with high baseline consumption. Not only for policymakers and program designers is it difficult to know which households are on the high use end of the spectrum of energy and water consumption in the shower: User self-assessment prior to the provision of feedback is also poor, as section 5.3 and previous research show (Beal et al. (2013)). Given that baseline consumption is by far the best predictor for subsequent savings, and give the strong correlation between age and baseline consumption, age could be used as a good proxy to identify high users and, by extension, households with a average high conservation potential. In the current dataset, restricting the intervention to participants below the age of 40 would raise the treatment effect by 51% and reduce the cost per kWh conserved by 34%. Moreover, one should keep in mind that the dataset at hand hardly contains any teenagers as participants. Yet this age group is reported to take by far the most resource-intensive showers (Gram-Hanssen (2007); Mayer et al. (1999)). This could potentially make teenagers a particularly interesting target group for shower feedback. Further research is necessary to confirm this.

To summarize, profiling could considerably further raise the treatment and cost-effectiveness of the intervention; while baseline consumption is the best single predictor of households to be targeted, from an implementation point of view, age might serve as a more accessible and yet effective proxy to identify households with expected high savings.

Chapter 7

Study limitations and research outlook

This study has generated many insights and raised many questions, but also has its limitations.

Validity for other behaviors - First of all, the finding that feedback on a very specific behavior, appliance, or domain can yield much higher savings requires further attention. The main question in this context is to what extent the high savings effect is specific to the particular characteristics of showering, or whether similar results can be achieved by providing real-time feedback with a targeted focus right at the point of use e.g., on the use of air conditioning, on hot water use at the tap, on doing laundry, etc.

Self-selection bias - Second, one of the key questions in the evaluation of the study results is to what extent the saving effects might have been affected by self-selection of the participants. In their classification of real-time feedback trials on electricity consumption, McKerracher and Torriti (2013) reports a weighted mean treatment effect of 2.6% for studies with a representative sample, compared to 4.5% for larger studies with opt-in design (as this one). While these results are comparable in their order of magnitude, the difference is not negligible.

One of the key issues for opt-in recruitment studies is a potential self-selection bias for individuals with more pro-environmental attitudes, compared to the general population. Yet as outlined in 6.3.3, participants' environmental attitudes are similar and even slightly less strong than the representative Swiss sample analyzed in Diekmann et al. (2008). Moreover, as outlined in section 6.3.3, the study indicates that these interventions can work equally well for a broad audience, as less motivated users tend to start out from a higher level of baseline consumption. Nevertheless, caution is still warranted with the generalization of the results. More research is needed to evaluate whether the finding of equal net treatment effect regardless of environmental attitudes is really applicable to the general population, or if it only holds true for samples above a certain threshold of minimal interest in these topics. As Costa and Kahn (2010) showed, defiers of such interventions might even slightly increase their usage as a response. One could also argue that despite their lower average score on the environmental attitudes-scale (relative to the Swiss Environment Survey 2007), the sample of households in this study must indeed have a minimum of interest in these topics, otherwise they would probably not opted into the study in the first place. Second, a certain interest and level of engagement is necessary in order to install the device in the first place and to keep it in their shower. The key question might thus boil down to what percentage of the general population would be willing to install the device in the shower (and to keep it). For households where this is the case, our study results indicate that the device is equally effective independent of user's environmental attitudes. Therefore, it might be interesting to implement the study in an opt-out setting to assess treatment effect, influence of environmental attitudes, and adoption rate of the device among a sample with a lower potential for self-selection bias across all user characteristics.

Another aspect that one should keep in mind in the discussion of self-selection bias is that the study sample of this study was recruited among a group of ewz customers who had previously participated in the *ewz Studie Smart Metering* (see Degen et al. (2013)). In that prior study, the average treatment effect of smart meters with in-home displays was 3.2%; this is in line with other electricity smart metering studies (Schleich et al. (2013)), but far below the 23% reduction achieved in this chapter. This is another strong indicator that at least the magnitude of the effect size can hardly be explained with particular characteristics of the sample recruited. Moreover, while the potential for selection bias in field experiments should not be neglected, randomized controlled trials are still the most reliable method to evaluate the impact of energy conservation programs: As Allcott (2011) and Allcott and Mullainathan (2012) showed, non-experimental estimators perform dramatically worse than experimental estimators.

Effect persistence - Third, although the treatment effect was sustained over the course of the study period, further research is necessary to determine the treatment effect over a longer period. On the one hand, recent literature on habit formation suggests that it takes on average two months for changed daily actions to become an automated process that no longer requires self-control (Lally et al. (2010)). That means that for the majority of the participants of this study, the new, less resource intense showering process should already have become a habit. Previous studies that investigated the persistence of effects had mixed results (see section 2.1). While some studies found evidence for a decay of the effect (Fielding et al. (2013)), others provide evidence for sustained savings throughout study periods of 7 months to two years (Ayres et al. (2009); Raw and Ross (2011)). A recent study by Allcott and Rogers (2014) shows that savings even from sporadic and paper-based feedback are sustained over much longer periods of time than previously accounted for. As outlined in 2.1, some researchers debate the topic of data "push versus pull" in this context (Boyd (2014); Foster and Mazur Stommen (2012); Froehlich et al. (2010). They argue that the savings are much more likely to be persistent for data-push systems, i.e., systems that don't require an additional layer of regular active user interaction to access the feedback information (e.g., login to a web portal, battery replacement, active screen activation by the user). Due to its energy supply from the water flow and its automatic screen activation, the *amphiro a1* shower meter represents such a data push system. A long-term study with additional *amphiro a1* users is still ongoing in order to evaluate the effect of the shower meters in the long run. Moreover, it would be interesting to evaluate whether participants slip back to their original shower habits when they don't receive real-time feedback any more (as during the baseline period).

Age and baseline consumption - Fourth, the strong (negative) correlation of baseline consumption with age is highly relevant for other domains. Given the magnitude of the difference in per-shower consumption within a single generation and the fact that a major transformation of habits has been observed for other daily routines, these findings clearly deserve more attention. More quantitative research in particular is needed to investigate this question with a careful research design, assuring that cohort effects can clearly be distinguished from age effects. In this context, it would also be interesting to include children and teenagers in a follow-up study, as they are reported to have the highest use of energy and water in the shower.

Positive Mechanism - Fifth, the results indicate that the treatment effect has been generated mainly by positive mechanisms, not peer pressure. The treatment effect was not stronger among individuals who are more susceptible to pressure (measured by the personality trait emotionality), nor on more conscientious individuals: On the contrary, less conscientious participants responded more strongly to the treatment. Yet the question arises whether the finding of positive mechanisms also applies to studies where peer comparison information is more prominent, e.g., in the *Opower* or *Ben Energy* home energy reports. As far as the additional information on the previous shower is concerned, we cannot exclude the possibility that this feature might have failed to produce any significant difference in the outcome due to shortcomings in the implementation. Maybe participants simply did not understand that feature or did not pay enough attention to the part of the display that communicated the additional information.

Goal setting - Sixth, the study also revealed that participants who set themselves a savings goal reduced their consumption significantly more than participants who did not. While this might be confounded with their level of interest of interaction with the device, it would be worthwhile to further explore the role of goals (self-set and provided externally) and to manipulate the level of target values in the device software.

Hawthorne Effect - Finally, the fact that control group participants slightly increased their per-shower consumption over time is also an interesting finding by itself. As discussed in section 4.6, the slight increase of the control group's consumption cannot be explained by changing weather conditions, as temperatures did not show any particular trend during the study period. A potential explanation could be what is referred to as *Hawthorne Effect* in the literature: That the feeling of being observed or merely participating in an experiment affects individuals' behavior. McKerracher and Torriti (2013) report that the effect is more likely to occur in studies with small sample sizes due to a higher level of interaction between study administrators and participants, increasing participants' awareness of being observed. Schwartz et al. (2013) for instance reports that households who received weekly postcards informing them that they were involved in a study on electricity usage reduced their consumption by 2.7%. Literature suggests that after a while, participants get used to or even forget that they are being monitored, reducing the influence of the Hawthorne Effect over time (Martinussen and Hunter (2009)).

In a similar vein, participants in the control group seemed to acquire a better sense for their water consumption per shower over the course of the study (see section 5.3). This indicates that the installation of the device has increased their awareness for their consumption, even without feedback information. The mere installation of the device and the awareness for participating in a study might have biased the consumption of the control group towards higher awareness and lower consumption. These aspects highlight the importance of using a difference-in-differences strategy to capture the causal effects unconfounded with these kind of time trends.

Overall, our work expands on the existing literature in several respects, answering and raising new questions at the same time. Moreover and equally important, the work shows a concrete way how energy can be saved and how emissions can be reduced at very large scale.

Appendices

Appendix A

Supplementary Tables

A.1 Treatment interaction effect with baseline usage

Table A.1: DiD estimates for energy: interaction with baseline consumption ("pre-kwh")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
RT info (=1)	-0.394***	-0.394***	-0.255**	-0.257**	-0.376***	-0.376***
	(0.060)	(0.060)	(0.119)	(0.118)	(0.054)	(0.054)
RT & past info (=1)	-0.369***	-0.368***	-0.151	-0.152	-0.400***	-0.399***
	(0.076)	(0.076)	(0.099)	(0.098)	(0.050)	(0.050)
RT × Trait	-0.301***	-0.292***	-0.229**	-0.213*	-0.400***	-0.394***
	(0.034)	(0.040)	(0.109)	(0.115)	(0.108)	(0.107)
RT & past × pre-kWh	-0.296***	-0.285***	-0.167***	-0.151**	-0.274***	-0.268***
	(0.100)	(0.109)	(0.053)	(0.058)	(0.049)	(0.056)
Trend × pre-kWh		-0.000		-0.000		-0.000
		(0.001)		(0.001)		(0.000)
Constant	1.637***	1.637***	1.543***	1.543***	1.611***	1.611***
	(0.065)	(0.064)	(0.095)	(0.095)	(0.059)	(0.059)
R^2	0.583	0.583	0.372	0.372	0.374	0.374
Obs	13298	13298	6711	6711	25027	25027
Clusters	255	255	102	102	269	269

* p<0.10, ** p<0.05, *** p<0.01

A.2 Treatment interaction effect with gender

Table A.2: DiD estimates for energy: interaction with fraction of females in household

	Single HH		Mixed HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
RT info (=1)	-0.446***	-0.464***	-0.413***	-0.427***	-0.301***	-0.251***
	(0.079)	(0.082)	(0.143)	(0.147)	(0.094)	(0.095)
RT & past info (=1)	-0.470***	-0.489***	-0.110	-0.122	-0.510**	-0.463**
	(0.139)	(0.142)	(0.189)	(0.189)	(0.211)	(0.215)
RT × female	0.193**	0.231**	0.486**	0.524**	-0.180	-0.290
	(0.094)	(0.099)	(0.241)	(0.257)	(0.178)	(0.179)
RT & past × female	0.163	0.204	-0.069	-0.033	0.207	0.102
	(0.150)	(0.153)	(0.287)	(0.285)	(0.423)	(0.436)
Constant	1.635***	1.636***	1.545***	1.546***	1.611***	1.609***
	(0.065)	(0.066)	(0.096)	(0.096)	(0.061)	(0.061)
R^2	0.578	0.578	0.371	0.371	0.370	0.371
Obs	13298	13298	6711	6711	25027	25027

* p<0.10, ** p<0.05, *** p<0.01

A.3 Treatment interaction effect with age

Table A.3: DiD estimates for energy: interaction with age

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
RT info (=1)	-0.426***	-0.422***	-0.201*	-0.201*	-0.380***	-0.379***
	(0.071)	(0.070)	(0.117)	(0.117)	(0.068)	(0.068)
RT & past info (=1)	-0.421***	-0.416***	-0.145	-0.145	-0.383***	-0.383***
	(0.096)	(0.096)	(0.111)	(0.111)	(0.055)	(0.055)
RT× age	0.165***	0.200***	-0.033	-0.036	0.049	0.045
	(0.046)	(0.048)	(0.070)	(0.074)	(0.050)	(0.050)
RT & past × age	0.039	0.075**	0.011	0.008	0.156***	0.153***
	(0.035)	(0.036)	(0.047)	(0.050)	(0.029)	(0.031)
Constant	1.618***	1.618***	1.488***	1.488***	1.625***	1.625***
	(0.067)	(0.067)	(0.092)	(0.092)	(0.062)	(0.062)
R^2	0.580	0.580	0.378	0.378	0.372	0.372
Obs	12802	12802	6363	6363	23925	23925
Clusters	246	246	97	97	257	257

* p<0.10, ** p<0.05, *** p<0.01

A.4 Treatment interaction effect with self-estimated savings potential

Table A.4: DiD estimates for energy: interaction with self-estimated savings potential ("sav-pot")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
	none	none	none	none	none	none
RT info (=1)	-0.381***	-0.387***	-0.220*	-0.221*	-0.389***	-0.387***
	(0.069)	(0.070)	(0.114)	(0.114)	(0.065)	(0.065)
RT & past info (=1)	-0.396***	-0.403***	-0.152	-0.151	-0.411***	-0.408***
	(0.092)	(0.094)	(0.101)	(0.101)	(0.059)	(0.059)
RT $\times$ sav-pot	-0.076***	-0.087***	-0.028	-0.035	-0.050*	-0.056*
	(0.029)	(0.030)	(0.073)	(0.075)	(0.030)	(0.031)
RT+ past $\times$ sav-pot	-0.098*	-0.110**	0.040	0.031	-0.032	-0.040
	(0.053)	(0.054)	(0.035)	(0.041)	(0.033)	(0.033)
Constant	1.636***	1.637***	1.542***	1.542***	1.612***	1.612***
	(0.066)	(0.066)	(0.097)	(0.097)	(0.061)	(0.061)
R^2	0.578	0.578	0.371	0.371	0.370	0.371
Obs	13298	13298	6711	6711	25027	25027
Clusters	255	255	102	102	269	269

* p<0.10, ** p<0.05, *** p<0.01

A.5 Treatment interaction effect with tendency to quantify oneself

Table A.5: DiD estimates for energy: tendency to monitor progress towards goals ("QS" for "Quantified Self")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
	none	none	none	none	none	none
RT info (=1)	-0.354***	-0.352***	-0.185	-0.184	-0.387***	-0.387***
	(0.072)	(0.072)	(0.138)	(0.138)	(0.070)	(0.070)
RT & past info (=1)	-0.392***	-0.389***	-0.151	-0.150	-0.420***	-0.420***
	(0.098)	(0.098)	(0.103)	(0.104)	(0.059)	(0.059)
RT × QS	-0.081	-0.096	0.079	0.104	-0.053	-0.051
	(0.068)	(0.068)	(0.166)	(0.167)	(0.060)	(0.061)
RT & past × QS	-0.117	-0.139*	0.153**	0.174**	-0.171***	-0.169***
	(0.078)	(0.076)	(0.074)	(0.081)	(0.056)	(0.060)
Constant	1.619***	1.619***	1.475***	1.475***	1.642***	1.642***
	(0.070)	(0.070)	(0.088)	(0.088)	(0.063)	(0.063)
R^2	0.566	0.566	0.384	0.384	0.369	0.369
Obs	12441	12441	6258	6258	23726	23726
Clusters	238	238	97	97	253	253

* p<0.10, ** p<0.05, *** p<0.01

A.6 Treatment interaction effect with tendency to compare oneself with the performance of others

Table A.6: DiD estimates for energy: interaction with tendency to compare with others ("comp")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
	none	none	none	none	none	none
RT info (=1)	-0.350***	-0.350***	-0.143	-0.136	-0.375***	-0.376***
	(0.072)	(0.072)	(0.123)	(0.124)	(0.068)	(0.068)
RT & past info (=1)	-0.394***	-0.393***	-0.142	-0.135	-0.414***	-0.414***
	(0.098)	(0.098)	(0.102)	(0.105)	(0.059)	(0.059)
RT × comp	-0.004	-0.015	-0.036	-0.011	-0.072	-0.077
	(0.044)	(0.046)	(0.081)	(0.082)	(0.049)	(0.049)
RT & past× comp	-0.090	-0.102*	0.148***	0.169***	-0.093**	-0.098**
	(0.055)	(0.055)	(0.054)	(0.055)	(0.042)	(0.045)
Constant	1.620***	1.620***	1.476***	1.476***	1.634***	1.634***
	(0.070)	(0.070)	(0.086)	(0.086)	(0.063)	(0.063)
R^2	0.566	0.566	0.385	0.385	0.371	0.371
Obs	12441	12441	6258	6258	23707	23707
Clusters	238	238	97	97	253	253

* p<0.10, ** p<0.05, *** p<0.01

A.7 Treatment interaction effect with happiness

Table A.7: DiD estimates for energy: interaction with happiness ("happy")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
	none	none	none	none	none	none
RT info (=1)	-0.357***	-0.359***	-0.207*	-0.207*	-0.386***	-0.382***
	(0.069)	(0.069)	(0.108)	(0.108)	(0.067)	(0.067)
RT & past info (=1)	-0.437***	-0.438***	-0.141	-0.141	-0.398***	-0.392***
	(0.117)	(0.117)	(0.103)	(0.103)	(0.058)	(0.058)
RT $\times$ happy	-0.119	-0.164**	0.058	0.054	-0.043	-0.077
	(0.076)	(0.082)	(0.156)	(0.164)	(0.057)	(0.060)
RT & past $\times$ happy	-0.128	-0.179	-0.053	-0.057	-0.160**	-0.198***
	(0.123)	(0.129)	(0.098)	(0.106)	(0.069)	(0.072)
Constant	1.631***	1.631***	1.482***	1.482***	1.613***	1.613***
	(0.066)	(0.066)	(0.089)	(0.089)	(0.062)	(0.062)
R^2	0.578	0.578	0.380	0.380	0.369	0.370
Obs	13047	13047	6508	6508	24665	24665
Clusters	250	250	100	100	265	265

* p<0.10, ** p<0.05, *** p<0.01

A.8 Treatment interaction effect with HEXACO personality trait honesty

Table A.8: DiD estimates for energy: interaction with HEXACO trait honesty

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
RT info (=1)	-0.297***	-0.297***	-0.201*	-0.200*	-0.366***	-0.366***
	(0.074)	(0.073)	(0.106)	(0.106)	(0.075)	(0.075)
RT & past info (=1)	-0.416***	-0.413***	-0.094	-0.093	-0.356***	-0.357***
	(0.109)	(0.108)	(0.106)	(0.106)	(0.063)	(0.063)
RT$\times$ honesty	-0.073	-0.116	0.158	0.208	0.016	0.033
	(0.073)	(0.078)	(0.188)	(0.192)	(0.094)	(0.093)
RT & past $\times$ honesty	0.273	0.236	-0.080	-0.020	0.066	0.083
	(0.199)	(0.199)	(0.077)	(0.081)	(0.087)	(0.091)
Constant	1.460***	1.460***	1.510***	1.510***	1.577***	1.577***
	(0.067)	(0.067)	(0.088)	(0.087)	(0.063)	(0.063)
R^2	0.561	0.561	0.384	0.384	0.361	0.361
Obs	10596	10596	5255	5255	20053	20053
Clusters	203	203	82	82	218	218

* p<0.10, ** p<0.05, *** p<0.01

A.9 Treatment interaction effect with HEXACO personality trait emotionality

Table A.9: DiD estimates for energy: interaction with HEXACO trait emotionality ("emo")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
RT info (=1)	-0.289***	-0.289***	-0.209*	-0.208*	-0.363***	-0.364***
	(0.073)	(0.073)	(0.113)	(0.113)	(0.075)	(0.075)
RT & past info (=1)	-0.426***	-0.426***	-0.087	-0.086	-0.351***	-0.352***
	(0.123)	(0.123)	(0.107)	(0.107)	(0.062)	(0.062)
RT $\times$ emo	0.151**	0.147**	0.253	0.239	0.045	0.055
	(0.064)	(0.066)	(0.239)	(0.250)	(0.091)	(0.094)
RT & past $\times$ emo	0.053	0.048	-0.015	-0.032	-0.018	-0.009
	(0.220)	(0.223)	(0.087)	(0.091)	(0.088)	(0.090)
Constant	1.462***	1.462***	1.505***	1.505***	1.577***	1.577***
	(0.068)	(0.068)	(0.087)	(0.087)	(0.063)	(0.063)
R^2	0.561	0.561	0.384	0.384	0.361	0.361
Obs	10596	10596	5255	5255	20053	20053
Clusters	203	203	82	82	218	218

* p<0.10, ** p<0.05, *** p<0.01

A.10 Treatment interaction effect with HEXACO personality trait extroversion

Table A.10: DiD estimates for energy: interaction with HEXACO trait extraversion ("extra")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
RT info (=1)	-0.301***	-0.299***	-0.212*	-0.204*	-0.369***	-0.369***
	(0.079)	(0.079)	(0.112)	(0.112)	(0.076)	(0.076)
RT & past info (=1)	-0.432***	-0.429***	-0.089	-0.078	-0.360***	-0.361***
	(0.129)	(0.129)	(0.108)	(0.109)	(0.062)	(0.062)
RT × extro	-0.060	-0.032	-0.193**	-0.263**	0.122	0.139
	(0.129)	(0.133)	(0.094)	(0.112)	(0.114)	(0.114)
RT & past × extro	-0.042	-0.011	0.032	-0.041	0.128	0.147
	(0.113)	(0.117)	(0.092)	(0.108)	(0.093)	(0.097)
Constant	1.459***	1.459***	1.511***	1.510***	1.575***	1.576***
	(0.068)	(0.068)	(0.088)	(0.087)	(0.063)	(0.063)
R^2	0.561	0.561	0.384	0.384	0.361	0.361
Obs	10596	10596	5255	5255	20053	20053
Clusters	203	203	82	82	218	218

* p<0.10, ** p<0.05, *** p<0.01

A.11 Treatment interaction effect with HEXACO personality trait agreeableness

Table A.11: DiD estimates for energy: interaction with HEXACO trait agreeableness ("agree")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
RT info (=1)	-0.293***	-0.293***	-0.181*	-0.195*	-0.369***	-0.369***
	(0.074)	(0.073)	(0.104)	(0.107)	(0.077)	(0.077)
RT & past info (=1)	-0.429***	-0.429***	-0.094	-0.107	-0.354***	-0.353***
	(0.113)	(0.113)	(0.104)	(0.107)	(0.062)	(0.062)
RT× agree	0.026	0.039	0.249	0.311	0.109	0.116
	(0.112)	(0.115)	(0.227)	(0.235)	(0.109)	(0.110)
RT & past × agree	-0.191*	-0.176	-0.267**	-0.191*	0.109	0.118
	(0.104)	(0.109)	(0.104)	(0.108)	(0.098)	(0.099)
Constant	1.451***	1.451***	1.511***	1.510***	1.575***	1.575***
	(0.067)	(0.067)	(0.089)	(0.089)	(0.063)	(0.063)
R^2	0.561	0.561	0.384	0.384	0.361	0.361
Obs	10670	10670	5255	5255	20053	20053
Clusters	204	204	82	82	218	218

* p<0.10, ** p<0.05, *** p<0.01

A.12 Treatment interaction effect with HEXACO personality trait conscientiousness

Table A.12: DiD estimates for energy: interaction with HEXACO trait conscientiousness ("consc")

	1-person HH		Unstable HH		2-person HH	
	w/o trend	with trend	w/o trend	with trend	w/o trend	with trend
RT info (=1)	-0.292***	-0.291***	-0.216*	-0.213*	-0.356***	-0.358***
	(0.073)	(0.074)	(0.113)	(0.114)	(0.069)	(0.070)
RT & past info (=1)	-0.407***	-0.406***	-0.091	-0.090	-0.351***	-0.352***
	(0.109)	(0.109)	(0.105)	(0.105)	(0.062)	(0.062)
RT × consc	-0.046	-0.049	-0.161	-0.193	0.453**	0.440**
	(0.075)	(0.078)	(0.116)	(0.136)	(0.178)	(0.174)
RT & past× consc	0.139	0.134	0.080	0.047	0.113	0.096
	(0.174)	(0.182)	(0.116)	(0.113)	(0.101)	(0.106)
Constant	1.453***	1.453***	1.508***	1.508***	1.573***	1.574***
	(0.067)	(0.067)	(0.088)	(0.088)	(0.062)	(0.062)
R^2	0.560	0.560	0.384	0.384	0.362	0.362
Obs	10670	10670	5255	5255	20053	20053
Clusters	204	204	82	82	218	218

* p<0.10, ** p<0.05, *** p<0.01

Appendix B

Bibliography

Alford, J. R., Hatemi, P. K., Hibbing, J. R., Martin, N. G., and Eaves, L. J. (2011). The politics of mate choice. *The Journal of Politics*, 73(02):362–379.

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

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

Allcott, H. and Mullainathan, S. (2012). External validity and partner selection bias. Working Paper 18373, National Bureau of Economic Research.

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

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press, Princeton.

Arroyo, E., Bonanni, L., and Selker, T. (2005). Waterbot: Exploring feedback and persuasive techniques at the sink. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 631–639, New York, NY, USA. ACM.

ASEW (2008). Warmes Wasser - Energieeffizient erzeugen und rationell verwenden. Technical report, Erlanger Stadtwerke, Erlangen. Broschürenreihe: Spar Energie – wir zeigen wie. <http://www.stwab.de/html/Service/Energie-sparen/Broschueren/Broschueren/Warmes_Wasser_AVG.PDF>.

Ayres, I., Raseman, S., and Shih, A. (2009). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. Working Paper 15386, National Bureau of Economic Research.

BAFU (2011). CO2-Emissionsfaktoren des Schweizerischen Treibhausgasinventars. Klimapolitik: Fragen und Antworten. <http://www.bafu.admin.ch/klima/09608/index.html?lang =de#sprungmarke0_20>.

BAFU (2013). Charts and evaluations: Households and consumption. Technical report, Bundesamt für Umwelt. <http://www.bafu.admin.ch/umwelt/10822/12896/index.html?lang=en>.

BDEW (2010). Energiedaten - Endenergieverbrauch in Deutschland. Technical report, Bundesverband der Energie- und Wasserwirtschaft e.V. <http://www.bdew.de/internet.nsf/id/DE_Energiedaten>.

Beal, C. D., Stewart, R. A., and Fielding, K. (2013). A novel mixed method smart metering approach to reconciling differences between perceived and actual residential end use water consumption. *Journal of Cleaner Production*, 60:116–128.

Biermeyer, P. (2001). Coming changes in the U.S. clothes washer market. Unpublished, Lawrence Berkeley National Laboratory, Berkeley.

Boyd, R. (2014). Visible metering - A literature review. Energy transition report, Edmonton. <http://www.edmonton.ca/environmental/programs/energy-transition.aspx>.

Bradley, J. (2013). A pill that tracks your health? The reality of the 'Quantified self' movement. Huffington Post Blog. 2013-09-18. <http://www.huffingtonpost.com/joseph-bradley/cisco-quantified-self-movement_b_3907545.html>.

Butler, D. and Memon, F. A. (2006). *Water demand management*. IWA Publishing, London, UK.

Campbell, H. E., Johnson, R. M., and Larson, E. H. (2004). Prices, devices, people, or rules: The relative effectiveness of policy instruments in water conservation. *Review of Policy Research*, 21(5):637–662.

Carabias-Huetter, V. (2013). The urgency of energy conservation: Required behaviour and social norm change for demand-side management. Technical report, SFOE (BFE), Nisyros, Greece.

Carragher, B. J., Stewart, R. A., and Beal, C. D. (2012). Quantifying the influence of residential water appliance efficiency on average day diurnal demand patterns at an end use level: A precursor to optimised water service infrastructure planning. *Resources, Conservation and Recycling*, 62:81–90.

Corral-Verdugo, V., Bechtel, R. B., and Fraijo-Sing, B. (2003). Environmental beliefs and water conservation: An empirical study. *Journal of Environmental Psychology*, 23(3):247–257.

Costa, D. L. and Kahn, M. E. (2010). Energy conservation. Working Paper 15939, National Bureau of Economic Research.

Darby, S. (2006). The effectiveness of feedback on energy consumption: A review for DEFRA of the literature on metering, billing and direct displays. Technical report.

Darby, S. (2012). Metering: EU policy and implications for fuel poor households. *Energy Policy*, 49:98–106.

Davis, M. (2011). Behavior and energy savings. Technical report, Environmental Defense Fund. <http://opower.com/uploads/library/file/5/edf_behavior_and_energysavings.pdf>.

De Oliver, M. (1999). Attitudes and inaction: A case study of the manifest demographics of urban water conservation. *Environment and Behavior*, 31(3):372–394.

Degen, K., Efferson, C., Frei, F., Götte, L., and Lalive, R. (2013). Smart Metering, Beratung oder sozialer Vergleich: Was beeinflusst den Elektrizitätsverbrauch? Final report to the Swiss Federal Office of Energy.

Diekmann, A., Meyer, R., Mühlemann, C., and Diem, A. (2008). Schweizer Umweltsurvey 2007: Analysen und Ergebnisse. Bericht für das Bundesamt für Statistik (BFS) und das Bundesamt für Umwelt (BAFU), Professur für Soziologie, ETH Zürich, Zürich.

Dolnicar, S., Hurlimann, A., and Gruen, B. (2012). Water conservation behavior in Australia. *Journal of environmental management*, 105:44–52.

Ehrhardt Martinez, K., Donnelly, K. A., and Laitner, S. (2010). Advanced metering initiatives and residential feedback programs: A meta-review for household electricity-saving opportunities. ACEEE Research Report E105.

eia (2013). Heating and cooling no longer majority of U.S. home energy use. Technical report. <http://www.eia.gov/todayinenergy/detail.cfm?id=10271>.

Elektrizitätskommission, E. (2013). Strompreise 2014 - Im Durchschnitt leicht steigende Tarife für Haushalte, keine Veränderung für mittlere Betriebe. Technical report. <http://www.admin.ch/aktuell/00089/index.html?lang=de&msg-id=50162>.

EPRI (2008). Utility business models for distributed energy resources in Massachusetts. *EPRI and Massachusetts Technology Collaborative Renewable Energy Trust Report*.

EPRI (2009). Residential electricity use feedback: A research synthesis and economic framework. Technical report.

Faruqui, A., Sergici, S., and Sharif, A. (2010). The impact of informational feedback on energy consumption: A survey of the experimental evidence. *Energy*, 35(4):1598–1608.

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.

Fetz, A. (2013). Swiss energy strategy and research projects concerning behavior change. IEA DSM Task 24 Workshop.

Fielding, K. S., Spinks, A., Russell, S., McCrea, R., Stewart, R., and Gardner, J. (2013). An experimental test of voluntary strategies to promote urban water demand management. *Journal of Environmental Management*, 114:343–351.

Fink, P.-C. (2012). Wasserversorgung: Schluss mit dem Wassersparen! *Die Zeit*. 2012-04-03. <http://www.zeit.de/2012/14/Wasserversorgung>.

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

Fischer, M. (2013). Privatkunden werden benachteiligt. *K-Tipp*, 2013(09). K-Tipp 2013(08). <www.ktipp.ch/artikel/d/privatkunden-werden-benachteiligt/>.

Foster, B. and Mazur Stommen, S. (2012). Results from recent real-time feedback studies. Technical report, ACEEE.

Fox, S. and Duggan, M. (2013). Tracking for health. Technical report, Pew Research Center's Internet & American Life Project, Washington, DC, USA.

Frischknecht, R., Itten, R., and Flury, K. (2012). Treibhausgas-Emissionen der Schweizer Strommixe. Bericht für das Bundesamt für Umwelt, BAFU, Uster.

Froehlich, J., Findlater, L., and Landay, J. (2010). The design of eco-feedback technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, page 1999–2008, New York, NY, USA. ACM.

Froes Lima, C. A. and Portillo Navas, J. R. (2012). Smart metering and systems to support a conscious use of water and electricity. *Energy*, 45(1):528–540.

Galbraith, K. (2009). Environmental studies enrollment soars. The New York Times Green Blog. 2009-02-24. <http://green.blogs.nytimes.com/2009/02/24/environmental-studies-enrollments-soar/>.

Gardner, G. T. and Stern, P. C. (2008). The short list: The most effective actions U.S. households can take to curb climate change. *Environment*, 50:12–23.

Geberit (2011). Planungshandbuch Sanitär - Trinkwasserhygiene. Technical report.

Gilg, A. and Barr, S. (2006). Behavioural attitudes towards water saving? Evidence from a study of environmental actions. *Ecological Economics*, 57(3):400–414.

Gram-Hanssen, K. (2007). Teenage consumption of cleanliness: How to make it sustainable. *Sustainability: Science, Practice, & Policy*, 3(2):15–23.

Hansen, P. G. and Jespersen, A. M. (2013). Nudge and the manipulation of choice. *Eur. J. Risk Reg*, 1:3–28.

Hauseigentuemerverband Schweiz (2012). HEV Schweiz Heizölpreise 2012. <http://www.hev-schweiz.ch/ vermieten-verwalten/heizoelpreise/heizoelpreise-2012/>.

Hauseigentuemerverband Schweiz (2013). HEV Schweiz Heizölpreise 2013. <http://www.hev-schweiz.ch/ vermieten-verwalten/heizoelpreise/heizoelpreise-2013/>.

Hay, T. (2013). The rise of the 'quantified self' in health care. The Wall Street Journal. Venture Capital Dispatch. <http://blogs.wsj.com/venturecapital/2013/08/13/the-rise-of-the-quantified-self-in-health-care/>.

Herrington, P. (1996). *Climate Change and the Demand for Water*. H.M. Stationery Office.

Hewlett, S. A., Sherbin, L., and Sumberg, K. (2009). How gen y & boomers will reshape your agenda. Harvard Business Review. 2009-07. <http://hbr.org/2009/07/how-gen-y-boomers-will-reshape-your-agenda/ar/1>.

Holzenergie Emmental (2014). Energieholzpreise 2013/2014. <http://www.holzenergie-emmental.ch/pdf/energieholzpreise.pdf>.

Houde, S., Todd, A., Sudarshan, A., Flora, J. A., and Armel, K. C. (2013). Real-time feedback and electricity consumption: A field experiment assessing the potential for savings and persistence. *The Energy Journal*, Volume 34(1).

Hurlimann, A., Dolnicar, S., and Meyer, P. (2009). Understanding behaviour to inform water supply management in developed nations: A review of literature, conceptual model and research agenda. *Journal of Environmental Management*, 91(1):47–56.

Inman, D. and Jeffrey, P. (2006). A review of residential water conservation tool performance and influences on implementation effectiveness. *Urban Water Journal*, 3(3):127–143.

Irvine, M. (2012). Young Americans less interested in the environment than previous generations. *The Washington Post*.

iwb (2012). Energie- und Wassertarife: Erdgastarife. <http://www.iwb.ch/de/privatkunden/kundenservice/tarife_preise/>.

Jenkins, J., Nordhaus, T., and Shellenberger, M. (2011). Energy emergence: Rebound and backfire as emergent phenomena. <http://thebreakthrough.org/blog/Energy_Emergence.pdf>.

Jorgensen, B., Graymore, M., and O'Toole, K. (2009). Household water use behavior: An integrated model. *Journal of Environmental Management*, 91(1):227–236.

Kappel, K. and Grechenig, T. (2009). "show-me": Water consumption at a glance to promote water conservation in the shower. In *Proceedings of the 4th International Conference on Persuasive Technology*, Persuasive '09, page 26:1–26:6, New York, NY, USA. ACM.

Kost, C., Schlegl, T., Thomsen, J., Nold, S., and Mayer, J. (2012). Studie Stromgestehungskosten Erneuerbare Energien. Technical report, Fraunhofer-Institut für Solare Energiesysteme ISE, Freiburg.

Kurz, T., Donaghue, N., and Walker, I. (2005). Utilizing a social-ecological framework to promote water and energy conservation: A field experiment. *Journal of Applied Social Psychology*, 35(6):1281–1300.

Kuznetsov, S. and Paulos, E. (2010). UpStream: Motivating water conservation with low-cost water flow sensing and persuasive displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1851–1860, New York, NY, USA. ACM.

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., and Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, 40(6):998–1009.

Laschke, M., Hassenzahl, M., Diefenbach, S., and Tippkaemper, M. (2011). With a little help from a friend: A shower calendar to save water. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, pages 633–646, Vancouver, BC, Canada. ACM.

Lee, M., Tansel, B., and Balbin, M. (2011). Influence of residential water use efficiency measures on household water demand: A four year longitudinal study. *Resources, Conservation and Recycling*, 56(1):1–6.

Makki, A. A., Stewart, R. A., Panuwatwanich, K., and Beal, C. (2013). Revealing the determinants of shower water end use consumption: Enabling better targeted urban water conservation strategies. *Journal of Cleaner Production*, 60:129–146.

Martinussen, M. and Hunter, D. R. (2009). *Aviation Psychology and Human Factors*. CRC Press.

Mayer, P. W., DeOreo, W. B., and AWWA Research Foundation (1999). *Residential end uses of water*. AWWA Research Foundation and American Water Works Association, Denver, CO.

McKay, L. (2010). Generation Green. *CRM Magazine*, (2010-04).

McKerracher, C. and Torriti, J. (2013). Energy consumption feedback in perspective: Integrating Australian data to meta-analyses on in-home displays. *Energy Efficiency*, 6(2):387–405.

Mitchell, D. L., Cubed, M., and Chesnutt, T. W. (2013). Evaluation of east bay municipal utility district's pilot of WaterSmart home water reports. Technical report.

Nieswiadomy, M. L. (1992). Estimating urban residential water demand: Effects of price structure, conservation, and education. *Water Resources Research*, 28(3):609–615.

Partridge, E. (2008). From ambivalence to activism. *Youth Studies Australia*, 27(2).

Power, S. (2011). A water fight over luxury showers. *The Wall Street Journal*. 2011-07-21. <http://online.wsj.com/news/articles/SB10001424052748704913304575371462611463490>.

Prognos AG (2012). Ex-Post-Analyse des Energieverbrauchs der schweizerischen Haushalte 2000 bis 2011 nach Bestimmungsfaktoren und Verwendungszwecken. Technical report, Prognos. <http://www.bfe.admin.ch/themen/00526/00541/00542/02721/ index.html?lang=de&dossier_id=02722>.

Prognos AG (2013). Ex-Post-Analyse des Energieverbrauchs der schweizerischen Haushalte 2000 bis 2012 nach Bestimmungsfaktoren und Verwendungszwecken. Technical report. <http://www.bfe.admin.ch/themen/00526/00541/00542/02721/index.html?lang= de&dossier_id=02722>.

Randolph, B. and Troy, P. (2008). Attitudes to conservation and water consumption. *Environmental Science & Policy*, 11(5):441–455.

Raw, G. and Ross, D. (2011). Energy demand research project: Final analysis - AECOM. Technical Report 60163857, Hertfordshire.

Salmond, K., Schoenberg, J., Girl Scouts of the United States of America, and Girl Scout Research Institute (2009). *Good intentions: The beliefs and values of teens and tweens today*. Girl Scouts of the USA, New York, N.Y.

Schleich, J., Klobasa, M., Goelz, S., and Brunner, M. (2013). Effects of feedback on residential electricity demand: Findings from a field trial in Austria. *Energy Policy*, 61:1097–1106.

Schlieben, M. (2009). Wähleranalyse (1): Wen die Jungen wählen. *Die Zeit*. 2009-09-17. <http://www.zeit.de/politik/deutschland/2009-09/jungwaehler-parteien>.

Schorsch, A. (2012). Die fünfeinhalb Wasserspar-Mythen. *n-tv Wissen*. <http://www.n-tv.de/wissen/Fakten_und_Mythen/Die-fuenfeinhalb-Wasserspar-Mythen-article5818821.html>.

Schwartz, D., Fischhoff, B., Krishnamurti, T., and Sowell, F. (2013). The Hawthorne effect and energy awareness. *Proceedings of the National Academy of Sciences*, 110(38):15242–15246.

SFOE (2013a). Kernenergie. Swiss Federal Office of Energy (BFE). <http://www.bfe.admin.ch/themen/00511/index.html?lang=de>.

SFOE (2013b). Schweizerische Gesamtenergiestatistik 2012. Technical report, Swiss Federal Office of Energy (BFE). Swiss Federal Office of Energy (BFE). <http://www.bfe.admin.ch/themen/00526/00541/00542/00631/?dossier_id=00763&lang=de>.

SFOE (2013c). Wasserkraft. Swiss Federal Office of Energy (BFE). <http://www.bfe.admin.ch/themen/00490/00491/index.html?lang=de>.

SFSO (2012). Familien, Haushalte – Daten, Indikatoren. Technical report, Swiss Federal Statistical Office. Swiss Federal Statistical Office. <http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/04/blank/key/ haushaltsgroesse.html>.

Shove, E. (2003). Converging conventions of comfort, cleanliness and convenience. *Journal of Consumer Policy*, 26(4):395–418.

Snyder, S. J. (2013). The quantified self evolves: A south by southwest (SXSW) vision of how gadgets will teach us about ourselves. *Time*.

Southerton, D., Chappells, H., and Vliet, B. v. (2004). *Sustainable Consumption: The Implications of Changing Infrastructures of Provision*. Edward Elgar Publishing.

Stadtwerk Winterthur (2014). Preise Erdgas für Haushalt. 2014-01-01. http://stadtwerk.winterthur.ch/angebot/erdgas/preise/erdgas-fuer-haushalt/.

Statistisches Amt des Kantons Basel Landschaft (2013). Wasser- und Abwassergebühren der Gemeinden 2013 - Kanton Basel-Landschaft.

Stewart, R. A., Willis, R. M., Panuwatwanich, K., and Sahin, O. (2013). Showering behavioural response to alarming visual display monitors: Longitudinal mixed method study. *Behaviour & Information Technology*, 32(7):695–711.

Taubinsky, D. (2013). From intentions to actions: A model and experimental evidence of inattentive choice.

Tiefenbeck, V., Tasic, V., Schoeb, S., and Staake, T. (2013). Mechatronics to drive environmental sustainability: Measuring, visualizing and transforming consumer patterns on a large scale. In *Proceedings of the IEEE IECON 2013, special section on energy informatics*, Energy Informatics 2013. IEEE.

Torres-Reyna, O. (2007). Panel data analysis. *Fixed & Random Effects*. <http://dss. princeton. edu/training/Panel101. pdf>.

Tschui, A. and Stadelmann, B. (2006). Studie Warmwasserverteilung - Kurzstudie über die Energieverluste der verschiedenen Verteilsituationen von Warmwasser im Mehrfamilienhaus. Technical Report 1100280.

Umweltbundesamt Deutschland (2013). Endenergieverbrauch der privaten Haushalte. Technical report. <http://www.umweltbundesamt.de/daten/private-haushalte-konsum/endenergieverbrauch-der-privaten-haushalte>.

United Nations (2013). International decade for action 'Water for Life' 2005-2015 - Focus areas: Water scarcity. <http://www.un.org/waterforlifedecade/scarcity.shtml>.

Voeroesmarty, C. J., McIntyre, P. B., Gessner, M., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E., Sullivan, C. A., Reidy-Liermann, C., and Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, 467(7315):555–561.

Willis, R. M., Stewart, R. A., Giurco, D. P., Talebpour, M. R., and Mousavinejad, A. (2013). End use water consumption in households: Impact of socio-demographic factors and efficient devices. *Journal of Cleaner Production*, 60:107–115.

Willis, R. M., Stewart, R. A., Panuwatwanich, K., Jones, S., and Kyriakides, A. (2010). Alarming visual display monitors affecting shower end use water and energy conservation in Australian residential households. *Resources, Conservation and Recycling*, 54(12):1117–1127.

Woodside, C. (2011). It isn't easy being green. *Nature Climate Change*, 1(1):13–15.

Zanzi, A. (2011). Vergleich der Wasser-, Abwasser- und Abfallgebühren in den 50 grössten Städten der Schweiz. Technical report, Eidgenössisches Volkswirtschaftsdepartement EVD, Bern.