DISS. ETH NO. 22054

Behavioral Interventions to Reduce Residential Energy and Water Consumption: Impact, Mechanisms, and Side Effects

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

VERENA TIEFENBECK

Dipl-Ing., Technische Universität München & Ecole Centrale Paris

born on 25.12.1983

citizen of Germany

accepted on the recommendation of

Prof. Dr. Elgar Fleisch, ETH Zurich (advisor)
Prof. Dr. Lorenz Goette, University of Lausanne
Prof. Dr. Thorsten Staake, University of Bamberg

2014

Acknowledgements

This dissertation is the result of my work in the period 2010 to 2014 at ETH Zurich, at MIT, and at the Fraunhofer Center for Sustainable Energy Systems in Boston. Throughout my doctorate, numerous people inspired and enriched my work and my life and supported me in many ways.

First of all, I would like to thank my supervisor Prof. Dr. Elgar Fleisch for his valuable support and for the motivating and rewarding work environment he created and maintains, encouraging his students to explore opportunities and to grow their potential. I am very thankful for his guidance and thought leadership, his encouragement and inspiration, his concise feedback, and his trust in me. Elgar helped me see and put issues into a broader context, focus on research questions that also have a high practical impact, and pursue them with scientific rigor.

Next, I would like to thank my two thesis co-advisors Prof. Dr. Lorenz Goette (University of Lausanne) and Prof. Dr. Thorsten Staake (University of Bamberg). As the Director of the Bits to Energy lab, Thorsten supported my work throughout all stages of my research, encouraged and inspired me through his example, his valuable ideas, his constructive feedback, and through his positive attitude. I am highly grateful for his steady support, his valuable ideas, his encouragement, for the theories and methodological skills he familiarized me with, and for his guidance in structuring and improving my thoughts as well as my sentences. The close collaboration with Lorenz in the second half of my doctorate further enabled me to learn and to grow. I am highly grateful for his experienced support and guidance through fruitful discussions, for the theories, models, and methodological skills he taught me, and for his example of a passionate researcher and teacher.

Furthermore, I would like to express my deep gratitude to Dr. Olga Sachs for her academic and personal support in the first half of my doctorate, and for being an incredibly valuable and patient mentor and friend also in difficult moments. Furthermore, I would like to thank Prof. Sanjay Sarma for hosting me at MIT, for his inspiring thoughts, feedback, and fruitful discussions. Particular thanks also go to Prof. Dr. Friedemann Mattern and Prof. Dr. Florian von Wangenheim for their advice, support, and encouragement.

Moreover, I would like to express my gratitude to Dr. Anne-Kathin Faust, Dr. Boris Krey, Dr. Nicole Mathys, and Dr. Michael Moser of the SFOE, as well as Dr. Romeo Deplazes and Dr. Lukas Küng of ewz for their trust in and support of our research from an early stage on. A big "Thank you" also goes to Mirjam Avdyli, Fabian Frei, Gian Andrea Jäger, and Valentin Peter of ewz for their highly competent management and skillful implementation of our joint field study.

In addition, I would also like to thank Dr. Kathrin Degen, Prof. Dr. Rafael Lalive, Dr. Kurt Roth, Dr. Olga Sachs, Vojkan Tasic, and Bryan Urban for their support in our collaborative publication efforts in the past years. I would also like to acknowledge Monica Heinz, Heather

Snyder, and Elisabeth Vetsch-Keller for their outstanding administrative management, organizational skills, and their always prompt, helpful, and friendly support.

To all my colleagues at ETH Zurich, at MIT, at the University of St. Gallen, at the University of Lausanne, and at Fraunhofer CSE, I would like to express my gratitude for the memorable time we spent together. In particular, I would like to thank Craig Akers, Christian Beckel, Dr. Kathrin Degen, Dylan Erb, Kristina Flüchter, Eric Graham, Dr. Tobias Graml, Dr. Gundula Heinatz, Dr. Stephen Ho, Flavius Kehr, Dr. Sumeet Kumar, Tobias Kowatsch, Simon Mayer, Dr. Jan Marckhoff, Jeff McAuley, Janelle Lamarche, Dr. Claire-Michelle Loock, Felix Lossin, Bojan Martin, Dr. Johannes Paefgen, Dr. Irena Pletikosa, Dr. Cordula Schmid, Samuel Schöb, Josh Siegel, Dr. Thomas Stiefmeier, Vojkan and Aleksandra Tasic, Bryan Urban, Dirk Volland, Dr. Markus Weiss, Dr. Erik Wilhelm, Dr. Ryan Williams, Dominic Wörner, and Dr. Michael Zeifman. In addition to innumerable fruitful discussions, I made friends which I hope to stay in touch with. I would also like to highlight the valuable work of the students who supported my projects, in particular Liliane Ableitner, Katherine Cheney, Caroline Duvier, Claudia Geiser, Angela Qin, once more Samuel Schöb, and Maxime Szybowski.

While I'm grateful and indebted to many people, some were particularly present, supportive, and influential for me over the past few years, including Dr. Christina Berchtold, Dr. Vera Blodow, David Cohen-Tanugi, Yves-Alexandre de Montjoye, Prof. Dr. Michael Eichmair, Kate Goldstein (1986 - 2014), Sabrina Sieber, and Silvia Stanglmayr. Finally, I want to express my deepest gratitude towards my parents, Marianne and Konrad, and my brother Christoph. I want to thank you for your unconditional support, your honest advice, and for your love and encouragement over the years. Thank you for being with me every step along the way.

Zurich, July 2014

Verena Tiefenbeck

Contents

List of Abbreviations

AG	Aktiengesellschaft (incorporation)
b	b coefficient
BFE	Bundesamt für Energie (see SFOE)
BTU	British thermal unit
$^\circ C$	Degree Celsius
c_p	Specific heat capacity
CHF	Swiss Franc (currency)
CO_2	Carbon dioxide
DiD	Difference-in-differences
EU	European Union
F	F-statistic
FB	Feedback
g	Gram
GWh	Gigawatt hour
HH	Household(s)
HVAC	Heating, ventilation, and air-conditioning
ICT	Information and communication technology
IHD	In-home display
ID	Identifier
IS	Information Systems
IT	Information technology
J	Joule
k	Kilo
kg	Kilogram
kJ	Kilojoule
kWh	Kilowatt hour
l	Liter
m	Meter
M	Mean value

MW	Megawatt
min	Minute
N	Sample size
Obs.	Observation
OLS	Ordinary Least Squares
p	p-value
PJ	Petajoule
r	Pearson’s correlation coefficient
RT	Real-time
s	Second
SD	Standard deviation
SFOE	Swiss Federal Office of Energy (BFE)
t	t-statistic
t	Ton
U.S.	United States (of America)
W	Watt

Abstract

Natural resources and the benefits derived from their utilization are fundamental for human life. Since the advent of industrialization, the global demand for energy and water in particular has been continuously growing. This development fuels a variety of well-known environmental and geopolitical problems (e.g., carbon emissions, energy security, access to water), creating a growing policy interest in resource conservation. Aside from technical parameters, behavior has been identified as the most important factor governing energy consumption. While the development and implementation of technological innovations is often slow, costly, and risky, pricing mechanisms and regulations often face acceptance issues and political opposition. Against this background, behavioral interventions are increasingly viewed as a politically feasible instrument to promote resource conservation and to increase energy efficiency quickly, at scale, and in addition to technological efficiency gains. In particular, feedback, i.e., providing information about one's own or other people's behavior, has been identified as a cost-effective, scalable, and relatively persistent tool to influence consumer choices. As a result, such behavioral interventions have been rolled out to millions of households. Among feedback programs, the most widespread approach consists in monthly or quarterly mailed reports that compare households' utility consumption with similar homes. While the average conservation impact between 1% and 3% may seem low at first glance, these programs are highly cost-effective and scalable (1% opt-out rate).

Given the scale of the programs that have been rolled out to date, it is all the more astonishing that key aspects of feedback interventions are still unclear. In particular, the underlying psychological mechanisms that drive people to adopt efficient technologies or to engage in curtailment behaviors are not well understood. Furthermore, evidence on the influence of household characteristics and demographics is also not conclusive. Little is known about potential side effects of behavioral interventions, whether the effect of such programs is contained to the target behavior(s), or whether these interventions also have positive or negative repercussions into other behavior domains, as studies conducted in laboratory settings suggest. Furthermore, while *behavior-specific real-time* feedback has been identified as a powerful instrument to influence consumer choices in a series of small-scale pilots, a demonstration of the cost-effectiveness and scalability of that approach is still missing.

To explore these questions in a real-world setting, two separate randomized controlled trials are designed and implemented as part of this thesis. In both studies, water and energy (resp. electricity) measurements of 697 (resp. 154) households over two (resp. three) months serve as the dependent variables. While the second study investigates side effects of a behavioral intervention, the first one explores the effectiveness of behavior-specific real-time feedback, the underlying psychological mechanisms, and the influence of household characteristics. To that end, extensive survey data are combined with granular resource consumption measurements related to a specific behavior (showering).

The results of the first study provide quantitative evidence for a growing intensity of resource use of daily behaviors due to changing norms and conventions: The data indicate that the resource intensity per shower increased to the 2.3-fold within a single generation. At the same time, the findings suggest that behavior-specific feedback can help address these issues. The results show that behavior-specific real-time consumption feedback can prompt substantial behavior changes, inducing an average reduction of 23% in both energy and water consumption for the target behavior. The effect is persistent throughout the duration of the study. Projected to one year, this amounts to energy savings of 443 kWh and a conservation of 8,500 liters of drinking water per household. Both in relative and in absolute numbers, the impact is substantially higher than that achieved by traditional feedback programs. Thanks to the large effect size, it is possible to disentangle the underlying psychological mechanisms and the role of household characteristics to explain the heterogeneity in the treatment effect. The results state that positive mechanisms drive the conservation effect, not psychological pressure. The findings also manifest that the savings are not driven by a small group of environmentalists: The net conservation effect is independent of environmental attitudes. In addition to its direct impact on behavior, real-time consumption feedback also appears to substantially increase knowledge about resource consumption.

However, the findings of the second study provide evidence that behavioral interventions can also cause measurable negative side effects. Moreover, the study shows that adverse outcomes can more than offset the benefits on the primarily targeted behavior if the program does not focus on the behaviors which are most relevant to households' resource consumption.

This thesis advances the theoretical foundation of behavioral interventions by contributing essential and far-reaching insights into underlying mechanisms and potential side effects, which are relevant for several research disciplines. At the same time, the findings entail important guidelines and recommendations for policymakers, utility companies, technology and program developers, and other practitioners. The insights into the underlying psychological mechanisms suggest that behavior-specific real-time feedback is highly scalable (without compromising conservation impact) and amenable to widespread adoption. Furthermore, profiling, e.g., targeting households with an above-average baseline consumption, can double the treatment effect, raising the cost-effectiveness of future deployments even further.

These insights can be vital for other innovative technologies and companies to overcome existing barriers to market entry. Furthermore, the outcomes of both studies emphasize the importance of behavioral programs achieving a) a high degree of effectiveness, by focusing on environmentally significant behaviors, combined with b) a high degree of efficiency through powerful instruments. That way, it is possible to maximize program impact, while reducing the potential of adverse outcomes. On the one hand, the example of real-time feedback on shower behavior demonstrates that this combination can have a substantial impact on resource consumption, also in absolute numbers. On the other hand, the second study shows that adverse outcomes can loom large: Focusing on the most relevant behaviors reduces the room for negative side effects. The mechanisms indicate that under these circumstances, behavioral interventions are able to unambiguously enhance social welfare (e.g., by reducing emissions) *and* individual utility: They help many individuals act in line with their preferences without decreasing utility for anyone, making such interventions a particularly normatively desirable policy instrument. From a policy perspective, the evidence for the massive growth in resource intensity for showering given in the first study, as well as the insights into adverse side effects presented in the second study, both point out the importance of adopting a comprehensive view in the evaluation of studies, behavioral programs, and resource demand projections.

This thesis also paves the way for follow-up research, e.g., on the long-term persistence of the effects. Altogether, these insights may equally be applied to other environmentally significant domains like space heating or electricity smart metering - and beyond, e.g., to fields like nutrition and exercise.

In conclusion, this thesis provides strong evidence that behavioral interventions are powerful, scalable, and cost-effective tools to promote resource conservation. When applied wisely, these programs can be beneficial for the individual and contribute to society. As a result, they hold the potential to play a significant role in the pursuit of environmental and geopolitical goals.

Kurzfassung

Natürliche Ressourcen und die mit ihrer Nutzung verbundenen positiven Aspekte sind essentiell für das menschliche Dasein. Seit dem Zeitalter der Industrialisierung ist der weltweite Energie- und Wasserverbrauch kontinuierlich angestiegen. Diese Entwicklung verstärkt eine Reihe wohlbekannter Probleme umwelt- und geopolitischer Natur (wie CO_2-Ausstoss, Energieversorgungssicherheit, oder Zugang zu Wasser), was zu einem wachsenden Interesse der Politik am sparsamen Umgang mit Ressourcen führt. Neben technischen Parametern wird der Energieverbrauch massgeblich von Verhaltensweisen bestimmt. Während die Entwicklung und Einführung technischer Innovationen oft langsam, kostspielig und mit Risiken verbunden ist, stossen Preismechanismen und gesetzliche Regulierung häufig auf Akzeptanzprobleme und politischen Widerstand. Vor diesem Hintergrund werden Massnahmen, die das Verhalten betreffen ("Verhaltensinterventionen") mehr und mehr als politisch durchsetzbares Instrument gesehen, um einen sparsameren Umgang mit Ressourcen zu erwirken und um Energieeffizienz zügig, in grossem Umfang, und zusätzlich zu technologiebedingten Effizienzgewinnen zu erhöhen. Insbesondere Feedback, d.h. das Zurverfügungstellen von Informationen zum eigenen Verhalten oder dem anderer Personen, wurde als kostengünstiges, skalierbares und relativ beständiges Instrument identifiziert, um das Verbrauchsverhalten zu beeinflussen. Infolgedessen wurden derartige Verhaltensinterventionen bereits bei Millionen von Haushalten durchgeführt. Der verbreitetste Ansatz bei solchen Feedbackprogrammen besteht darin, Haushalten monatlich oder quartalsweise Berichte zu senden, die den jeweiligen Ressourcenverbrauch mit demjenigen ähnlicher Haushalte vergleichen. Wenngleich der durchschnittliche Einspareffekt von 1 bis 3% auf den ersten Blick gering anmuten mag, sind diese Programme höchst kostengünstig und skalierbar (1% Ausstiegsquote).

Angesichts des Umfangs der bis dato durchgeführten Programme ist es umso erstaunlicher, dass zentrale Aspekte von Verhaltensinterventionen noch nicht ergründet sind. Insbesondere die zugrundeliegenden psychologischen Mechanismen, die Menschen dazu bewegen, effiziente Technologien zu nutzen oder sich ressourcensparsam zu verhalten, sind wenig ergründet. Aussagen zum Einfluss von Haushaltscharakteristika und demographischer Aspekte sind widersprüchlich. Auch über das mögliche Auftreten von Nebeneffekten bei Verhaltensinterventionen ist wenig bekannt: Nämlich ob die Wirkung auf die Verhaltensbereiche beschränkt ist, auf die das Programm (eigentlich) abzielt, oder ob diese Programme darüber hinaus posi-

tive oder negative Nebenwirkungen auf andere Verhaltensbereiche ausüben, worauf Laborstudien hinweisen. Darüber hinaus hat eine Reihe kleinerer Pilotstudien ergeben, dass *Echtzeit-Feedback zum Ressourcenverbrauch bei einer konkreten Handlung* ein besonders wirksames Instrument ist, um Verbrauchsverhalten zu beeinflussen; ein Nachweis der Kosteneffizienz und der Skalierbarkeit dieses Ansatzes wurde jedoch bislang noch nicht erbracht.

Um diesen Fragen in der Praxis auf den Grund zu gehen, wurden im Rahmen dieser Dissertation zwei voneinander unabhängige randomisierte kontrollierte Studien konzipiert und durchgeführt. In beiden Studien dienen Wasser- und Energie- (bzw. Strom-)verbrauchswerte von 697 (bzw. 154) Haushalten über zwei (bzw. drei) Monate hinweg als abhängige Variablen. Während die zweite Studie die Nebenwirkungen von Verhaltensinterventionen untersucht, ermittelt die erste Studie die Wirkung von verhaltensspezifischem Echtzeit-Feedback, sowie die psychologischen Mechanismen, die diesem zugrunde liegen und den Einfluss von Haushaltscharakteristika. Hierzu werden umfangreiche Umfragedaten mit detaillierten Messwerten zum Ressourcenverbrauch bei einem konkreten Verhalten (Duschen) kombiniert.

Die Ergebnisse der ersten Studie liefern einen quantitativen Nachweis für die wachsende Ressourcenintensität alltäglicher Verhaltensweisen, vermutlich hervorgerufen durch sich wandelnde Normen und Konventionen: Die Daten zeigen, dass die Ressourcenintensität pro Duschvorgang innerhalb von nur einer Generation auf das 2,3-fache angestiegen ist. Die Ergebnisse weisen jedoch darauf hin, dass von verhaltensspezifisches Echtzeit-Feedback dazu beitragen kann, dies zu adressieren. Die Resultate zeigen, dass verhaltensspezifisches Echtzeit-Feedback erhebliche Verhaltensänderungen bewirken kann: Die durchschnittliche Einsparung an Energie und Wasser beim anvisierte Verhalten beträgt 23%. Der Effekt ist während der gesamten Studiendauer stabil. Auf ein Jahr hochgerechnet ergeben sich somit Einsparungen von 443 kWh an Energie und von 8.500 Liter Trinkwasser. In relativen wie absoluten Zahlen ist die Wirkung somit massgeblich höher als bei konventionellen Feedbackprogrammen. Dank der grossen Effektstärke ist es möglich, die zugrunde liegenden psychologischen Mechanismen und den Einfluss von Haushaltscharakteristika zu analysieren, um die Varianz im Behandlungseffekt (treatment effect) zu erklären. Die Ergebnisse offenbaren, dass *positive* Mechanismen den Einsparungseffekt bedingen, nicht psychologischer Druck. Zudem zeigen die Resultate, dass der Einspareffekt nicht nur von einer kleinen Teilgruppe besonders umweltbewusster Teilnehmer getrieben werden: Die Nettoeinsparung ist unabhängig von der Umwelteinstellung. Abgesehen von der direkten Wirkung auf das Verhalten scheint Echtzeit-Feedback auch das Wissen zum Ressourcenverbrauch massgeblich zu steigern.

Die Ergebnisse der zweiten Studie liefern jedoch Indizien dafür, dass Verhaltensinterventionen auch messbare negative Nebenwirkungen verursachen können. Darüber hinaus zeigt die Studie, dass die negativen Nebenwirkungen die positive Wirkung des primär anvisierten Verhaltens mehr als zunichte machen können, wenn das Programm nicht auf die Verhaltensbereiche abzielt, die den Ressourcenverbrauch im Haushalt am massgeblichsten beeinflussen.

Diese Dissertation erweitert die theoretischen Grundlagen zu Verhaltensinterventionen, indem sie essentielle und weitreichende Erkenntnisse zu den zugrunde liegenden Mechanismen und zu möglicherweise auftretenden Nebeneffekten beiträgt. Zugleich leiten sich wichtige Leitfäden und Handlungsempfehlungen für Politik, Energieversorger, Technologie- und Projektentwickler sowie weitere Interessensgruppen ab. Die Erkenntnisse bezüglich der psychologischen Mechanismen legen nahe, dass verhaltensspezifisches Echtzeit-Feedback in hohem Masse skalierbar ist (ohne dabei den Einspareffekt zu schmälern) und sich für einen grossflächigen Einsatz eignet. Zudem kann eine intelligente Segmentierung ("Profiling") - also eine gezielte Ansprache von Haushalten mit einem Verbrauch oberhalb des Durchschnitts - die Effektgrösse verdoppeln und somit die Kosteneffizienz künftiger Programme noch weiter erhöhen. Diese Erkenntnisse können auch für andere innovative Technologien und Unternehmen überlebenswichtig sein, um vorhandene Markteintrittsbarrieren zu überwinden.

Darüber hinaus betonen beide Studien, dass Verhaltensinterventionen zugleich ein hohes Mass erzielen müssen a) an Effektivität, indem sie auf besonders umweltrelevante Handlungsbereiche abzielen und b) an Effizienz durch besonders wirksame Interventionen. Dies ermöglicht es, die positive Wirkung der Programme zu maximieren und gleichzeitig das Potential von negativen Nebenwirkungen zu reduzieren. Zum einen veranschaulicht das Beispiel von Echtzeit-Feedback zum Duschverhalten, dass diese Kombination - auch in absoluten Zahlen - einen massgeblichen Effekt auf den Ressourcenverbrauch haben kann. Wie jedoch die zweite Studie zeigt, können negative Nebenwirkungen eine grosse Rolle spielen. Indem man auf die umweltrelevantesten Verhaltensbereiche abzielt, reduziert man gleichzeitig die verbleibende Einflussmöglichkeit negativer Nebenwirkungen. Die psychologischen Mechanismen deuten darauf hin, dass Verhaltensinterventionen unter diesen Umständen gleichzeitig sowohl das gesellschaftliche Wohl (z.B. durch reduzierte Emissionen), als auch den Nutzen des Einzelnen steigern: Sie können vielen Einzelpersonen helfen, im Einklang mit ihren Präferenzen und Zielvorstellungen zu handeln, ohne dies auf Kosten anderer Menschen zu tun; solche Verhaltensinterventionen sind infolge dessen ein normativ besonders anzustrebendes Politikinstrument. Beide Studien zeigen, wie wichtig eine ganzheitliche Betrachtungsweise ist bei der Auswertung von Verhaltensinterventionen und bei Bedarfsprognosen bezüglich des Ressourcenverbrauchs: Dies zeigen sowohl der in der ersten Studie erbrachte Nachweis der massiv wachsenden Ressourcenintensität beim Duschen, als auch die Erkenntnisse aus der zweiten Studie zu negativen Nebenwirkungen von Verhaltensinterventionen.

Diese Arbeit ebnet auch den Weg für weitere Forschungsaktivitäten, beispielsweise zur Langzeitstabilität der Effekte. Zudem lassen sich die Erkenntnisse auf andere umweltrelevante Bereiche wie Raumheizung oder Strom-Smart-Metering übertragen - auch darüber hinaus auf Bereiche wie Ernährung und Bewegung.

Insgesamt zeigt die vorliegende Arbeit, dass Verhaltensinterventionen wirksame, skalierbar und kosteneffektive Instrumente sind, um den sparsameren Umgang mit Ressourcen massgeblich voranzutreiben. Sinnvoll ein- und umgesetzt, können Verhaltensinterventionen sowohl das gesellschaftliche Wohl, als auch das Wohl des Einzelnen erhöhen. Infolgedessen haben sie das Potential, einen entscheidenden Beitrag zur Erreichung umwelt- und geopolitischer Ziele zu leisten.

Motto and Work Published Earlier

"Nuanced research into human behavior and energy-use decisions is not new, nor is the idea that energy efficiency may be generally cost-effective. 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."

Hunt Allcott & Sendhil Mullainathan, Science Magazine (327), 2010

Previously Published Work Parts of this dissertation have already been published previously by myself and colleagues as scientific articles in peer-reviewed journals or in conference proceedings; other parts are currently being prepared for publication in additional outlets. While I am the first author of all of these documents and hereby declare that the majority of the content that has been integrated into this thesis has been written by myself, other co-authors have contributed to these documents with their reviews, changes, suggestions and edits. As a result, some sections of this thesis literally correspond to parts of work published earlier or have strong similarities with own work published earlier.

In particular, in the related work chapter 2, section 2.2 on side effects of feedback interventions is an adaption of the literature review of the article published in Energy Policy (Tiefenbeck et al. (2013a)) along with my colleagues Olga Sachs, Kurt Roth, and Thorsten Staake. Parts of section 2.3.3 have been integrated from the conference proceedings paper Tiefenbeck et al. (2013b), co-authored by Vojkan Tasic, Samuel Schöeb, and Thorsten Staake. Although parts of section 2.1.2 correspond to paragraphs of the report to the Swiss Federal Office of Energy (SFOE, Tiefenbeck et al. (2014)), one should note that these parts originate from this thesis and were subsequently integrated to a later version of the report.

In chapter 3, the motivation and methodology sections 3.1 and 3.2 and have been modified and integrated from the report to the Swiss Federal Office of Energy (SFOE, Tiefenbeck et al. (2014)), co-authored by Lorenz Goette, Kathrin Degen, Vojkan Tasic, and Thorsten Staake. Parts of 3.3 have also been adapted from the results chapter of that report, parts of which had originally been written by Lorenz Goette. While most of the corresponding text has been re-written or modified for this thesis, several tables, equations, and figures have been imported (with minor modifications) from that report. This also applies to the supplementary tables A. Some parts of 3.4 have also been integrated from the report to SFOE; yet some of those parts originate from this thesis and were then integrated into the report to SFOE, hence the correspondence of these paragraphs.

Chapter 4 is by far the one that has the highest content of work by myself with colleagues that has already been published, in this case in Energy Policy (Tiefenbeck et al. (2013a)). While the structure of the chapter has been adapted to mirror the structure of chapter 3, the content of these sections has already been published in that paper.

Chapter 1

Introduction

This chapter outlines the general motivation and objectives of the thesis and gives a brief overview of the methodological approach. It further highlights the contributions to theory and practice and closes with the outline of this thesis.

1.1 Motivation

Natural resources and benefits derived from their utilization are fundamental for human life (WTO (2010)), yet their availability is limited and unequally distributed across countries and regions. World primary energy use has doubled over the past 40 years, with over 80% being provided by fossil fuels (IEA (2013)); during the same period, global water use has increased by 75% (Shiklomanov (1999)). The growing demand for natural resources and their consumption fuels a variety of well-known problems. This is particularly true for the water-energy nexus. From an environmental perspective, this includes problems associated with carbon emissions; water, soil, and air pollution; and depletion of resources. From a geopolitical point of view, access to clean water and energy security is crucial for modern economies and prosperity of nations; therefore, the uneven distribution of resources creates political dependencies. As a result, conserving energy and water is a key policy goal in many countries. Ambitious goals have been put forward and moved from pure visions to actual policy guidelines. One prominent example is the 20-20-20 targets of the European Union, whose key objectives for 2020 aim at a 20% reduction in EU greenhouse gas emissions (from 1990 levels), a 20% share of renewable energy sources, and a 20% improvement in energy efficiency. Switzerland has decided on even more ambitious goals with the adoption of its long-term plan "Energiestrategie 2050". In order to reach these goals, both the energy supply and demand side will have to undergo massive changes (Lester and Hart (2012)). Not only does it imply a timely decarbonization of energy production (e.g., through the integration of renewable energy sources), but also requires efficiency gains across all sectors.

On the demand side, households have 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 carbon emissions trading, fuel economy standards or changes on the energy supply side (Gardner and Stern (2008)). The residential sector accounts for 20% of the CO_2 emissions from fossil fuel combustion (eia (2013b)) and for approximately 25% (resp. 22%) of total primary energy consumption in the EU-27 (resp. the U.S.) (European Environment Agency (2012); U.S. Department of Energy (2012)). Although considerable technological progress has been made resulting in 25% improvement of energy efficiency for households in the period 1990-2010 (European Environment Agency (2013)), residential energy and electricity demand have continued to rise: In Europe (EU-27 and other EEA states), per capita energy consumption has increased by 6.6% between 1990 and 2008, partly owing to an increase in per capita electricity consumption by 32% (European Environment Agency (2011)). Against this backdrop, the reduction goals put forward are very ambitious and it will not be possible to attain them without a significant contribution from the household sector.

Aside from technical parameters, residential resource consumption is substantially governed by human behavior (Haas et al. (1998)). Traditional efforts of economists and policymakers to encourage resource conservation (through curtailment or the uptake of more efficient technology) reflect a rational-economic model of behavior, relying either on regulations or pricing mechanisms. Regulations include technology standards (e.g., ban on incandescent light bulbs) and rationing (e.g., outdoor water-use restriction, rolling electricity blackouts), while pricing mechanisms include e.g., carbon tax, tiered pricing schemes, tax credits (Samuelson (1990); Allcott (2011b)). Yet these structural instruments often face policy resistance, or are slow and costly to implement. Moreover, at least in the past, these instruments were not able to revert or to stall the continuous growth of energy demand.

The second traditional approach to promote resource conservation is based on attitude-behavior models developed by psychologists. The resulting environmental campaigns seek to change people's *attitudes* with pro-environmental appeals - the underlying rationale is that a change of attitudes will also bring about behavior change. Yet these campaigns often reach only a small minority of households. While attitudes may predict behavior to some extent, people do not necessarily act in line with them and *"changing attitudes toward conservation may never be as effective as operating on conservation behavior more directly"* (Baca-Motes et al. (2013)).

In contrast, in the past several years, so-called *behavioral instruments* have drawn a lot of attention as a cost-effective and political feasible tool to increase energy efficiency and to reduce resource consumption quickly, at scale, and in a relatively persistent way (Allcott (2011b); Allcott and Rogers (2014)). These approaches are built on the insight that behavior is governed by a complex interplay of a variety of intrinsic (altruistic), extrinsic (material self-interest), and reputational (social or self-image concerns) considerations (Bénabou and Tirole (2006)). As a result, behavioral interventions generally integrate s messages, thus appealing to different resource conservation motives simultaneously. In particular, programs using feedback, i.e., providing information about one's own or other people's behavior, have been rolled out to millions of households over the past years. Among feedback programs, the most widespread approach consists in monthly or quarterly mailed reports that compare household's utility consumption with similar homes as a frame of reference. This implies that the majority of these interventions are still delivered in the form of paper-based utility bills, thus with a substantial time lag and aggregated over long time periods and to the entire household. These programs generally yield an average conservation impact between 1% and 3% (Ayres et al. (2009); Allcott and Rogers (2014)). While this may seem low at first glance, these programs are highly cost-effective and scalable (1% opt-out rate).

While existing paper-based behavioral interventions demonstrate that feedback can cost-effectively influence consumer behavior on a large scale, it has also been shown that feedback works best when it is delivered frequently, timely, clearly, and on specific actions which individuals can easily influence. So far, large-scale implementations for that type of behavioral intervention are still missing. Furthermore, as section 1.2 will elaborate, the underlying mechanisms that drive people to adopt efficient technologies or to engage in curtailment behaviors are not yet well understood, nor is the question of potential side effects of behavioral programs. A better understanding of underlying mechanisms and side effects could make behavioral interventions more effective and help to mitigate adverse effects.

1.2 Objectives and Contributions

This thesis aims to shed light on existing caveats in the research on behavioral interventions, with the goal of contributing to both theory and practice. On the theoretical side, the thesis is built mainly on theories from behavioral economics, social psychology, marketing, environmental policy and information systems research. It aims to contribute to the existing knowledge of these disciplines by applying insights from these disciplines to real-world settings. The key goals of this thesis are threefold: a) Investigating the effectiveness of behavior-specific real-time feedback, b) a better understanding of the underlying psychological mechanisms and of the role of household characteristics, and c) to explore potential side effects of feedback interventions.

As of today, most behavioral programs provide feedback with substantial time lag and a high level of aggregation (e.g., monthly utility bills or web a portal with information on the household level). This creates a disconnect between individuals' behavior (e.g., showering) and the associated resource consumption. Although information and communication technologies

(ICT) could open novel avenues to overcome these issues, research on real-time feedback is still in its infancy and little is known about its effectiveness when implemented on a larger scale or for a specific target behavior. The study presented in chapter 3 of this thesis is one of the first studies to evaluate the feasibility and effectiveness of behavior-specific real-time feedback on a larger scale. That way, this thesis also contributes to Information Systems (IS) research, by showcasing a practical solution to investigate the response of individuals to such feedback systems in a controlled field trial and by evaluating the cost-effectiveness of this approach.

In spite of the large-scale deployment of behavioral interventions that has been realized, the psychological mechanisms that drive people to respond to behavioral interventions are not well understood to date. Of particular interest here is whether feedback interventions coerce individuals into curtailment behaviors by putting negative pressure on them, or whether these programs resonate with individuals' preferences and provide people with the tools necessary to act in line with their goals. While the first would both benefit the individual *and* contribute to society, the latter would be at the expense of individuals' utility. This would make these interventions less desirable and might potentially undo all the welfare gains of reduced resource consumption. The question then arises whether individuals with certain personality traits or household characteristics are more responsive to feedback interventions. Psychologists in particular have put forward a number of behavioral frameworks to explain what drives people to engage in conservation behaviors or in the uptake of less resource-intensive technology. Yet few studies are available to validate the proposed models in an empirical setting. This thesis will contribute to this literature by quantifying the impact of psychological variables on real-world outcomes (e.g., utility consumption).

In a similar vein, researchers from various disciplines have investigated how household characteristics influence residential utility consumption and households' response to behavioral interventions, with very mixed results. Yet most empirical studies are limited to a few selective variables, which might produce spurious correlations and biased results. The study presented in chapter 3 seeks to provide a more comprehensive evaluation, by combining detailed measurement data with extensive survey data that cover psychological factors and household characteristics. A better understanding of the interplay of psychological variables and household characteristics paves the way for making behavioral interventions more (cost-)effective. In addition, these insights also help understand to what extent real-time feedback programs are scalable.

Finally, it is still unclear whether behavioral interventions have side effects: Do they trigger cross-domain adoption of additional environment-friendly behaviors (positive spillover) or reduced engagement elsewhere, as studies from other fields suggest? Although there is ample evidence for moral licensing from various domains, most of the research on behavioral spillover has been carried out in a laboratory setting, and the behaviors exhibited there may not be reflective of typical behaviors outside the laboratory. As a consequence, these repercussions are currently not taken into account in the design and analysis of behavioral intervention in the sustainability context. The study presented in chapter 4 of this thesis investigates side effects of feedback interventions on other resource consumption behaviors. The findings might fundamentally alter the net performance of behavioral programs. Empirical insights into side effects subsequent to positively perceived actions are also relevant to other domains, e.g., to healthcare, discrimination, or education.

For most of these aspects, it is difficult to draw a line between theory and practical relevance. The resulting implications, which are described in the discussion part of the two empirical studies (sections 3.4 and 4.4) and consolidated in chapter 5, are directly relevant for various practitioners, including policymakers, utility companies, technology providers, system developers and, ultimately, consumers alike. The insights of this thesis can be useful to craft guidelines and recommendations for designing behavioral interventions which are more effective and minimize adverse outcomes. For instance, insights into the influence of particular personality traits or household characteristics can help target households with higher expected conservation impact: This approach could help increase the cost-effectiveness of such programs. Insights into psychological mechanisms and household characteristics are also essential for the cost-benefit analysis of large-scale rollouts of similar interventions: The external validity of pilot studies is often questioned, thus it is unclear to what extent pilot results can be extrapolated to a broader audience. Regarding risks associated with large-scale implementations, this thesis further sheds light on the question whether the benefits for society of reduced resource consumption come at the expense of utility loss for individuals (e.g., those particularly susceptible to psychological pressure), which would make these instruments less appealing from a welfare perspective.

To explore these questions in real-world settings, two separate randomized controlled trials are designed and implemented as part of this thesis. They combine measured data with survey data, resulting in unique and - by comparison to existing research - large datasets.

To summarize, while feedback interventions show a very promising potential, three aspects in particular require further investigation: a) field studies that show the feasibility and effectiveness of more advanced (i.e., real-time and behavior-specific) feedback, b) a better understanding of the psychological mechanisms driving the behavior change and of the role of household characteristics, and c) an investigation of potential side effects that might affect the net outcome of these programs in a positive or in a negative sense.

1.3 Approach

To examine these topics in real households, two large-scale field experiments are carried out. Both studies are set up as randomized controlled trials with 154 resp. 697 apartments. In both cases, utility consumption data (water and energy/electricity) are measured and used as the dependent variables. In the first study, the measurement data collected characterize the outcome of a specific behavior (showering), while the second study collects household-level data on electricity and water use.

The first study investigates the effectiveness of behavior-specific real-time feedback, the underlying psychological mechanisms that drive people to change their behavior, and the influence of household characteristics in the context of showering. The study is funded by the Swiss Federal Office of Energy (SFOE) and conducted in cooperation with the University of Lausanne, the ETH Zurich spinoff company Amphiro AG and the local utility company ewz. Participants are recruited from a larger pool of households that had previously participated in a smart metering study. Each participating household receives a smart shower monitor that displays real-time feedback to the user in the shower. The device stores time series data on each shower. The devices are installed by the participants and remain deployed in their shower for two months (December 2012 through February 2013). Participants are randomly assigned to one of three display conditions to ensure the internal validity of the experiment. Online surveys are administered prior to and after the field data collection phase. The data are analyzed in close collaboration between the researchers of ETH Zurich and the University of Lausanne. In order to obtain the causal effects unconfounded with time trends that affect all treatments alike, a difference-in-differences strategy is used. Furthermore, a fixed-effects model is estimated using ordinary least squares (OLS) to fully take advantage of the experimental setup.

The second study, which investigates side effects of feedback interventions, is carried out in the Greater Boston area. The study site is a multifamily building complex with 200 apartments with a similar building structure and appliances. Apartment water consumption is measured daily (automatically, by a submetering system) and electricity consumption weekly (meters read by members of the research team). The study is developed and conducted in a collaboration between ETH Zurich and the Fraunhofer Center for Sustainable Energy Systems (Cambridge, MA, USA). The implementation is supported by Corcoran, the managing company of the study site.

1.4 Thesis Outline

The structure of this thesis is as follows: Chapter 1 describes the general context and gives a brief overview of the issues addressed in this thesis. Chapter 2 provides an overview of the related work on existing behavioral feedback interventions promoting resource conservation and their underlying psychological mechanisms, on spillover processes of behavioral interventions and on Green Information Systems; this is followed by the presentation of the research questions guiding this thesis. Chapter 3 and 4 contain the empirical evaluation of the two field studies carried out to answer these research questions: Chapter 3 presents a study on the effectiveness of behavior-specific real-time feedback and investigates the underlying psychological mechanisms. Chapter 4 describes an experiment to investigate side effects of behavioral interventions. Chapter 5 concludes with a general discussion of the key findings of this thesis, its limitations and contributions to theory and practice.

Chapter 2

Theoretical Background

This chapter outlines the theoretical background of this thesis. The first section 2.1 provides an overview of feedback interventions: It puts feedback interventions into the larger context of strategies for resource demand side management, then presents a classification framework for feedback interventions, followed by an overview of mechanisms and factors that have been identified as potential drivers of the effectiveness of these programs. Section 2.2 introduces the existing body of literature on side effects of pro-environmental (and, more general, pro-social) behavior, outlining both existing work on positive spillover effects and evidence for negative side effects. Finally, section 2.3 shows opportunities for more advanced feedback interventions using modern technology. The section depicts the general potential ascribed to the combination of feedback interventions and ICT, followed by an overview of existing large-scale studies using ICT and of previous applications providing behavior-specific real-time feedback.

2.1 Behavioral Feedback Interventions: Primary Effects and Psychological Mechanisms

As outlined in section 1, feedback - providing information about one's own or other people's behavior - has been identified as an effective policy instrument to influence resource consumption. Feedback interventions thus fall under the umbrella of *demand side management strategies*. This context is described in section 2.1.1. Among feedback interventions, a variety of different implementation approaches have been developed, mainly in the field of electricity consumption. Section 2.1.2 gives an overview of existing large-scale programs: It structures existing large-scale programs in the electricity sector based on a common classification scheme, subsumes their insights regarding impact and effect persistence, and presents related work on large-scale feedback interventions for water demand management. Finally, section 2.1.3 gives an overview of behavioral models, psychological mechanisms and household characteristics that have been identified as relevant to residential resource consumption and for the effectiveness of feedback interventions.

2.1.1 Feedback as Strategy for Resource Demand Side Management

Numerous reports of governments and international organizations repeatedly emphasize the potential of energy efficiency in particular to cost-effectively address questions of energy security and environmental issues (e.g., Choi Granade et al. (2009); IEA (2012); European Environment Agency (2013)). Behavior plays a crucial role in resource consumption and *"navigating the interface between policymaking and human behaviour is key to achieving sustained reductions in energy consumption"* (European Environment Agency (2013)). However, in practice, individuals often fail to engage in resource-efficient behaviors. As a consequence, policymakers and scientists alike seek to identify the economic, institutional, political, and social barriers to resource conservation. To address these barriers, they have set up a variety of demand-side management programs that seek to influence individuals' resource consumption activities (Carragher et al. (2012)).

Traditional efforts to encourage resource conservation (through curtailment or the uptake of more efficient technology) typically either rely on a rational-economic model of behavior or on an attitude-behavior-choice model (aiming at attitude change to alter behavior) (Samuelson (1990)). Based on the rational-economic model, economists have focused on monetary incentives and the profitability of efficiency investments. Yet consumers and firms are often not undertaking privately profitable investments in energy efficiency (e.g., Jaffe and Stavins (1994); Allcott and Greenstone (2012)), a phenomenon which is referred to as *energy efficiency gap*. To address this gap, policy interventions traditionally employ either financial strategies (e.g., pricing, tax credits, subsidies), regulations (e.g., ban on incandescent lightbulbs), or voluntary industry standards to stimulate resource conservation (e.g, EnergyStar program). Yet these approaches tend to be slow and costly to implement and to face strong political resistance (Fielding et al. (2013)). Moreover, the most prevalent forms of prescriptive measures (i.e., rationing or technology standards) are among the least flexible command-and-control approaches and generate significant economic losses (Olmstead and Stavins (2009)).

Social scientists, and psychologists in particular, on the other hand, have originated campaigns that focus on attitude change. The underlying model is that attitudes are important determinants of intentions, which in turn are key drivers of behavior (see section 2.1.3.3). Yet campaigns seeking to raise awareness for issues related to resource consumption or to change people's attitudes, do often not result in successful results in terms of actual behavior changes (McKenzie-Mohr (2013); Chong et al. (2013)).

In recent years, however, behavioral economists and psychologists alike have revealed a plethora of cognitive biases and phenomena that are neither consistent with a rational-economic model, nor an attitude-behavior-choice model. These insights open up new possibilities for behavioral interventions. One of the key insights is that decision-making contexts may lead individuals systematically to fail in acting on their intentions or to achieve their preferred ends (Hansen and Jespersen (2013)). As a consequence, (Thaler and Sunstein (2009)) suggested that policymakers should act as choice architects by influencing citizens' behavior with behavioral *nudges*, without recurring to injunctions (restrictions) or financial incentives. They refer to this as *libertarian paternalism*. One of the key strategies to alter choice architecture is to make behavior - or its outcomes - more visible (Kinzig et al. (2013)). In the context of resource consumption, this is realized by giving households feedback on their electricity, gas, or water consumption. Another key insight which stands in contrast to classic economic theory is that individuals respond to social comparisons. The work of Robert Cialdini and his colleagues in particular has highlighted the practical relevance of these insights (Schultz et al. (2007); Goldstein et al. (2008); Nolan et al. (2008); Cialdini (2009)). As a result, social comparison elements are a recurrent feature of many behavioral programs (see section 2.1.3.4).

Over the past few years, the number of feedback programs has dramatically increased (Allcott (2011b)). Relatively quick to implement and easy to scale up, these programs have shown high economic returns and even negative net carbon abatement costs (Allcott and Mullainathan (2010)).

Overall, feedback interventions are increasingly viewed as a politically feasible, effective, and quickly scalable alternative to traditional demand side management tools to promote resource conservation.

2.1.2 Classification of Feedback Interventions and Effect Persistence

While the vast majority of feedback interventions have been carried out in the electricity domain, several programs and studies also focus on gas, oil, or water consumption. Accordingly, the majority of this section will present insights from feedback studies on electricity usage, followed by an overview of existing interventions in the water domain.

Classification of Feedback Interventions and Effect Sizes Feedback on resource consumption can be provided in a variety of forms, which strongly differ with respect to cost of implementation and level of information provided. To facilitate the comparison between different approaches, EPRI developed a classification scheme which distinguishes six types of feedback interventions regarding their immediacy and frequency of information and their level of data aggregation (EPRI (2009)). This scheme was also adopted by the meta-analysis of 57 feedback studies by Ehrhardt Martinez et al. (2010).

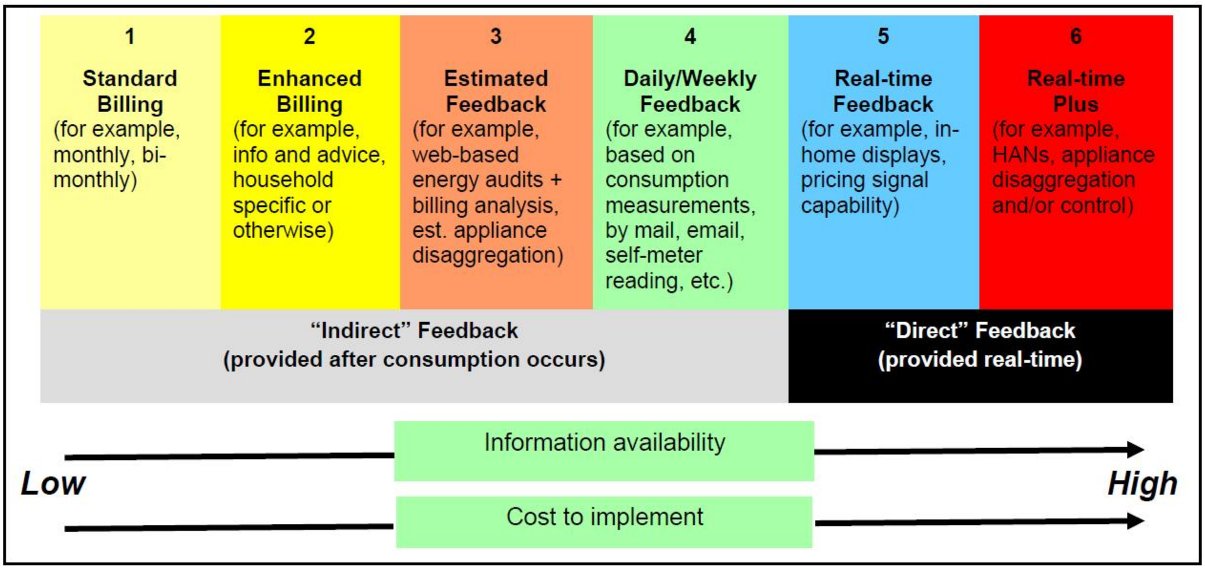


Figure 2.1: Classification of Feedback Types Developed by EPRI. Source: EPRI (2009)

Figure 2.1 shows the classification scheme developed by EPRI. On the left side of the spectrum, it distinguishes four types of indirect feedback (ordered by their typical level of information level and cost of implementation): standard billing, enhanced billing (e.g., Opower home energy reports), estimated feedback (e.g., web-based appliance disaggregation) and daily/weekly feedback (e.g., self-meter reading). In the upper range both in terms of information availability and cost of implementation, the scheme distinguishes two categories of direct feedback: real-time feedback (e.g., IHDs) and "real-time plus" feedback (e.g., behavior-specific feedback or appliance control). Ehrhardt Martinez et al. (2010) report median electricity savings of 3.8% for enhanced billing (e.g., including energy conservation advice), of 6.8% for estimated feedback (e.g., typical break-down by appliances), of 8.4% for daily/weekly feedback (e.g., meter self-reading in combination with a web portal, as in the case of the Swiss company BEN), of 9.2% for real-time feedback (e.g., in-home portals) and of 12% for "real-time plus" feedback. "Real-time plus" feedback in this taxonomy refers to appliance-specific or behavior-specific real-time feedback.

Altogether, these numbers suggest an increasing effectiveness of feedback as the level of timeliness, frequency, and detail increases, with behavior-specific real-time feedback being particularly effective. The savings numbers presented by Ehrhardt Martinez et al. (2010), 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 and a proper research design (see below).

The currently most prevalent form of feedback on residential utility consumption today falls into the EPRI category enhanced billing. The U.S.-based company Opower in particular uses this channel, mailing *home energy reports* with personalized energy use feedback, social comparisons, and energy conservation information to millions of households every month or every few months (Allcott and Rogers (2014)). In a meta-analysis comprising 14 field experiments and a total of 550,000 households, Allcott and Mullainathan (2012)) report an average treatment effect in the range of 2%. They report that monthly instead of quarterly report frequency increases the treatment effect by 0.5%. These numbers are consistent with the effect size found by previous studies: Ayres et al. (2009) report 1.2% and 2.1%, respectively, for two pilots comprising a total of 75,000 households; Davis (2011) reports an average reduction of 1.8% (N=750,000).

Overall, two aspects are noteworthy in this context: On the one hand, these programs demonstrate that feedback can be highly cost-effective and scalable (1% opt-out rate). On the other hand, treatment effects ranging between 1 and 3% also indicate that these programs do not prompt fundamental behavior changes. Given the low frequency of the feedback (monthly or quarterly), the lack of timely information and the low level of disaggregation, the contribution of individual behaviors to the total resource consumption remains opaque to consumers (Faruqui et al. (2010)). Moreover, the consequences of specific actions are not salient to consumers at the moment when they could actively influence them (Houde et al. (2013)).

Moving further to the right in the EPRI classification scheme, a series of recent studies have evaluated the effectiveness of daily/weekly feedback and real-time feedback on a large scale (Loock et al. (2013); Schleich et al. (2013); intelliekon (2011)). In these studies, households were able to access feedback information regarding their electricity use on web portals, phone apps or dedicated in-home displays (IHDs). Faruqui et al. (2010) presents a meta-study on IHDs and McKerracher and Torriti (2013) gives an overview of 27 completed trials on real-time electricity consumption feedback. Overall, existing feedback studies show a large variety with respect to the data they collect and present to users, frequency of feedback, content, level of aggregation / breakdown to end-uses, medium of presentation, inclusion of comparisons, and combination with additional information and other instruments (Fischer (2008)). Moreover, these studies largely vary with respect to sample size, duration, recruitment method, household data collected, and treatment effect (outcome) observed.

While earlier reviews report savings in the range from 5 to 15% (Darby (2006); Ehrhardt Martinez et al. (2010)), more recent large-scale trials measuring savings against a control group find more modest effects for in-home displays and web portals in the order of 1 to 6% (Schleich et al. (2011); Darby (2012); Schleich et al. (2013); Degen et al. (2013); Carabias-Huetter (2013)). A meta-study by McKerracher and Torriti (2013) analyzes 27 completed and 7 upcoming IHD trials with respect to treatment effect, sample size, recruitment method, year, peer-reviewed publication and type of IHD used. They find that earlier studies were characterized by smaller sample sizes, unrepresentative samples, a higher involvement by the study administrators, and more prone to Hawthorne effects than more recent studies. In contrast to the higher savings reported by earlier metastudies, McKerracher and Torriti (2013) conclude that *"3-5% is a more accurate expected conservation figure for a large-scale roll-out of*

IHDs". To evaluate the effect of recruitment method and sample size, the meta study classifies existing trials into four classes and weighs treatment effects by sample size: While the weighted mean conservation effect for studies with a representative sample ("class A") is 2.6%, for "class B" studies, i.e., studies with opt-in design, 100 or more participants and a low degree of involvement by the administrators, it is 4.5%; for "class C" studies, i.e. 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%. The forth category contains studies lacking information on sample selection or recruitment methodology.

The British research project AECON is a recent example for a large-scale study with opt-out ("class A") design. The study covers 60,000 households, including 18,000 with smart meters. It finds that the provision of an IHD generally reduces consumption by 2-4% and that the savings are persistent to the end of the trial (Raw and Ross (2011)). In the trial described by Schleich et al. (2013) with 1500 Austrian households (opt-in participation, "class B"), participants had the option to choose between access to a web portal and monthly paper-based feedback mailed to their household. 46% of the participants chose the web portal option; the two feedback options were equally effective (4.5% reduction on average). In a trial involving Swiss 5,000 households, Degen et al. (2013) find a treatment effect in the range of 3% for households with an IHD compared to the control group. By contrast, households who receive a mailed report comparing household's electricity use with a similar household do not reduce their consumption significantly in that study.

On the whole, while a number of earlier, smaller studies indicates a very promising potential for real-time feedback, these results may be subject to sampling bias. More recent studies with larger, more representative samples report treatment effects that are more modest, yet still higher than for enhanced billing. Existing large-scale studies on the impact of real-time feedback do not support its cost-effectiveness for large-scale deployments. Most studies report an initial user interest in the information, with many users exploring the information provided on web portals (resp. in-home display); yet this is followed by a strong decay in portal / IHD views in subsequent weeks (Goelz (2011); Degen et al. (2013)). Large-scale studies on behavior-specific real-time feedback ("real-time plus" in the EPRI classification scheme) are completely missing so far.

Effect Persistence 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) find that once the intervention ends, the effect eventually dissipates and households return to pre-intervention consumption levels. Similar, a Dutch study with 300 households who received a

IHD find that the savings persist neither in the households who return the monitor after the initial four-month study period, nor in those who keep it (van Dam et al. (2010)). By contrast, Ayres et al. (2009) report sustained savings throughout a seven- resp. twelve-month study duration. Raw and Ross (2011) also report persistent effects for electricity smart meters to the end of the AECON trial (depending on the group between one and two years). 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 show persistent savings over the course of the pilots (up to 21 months). In a study with 150,000 households receiving different types of home water use reports, Ferraro et al. (2011) provide evidence that the effect of social comparisons is more persistent than for simple appeals to pro-social preferences. The most comprehensive analysis on long-term effects available so far is carried out by Allcott and Rogers (2014), evaluating electricity consumption data of 234,000 households over four to five years. They conclude that savings are much more persistent than previously generally assumed: While the immediate response to the first home energy report is followed by a relatively quick decay, the authors observe 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 an 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.
Given the discrepancy in the findings regarding effect persistence, 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), which may represent a barrier to longer effect persistence. As a result, they conjecture that the future of real-time feedback lies in systems with data push.

Feedback Studies on Residential Water Use In the domain of residential water consumption, less feedback studies are available, but the conservation effects yielded so far are encouraging, also compared to other water demand management strategies. Three larger studies recently investigated the impact of feedback on residential water consumption (Ferraro and Price (2013); Fielding et al. (2013); Mitchell et al. (2013). All three of them provided mail-based feedback to households in a randomized controlled trial (similar to the home energy reports for electricity use). Ferraro and Price (2013) analyze the outcome of a large-scale mail-based residential customer conservation education program, a randomized experimental design with 100,000 households. They compare the effectiveness of social comparison messages with simple pro-social messages and technical information alone. They report that the strong social norm message a) had a much higher impact than technical advice, b) yielded a

4.8% reduction, about twice the impact as the effect achieved by similar programs on electricity conservation (Ayres et al. (2009); Allcott (2011b)), and c) that the program 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 and 13% relative to the control group. Mitchell et al. (2013) evaluates a 12-month pilot study with 10,000 households. All households in the treatment group received between four and seven home water reports with social comparison modules, similar to the Opower home energy report described in section 2.1.2. The reports were either delivered electronically (if an email address was available), or mailed otherwise. The study consists of two experiments: While the first experiment evaluates the impact of home water reports on a representative sample of households, the second experiment focuses on a more homogenous group of households, which - based on their household characteristics - had upfront been identified as good candidates for such interventions. The authors report a treatment effect of 4.6% for the representative sample and of 6.6% for the sample that had been pre-selected based on household characteristics. Notably, despite the campaign's effect on water conservation, household's knowledge on how much water they were consuming did not improve compared to the control group.

Inman and Jeffrey (2006) conclude in their review on water demand-side management programs that such campaigns could 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. Lee et al. (2011) analyzes 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 year; yet these figures are based on a comparison with the previous year, not an actual control group. Based on monthly data from 19,000 households, Campbell et al. (2004) investigates 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 might help overcome such offsetting issues. Fielding et al. (2013) lists several drawbacks for pricing mechanisms and mandatory approaches like equity issues involved with their implementation, limits to price elasticity, public resistance, the political will required for their implementation, and evidence that they do not necessarily result in long-term change. In their literature review on environmental behaviors, Kurz et al. (2005) only identifies five studies related to water conservation (out of 87 reviewed); in another meta-study on water demand-side management, Hurlimann et al. (2009) states 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.

Overall, while demand management using feedback has been less explored in the water domain than for electricity use, the results are encouraging: Treatment effects reported for feedback regarding water consumption are generally higher than for electricity. This may be partially due to the fact that water, unlike electricity, is visible and tangible, which makes it easier for users to identify high-impact behaviors (Darby (2006)). Real-time feedback on water consumption (aggregated and behavior-specific) is even more scarce than for electricity and practically nonexistent on a large scale.

All in all, feedback interventions have already established themselves as cost-effective, scalable, and relatively persistent tool to promote resource conservation and to increase energy efficiency quickly, at scale, and in addition to technological efficiency gains. While the conservation effect of the most widespread approach (i.e., enhanced billing) is rather modest (yet cost-effective), a number of smaller studies have indicated a promising potential for behavior-specific real-time feedback interventions. Yet so far, no study has demonstrated the cost-effectiveness of this approach on a larger scale.

In order to evaluate cost-effectiveness and scalability of these novel feedback approaches, it is also necessary to understand how these interventions work, on whom they work, and under which conditions. The following section will provide an overview of the related work from several disciplines that has approached these questions from different angles.

2.1.3 Factors and Mechanisms Driving Resource Conservation

Research on resource consumption is abundant, and so is the variety of answers on what drives and facilitates conservation behavior. Most existing research can broadly be grouped into two categories: On the one hand, studies collecting utility consumption data along with a basic set of household characteristics; on the other hand, survey-based studies collecting an extensive set of psychological variables (attitudes, beliefs, preferences, etc.), but only self-reported behaviors.

The first type of studies consists in a mainly quantitative approach, generally pursued by economists, to explain the variance in resource consumption data. Typically, these studies do not seek to develop a deeper understanding of underlying psychological motives. Based on large datasets of utility records, those studies seek to carve out observable or easily retrievable variables that can be used to predict households' response to utility programs. They mainly use information on demographics and similar characteristics that are relatively easy to obtain; creating personality profiles of consumers is more of a challenge with respect to data availability.

The second type of studies serves social scientists to build behavioral models to explain how the interplay of psychological variables drives human behavior behind the scenes. These models typically seek to explain (environmental) behavior as a result of problem awareness, beliefs, attitudes, and situational factors. Examples include the theory of reasoned action (Fishbein and Ajzen (1975)), Ajzen's theory of planned behavior (TPB, Ajzen (1991)), Schwartz' norm activation model (NAM, Schwartz (1977)) or Stern's value-belief-norm theory (Stern (2000)). Traditionally, pro-environmental behavior is described as a mixture of self-interest and altruism (Bamberg and Moeser (2007)). Based on an meta-analysis of 57 studies, Bamberg and Moeser (2007) developed a framework that combines rational choice models like Ajzen's theory of planned behavior with altruistic motives (as described by Schwartz' norm-activation model). Most of these models were developed based on self-reported behaviors.

Recently, social psychology and economics literature have converged in their answer for what drives pro-social behavior: Behavioral economics have augmented the neoclassical utility function (the willingness to pay for a need) by factors beyond wealth (Levitt and List (2007); Akerlof and Kranton (2010)), and both disciplines have integrated social norms (customary rules governing behavior in groups) into their models. As a result, both disciplines consider pro-social behavior to be governed by three motives: extrinsic (material self-interest), intrinsic (altruistic), and reputational (social or self-image concerns) considerations (Bénabou and Tirole (2006); Ariely et al. (2009)). More specifically, pro-environmental behavior is thus affected by rational-economic motives, intrinsic environmental concerns, and social-reputational aspects (Nolan et al. (2008); Goldstein et al. (2008); Griskevicius et al. (2010); Ek and Söderholm (2010); Allcott (2011b); Viscusi et al. (2011)).

However, both presence and strength of these conservation motives hinge upon individual and personal characteristics (Naderi (2011)). Therefore, this section will first provide an overview of household characteristics that are generally considered as relevant to resource consumption and conservation behavior, followed by a brief description of the three conservation motives identified by both behavioral economists and social psychologists.

2.1.3.1 Household Characteristics as Predictors for Resource Conservation

A variety of studies have investigated correlations between residential resource consumption or household's response to behavioral interventions on the one hand and situational factors, in particular household characteristics, on the other hand. As a result, a whole string of socioeconomic, demographic, and physical environment variables have been identified as important determinants of environmental behavior (Gregory and Leo (2003)). These typically include household size, income, education, age, gender, square footage, age of the dwelling, and ex-ante (i.e., baseline) consumption levels. The findings are very mixed; for cases like income or education, they even point into opposite directions. To give a few examples: Gregory and Leo (2003) reports a positive correlation between income and environmental activities.

Jorgensen et al. (2009), on the other hand, identifies a positive correlation between income and water consumption. Brandon and Lewis (1999) finds that income is positively correlated with electricity consumption; and despite being correlated with higher environmental awareness, higher income does not necessarily translate into a higher response to feedback programs. De Oliver (1999) states that substantial documentation indicates a positive relationship for both income and education with water conservation. Both Gatersleben et al. (2002) and Poortinga et al. (2004) find that higher levels of income are positively correlated with home energy consumption. While Poortinga et al. (2004) report that a higher level of education is related to lower home energy consumption, Gatersleben et al. (2002) find no correlation between level of education and home energy use. Gilg and Barr (2006) report that members of the following groups tend to engage more in conservation behaviors: elder people, females, home owners, people with lower income, with less education, and those living in small household sizes. By contrast, Raw and Ross (2011) find larger electricity savings for higher levels of income and education. This is just a small and arbitrary extract from a wide collection of papers to illustrate that the existing body of literature provides very mixed results on the correlation of demographics and other contextual factors with residential resource consumption and with household's response to resource conservation programs.

Although most of these factors are highly correlated among one another, in many cases, the factors are analyzed one by one, with the result that the findings are not conclusive.

2.1.3.2 Rational-Economic Motives for Resource Conservation

Traditional economics are based on a rational-economic model of behavior, predicting that consumers will engage in conservation motives when this serves their economic self-interest (Samuelson (1990)). Based on these models, primary instruments to influence resource consumption are pricing mechanisms, rebates, or tax credits for energy-efficiency improvements. In line with traditional micro-economic theory of environmental economics (Kolstad (2010)), Olmstead and Stavins (2009) describe price-based approaches as superior strategies to manage residential water demand compared to the most prevalent prescriptive or command-and-control regulatory approaches. Yet people are often fail to make privately profitable investments in energy efficiency, a phenomenon which is referred to as energy efficiency gap (e.g., Jaffe and Stavins (1994); Allcott and Greenstone (2012)). However, Jessoe and Rapson (2014) report that the combination of feedback information and price incentives can be used as an efficient behavior change tool. In their experiment, households who receive an IHD to monitor their electricity consumption in real time are three standard deviations more responsive to notifications for demand response events (temporary price increases) than the group that only receive notifications about the price increase The authors conclude that feedback can help to surmount the major obstacle to price-based solutions: imperfect information (i.e., poorly informed consumers).

However, price-based approaches have been criticized for being slow, costly, and politically difficult to implement (Allcott (2011b); Fielding et al. (2013)). Moreover, monetary incentives may change how tasks are perceived, with perverse outcomes: Financial incentives may weaken intrinsic motivation; once they are discontinued, agents might be even less motivated to pursue the intended outcomes (Bénabou and Tirole (2006); Gneezy et al. (2011)). Research on health behaviors suggests that monetary incentives can produce short-lived effects and be effective in limited circumstances for short and simple tasks, but not for behaviors that need be maintained over longer periods of time (Jochelson (2007)). Furthermore, rebound effects have been found to be able to substantially undermine the ability of price-based interventions to drive emissions reductions (Greening et al. (2000); Jenkins et al. (2011)).

Overall, while economic factors certainly play an important role in consumer decisions, price incentives tend to be costly or politically difficult to implement. In addition to that, they do not necessarily work as standard economic models predict (cf. "energy efficiency gap"), and are associated with potential drawbacks like rebound effects, crowding out of intrinsic motivation, and short-lived effects.

2.1.3.3 Intrinsic Motivation to Conserve Resources

Intrinsic motivation is defined as complying with personally held and deeply engrained ideals or as *"doing something because it is inherently interesting or enjoyable"* (Ryan and Deci (2000)): Curiosity, eagerness to explore, being active are part of human nature and do not require extraneous stimulation (Ryan and Deci (2000)). Individuals can directly derive personal satisfaction from comparing their behavior against personal goals (Harding and Hsiaw (2014)). Likewise, most individuals are generally motivated to engage in pro-environmental behavior as they inherently care about the environment (Naderi (2011)). Thus in the context of feedback, individuals may benefit from the pro-social component of living up to their environmental attitudes; or, they may generally enjoy measuring progress towards goals and tracking metrics about their life. This section will briefly outline the related work for both of these aspects. This is followed by a short overview of the role of intention and self-efficacy, which prior work has identified as relevant in this context.

Environmental Attitudes Research on the relationship between environmental attitudes and resource consumption or conservation is abundant. As a result, the goal of this section cannot be to give a complete picture, but to provide an overview of aspects that are particularly relevant to feedback interventions. As outlined in section 2.1.1, social sciences have for a long time emphasized the pivotal role of attitudes for behavioral outcomes. Numerous studies have shown that individuals can derive utility from resource conservation and pro-social behavior in general (De Young (1986); Bamberg (2003); DellaVigna et al. (2012)). Against this backdrop, intrinsic motivation has been identified as an essential driver of pro-social and pro-environmental behavior (De Young (1986); Bénabou and Tirole (2006); Ariely et al. (2009)). Environmental concern and pro-environmental attitudes are widespread: The vast majority of the Swiss population indicate that they are willing to protect the environment even if this implies higher personal costs and efforts for themselves (Diekmann et al. (2008)). Even 83% of Americans say that they agree with the statement that stricter laws and regulations are needed to protect the environment (Pew Research Center for the People and the Press (2009)).

Yet a number of studies found evidence that while attitudes towards energy conservation may be correlated with conservation knowledge and reported behaviors, they are not necessarily good predictors of residential energy and water consumption (Heslop et al. (1981); Neuman (1986); Gatersleben et al. (2002); Gregory and Leo (2003)).

Thus the role of environmental attitudes in the context of resource conservation is somewhat ambiguous. To some extent, this may be due to data availability: Most studies that collect detailed information on attitudes, preferences, and personality do not incorporate measurement data. In contrast, studies that investigate behavioral programs using utility consumption data typically don't have access to information on environmental attitudes and the like.

Tendency to Monitor Progress Towards Goals One of the central tenets of behavioral economics is the observation that people often fall short of their own expectations and don't act in line with their long-term preferences. Environmental behavior is one of the areas where this is particular true: *"Although, many people care about the environment and adopt goals to act in a sustainable and pro-environmental way, only a few people actually follow through with their good intentions"* (Gutsell et al. (2012)). A key concept in this context is the notion of self-control, individuals' mental effort to align her behavior with a preferred state. According to models of self-control, individuals pursue goals by using them as reference points to which they compare the current state (Gutsell et al. (2012)). In the absence of such goals and of indicators for the present state, people are often lead astray in their choices to behave in ways that are not in their best interest (Thaler and Sunstein (2009); Hansen and Jespersen (2013)). One of the key ingredients to self-control is the ability to track one's behavior: Failure to do so renders control difficult (Baumeister (2002)). Previous research also established that feedback is essential for goal pursuit (Fishbach and Finkelstein (2012)) and that individuals' goal orientation shapes feedback-seeking behavior (Ashford (1986); VandeWalle (2003)).

Technological progress - in particular, the ubiquity of smart phones and sensors - has made it increasingly easy to collect and monitor metrics about one's life. As of today, 69% of U.S. adults track a health indicator like weight, diet, exercise routine, or symptoms (Fox and Duggan (2013)). The broad and voluntary adoption of such technologies reflects individuals' preference to monitor their progress towards certain goals. Examples include watches to sense the heart rate, wristbands to monitor workout schedules and sleep habits, smartphone apps that keep track of the way individuals spend their money, etc. A common thread among all these technologies is that they provide people with data about themselves, evaluate their performance, and influence their subsequent decisions. Various media reports have described the "quantified self" movement as a major trend that has reached mainstream popularity (Bradley (2013); Hay (2013); Snyder (2013)). This tendency is not only in line with Peter Drucker's widely cited management philosophy *"If you can't measure it, you can't manage it"*, but also in line with social science theory on self-control and goal setting (Loock et al. (2013)) and with the nudging approach described by Thaler and Sunstein (2009), which will be discussed in section 3.4.3.1.

Overall, these insights show that people may derive personal satisfaction from tracking metrics about their life and that a majority of people already engages in such behavior. As a result, this aspect may play an important role for the effectiveness of feedback on resource consumption.

Intention to Conserve and Self-Efficacy Various social-cognitive models regard *intention* as an immediate antecedent of behavior (Ajzen (2002)) and as the best predictor of behavior change (Schwarzer (2008)). Examples include the theory of reasoned action (Fishbein and Ajzen (1975)), Ajzen's theory of planned behavior (TPB, Ajzen (1991)), or the behavioral model developed by Bamberg and Moeser (2007), which states that behavioral intention ultimately mediates the impact of all other psycho-social variables on pro-environmental behavior. Intentions are also considered to be highly correlated with behavior (Ajzen and Fishbein (1969); Eccles et al. (2006)). However, people often do not act in line with their intentions and various researchers have investigated the lack of consistency between intentions and actual behaviors (*intention-behavior gap*, Festinger (1957); Sheeran (2002)).

To explain the discrepancy between intentions and behavior, social scientists have supplemented the existing behavioral models with post-intentional factors. In particular, self-efficacy, the belief in one's own ability to complete tasks and to reach goals, has been identified as an important factor to bridge the intention-behavior gap (Schwarzer (2008)). The notion of self-efficacy was first introduced by Bandura (1977). As Judge et al. (2002) points out, it describes the same concept as the strongly related measures of self-esteem and locus of control. Bandura qualifies self-efficacy as the most important precondition for behavioral change and shows its strong influence on peoples' behavior (Bandura et al. (1980); Bandura (1997)). In the context of pro-environmental behavior, a meta-study by Hines et al. (1987) finds a moderate relationship for self-efficacy and pro-environmental behavior (r=.37).

Given the prominent role that intentions and self-efficacy play in many behavioral models and their correlation with other factors governing behavior, these two variables should be included in the analysis of psychological mechanisms driving behavior.

2.1.3.4 Social Norms and Social-Reputational Aspects of Resource Conservation

Social norms are widely held ideals that govern behavior (Bicchieri and Muldoon (2014)). They serve as customary rules that convey what is generally expected by society (or a subgroup thereof) and whether this norm is generally obeyed by other group members (Bicchieri and Muldoon (2014)). If both normative expectations and widespread compliance are in place, individuals generally seek to comply with the norm social, as compliance with social norms can bring rewards and positively enhance reputation, while transgressions may lead to informal sanctions and feelings of guilt (e.g., Heywood (2002); Garcia and Wei (2013)).

In both social psychology and behavioral economics literature, social-reputational aspects have been established as a central motive for energy conservation (Bénabou and Tirole (2006); Griskevicius et al. (2010); Allcott (2011b)). Levitt and List (2007) state that individuals derive moral utility from energy conservation, yet this is conditional on individuals' beliefs about the social norms. Yet just as individuals typically understate their resource consumption (Gardner and Stern (2008); Attari et al. (2010)), most individuals also tend to overrate their performance relative to the social norm (Kruger and Dunning (1999); Epley and Dunning (2000)). As a consequence, feedback can help in particular high consumers to correct their misconceptions downwards.

Social norms have been proven as powerful motives for towel reuse in hotels (Goldstein et al. (2008)), electricity conservation (e.g., Nolan et al. (2008)), the diffusion of solar panels (Bollinger and Gillingham (2012)), and the adoption of hybrid cars (Griskevicius et al. (2010)). Social norm-based feedback is also the most prominent feature of the "home energy reports" (resp. home water reports) sent to millions of U.S. households. The social comparison element typically consists of two parts: A *descriptive social norm*, stating how similar households in the neighborhood do as a frame of reference, and an *injunctive social norm*, which conveys a judgment on whether the household's behavior relative to the norm is socially approved or disapproved of.

Overall, social norms are key determinants of human behavior; accordingly, social normative feedback has been identified as an effective strategy to influence consumer decisions. Yet little is known whether these interventions affect individuals' private utility in a negative way (Allcott (2011b)). Both anticipated studies will cover social norms either in survey questions (e.g., participant's beliefs about other people's shower behavior, chapter 3) or as key feature of the feedback intervention itself (chapter 4).

All in all, human behavior is governed by a complex interplay of psychological variables and situational factors. Insights from sociology and psychology in particular constitute the foundation of behavioral interventions and are valuable sources for the design of feedback programs and technologies in order to maximize the impact of these instruments.

While the importance of particular factors may still be debated in the literature, there is at least consent that behavioral interventions can have an impact on human behavior. What is less clear, however, is whether these primary effects operate in isolation for the target behavior, or whether they entail repercussions on subsequent actions. The following section will present work related to that topic.

2.2 Behavioral Feedback Interventions: Side Effects

Despite strong evidence for the influence of behavior change in one area on consumer choices in other environmental domains (e.g., Thøgersen (1999b)), most studies investigate effects on the target behavior only (e.g., Schultz et al. (2007); Goldstein et al. (2008); Ayres et al. (2009); Ehrhardt Martinez et al. (2010); Ferraro and Price (2013)). Those studies that did investigate effects of an intervention on both a target outcome and side effects on other behaviors can be broadly grouped into two categories: positive and negative side effects. This section will first present related work on positive behavioral spillovers, followed by work on negative side effects (moral licensing).

2.2.1 Positive Spillover Effects

The concept of positive side effects of environmental campaigns is built on individuals' desire for consistency in their actions or at least the appearance of consistency (Festinger (1957)). Thus, many environmental campaigns are motivated by the assumption that "simple and painless" behavioral changes (such as turning off the computer monitor or printing double-sided) will lead to the adoption of higher-impact changes in environmental behavior (Thøgersen and Crompton (2009)). For instance, the UK government's Department for Environment, Food and Rural Affairs (DEFRA) recommends that *"[w]e need to promote a range of behaviours as entry points in helping different groups to make their lifestyles more sustainable - including catalytic (or 'wedge') behaviours if identified through research"* (DEFRA (2008)). Similar, the UK's Sustainable Consumption Round Table suggests that the best way to promote pro-environmental behavior *"is to drop new tangible solutions into people's daily lives, catalysts that will send ripples, get them talking, sweep them up into a new set of social norms, and open up the possibility of wider changes in outlook and behaviour"* (National Council and Sustainable Development Commission (2006)). The underlying idea of this positive spillover of environmental behavior is that the *"adoption of a particular behavior increases the motivation for an individual to adopt other, related behaviors"* (Thøgersen and Crompton (2009)), based on environmental values that foster feelings of moral obligation (Thøgersen (1999a)). Kotchen and Moore (2007) report a decrease in energy consumption by participants in a green electricity program who pay a price premium for each unit of electricity consumed. The magnitude of the effect, however, is within the range of the estimated price elasticity for electricity consumption;

hence the study cannot determine whether the response was due to the voluntary price premium or to positive spillover effects. A Danish study based on phone survey data (N=1,002) finds a positive spillover from recycling on packaging waste prevention (Thøgersen (1999b)). A later study with Danish consumers reveals cases of transfer of environment-friendly conduct between behavioral categories only in a limited number of possible instances and only of modest size; at the same time, they also identify a limited number of negative cross-lagged effects (i.e., two sets of correlations separated by a time interval) (Thøgersen and Ölander (2003)). A common theme among the majority of the existing studies supporting positive spillover is that they are based on self-reported survey data. Yet self-reported data are often criticized for their limited reliability and the limited insights they provide into real behavior and decision making (e.g., Krampf et al. (1993); Webb et al. (2003)). More recently, the concept of positive spillover from one simple environmental "entry point" behavior to a wider range of conservation efforts has generally become quite controversial (Thøgersen and Crompton (2009)).

2.2.2 Negative Spillover Effects: Moral Licensing

The first paragraph of this section will present the general context and an overview of negative behavioral spillover (moral licensing) in various domains, while the second paragraph focuses on evidence for these effects in the particular context of residential resource consumption.

Moral Licensing in a Variety of Behavioral Domains While section 2.2.1 described the related work on positive side effects of behavioral interventions, this part presents an overview of the growing body of research which suggests that on the contrary, the behavioral spillover may be negative: The adoption of a more environment-friendly choice in one domain may actually increase the likelihood of less environment-friendly behavior in other areas. In their meta-study on environmental behavior, Steg and Vlek (2009) report that *"factor analysis reveals that individuals are fairly inconsistent in their environmental behavior."* In general, although most individuals strive to see themselves as moral actors (Jordan et al. (2011)), they are tempted to act in ways that make them feel immoral (Merritt et al. (2012)). Moral licensing is defined as the phenomenon whereby *"people can call to mind previous instances of their own socially desirable or morally laudable behaviors,"* making them *"more comfortable taking actions that could be seen as socially undesirable or morally questionable"* (Miller and Effron (2010)). To study this phenomenon, Sachdeva et al. (2009) conduct three experiments looking into the effect of previous actions on donations and environmental decision making. The authors suggest that affirming moral identity leads people to feel licensed to act immorally, and they propose a framework of self-regulation that balances moral self-worth and the cost inherent in altruistic behavior. To investigate the behavioral antecedents of the moral licensing, Miller and Effron (2010) review previous studies on psychological licensing and suggest that three major conditions are associated with activating moral licensing: 1) the behavior is relatively unimportant to one's identity, 2) the behavior is framed as progress rather than commitment to a goal and 3) avoiding hypocrisy is of minor concern. All three conditions apply to pro-environmental behaviors and the way environmental campaigns are perceived by the public (see Crompton and Kasser (2009); Cornelissen et al. (2008); Thøgersen (1999a)). People's tendency to morally "trade" one environmentally friendly action for other less pro-environmental behaviors might even be reinforced by current environmental programs that frame environmental behaviors as interchangeable actions, e.g., the "Pick 5" campaign of the U.S. Environmental Protection Agency (EPA (2011b)), in which participants pledge pro-environmental actions that they pick from a list of items.

Evidence for moral licensing has been found in various domains of human behavior. The majority of studies that investigate moral licensing focus on racism (Monin and Miller (2001); Effron et al. (2009); Bradley-Geist et al. (2010); Merritt et al. (2012)), disclosure of conflicts of interests (Cain et al. (2005b,a)), donations (Strahilevitz and Myers (1998); Khan and Dhar (2006)), sexism (Monin and Miller (2001)), nutrition (Khan and Dhar (2007); Wilcox et al. (2009)), choices with different levels of cultural sophistication (Khan and Dhar (2007)), or the purchasing of luxury goods (Kivetz and Simonson (2002)). Moral licensing is not confined to related actions within the same behavioral domain, but has also been observed between behaviors that are not closely related (cross-domain moral licensing). Khan and Dhar (2006) find that the hypothetical choice of volunteering for one community service organization or another licenses participants to express a preference for a luxury good over a utilitarian one. Chiou et al. (2011) report increased smoking among participants who believe that they are taking a dietary supplement. Mazar and Zhong (2010) demonstrates in a series of laboratory experiments that individuals who are given the opportunity to purchase green goods are more prone to negative behaviors in other domains (in their study, stealing and lying). Kruger and Gilovich (2004) as well as Wilcox et al. (2009) find that the pure anticipation of a positive behavior can be sufficient to license morally less laudable behavior. They conclude that people are willing to give themselves credit for their good intentions, even without acting on them. Clot et al. (2011) investigates how intrinsic motivation affects participants' willingness to donate money to an environmental organization after a primary virtuous act (dedicating time to an environmental program) that was framed either as voluntary or mandatory. They find that moral licensing occurs among intrinsically motivated individuals facing mandatory conditions as well as among non-intrinsically motivated individuals under voluntary conditions. With the exception of two papers (Chiou et al. (2011); Conway and Peetz (2012)), all of these studies are laboratory experiments and have the shortcoming that behaviors exhibited there may not be reflective of typical behaviors outside the laboratory. In their reviews on environmental behavior and household energy consumption, Abrahamse et al. (2005); Wilson and Dowlatabadi (2007), and Steg and Vlek (2009) advocate the importance of examining real data and actual energy use. As discussed by Levitt and List (2007), experimental findings can only be extrapolated beyond the lab to a limited extent, since important factors influencing human behavior are fundamentally biased by the nature of laboratory experiments: scrutiny by others, the particular context of a decision, and how participants are selected. In line with Levitt and List (2007), Allcott and Mullainathan (2010) make the case *"to do the 'engineering' work of translating behavioral science insights [...] from the laboratory to the field of practice,"* arguing that this missing step would have high economic returns.

Similarly to the moral licensing effect, the term "rebound effect" is often used in the economic literature to describe net negative outcomes of energy efficiency increases. In contrast to moral licensing, rebound effects are rooted in neoclassical economic theory (see Greening et al. (2000); Jenkins et al. (2011) for extensive reviews). They describe phenomena that can be ascribed to substitution effects, price effects, and income effects (see Madlener and Alcott (2009) for a more recent overview of discussions and context). According to these effects, lower energy consumption (e.g., resulting from more energy-efficient appliances) results in a reduced cost of living and thus higher disposable income, allowing individuals to increase their consumption of these products or other ones that also require energy for their production or operation. These microeconomic mechanisms are driven by changes in supply and demand rather than by non-monetary, psychological mechanisms influencing individuals' decision making processes as in moral licensing.

Before turning into a "hot" topic of economic analysis today, rebound mechanisms were - and still are today - not taken into account in energy and emissions forecasting and analysis. Similar, despite the broad evidence for moral licensing in other domains, the topic is still absent in today's program evaluation and resource demand projections.

Moral Licensing in the Domain of Residential Resource Consumption Very recently, a number of field studies have investigated side effects in the domain of household utility use. Based on smart metering data from 2,500 Irish households (12 months), McCoy and Lyons (2014) report that exposure to time-of-use pricing and feedback information can lead to reduced investments in energy efficiency measures. While the authors do not settle on a specific mechanism, they conjecture that the phenomena observed might be due to negative spillover processes, namely that conservation efforts might give people an alibi to forgo investments. Jacobsen et al. (2012) analyze changes in electricity consumption in response to enrollment in a green electricity program with 910 participating households. They find that households that enroll at the minimum level increase electricity consumption by 2.5% (before-after difference between participants and nonparticipants). In a recent pilot carbon offset program with 30,000 customers, Harding and Rapson (2013) report evidence for increased electricity consumption after the adoption of a carbon-offsetting program and demonstrate the importance of careful program framing to avoid negative side effects. All three of these studies use field data, but have two essential limitations: First, they are restricted to within-domain licensing (e.g., reducing negative externalities of electricity consumption) and second, they both ignore or cannot exclude income effects as an explanation of their findings. Bento et al. (2010) study cross-commodity effects and the role of culpability in the willingness to prevent environmental harm. They use web-based contingent valuation in a framed field experiment, supplemented by real-money laboratory experiments. One of their main findings is that the moral licensing effect dominates guilt effects (moral cleansing). The findings of Bento et al. (2010), however, are limited by being based on a hypothetical scenario.

Altogether, the evidence from these studies implies that environmental programs targeting a specific behavior might actually yield a much smaller net CO_2 reduction than those reported by program evaluators who focus solely on the change in the target behavior. Even worse, such campaigns might result in a net negative CO_2 outcome, in which the CO_2 reduction of the target behavior is more than offset by higher CO_2 emissions in other environmental domains due to moral licensing.

So far, studies that empirically investigate cross-domain effects with respect to behavioral spillover and moral licensing are missing - not only in the resource consumption context, but also in other domains.

In conclusion, there is still an ongoing debate whether engagement in pro-environmental entails measurable positive or negative side effects. So far, empirical evidence from real-world settings is scare. As a consequence, this topic is still a blind spot in today's program evaluation and in resource demand and emissions forecasting analysis.

2.3 Green Information Systems

The previous two sections 2.1 and 2.2 outlined the concept of feedback interventions, presented a common taxonomy along with typical conservation effects and findings on the persistence of effects for existing large-scale studies, followed by a literature review on side effects. Overall, existing feedback studies reveal a trade-off between specificity of the feedback on the one hand, and program reach on the other: Enhanced billing, for instance, has been found to be very cost-effective and scalable - yet it is limited to utility consumption data at the household level, aggregated over one or several months. Real-time feedback on specific behaviors or appliances, on the other hand, addresses these issues - however, so far, pilot studies are limited to small trials and no large-scale program has been implemented so far, mainly for reasons of cost-effectiveness. Yet technological progress in the ICT domain could help to close the gap between program reach and specificity in the future, as the following sections 2.3.1 and 2.3.2 describe.

2.3.1 The Potential of Green Information Systems

The role of ICT in the sustainability context is often reduced to its negative effects and possible ways to reduce its ecological footprint (e.g., by increasing the energy efficiency of data centers). This is true both for the public debate and for the "Green IT" domain in Information Systems research (Watson et al. (2010)). ICT, however, is not only part of the problem, but could be an essential part of the solution to address the environmental challenges the today's society faces by *"informing beliefs, enabling actions, and transforming outcomes"* (Melville (2010)).

As of today, for the vast majority of consumers, feedback on their utility usage is still limited to yearly, quarterly or, at best, monthly bills aggregated to the household level. As a result, consumers are not able to make the link between the individual action its outcome on resource consumption. The opacity of this connection, aggravated by people's misconceptions about their resource use (Gardner and Stern (2008); Attari et al. (2010)) make it nearly impossible for consumers to focus on environmentally significant behaviors and take action to make an actual difference.

ICT can help to bridge this *"environmental literacy gap"* (Froehlich et al. (2010)) by making the impact and relevance of specific appliances and activities on resource consumption salient (Stewart et al. (2013)). Meta-studies on behavioral interventions consistently report that feedback is particularly effective when it is provided 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)). That way, processes of resource consumption become more transparent and controllable for the user ((Faruqui et al. (2010)). Despite numerous research reports which highlight in unison the potential of Information Systems to promote a cleaner future - see, for example, McKinsey's widely cited report (Boccaletti et al. (2008)), the report of the Economist Intelligence Unit (Unit (2008)), a study conducted by Cisco (Cisco (2008)), or a speech of the president of the European Commission (Barroso (2008)) - the vectors to realize this potential are still fundamentally underexplored (see sections 2.3.2 and 2.3.3).

The progress of information technology and the ongoing large-scale deployment of smart utility meters open up new possibilities to effectively bring resource consumption information to the attention of individuals in ways that effectively influence consumer decisions towards more sustainable behaviors (Froehlich et al. (2010)). ICT makes it possible to collect and to process high-resolution data on a large scale, which can then be used to provide tailored and specific feedback to end users on a large scale (Ehrhardt Martinez et al. (2010)). Compared to traditional billing or feedback that involves more basic technology, these measures are described as more costly (EPRI (2008); Fischer (2008); Schleich et al. (2013)) and as not very cost-effective (McKerracher and Torriti (2013)). Thus, ICT-supported feedback still needs to validate its superior ability to reduce (or shift) consumption (compared to traditional feedback tools) and demonstrate its scalability and cost effectiveness in the field (Darby (2010)). Aside from technological, legal and political challenges, this requires in particular a better understanding of user interaction and engagement ((Darby (2010)). Or as Williams (2011) phrased it in his Nature article, *"understanding the interaction of ICTs with economic and social systems presents significant and interdisciplinary methodological challenges. Grappling with such complexity is at the heart of modern society's emerging concern over sustainability"*.

Given the success of behavioral interventions even with traditional methods, coarse data, and substantial time lag, one can easily imagine that the combination of behavioral feedback and ICT - either with smart metering as a generic enabler technology or with novel ICT devices for dedicated applications - clearly hold the potential to play a much more significant role in the pursuit of environmental and geopolitical goals.

2.3.2 ICT as Facilitator of Advanced Feedback on a Large Scale

Utility bills reveal the true costs of individuals' actions only intermittently and with substantial time lag (Gilbert and Zivin (2013)). As a result, prices and quantities are opaque to the agent when the consumption is happening (Gilbert and Zivin (2013)). This lack of salience facilitates inattention and incomplete information processing (DellaVigna (2009)). This lack of attention in combination with agents' present-biased preferences leads to overspending (Harding and Hsiaw (2014)). Smart meters can help to overcome the salience problem in household electricity spending by making information on prices and quantities available to households in real time (Gilbert and Zivin (2013)).

Over the past few years, strong national efforts have been made across the globe to deploy smart meters: More than 50 million smart meters have been deployed in the EU alone, and industry expects 155 million devices by 2017 (BergInsight (2012)). While smart meters are also used for gas and water metering, they are particularly employed to collect electricity consumption data. Although they typically collect data in the range of 10 to 60-minute-intervals, most households continue to receive only monthly, quarterly or yearly information on their utility consumption. Among those households who do have access to the information collected by their smart meter (typically on a web portal or in-home display), many fail to use it (Raw and Ross (2011)) or use it less and less after an initial period of interest (Ehrhardt Martinez et al. (2010); intelliekon (2011); Degen et al. (2013)). Larger field studies on feedback using smart meters (or ICT in general) to reduce residential energy consumption are still quite rare. Jessoe and Rapson (2014) test the effect of smart metering information displayed by an in-home display on price elasticity of demand. In a framed field trial with 437 households and three conditions, they analyze the response of households to temporary price increases by 200-600 percent. 130 households ("price-only treatment") experience pricing events with prior notification; another 100 households received the same pricing program, plus real-time information about their electricity use provided via an in-home display ("price + IHD treatment"); another 207 households served as control group. While the price-only group reduce their consumption by 0-7% on average during pricing events, households in the group with IHDs cut their consumption by 8 to 22%. The study also finds evidence for persistent reductions in energy use beyond pricing events, which the authors attribute to household learning about disaggregated uses and habit formation. Allcott (2011a) evaluates customer response to price signals displayed by ambient orbs that change colors to indicate the current electricity price (N=693). The treatment group shave their peak energy consumption by 5 to 14%, without substantial demand shifting into off-peak hours, resulting in an overall reduction in energy consumption of roughly 5% compared to the control group.

Although smart metering opens up a series of new avenues which researchers and practitioners have only started to explore, industry's initial enthusiasm has considerably suffered from a number of not cost-effective pilot projects with rather disappointing results. Allen and Janda (2006) question whether direct real-time feedback presents a significant improvement over weekly or monthly data; at the same time, their findings are based on a small dataset of 10 households. Those households received a *"simple, commercially available"* electricity monitor, which the authors themselves contrast with *"technologically innovative and graphically stimulating feedback"* options. In a study with 5,000 Swiss households, Degen et al. (2013) analyzed the impact of an IHD displaying smart meter information to a randomly selected subgroup of 1,000 households. Households with a smart meter and IHD reduced their consumption by 3.2% reduction throughout the study duration (12 months). In their review of 26 existing feedback studies, McKerracher and Torriti (2013) report 4% as the best estimate of conservation effect from real-time feedback in the Australian context and conclude that *"providing real-time feedback may not be a cost effective strategy for reducing carbon emissions in Australia".*

At the same time, real-time feedback - and even more so feedback on the appliance level - is still in its infancy (Darby (2010)). Researchers have only started to develop more powerful methods and are increasingly able to identify household characteristics, characteristic appliances, or maintenance issues with good accuracy (Beckel et al. (2013)). This may pave the way for customized feedback and novel applications for user engagement which might considerably increase program impact, reach, and cost-effectiveness in the future.

2.3.3 Behavior-Specific Real-Time Feedback on Resource Consumption

There is broad evidence in the literature that ideal feedback should be provided at the point and time of consumption and disaggregated to individual behaviors and appliances (e.g., Ehrhardt Martinez et al. (2010); Darby (2006)). Yet as of today, real-time feedback at the appliance (or behavior) level is still in its infancy. While numerous prototype studies have been carried out in this area (see following paragraphs), only a few of these projects have made their way into a commercial and cost-effective product. Examples include plug-level electricity metering systems (e.g., by *Plugwise* or *Thinkeco*)and circuit-level metering systems (e.g., by *PowerHousehDynamics* or *TED*). Yet none of these products have been rolled out on a large scale.

The vast majority of ICT artifacts presented in academic literature still has not overcome prototype status. Many of the pilot studies presented lack a research hypothesis, a meaningful sample size, or a clean research design. In the electricity domain, as one of the earliest studies on disaggregated real-time feedback with randomized group assignment, Dobson and Griffin (1992) conducted a 60-day field trial with 100 Canadian households, 25 of which received energy- and cost-related real-time feedback on cost and energy consumption of various electric end uses. They find a 13% reduction for households receiving real-time feedback, compared to the control group. Ueno et al. (2006) and Ueno et al. (2003) provide an on-line energy-consumption information system to 10 resp. 8 households. The system visualizes 30-minute consumption data of individual appliances and, in Ueno et al. (2006), information on space heating. In the earlier study (without feedback on space heating), households reduce their energy consumption by 9% (Ueno et al. (2003)), in the 9-month follow-up study (incl. space heating) by 12%. Wood and Newborough (2003) compare the impact of appliance-specific real-time energy consumption feedback, information (e.g., energy saving tips), and a combination thereof in the context of domestic cooking. They find an average 15% reduction for households with a dashboard displaying information on current and previous energy use for cooking (N=19), compared to a 3% reduction for those who only receive an information package (N=9). While the magnitude of the effects is encouraging, these results warrant caution due to their small sample size and in particular due to the lack of a clean research design.

Studies in the water domain using ICT artifacts to promote sustainable behavior still concentrate on establishing a proof of concept / operation and on interface design. They generally describe 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) designed "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) present "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. Other studies are limited to surveys that evaluate respondents' intention to use systems with certain characteristics or compare participants' initial reactions to different forms of feedback information. Froehlich et al. (2012) for instance evaluates concepts of water eco-feedback displays in an online survey with 651 respondents and conducts semi-structured in-home interviews with 10 families. They find a strong preference for specific, detailed information about water usage at the individual fixture level, for a breakdown by hot / cold, and for comparison data for contextualization (with self-comparison as the most preferred). The study, however, is limited to hypothetic intention-to-use questions and does not include real-world usage data.

Despite a plethora of innovative concepts to visualize feedback on shower behavior, all these studies share the limitation of a very limited 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 participating in the second part of the study with an alarming visual display device 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.

Overall, from a theoretic point of view, there is a compelling case for behavior-specific real-time feedback, as presented in the literature on nudges (Thaler and Sunstein (2009)) or the model of inattentive choice by Taubinsky (2013). Small-scale pilots with feedback on specific actions or appliances also indicate a large potential for that kind of disaggregated feedback. In practice, however, these applications still need to prove their cost-effectiveness, scalability and effectiveness on a large scale and in clean experiments.

2.4 Research Gaps

Numerous studies have been carried out on residential resource consumption and on behavioral feedback interventions in this domain. Based on the existing literature, this thesis will particularly address research gaps in the following four areas:

Evidence for the Feasibility and Cost-Effectiveness of Real-Time Feedback Several meta-studies have concluded that feedback works best if it is delivered frequently, timely, clearly, and on specific actions which individuals can easily influence (Darby (2006); Fischer (2008); EPRI (2009)). The progress of information technology has repeatedly been hailed as a game-changer. In particular systems that provide feedback on specific behaviors in real time could make processes of resource consumption more salient, transparent and controllable for the user (Faruqui et al. (2010); Ehrhardt Martinez et al. (2010)). Yet so far, evidence for the feasibility, scalability, and cost-effectiveness of this kind of feedback on a larger scale is still missing.

The study in chapter 3 seeks to address this gap by deploying a behavior-specific real-time feedback system in a larger number of households. Beyond representing the largest study on behavior-specific real-time feedback to date, the study also collects detailed survey information to evaluate to what extent the results can be extrapolated from the sample recruited to the general population.

The Psychological Mechanisms Driving Resource Conservation While behavioral interventions have been rolled out to millions of households, the underlying psychological mechanisms that drive people to adopt efficient technology or to engage in curtailment behavior are not well understood (Allcott (2011b); Allcott and Mullainathan (2012); Ferraro and Price (2013); Allcott and Rogers (2014)). According to Gregory and Leo (2003), *"Although research in environmental behavior is abundant, past studies attempting to link psychological variables to conservation behavior are thought to have produced mixed findings and are considered inconclusive"*. For instance, there is a long and still ongoing debate in the literature whether pro-environmental attitudes, a priori intentions and self-efficacy are good predictors - or even prerequisites - for individuals' response to behavioral interventions. Existing studies either lack access to measured data on behavior (relying on self-reported data instead), or lack access to detailed information on individuals' preferences, attitudes, and personality.

One of the key questions in this context is through which psychological channels behavioral interventions operate. Based on the existing literature, two hypothesis are of particular interest: The first hypothesis is that individuals are coerced into conservation behaviors by psychological pressure, be it intrinsic or social. In that case, negative sensations, in particular the anxiety to fulfill intrinsic or social expectations, would create the impetus to saving. As a consequence, individuals who are more susceptible to pressure should be particularly responsive to this kind of intervention. The second and alternative hypothesis is that feedback can operate through positive channels, that it helps people to act on their goals and preferences: Behaviors like showering are to a large extent governed by automatic processes and not subject to individuals' conscious decision making. By making behavioral outcomes salient, people are enabled to evaluate this information against their goals and interests and to make deliberate decisions in line with their preferences. In that case, the intervention should be particularly effective on individuals with strong pro-environmental preferences or on individuals who in general tend to monitor progress towards goals. Another open question is to what extent users actively process the information provided by behavioral interventions, and whether feedback helps individuals develop a better sense for their resource consumption.

Depending on the underlying mechanisms, behavioral interventions may thus increase or decrease individual utility: They may increase individuals' utility if they help individuals to act in line with their preferences; or they may potentially undo all of the welfare gains of reduced resource consumption if they coerce individuals into conservation behavior through psychological pressure.

The importance of identifying these mechanisms has been expressed in a number of recent studies (e.g., Allcott (2011b); Ferraro and Price (2013)). This thesis explores these questions with the study in chapter 3. This can support the development of more effective strategies for the promotion of resource conservation behaviors and to minimize perverse outcomes. Moreover, it helps evaluate to what extent these programs are cost-effective on a large scale and amenable to wide-spread adoption.

The Role of Household Characteristics Evidence on the influence of household characteristics and demographics is not conclusive. Despite a broad body of research, the findings are very mixed and even point into different directions (McKerracher and Torriti (2013)). While some studies only analyze self-reported behaviors, others only collect very limited and incomplete information on participating households and individuals. Yet many of these variables are highly correlated among one another; as a consequence, the relationships reported might be biased and it is not clear which of the (unobserved) variables genuinely moderates resource consumption or the conservation effect. Depending on which subset of variables has been analyzed, spurious correlations may have been reported. The extensive literature review carried out for this thesis was not able to identify a single study so far that both measures behavioral outcomes and collects survey data on a larger set of relevant key variables that generally help explain the variance in environmental behavior or in the response to behavioral programs.

The study in chapter 3 of this thesis will collect both fine-grained utility consumption measurements and extensive survey data: This combination makes it possible to analyze the variables identified as important predictors of resource consumption and response to feedback interventions in one comprehensive model. The anticipated results can be used e.g., for profiling to optimize the cost-effectiveness of feedback programs.

Side Effects (Behavioral spillover Effects) The evaluation of behavioral programs is generally limited to the target behavior addressed by the program. Yet past research has shown that changes in one domain can also affect other behaviors. Many environmental campaigns are even motivated by the assumption that "simple and painless" behavioral changes will catalyze positive changes in other, more environmentally significant areas. On the other hand, resource conservation in one area might make people more careless or wasteful elsewhere. So far, it is still unclear whether a pro-environmental behavior change in one area is a) contained to that target area, b) whether it leads to a cross-domain adoption of additional environment-friendly behaviors (positive spillover) or c) whether it induces reduced engagement elsewhere (negative spillover, in particular moral licensing). Although a growing body of literature has found broad evidence for moral licensing, most of these studies have been carried out in a laboratory setting or are based on self-reported behaviors in surveys. Also, cross-domain effects (e.g., from water use to electricity consumption) have so far received relatively little attention. Chapter 4 of this thesis addresses this gap by investigating side effects of a behavioral intervention on resource consumption in other domains in real households

To explore these research gaps in a real-world setting, two separate randomized controlled trials are designed and implemented. In both studies, water and energy (resp. electricity) measurements of 154 resp. 697 households serve as the dependent variables. These two studies will be described in detail in chapter 3 and 4 of this thesis.

Chapter 3

Field Study with Amphiro Shower Meters

This study explores the effectiveness of behavior-specific real-time feedback on energy and water use in the context of showering. It further investigates the underlying psychological mechanisms and the influence of household characteristics.

3.1 Motivation

One of the end-uses that is clearly underrepresented in the promotion of energy conservation measures is water heating. Yet water heating is the second-largest energy end use in households, accounting for 12-18% of residential energy consumption (BDEW (2010); Umweltbundesamt Deutschland (2013); eia (2013a); Prognos AG (2013a); BAFU (2013)). As figure 3.1 shows, this is equivalent to 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 (2013a)), or 2,500 kWh per home.

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 3.2).

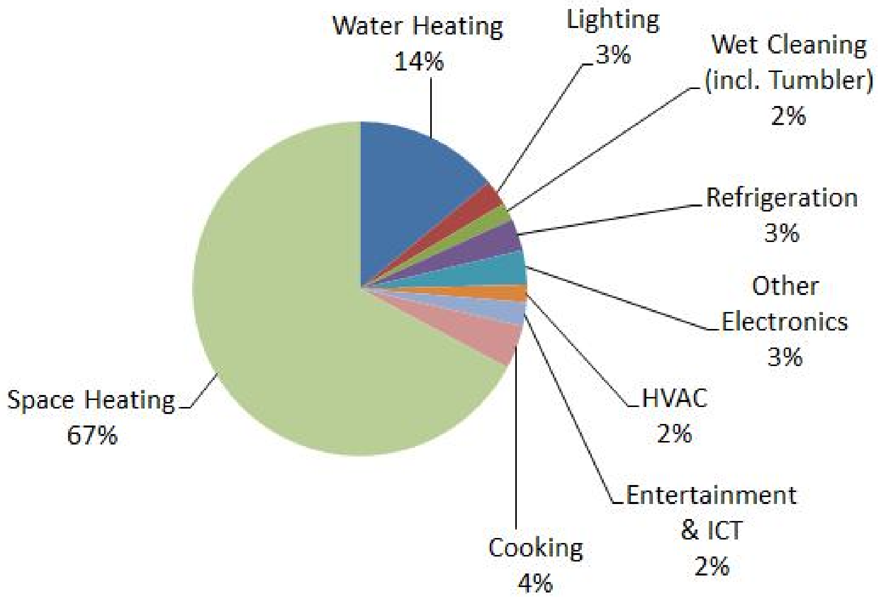


Figure 3.1: Energy end uses in Swiss households in 2011. Source: SFOE data, BAFU (2013)

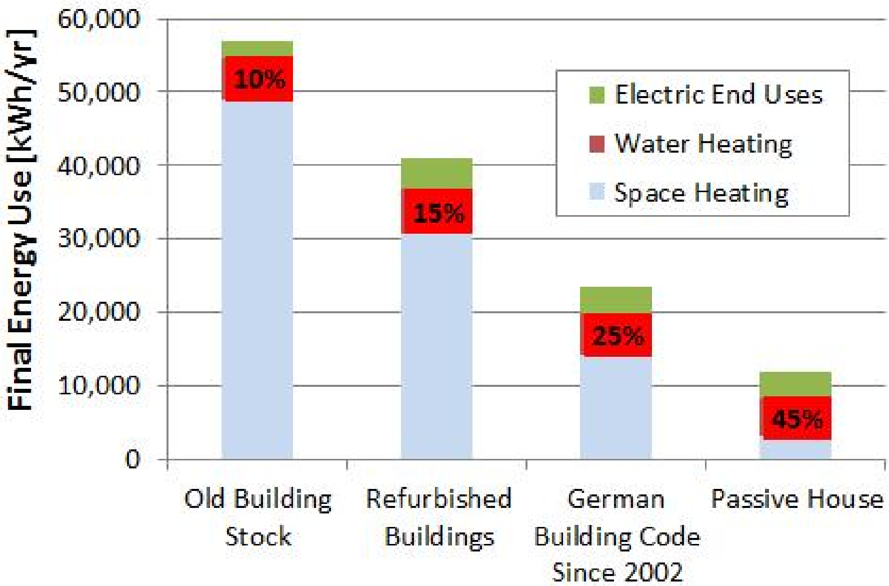


Figure 3.2: Residential energy end use by building type. Source: ASEW (2008)

As the breakdown by energy sources in figure 3.3 shows, the majority of Swiss households rely on fossil fuels for water heating: 40% use fuel 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 (2013a)). This implies that the carbon footprint of water heating in Switzerland (per kWh of thermal energy) is nearly twice as high as per kWh of electricity: Electricity in Switzerland is mainly produced with fossil-free resources - 56%

hydro power, (SFOE (2013b)) 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)).

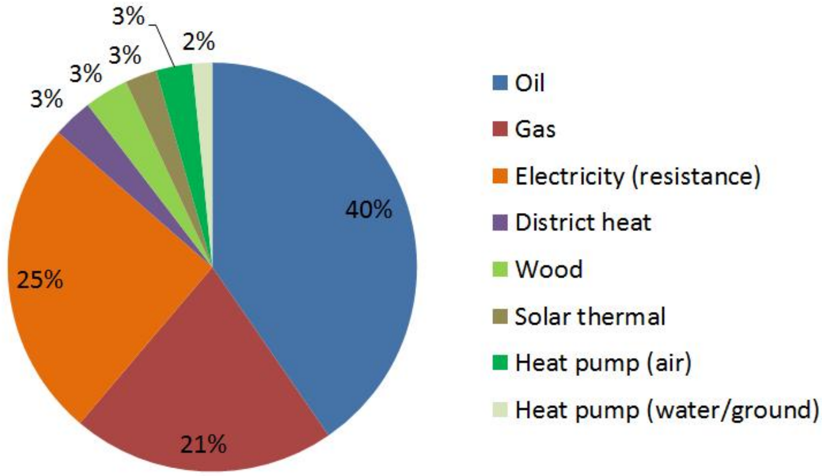


Figure 3.3: Hot water generation in swiss households by fuel type. Source: Prognos AG (2013a)

The diversity of fuels used for water heating might be one of the reasons why so far, water heating has been rather neglected in behavioral interventions compared to programs that target electricity consumption alone. It could also be the fact that hot water consumption is barely quantified as of today. One of the main barriers to this are 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 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 it is argued that water conservation in Central Europe leads to congestion issues in sewage systems and to higher water tariffs. Although this may be true for cold water, the conservation of hot water clearly makes sense given the large amount of incorporated energy.

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

The combination of favorable conditions for feedback and the large amount of energy that water heating consumes has motivated this study. In this context, the study investigates how

Characteristic	Explanation
Frequency	Feedback is provided to users every time they take a shower
Control	Users can easily influence their energy and water consumption in the shower. Whereas energy consumption of e.g., a refrigerator largely depends on technical parameters, user behavior is key for resource 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
Personality	Showering is typically carried out by a single individual
Visibility / Tangibility	In contrast to electricity use, water consumption can generally be perceived with human senses
Simplicity	Consumption processes are relatively straightforward (no standby loads) and in metrics that are easy to understand (liters in particular)
Fewer distractions	No cell phones, etc. in the shower
User action required	In the case of ampiro a1, the device requires user action only once (installation). Subsequently, the feedback is provided automatically, users do not need to actively log in to a web portal or make sure that the device is powered

Table 3.1: Particular characteristics of showering making it an excellent candidate for real-time feedback

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.

3.2 Methodology

This section describes the implementation and data collection procedure of this study. The study collected both detailed survey data (before and after the intervention) and detailed resource consumption measurements in the shower. Altogether, 697 participating households were recruited among a larger sample of 5,000 ewz-customers who had previously completed the *ewz Studie Smart Metering*. Households were randomly assigned to three 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 displayed different feedback content in the shower. 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 3.4 shows a snapshot of the shower meter display and the device in its position between the shower head and the shower hose; a more detailed device description follows in section 3.2.3.

Figure 3.4: On the left: display and body of *amphiro a1* - on the right: device installed between showerhead and showerhose

3.2.1 Partners and Collaborators in the ewz-Amphiro Study

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.

3.2.2 Timeframe and Recruitment

Timeframe The field deployment phase of the study lasted from early December 2012 to early February 2013. In the preceding months, the researchers adapted the user manual and the website for each feedback condition (see section 3.2.4) 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 helped the research team from ETH Zurich address study participants' questions and issues.

Recruitment for the study took place as the *ewz Studie Smart Metering* phased out, in staggered (bi-)weekly batches between September and November 2012. All devices were shipped on November 29/30, 2012. The packages contained a return envelope with prepaid postage and shipping address for the readout at the end of the study.

Subsequent to the two-month field deployment phase, all study participants were asked by email 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. The data readout, device reconfiguration and reshipping procedure was completed in April 2013. Thereafter, the individual datasets were merged, anonymized and analyzed and prepared for dissemination in peer-reviewed journals, international conferences, and local workshops.

Recruitment Participants of this study were recruited among 5,000 participants of the *ewz Studie Smart Metering* (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 the survey, 697 households were selected. Households with more than two members could not be admitted: As the study devices could only store data of up to 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 one-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 one-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: The device is designed for handheld shower heads; it cannot be installed in wall-mounted showers or body sprays
- Approval of the conditions of data privacy protection statement
- Stated willingness to (temporarily) ship the device back after two months for the data readout

The 697 selected 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 participate in a (separate) long-term study instead.

3.2.3 Description of the Feedback 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 feedback directly 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 batteries and allows tracking behavior over extended periods of time. Showers can be interrupted to three minutes (e.g., for lathering up) to be stored as one coherent shower (otherwise the device restarts from zero); water extractions below five liters are not considered as showers and are not stored: The underlying assumption is that most of these occurrences serve other purposes (e.g., for flower watering or bathtub cleaning). During each shower, the device continuously measures the water temperature and generator speed. Based on these data, the water consumption, energy consumption and energy efficiency class of the current shower are permanently calculated. The standard device displays real-time feedback on water and energy consumption since the begin of the shower, water temperature, and an energy efficiency class; the latter is visualized by a letter ranging from A to G and a polar bear animation. Standard devices can store 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 3.2.4 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. (2013b). 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 the temperature, see section 3.2.4
- 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 3.2.4.

3.2.4 Research Design

In order to answer the research questions outlined in section 2.4, a 2 (household size) x 3 (feedback content) randomized controlled trial was carried out over two months. Participants were randomly assigned to one of the following three display content conditions: In a control condition, the device was installed in the shower, but only displayed water temperature, with no information on the volume of water used or duration of the shower. In the two experimental groups (*real-time information* and *real-time plus past information*, respectively), information about the current water and energy use was provided. One of the two experimental groups (*real-time plus past information* condition) was additionally exposed to information about the water use in the previous shower. In both experimental groups, during the first 10 showers, the device displayed only the temperature, as in the control group; the full information display was activated from the 11th shower to obtain a baseline measure during the first ten showers. In the user manual of the treatment, 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. Control group participants were informed that different display contents were being evaluated as part of the study, allowing participants to test different display versions during and after the study.

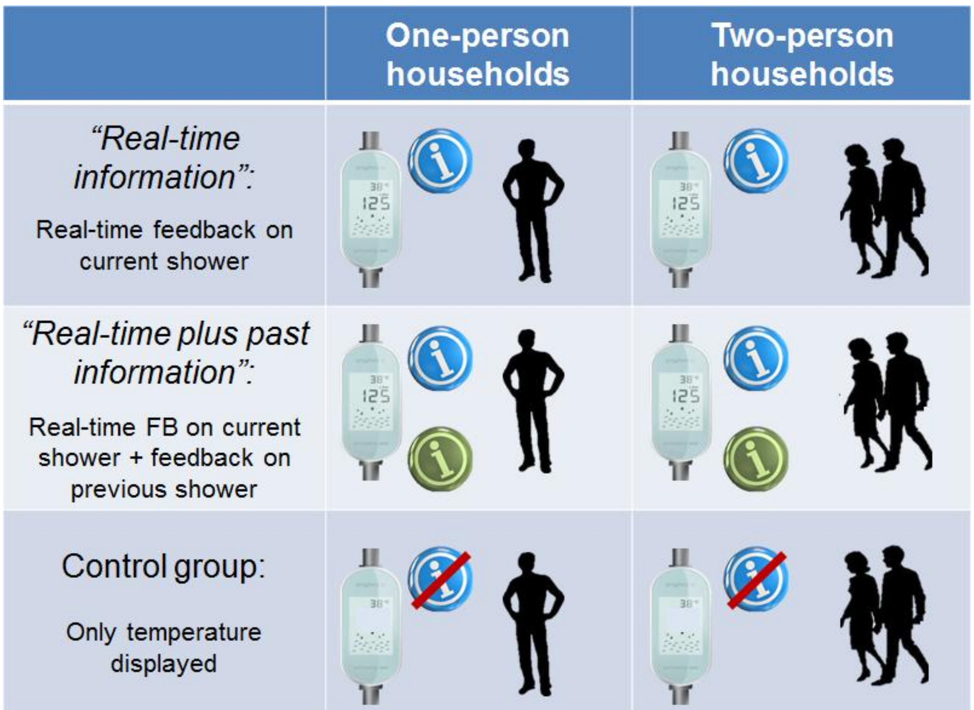


Figure 3.5: 3 x 2 research design

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.

3.2.5 Participation Rate

Overall, 5,919 households were invited to participate in the study. 1,348 filled out the initial 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 one-person households (the initial goal being to have 50% one-person households in the study).

At the end of the study, 685 households shipped their device back (98.3%). 636 devices 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.

3.2.6 Data Collected and Data Handling

This section describes which and how data were collected and processed for this study. It outlines the key terms of privacy protection and gives an overview of the shower measurement and survey data collected. This section does not yet describe in detail the different steps undertaken for the statistical analysis: This will be done in section 3.3.1, which outlines the empirical strategy for the data analysis.

3.2.6.1 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 covered the scope of data collection, the concept for data storage, data processing, and data deletion. The survey also explained that 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.

3.2.6.2 Survey Data

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 3.2.2. Two questions covered whether the household pays for water and heat energy based on its consumption, or whether they paid a fixed rate or rent independent or 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 numbers regarding their own shower behavior (water volume per shower, water temperature, duration), also relative to other study participants.

The final survey inquired about extended periods of absence by study participants and the use of the shower by guests (to check whether the classification of a household as one-person- or two-person-user-household was valid). Otherwise, the survey mainly focused on participants' experience with the smart shower meter: readability, comprehensibility, and information content of the display elements; another self-estimate of their shower 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. This will be discussed in section 3.4.4.2.

In addition to the two surveys designed specifically for this study, it was also possible to use data that had been collected previously in the *ewz Studie Smart Metering* by the University of Lausanne: In the entry survey for the ewz-Amphiro study, respondents were asked whether they agreed that their data from the *ewz Studie Smart Metering* 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*. Questions of particular relevance for the ewz-Amphiro study included personality traits (measured with the HEXACO model of personality structure), environmental attitudes, preferences regarding goal pursuit, comparisons with other, among other topics.

3.2.6.3 Shower Data

In addition to the data recorded by standard devices for every shower taken (water consumption and average temperature), study devices additionally stored the duration of each shower, as wells as the duration and number of interruptions. This reduced 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 Tiefenbeck et al. (2013b). Energy consumption was calculated under the assumption that no energy losses occur.[1].

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 regarding temperature, volume, and flow rate, shower #1 substantially deviated from the typical pattern both with respect to temperature and volume: An unusually high number of households had only extracted between 5 and 10 liters and at lower temperatures. Probably, a rather large fraction of participants simply turned on the water after the installation of the device to check

[1]Section 3.4.1.1 contains a more detailed calculation of the energy consumption per shower, including an assessment of the energy savings that take into account the energy losses.

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.

3.2.6.4 Logistics and Data Matching

In order to be able to match the different datasets of each participating household (see section 3.2.6.2), survey respondents had to enter 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 the devices and shipped them to the participants, each parcel containing a prepaid reply envelope and a treatment-specific user manual.

After two months of deployment in the field, participants received an email asking them to temporarily return their device for the data readout. 685 households shipped their device back for the data readout. Members of the research team read out each device individually using a readout terminal with a webcam (for details see Tiefenbeck et al. (2013b)). In the course of that process, the serial number of the device was scanned and linked to the corresponding household's study ID (see section 3.2.6.1). The read-out process included the following tasks:

- Visual data read-out in the read-out terminal (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, equipment check esp. for missing small parts (o-ring seals, sieve)
- Repackaging for reshipping (new envelope with the correct address label)

3.2.6.5 Filtering

In order to ensure that the results were not driven by measurement errors or extreme outliers, several measures for data quality assurance were carried out. 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), defective devices were easy to detect.

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 with the device was used only by one of the household members (separate bathrooms). Conversely, many one-person households in particular reported frequent visits by partners, friends, or family. As a consequence, shower users changed over time, which was violating the assumption of a single person not interacting with other household members. Altogether, 102 households with similar inconsistencies were flagged. While they were included in the overall assessment of the treatment effects, they were excluded them from the analysis of the psychological mechanisms. The remaining dataset contained 524 households and data of 39,024 showers.

A third measure was to analyze the influence of outliers on the results. For that purpose, averages and standard deviations both of shower temperatures and of water volumes were calculated 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. However, 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 was 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.

Weather As seasonal fluctuations might also affect shower behavior (e.g., a severe temperature outside drop might lead to more extended showers), weather data from Zurich for the study period were collected from the publicly available website www.freemeteo.com. As figure 3.6 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.

3.3 Results

The main goals of this section are to quantify the effect of the intervention on shower behavior, to evaluate how individuals realize these savings and to understand the underlying psychological mechanisms. For that purpose, the set of shower panel data was combined with an extensive set of survey questions. The structure of this section is the following: The section first gives overview of the overall empirical strategy pursued and present participants' evaluation of the shower meter. This is followed by a description of participant's shower behavior before the onset of the treatment and by tests that verify whether the randomization process has successfully produced balance on observable key characteristics. Section 3.3.5

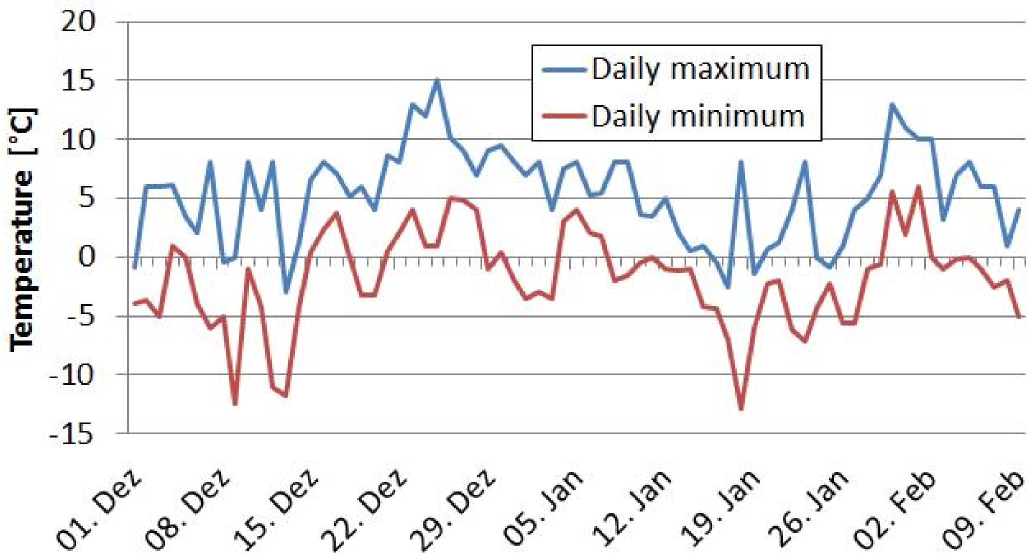


Figure 3.6: Ambient temperature in Zurich during period of study

then quantifies the main treatment effect before section 3.3.6 analyzes the underlying psychological mechanisms.

3.3.1 Empirical Strategy

The first step consisted in assessing participants' overall evaluation of the device (section 3.3.2). The second step analyzed energy and water consumption before the onset of the intervention (section 3.3.3), presenting descriptive statistics on key variables. Moreover, the analysis assessed correlations between shower behavior in the baseline period and a range of variables of interest that previous studies had associated with heterogeneity in residential resource consumption. In a third step, randomization checks were performed with these key variables to ensure that all groups exhibited the same shower behavior and did not differ in any other key trait observed (section 3.3.4).

The forth step investigated the main treatment effect. The corresponding section 3.3.5 first provides a visual impression of the shower-to-shower means by treatments (section 3.3.5.1). Then a difference-in-differences strategy was used to obtain the causal effects unconfounded with time trends that affected all treatments alike (section 3.3.5.2). A fixed-effects model was estimated using ordinary least squares (OLS). Furthermore, the study evaluated how the participants put these changes into practice by assessing which shower parameters they changed (duration, temperature, flow rate, etc.).

Finally, the study investigated the underlying psychological mechanisms in several steps (section 3.3.6). As a first step, in order to evaluate whether the treatment induced active information processing and learning, the association between participants' estimated water use per shower and their actual consumption was evaluated. The analysis thereafter examined how the treatment interacted with a set of variables of interest, at first separately for each variable of interest (for comparison purposes) and then in one single joint model.

3.3.2 Device Evaluation by the Participants

As figure 3.7 shows, the device was generally rated very favorably by the study participants.

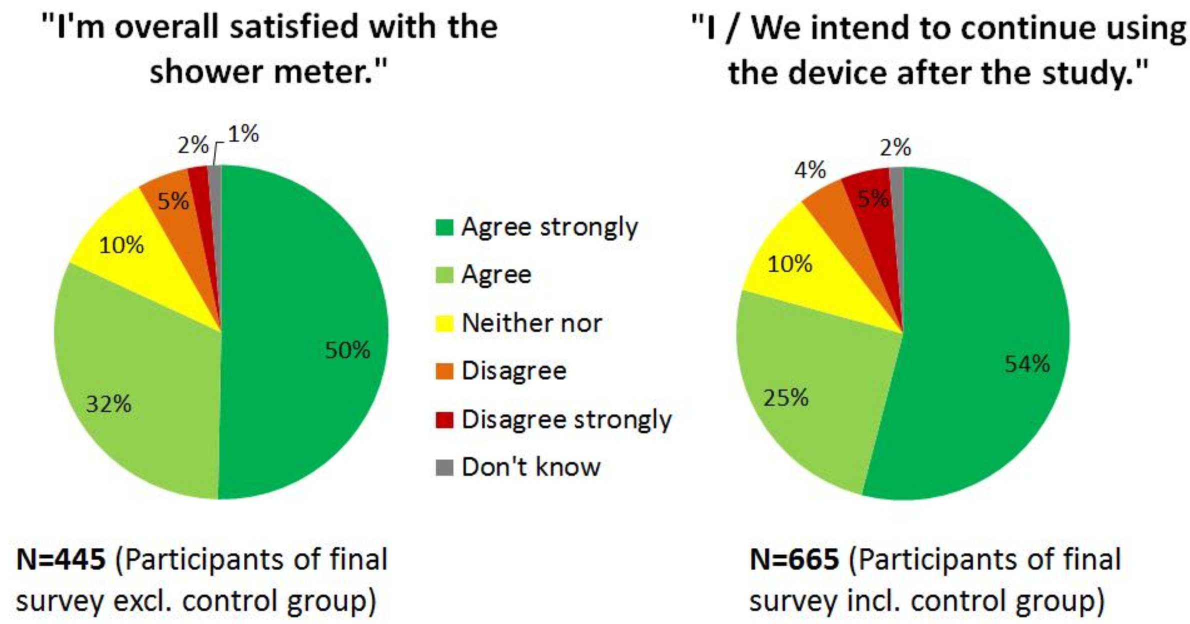


Figure 3.7: Overall user satisfaction and intention to continue using the device after the study

Overall, 82% of the treatment group (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. In addition to questions on overall device evaluation and intent to continue using the device, participants of the final survey were also confronted with a list of semantic scales. The semantic differential method is widely used in psychology to assess attitudes (Dickson and Albaum (1977)). Each scale consisted of two polar opposite adjectives that was separated by a seven-point rating scale, e.g., "boring - entertaining", "emotional - rational". For each bipolar pair, participants were asked to indicate to what extent they rather associated the shower meter rather with the adjective on the left or on the right, or if they equally associated it with both characteristics (neutral level). The location of positive and negative characteristics was varied; while internally, integer levels ranging from -3 to +3 were used to distinguish the seven choice options in the analysis, these numbers did not appear on the survey, as this might bias participants towards choosing the side with the positive anchors (Heise (1969)).

Table 3.2 shows the results of the individual regressions of each semantic scale on the experimental group. In the first column with the semantic pairs, the semantic pole towards which both groups leaned is marked in bold; the second column contains the control group mean, the constant of the regression; the third column shows the correlation coefficient and the forth column the corresponding standard error; the fifth column contains the corresponding p-value. Across all dimensions, both treatment and control group always rated the device on the same side of the neutral pole: rather comprehensible than confusing, rather clear than fuzzy, rather precise than inaccurate, rather effective than ineffective, rather helpful than useless, rather pleasant than unpleasant, rather relaxing than stressful, rather inventive than ordinary, rather interesting than boring, rather inspiring than annoying, rather novel than familiar, rather rational than emotional, and rather entertaining than unexciting. While some of these results might be subject to some extent of social desirability bias and should be treated with caution, they still reveal first tendencies: Treatment group members perceived the device as significantly more effective, helpful, inventive, interesting, and entertaining; they also rated it as significantly less rational and less relaxing. The less rational aspect is probably due to the display of the polar bear, adding a more emotional component to the device. While the treatment group perceived the device as less relaxing than the control group (probably due to the constantly increasing consumption figures on the display), they did not rate the device on the stressful side of the spectrum. This is a first indicator that the device did overall not inspire stress or pressure. Also, the treatment group rated the devices as just as pleasant (in the sense of comfort, "angenehm") as the control group.

Given the recruitment strategy (see section 3.2.2), 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 receiving such a device: They unconditionally received it as a thank-you gift, independent of their participation in the *ewz-Amphiro study*. Therefore, 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 have purchased the device on their own, would look like.

3.3.3 Baseline Consumption

This section provides descriptive statistics for showers during the baseline period, i.e., before the onset of the treatment. It also describes correlations between resource consumption during the baseline period and variables that had previously been associated with heterogeneity in household resource consumption in the literature. Table 3.3 presents a set of descriptive

Semantic pair	**Constant**	**Regr. coef.**	**Std. error**	***p*-value**
Comprehensible - Confusing	-2.27	+0.16	0.09	0.08*
Fuzzy - **Clear**	+1.82	-0.01	0.12	0.91
Precise - Inaccurate	-1.48	-0.17	0.10	0.09*
Effective - Ineffective	-0.56	-0.75	0.13	0.00***
Useless - **Helpful**	+1.11	+0.57	0.12	0.00***
Pleasant - Unpleasant	-0.66	-0.08	0.13	0.54
Stressful - **Relaxing**	+0.50	-0.37	0.10	0.00***
Ordinary - **Inventive**	+1.05	+0.47	0.11	0.00***
Boring - **Interesting**	+1.50	+0.39	0.10	0.00***
Annoying - **Inspiring**	+0.69	+0.12	0.11	0.24
Novel - Familiar	-1.01	-0.20	0.13	0.12
Emotional - **Rational**	1.19	-0.48	0.12	0.00***
Unexciting - **Entertaining**	+0.71	+0.41	0.10	0.00***

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

Table 3.2: Semantic differential to assess perceived characteristics of the shower meter. A negative (resp. positive) mean indicates stronger support of the adjective on the left (resp. right)

statistics of the dataset. The dataset on its own is already of interest: To the best of our knowledge, no other study has ever collected so fine-grained shower data from such a large number of households. Baseline data are of particular interest, as they are not "disturbed" by the intervention yet. However, as section 3.4.1.4 will point out, this period might be most subject to Hawthorne effects.

As the first column of table 3.3shows, the average baseline shower lasted slightly more than four minutes, it used roughly 44 liters of water per shower and 1.6 kWh of energy (without accounting for losses, see section 3.4.1.1). For each shower parameter, the table also provides the standard deviation values for the entire dataset, between households, and within households.

The histograms in figure 3.8 show key shower characteristics for the baseline period (i.e., up to shower #10) over all experimental conditions: a) the distribution of water volume used per shower, b) average water temperature, c) implied use of energy, and d) water flow rates. As subfigure 3.8a and subfigure 3.8c illustrate, distributions of water and energy consumption vary widely between showers. While some individuals used as little as 10 liters of water per shower, others consumed over 100 liters. Ranked by their per-shower consumption, high users (90th percentile and above) used over seven times as much energy and water as low users (10th percentile and below). As water and energy consumption are highly correlated ($\rho = 0.993$ - see section 3.3.5), the variance in energy used per shower is equally high. By

Variable	**Mean**	***SD* overall**	***SD* between**	***SD* within**
Volume [l]	43.9	36.6	26.5	25.4
Energy [kWh]	1.6	1.4	1.0	1.0
Temperature [°C]	36.2	4.4	2.8	3.4
Flowrate [l/s]	11.0	2.5	2.3	1.0
Showertime [s]	246.5	196.7	137.6	144.8
Breaktime [s]	33.7	58.2	36.1	45.8
Showerstops [·]	.5	.8	.5	.6
No. of showers [·]	74.6	40.2	40.3	0

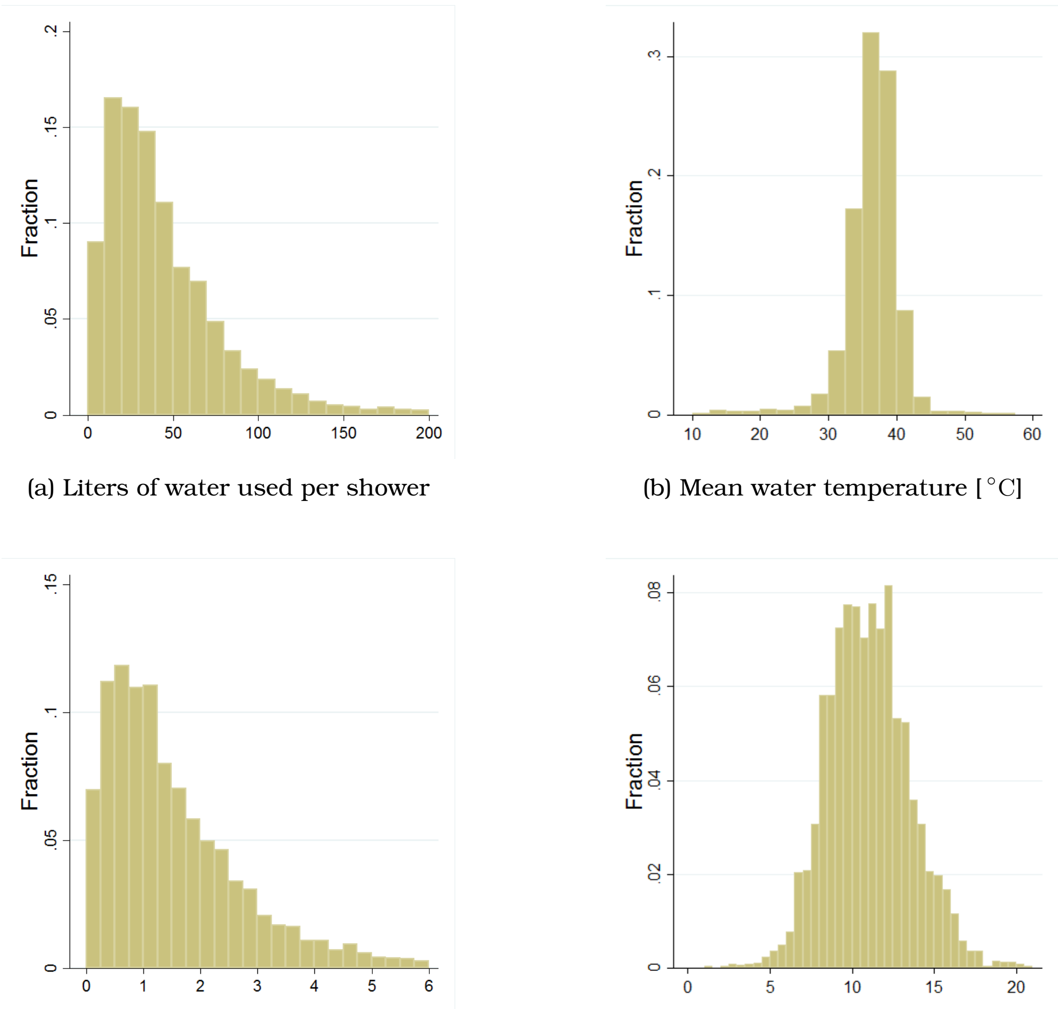
Table 3.3: Descriptive statistics of baseline consumption panel data

contrast, 89% of showers fall into the temperature band between 31 and 41 °C, and 91% of showers were taken with flow rates between 7 and 15 liters/minute.

Table 3.4 presents correlations between baseline energy consumption per shower and a range of variables that had previous studies had associated with variance in household resource consumption. Most importantly, the table shows that age was highly correlated with baseline water consumption. The correlation between age and energy consumption is negative: Older individuals tended to use less energy per shower. The simple correlation is equal to -0.318, and highly statistically significant. Using the underlying regression from the reported correlation, this implies that raising age by 10 years reduces predicted energy consumption by 0.25 kWh per shower (and water consumption by 6.3 liters). Figure 3.9 visualizes the correlation between the baseline consumption for water and age.

The effect is quantitatively important and relevant for several reasons; sections 3.4.2.9 and 3.4.3.6 will return to this point. Age was also highly correlated with participants' agreement with the statement in line (5) in table 3.4 that measured environmental attitudes: The correlation is statistically highly significant and positive, implying that older people rated themselves as more willing to protect the environment even at high personal costs and efforts. On the other hand, age is highly negatively correlated with the tendency to quantify one's performance or to compare it with other people. The effect is highly significant in both cases, as section 3.4.3 will discuss.

Strikingly, individuals' self-reported environmental attitudes are de facto strongly correlated with shower behavior: The stronger the self-reported pro-environmental attitudes, the less water and energy was used per shower. The correlation is equal to -0.207 and statistically highly significant. Again, the effect is also economically large: A one-point increase on the 1-to-5 Likert scale for environmental attitudes is associated with 0.29 kWh less energy consumption, and 7.7 liters less water use. On the other hand, individuals with a high baseline consumption did seem to realize that they could save water and energy: There is a strong



(a) Liters of water used per shower

(b) Mean water temperature [°C]

(c) Energy used per shower [kWh]

(d) Flow rate [liters/min]

Figure 3.8: Frequency distribution of four key shower characteristics (baseline period)

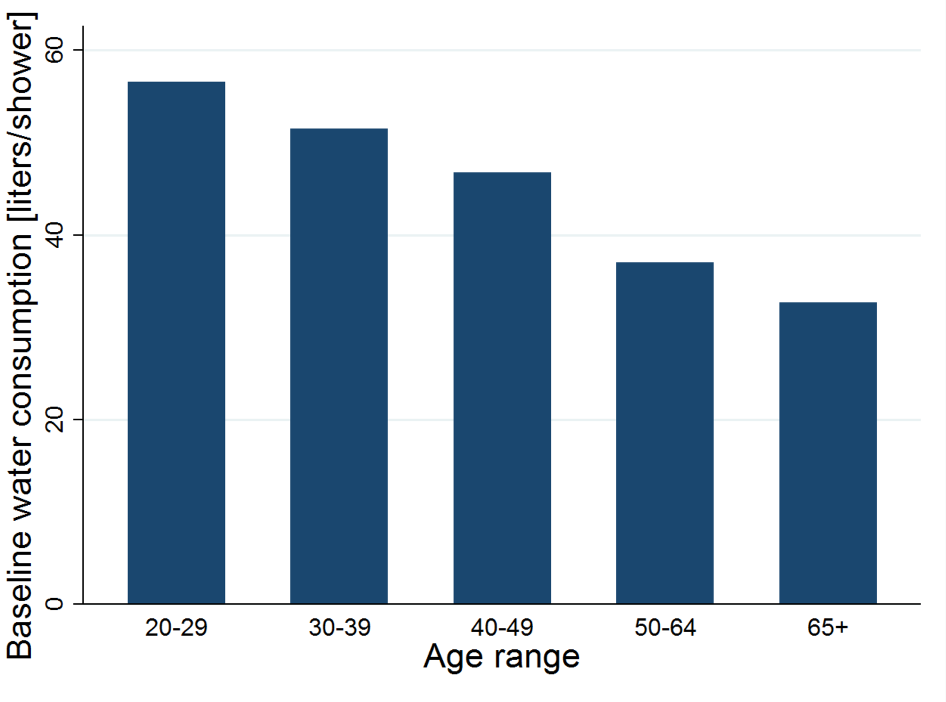


Figure 3.9: Water consumption per shower by age group (baseline means)

positive correlation between the degree of agreement with the statement that the household could save water and baseline consumption.

Table 3.4 also shows that household structure (one-person vs. two-person household) hardly affected shower behavior. Counter-intuitively, the fraction of female inhabitants was negatively correlated with water and energy consumption per shower; the correlation is significant, but the effect is small. Remarkably, the personality trait conscientiousness was neither correlated with age, nor with environmental attitudes, nor did it affect baseline shower behavior. This point will be picked up again in section 3.3.6 in the analysis of the psychological mechanisms.

3.3.4 Randomization Checks

It is only possible to establish a causal link between the intervention and the behavioral outcome if the random assignment of the participating households has successfully produced balance on various characteristics across the three experimental conditions. Therefore, a series of randomization checks was performed to verify if observable characteristics were identical across the three conditions before the onset of the intervention. The following equation was estimated:

$$y_i = \alpha_i + \beta_1 T_{1i} + \beta_2 T_{2i} + d_t + \epsilon_i \tag{3.1}$$

	Baseline consumption (in kWh)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age (1)	-0.318							
	(0.000)							
Number of household members (2)	-0.018	-0.043						
	(0.659)	(0.295)						
Fraction female 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 a high personal cost (5)	-0.207	0.236	0.087	0.169	0.046			
	(0.000)	(0.000)	(0.031)	(0.000)	(0.251)			
Propensity to monitor 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)	
Conscientiousness (8)	-0.0654	-0.0346	-0.0334	0.1255	-0.0656	-0.0415	0.2500	0.0876
	0.1417	0.4429	0.4528	0.0071	0.1405	0.3576	0.0000	0.0528

Table 3.4: Empirical correlations of energy usage and socio-demographic variables

where the dependent variable y_i stands for the variables from table 3.4, namely mean energy use during the baseline phase and various individual-level traits that are potentially important for this analysis. T_1 and T_2 are indicators for the *real-time information* and *real-time plus past information* conditions. A shower fixed effect d_t was included to capture trends in the best possible way.

Table 3.5 displays the regression coefficients for the individual regressions along with the p-value of the (two-sided) hypothesis that the correlation is zero (in parenthesis). It shows are only minimal differences between the means of the three treatments; for none of the variables examined, the mean is significantly different between the three groups. This means that the random assignment to the groups has had its desired effect: Most importantly, there were no differences with respect to energy use per shower during the baseline phase. Thus, before the onset of the intervention, all groups had the same shower behavior. As can be seen in the row displaying the p-values of the regression model, one cannot reject $\beta_1 = 0$ and $\beta_2 = 0$ for any of the variables. The constant term of each regression is also of interest, it represents the mean value of the control group for each dependent variable.

3.3.5 Main Treatment Effect

This section presents the main treatment effects in several steps: It first provides a qualitative impression of the treatment effect by illustrating shower-to-shower means of energy consumption in each of the three conditions. Then a difference-in-differences method was used to compare the change of the mean consumption from the pre- to the post-intervention period in the treatment groups to that in the control group. In a next step, a fixed effects model was set up and estimated using OLS. The models was also used to analyze user strategies to reduce their energy and water consumption.

3.3.5.1 Qualitative Analysis: Shower-to-Shower Means

As a first step to evaluate the overall treatment effect over time, figure 3.10 illustrates the shower-to-shower means of energy consumption for each of the three conditions.

Up to the 10th shower, the device in all groups displayed only water temperature. As figure 3.10 shows and as described in the previous section 3.3.4, the three groups did not differ significantly in their baseline energy use per shower any other observed characteristic. At shower 11, the display was activated in the two treatment conditions. With the onset of the real-time feedback, energy (resp. water) consumption of the experimental groups instantly dropped - by roughly 0.343 kWh (resp. 9.9 liters of water) per shower, as section 3.3.5.2 will show. Relative to the baseline mean of 43.9 liters and 1.6 kWh, this represents a 22% (resp. 23%) reduction. The figure also suggests that the effect - the overall distance between the control group and the treatment groups in the graph - was remarkably stable over time and did not show any tendency to decrease throughout the duration of the study. However, the figure also indicates

	Baseline consumption	Age	Fraction female	Household structure	Environmental attitudes	Monitor goals	Compare with others
Real-Time info (=1)	0.000	-0.011	-0.017	-0.028	0.079	0.019	0.173
	(1.000)	(0.935)	(0.686)	(0.544)	(0.396)	(0.823)	(0.121)
Real-time and past info (=1)	0.092	-0.072	0.008	-0.013	0.148	0.030	0.111
	(0.391)	(0.580)	(0.848)	(0.783)	(0.108)	(0.725)	(0.318)
Constant	1.597***	4.037***	0.683***	0.530***	3.411***	3.328***	2.591***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	0.002	0.001	0.001	0.001	0.004	0.000	0.004
Observations	596	581	535	596	579	565	572

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

Table 3.5: Randomization checks

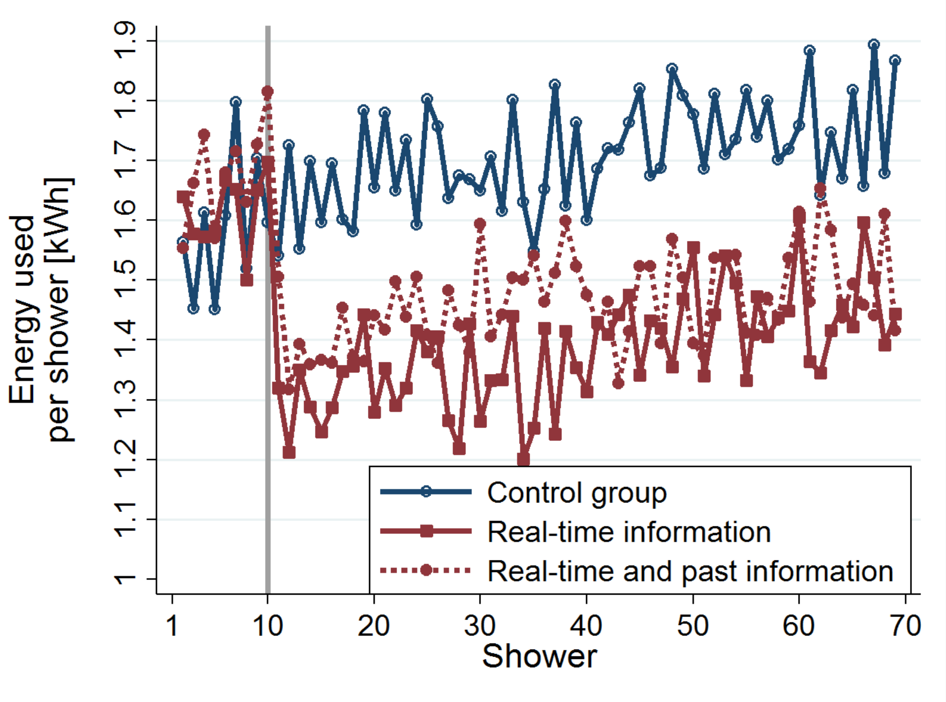


Figure 3.10: Shower-to-shower means for energy usage by experimental condition.

that the per-shower consumption of the control group slightly increased over time. This might be attributed to what is referred to as *Hawthorne effect* in the literature; section 3.4.1.4 will return to this point. In any case, the fact that the control group's consumption seemed to drift over time suggested that a difference-in-differences strategy would be a better approach to assess the true treatment effect. This was the next step of the analysis.

3.3.5.2 Difference-in-Differences Strategy

A difference-in-differences method was used to compare the change of mean consumption from the pre- to the post-intervention period in the treatment groups to that in the control group by calculating

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

where for each household i, $\overline{y_{i0}}$ is the mean energy (resp. water) consumption during baseline period, and $\overline{y_{i1}}$ is the mean value of shower 11 until the last observed value. In this approach, between-subject heterogeneity does not affect the standard error of the variables, as it is canceled out in the comparison, allowing for a valid estimate of the treatment effect.

Figure 3.11 confirms the qualitative findings of figure 3.10. It visualizes the difference-in-difference estimates by showing the mean change from the pre- to the post-intervention separately for each group; one-person-households, two-person-households, and households with an unstable household composition (described in section 3.2.6.5) are displayed separately.

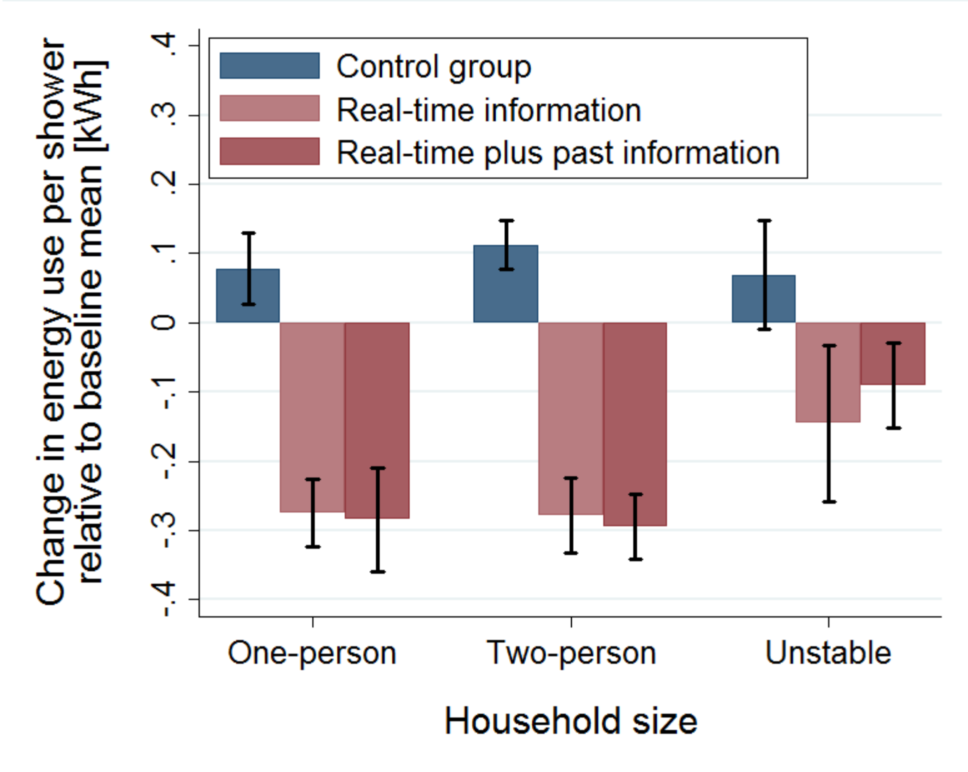


Figure 3.11: Difference-in-differences estimates of the main treatment effect. Black error bars indicate 95% confidence intervals.

The figure visualizes a modest, yet mostly insignificant increase in the energy consumption of the control group over the study period. By contrast, it shows a strong decrease in both treatment groups, for of each type of household. The true treatment effect is the difference between the increase in the control group and the decrease in the treatment groups, and the standard error bars around the means already indicate that the reduction due to the treatment is highly significant. As the figure also shows, the treatments worked approximately equally well for one- and two-person households, and equally well for the *real-time information* and the *real-time plus past information* condition. However, the graph also indicates that the treatment appears not to have worked so well in the unstable household group (see section 3.2.6.5).

In order to quantify and to properly test the indications of the treatment strengths that the visualizations above indicate, a statistical model was set up that makes it possible to fully take advantage of the experimental setup. A fixed effects model was used to control for time-invariant characteristics that were unique to the individual household (Torres-Reyna (2007)). The participating household had, for instance, different kinds of shower heads which equally affected the shower flow rate (and thus water and energy consumption) throughout the study. These characteristics were unobserved, time-invariant, and unique to every household. Thus, a fixed-effects model of the following form was estimated

$$y_{it} = \alpha_i + \beta_1 T_{1it} + \beta_2 T_{2it} + d_t + \epsilon_{it} \tag{3.2}$$

where y_{it} is the dependent variable (e.g., water consumption in liters, energy consumption in kWh, or shower time in seconds) of shower number t in household i. The individual fixed effect α_i eliminates all variance stemming from time-invariant unobserved characteristics (e.g., different types of shower heads). T_{1it} and T_{2it} are binary (dummy) variables, which 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* condition, respectively. A shower fixed effect d_t was also included to capture time trends that equally affect all households and treatments alike. Equation 3.2 was estimated using OLS; in that estimate, the residuals were allowed to be correlated within a household in arbitrary ways. This was corrected for by reporting standard errors clustered at the household level. The treatment effects were estimated separately for each category of households: one-person households, two-person households, and households with an unstable / unclear household composition.

Table 3.6: 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

The results are provided in table 3.6. It presents treatment effects on water volume (in liters, columns 1-3) and on energy (in kWh, columns 4-6) used per shower. As the table shows, the activation of the display had a strong and significant impact on the amount of water and energy used per shower, as was already to be suspected from the graphs above. However, in this case, the standard errors were calculated in a reliable way and are valid for inference. Again, the results show that the treatment effect was similarly high for the *real-time information* and the *real-time plus past information* condition. This is true both for one- and two-person households. This is notable, as in many cases, the information on the previous shower contained information about the behavior of the other household member. One could imagine that this might have fostered competition among household members or induced a larger reduction effect among two-person households, as both members knew that their consumption would be visible to the other household member. Yet apparently, the provision of information on the previous shower did not have an additional effect, neither in one-, nor in two-person households. The 2nd and 5th column show the outcome for the unstable households, for which the treatment appears less effective than in genuine one- or two person households, but nevertheless jointly significant (p=0.07 for water, p=0.07 for energy, both not reported in the table). It is difficult to interpret the weaker results for this group. 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 had different shower patterns, this strongly affected the outcome of that period and of the intervention overall: While some periods of the dataset (e.g., the baseline period) only reflect the shower behavior of one household member, later (or earlier) periods 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 moment and duration of their visit, their shower behavior affected the outcome in irregular ways. In all of these cases, the occurrence of visitors/absent household members was erratic. As these erratic events take place both in the treatment and control group, as well as during both baseline and intervention period, they add "noise" to the dataset and reduce the treatment effect. It might also be the case that more "marginal" members of the household did not really care about the device or exhibited some other characteristic behavior that made them less responsive to the treatment. As no other evidence supports that water consumption differs between household types, the first explanation seems more plausible (see table 3.4).

The results confirm once more the considerable magnitude of the effect: As can be seen from the constant, the average water (resp. energy) use per shower in the control group was 43.9 liters (resp. 1.6 kWh). The intervention reduced water use per shower by almost 10 liters (resp. 0.36 kWh), i.e., almost by a quarter. One should keep in mind that the figures given for energy are just a product of the total volume, the temperature gradient and the heat capacity, thus not yet taking into account any heating, transportation, or storage losses. Thus, they only represent the lower bound of actual energy savings (section 3.4.1.1 will provide a more comprehensive calculation). But even these figures show that the shower meter can lead to behavioral changes that amount to roughly 5 percent of a household's daily energy consumption.

User Strategies to Reduce Energy and Water Consumption The treatment merely displayed information on energy and water consumption in the shower, yet it did not prescribe or recommend in any normative way which consumption goals users should achieve. Even more, users were not given any recommendation on how to adapt their behavior. In order to reduce water their consumption, they could employ three basic strategies (or any combination thereof): a) shortening their shower time (by reducing time spent in the shower and by not turning on the water before actually getting into the shower), b) reducing the flow rate (manually or by installing a mechanical flow restrictor valve or low-flow shower head), while possibly maintaining the same shower duration, and c) turning off the water during the process, while they are soaping or shaving. In order to reduce their energy consumption, they could engage in any (or all) of those, plus d) reduce water temperature.

The most commonly chosen path of policy intervention focuses on the second strategy: Imposing standards and regulations to shower head manufacturers in order to reduce or limit the shower flow rate. Therefore, the question arises which strategies individuals mainly adopted to reduce their consumption by nearly a quarter on average, when given the choice and when not confronted with any prescriptions.

The results are displayed in table 3.7. Panel A, columns 1-3 contain the regression results with shower duration as the dependent variable. The average shower duration in the control group (constant term) was 246 seconds. The treatment lead to a sharp reduction in time between 45 and 55 seconds (roughly -20%), depending on the experimental group and household type (again, the effect was weaker for households with an unstable composition). Columns 4-6 in panel A show the treatment effect on water flow rate, measured in liters per minute (only when water was running). Here, the control group mean (constant term) is roughly equal to 11 l/min. The regression results indicate a modest, yet statistically significant, reduction in the flow rate of roughly 0.3 l/min (-2 to 3%) in one-person households. While the regression coefficients are also negative in the other household types, one cannot reject the null hypothesis of constant flow rate in those groups. In any case, the effect size for the flow rate is at best very small.

Panel B of table 3.7, columns 1-3, shows the treatment effect on mean shower temperature. A lower temperature - at constant water volume - would reduce the energy consumption of showering. The constant terms shows that the control group average was 36 °C. Again, there is a small, yet significant, reduction for single person households by 0.4 to 0.7 °C. For two-person households, the point estimates are also negative, yet fail to be significant. Overall, neither flow rate nor water temperature appear to be important margins of reduction. Columns 4-6 in panel B of table 3.7 examine how the intervention affected interruptions of the water flow during the shower. Turning off the water while applying soap or shampoo or while shaving could also reduce water and energy consumption. The constant shows that the control group interrupted their showers for 30 and 35 seconds on average. This time, the largest effect appears in the group of two-person households, (reduction between 5 and 9s); the effect for one-person households is also positive, yet fails to be significant. But also this margin of adjustment does not appear to be particularly important.

To summarize, the activation of the display showing information on the water and energy use during the shower dramatically affected user behavior, reducing energy and water consumption in both treatments by nearly a quarter. The results show that the vast majority of the reductions was achieved by a large and significant reduction of shower time. While the point estimates for flow rate, water temperature, and shower interruptions all point in the direction of conservation efforts, their effects are small and in many cases fail to make a clear and important contribution to the overall outcome observed. As a result, water and energy are very highly correlated ($\rho = .993$).

3.3.6 Psychological Mechanisms and Household Characteristics

In addition to evaluating the main treatment effect of the intervention, this study also investigated the psychological mechanisms underlying the conservation effects and how household-level and individual-level factors affect the effectiveness of the treatment.

3.3.6.1 Psychological Mechanisms: Motivation and Implementation

Motivation Section 3.3.5 showed that providing users with real-time information on their water and energy use in the shower leads to large and significant behavior changes and conservation effects. The key question is which psychological mechanisms made people change their shower habits so dramatically. From a policy perspective, this is crucial for a comprehensive evaluation of the overall welfare implications of the intervention: If the intervention operates through positive mechanisms and simply enables individuals to act in line with their personal values and preferences (Thaler and Sunstein (2009)), then individual consumers and society as a whole experience unambiguous welfare gains (Allcott (2011b)). If, on the other hand, the device coerces individuals into conservation efforts through negative psychological pressure and increased moral cost of resource consumption (DellaVigna et al. (2012)), then these interventions might be less desirable from a welfare perspective. Moreover, from a practitioner's point of view, a better understanding of the psychological mechanisms and an identification of particular household characteristics could allow to further improve these kinds of interventions along those dimensions. It also makes it possible to segment households into different groups and to target specific population groups, in order to maximize program effectiveness. Finally, for questions of external validity, it is important to understand to what extent the findings might be specific to people who tend to self-select into these kind of pilot studies, and to what extent the findings can be generalized to the overall population.

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

Table 3.7: DiD estimate of baseline effects by household types on other outcomes

Implementation Susceptibility to pressure was measured with the personality traits 'emotionality' and 'conscientiousness' in the HEXACO personality inventory (Lee and Ashton (2004)): If the treatment effect operates through psychological pressure, one can expect a stronger behavioral reaction for more conscientious and for emotionally unstable individuals. If, on the other hand, the device appeals to individuals' personal preferences and values in a positive way, it should be possible to detect a correlation with those variables. These preferences are operationalized with participants' environmental attitudes and their tendency to monitor progress towards goals; furthermore, all remaining HEXACO personality dimensions were included in the analysis and a set of unspecific household characteristics that are generally associated with variance in residential resource consumption. Aside from these survey data, the study design (one- vs. two-person households, *real-time information* vs. *real-time plus past information* condition) can also contribute to draw conclusions on the psychological mechanisms (see section 3.3.6.3 and 3.3.6.4): If the behavior change is moderated by peer pressure, the effects should be stronger for two-person households and for households in the *real-time plus past information* condition, respectively.

For the evaluation of psychological mechanisms, the analysis took advantage of the fact that the dataset collected contains both granular measurements of behavioral outcomes (shower data) and high-quality survey data on household characteristics, attitudes, preferences, beliefs, and personality traits. Given the magnitude of the effect size, it should be possible to detect interactions of particular user characteristics with the treatment effect despite confounding factors that are characteristic for data collected in field studies.

The first step of this analysis evaluated whether the intervention made users develop a better sense for their resource consumption, which would mean that the information was actively being processed by users (section 3.3.6.2). Second, the statistical model from section 3.3.5.2 was augmented by examining the interactions of a list of variables of interest with the treatment effect. Given that most previous studies only had access to a limited subset of the variables collected in this study, the interaction of each variable of interest with the treatment will first be analyzed separately (section 3.3.6.3), for the purpose of comparison. Finally, to tease out which of the variables genuine moderate the treatment effect, all candidates for interaction effects are included in one large joint regression model (section 3.3.6.4). The results of the separate interaction estimates and of the joint model were illustrated by visualizing the interaction effect of selected variables of interest for the 10th and 90th percentile.

3.3.6.2 Association between Estimated and Actual Use

The first step in the analysis of the mechanisms evaluated whether the feedback information was actively being processed and remembered by the users. If this was the case, then households exposed to the consumption information should have acquired an improved knowledge on how much water they typically use per shower. 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. Both the initial and the final survey contained questions asking participants to estimate their water consumption per shower. As figure 3.12 illustrates, prior to the intervention, the majority of participants considerably underestimated their water consumption per shower (b=0.41). At the end of the study, the accuracy of the control group estimate remained poor (b=0.49), while the two experimental conditions had developed a good sense of their average water consumption (b= 0.88). This indicates that the information is actively being processed. This is in contrast to the study by Mitchell et al. (2013) mentioned above, which found no evidence for improved knowledge and quantification abilities after an intervention using paper-based home water reports.

The slight increase of the estimation accuracy among the control group indicates that the installation of the device has also raised that group's awareness. Thus, 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 their participation in the study.

Two aspects should be pointed out concerning figures 3.12 and 3.13: First, the estimated values in both figures 3.12 and 3.13 are based on the estimate of the survey respondent for his personal consumption, whereas in two-person households, actual usage data are generated by two persons. Second, figure 3.12 compares consumption estimates from the initial survey with baseline use data, while figure 3.13 compares estimates from the final survey with consumption data after the begin of the intervention. The reason is that upfront estimates should be compared with measurement data, before the intervention influenced the shower behavior. On the other hand, ex-post estimates of the treatment group were affected by the information displayed on their screens; therefore, ex-post estimates were compared with the study period in which (treatment group) participants were able to monitor their consumption.

With respect to relative consumption, participants seemed 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 3.14 illustrates, participant's perception of their relative position is very poor: While low-users tended to overestimate their usage relative to others, high-users tended to underestimate it.

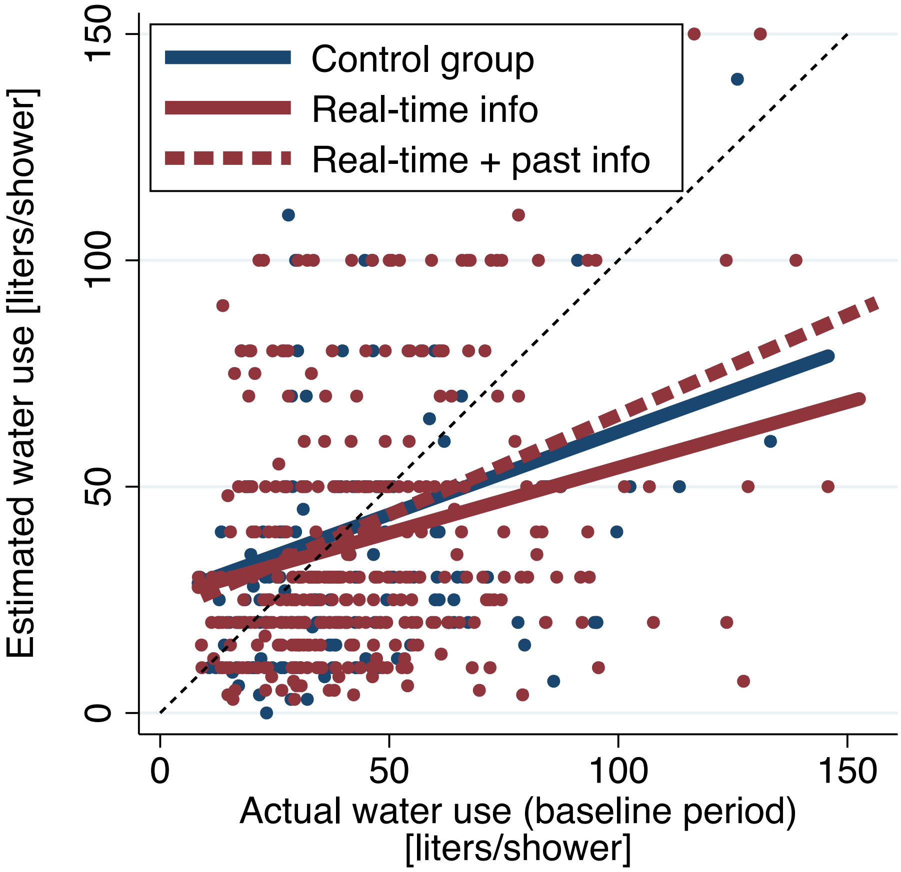


Figure 3.12: Actual water usage per shower during baseline period vs. upfront estimate (initial survey)

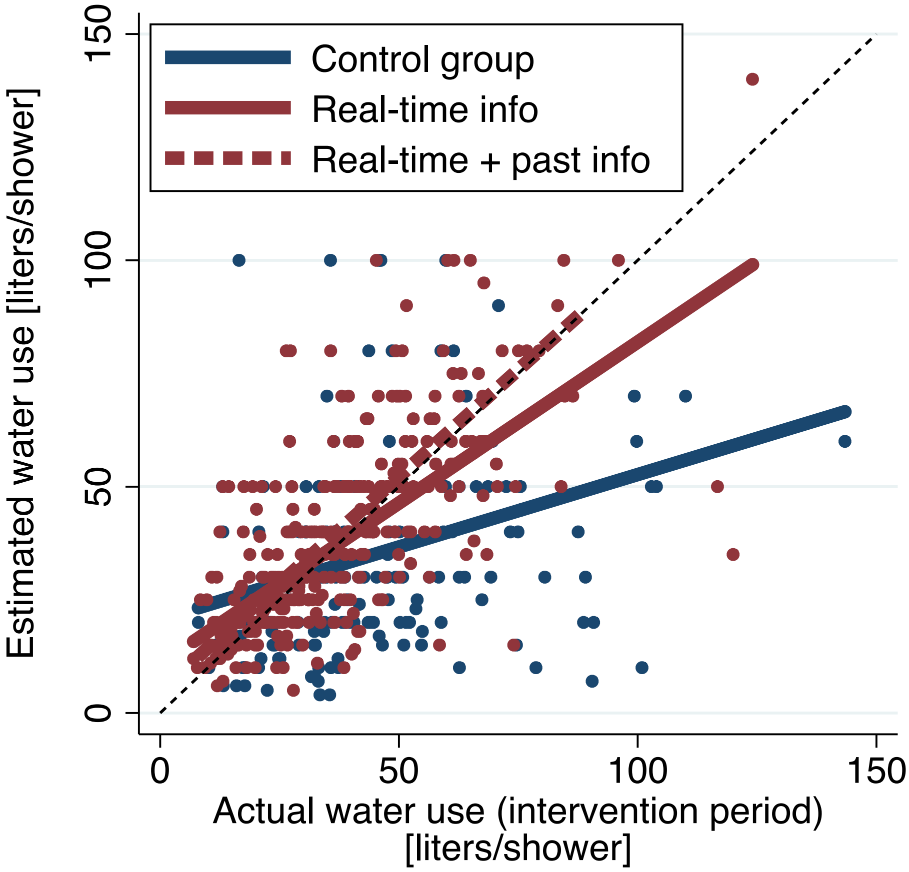


Figure 3.13: Actual water usage per shower during intervention period vs. ex-post estimate (final survey)

This means that participants had a very limited understanding of what an average shower is: There is no clear social norm with respect to 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".* An analysis of the final survey revealed that the intervention did not improve participants' sense for their ranking relative to others. At the same time, the intervention did not contain any explicit social comparison element that would have allowed them to update their beliefs about their position relative to the norm.

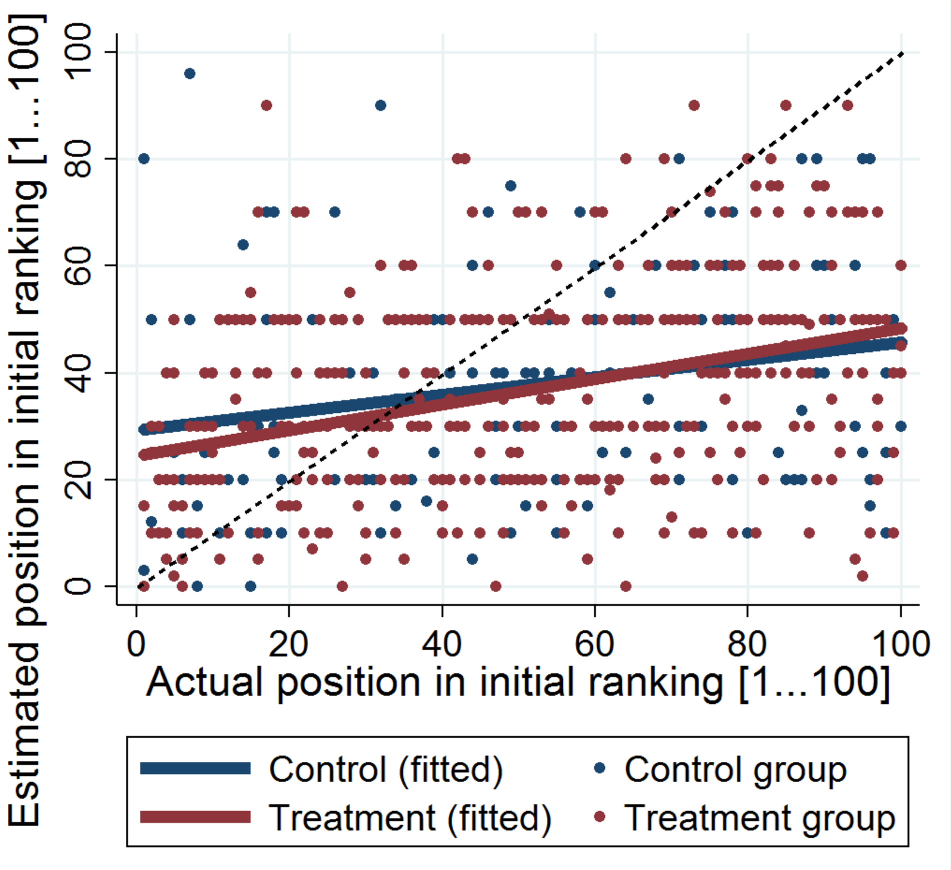


Figure 3.14: Ranking of participating households by water consumption per shower: estimated position in initial survey vs. actual rank during baseline period

3.3.6.3 Separate Analysis of Interaction Effects

The results so far show showed a large behavioral effect (section 3.3.5) and that the feedback information was actively being processed by the participants (section 3.3.6.2). Moreover, the intervention seemed to work approximately equally well for one- and two-person households, and equally well for the *real-time information* and *real-time plus past information* conditions. The following analysis investigated the psychological mechanisms behind the observed behavioral response. This is not only relevant to the external validity of the study, but is also crucial for an evaluation of welfare implications of such interventions. As outlined in section 3.3.6.1 and in the related work chapter 2.1.3, for psychological pressure, the two HEXACO traits emotionality and conscientiousness were analyzed; with respect to attitudes and preferences, individuals' tendency to monitor progress towards goals and environmental attitudes were of particular interest. In addition to these psychological variables, the analysis also incorporated a set of household characteristics that previous studies had identified as relevant to resource conservation. As a result, the analysis covered the following set of variables: All six HEXACO personality traits (openness to experience, conscientiousness, extraversion, agreeableness, emotionality, and honesty-humility), a set of personal preferences, attitudes, and beliefs (self-estimated savings potential, tendency to monitor progress towards goals, tendency to compare oneself with others, satisfaction with life / happiness), and a set of unspecific household characteristics which are typically associated with the heterogeneity in the response to conservation campaigns: age, gender, income, and ex-ante (baseline) consumption.

Given that most previous studies that investigating the effectiveness of behavioral interventions on resource consumption only had very limited access to information on household characteristics and psychological variables, they evaluated the correlation of selective variables individually. Therefore, in line with previous research, the interaction of each variable with the treatment was first measured separately.

The statistical model from section 3.3.5.2, which estimated the overall treatment effect, was thus augmented by adding a separate interaction term for each of the two treatment conditions:

$$y_{it} = \alpha_i + \beta_1 T_{1it} + \gamma_1 T_{1it} \cdot z_i + \beta_2 T_{2it} + \gamma_2 T_{2it} \cdot z_i + d_t + \epsilon_{it} \tag{3.3}$$

In this model, the coefficients γ_1 and γ_2 measure how the variable of interest z_i interacts with the treatment T_{1it} resp. T_{2it}. All interacting variables were centered at their sample mean: That way, the coefficients β_1 and β_2 preserved their meaning as estimates of the treatment effect for a household i who has an average score on the variable of interest z (in the somewhat special case of gender, the average ratio of females in the participating households was taken).

As an additional analysis, the possibility that the time trend in the control group depends on the current variable of interest z_i was included. The model was thus extended to

$$y_{it} = \alpha_i + \beta_1 T_{1it} + \gamma_1 T_{1it} \cdot z_i + \beta_2 T_{2it} + \gamma_2 T_{2it} \cdot z_i + d_t + \delta_t \cdot z_i + \epsilon_{it} \tag{3.4}$$

This is particularly relevant in the light of a possible presence of Hawthorne effects (as indicated in section 3.3.5.1 and discussed later in section 3.4.1.4): It could be the case that a particular trait that moderates the treatment effect also affects the behavior of the control group: For instance, it is possible that individuals who are more conscientious or more susceptible to pressure are also more subject to Hawthorne effects (i.e., they modify their behavior because they are aware of being studied - see discussion in section 3.4.1.4). This could induce different time trends in the control group depending on the personality trait z_i. The addition of the term $\delta_t \cdot z_i$ addressed this issue by allowing time trends to differ with the variable of interest.

In addition to analyzing the interaction of the variables of interest with each treatment separately, the equation was also estimated separately for each household type. Table 3.8 shows an example of the estimation results for the univariate interaction effects, in this case for environmental attitudes as the independent variable z. In column 1 of this table, the regression coefficients of equation 3.3 (i.e., without allowing the time-trend to depend on z_i) are estimated for one-person households; column 3 contains the estimates of this equation for the group of unstable households and column 5 for two-person households. Likewise, columns 2, 4, and 6 show the estimates for equation 3.4. The first and second entries (*"RT info"* resp. *"RT & past info"*) contain the regression coefficients β_1 resp. β_2, along with their standard errors. The following two entries contain the (in this context more relevant) regression coefficients γ_1 resp. γ_2 of the interaction terms. The coefficients can be interpreted the following way: While the constant estimates the control group average (of roughly 1.6 kWh), the entries *"RT info (= 1)"* resp. *"RT & past info"* contain the kWh-estimates for the mean treatment effects. The next two entries *"RT info × Trait"* resp. *"RT & past × Trait"* show to what extent a variation of the variable of interest ("trait" -in this case, environmental attitudes) by one unit affects energy consumption. In the example of environmental attitudes presented in the table, the interaction is not significant, which is remarkable: While environmental attitudes were highly correlated with baseline consumption (b=-.21, p<.0001), they do not explain any of the heterogeneity in the treatment effect. The following analysis will return to this point.

The detailed outcomes of the estimates of the other variables of interest are provided in the supplementary tables section at the end of this dissertation (A.1-A.12). They show significant interaction effects for the following variables: age, baseline consumption, self-estimated savings potential, tendency to monitor progress towards goals, and conscientiousness. Notably, conscientiousness interacts negatively with the treatment effect, which means that less conscientious individuals respond more to the treatment. Neither emotionality, nor envi-

Table 3.8: Difference-in-differences estimates for energy: interaction with environmental attitudes

	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	-0.350***	-0.352***	-0.194*	-0.192	-0.382***	-0.382***
	(0.069)	(0.068)	(0.116)	(0.116)	(0.066)	(0.066)
RT & past info	-0.408***	-0.410***	-0.145	-0.153	-0.424***	-0.424***
	(0.097)	(0.097)	(0.102)	(0.102)	(0.058)	(0.058)
RT info × Trait	0.105**	0.093	-0.102	-0.058	-0.029	-0.038
	(0.053)	(0.057)	(0.140)	(0.148)	(0.077)	(0.082)
RT & past × Trait	-0.146	-0.158	0.100	0.152*	0.176*	0.167
	(0.101)	(0.101)	(0.087)	(0.090)	(0.095)	(0.103)
Constant	1.632***	1.632***	1.487***	1.489***	1.615***	1.615***
	(0.066)	(0.066)	(0.087)	(0.087)	(0.061)	(0.061)
R^2	0.578	0.578	0.380	0.381	0.370	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

ronmental attitudes or individuals' intention to reduce their consumption interact with the treatment. The following sections will return to this point.

Figure 3.15 provides an illustration of the results for a selected set of variables of interest. As the overarching goal of this analysis was the detection of interaction effects, not any more the magnitude of the treatment effect, the two treatment groups, as well as one- and two-person households were collapsed and the group of unstable households was excluded from this analysis. Figure 3.15 illustrates these interactions for the 10th (purple bars) and the 90th (pink bars) percentile of each variable of interest.

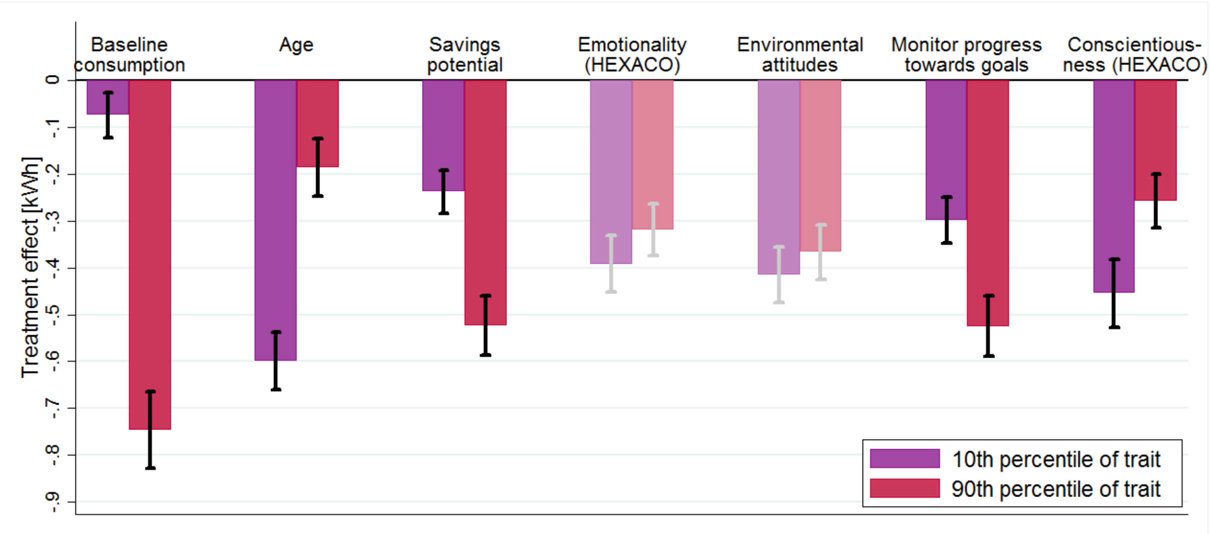


Figure 3.15: Individual linear interaction of selective variables with the treatment effect, illustrated here for the mean of the 10th and 90th percentile of each variable. Black (resp. grey) error bars indicate 95% confidence intervals.

In line with the regression estimates (tables 3.8, A.1-A.12), the illustration suggests that the impact of the treatment is stronger for households with a high baseline consumption. Younger households and those with a higher self-estimated savings potential appear to react much more to the treatment, as well as participants with a tendency to quantify their performance. As in the regression tables, participants with a lower level of conscientiousness exhibit a stronger reaction to the feedback device, while environmental attitudes and the emotionality trait do not interact with the treatment. However, an interpretation of these outcomes is difficult, as the interaction could be biased by omitted variables. As shown in table 3.4, many variables are highly correlated with the baseline consumption. These issues will be discussed in the following sections.

3.3.6.4 Joint Evaluation of Interaction Effects

One of the strengths of the dataset collected is that it comprises an exhaustive list of variables which most previous studies could only analyze in isolation. The baseline correlations in table 3.4 showed a many cross-correlations; in particular, many candidates for interaction with the treatment were correlated with baseline consumption. Therefore, to goal of the next step was to disentangle the specific contributions of each variable. To that end, all candidates for interaction effects were evaluated jointly by including all of them in linear combination in the following regression model:

$$y_{it} = \alpha_i + \beta T_{it} + \gamma' T_{it} \cdot \mathbf{z_i} + d_t(+\delta \cdot t \cdot \mathbf{z_i}) + \epsilon_{it} \tag{3.5}$$

Here, T_{it} is the indicator of the (collapsed) treatments; the vector γ contains the estimates for interactions between the treatment and the set of variables contained in vector $\mathbf{z_i}$; the parentheses around $+\delta \cdot t \cdot \mathbf{z_i}$ indicate the option of including dependencies of the control group time trend on $\mathbf{z_i}$.

As table 3.9 shows, the results are fundamentally different from the estimates of the separate interaction analyses. Column (1) presents the results of the joint regression when the all variables of interest whose interaction terms had turned out to be significant in section 3.3.6.3 (see tables 3.8, A.1-A.12) are included in the model. Four variables interact with the treatment: baseline consumption, the tendency to monitor progress towards goals, conscientiousness, and environmental attitude. The latter is of particular interest, as it had not been significant in the individual regression. By contrast, age and self-estimated savings potential did not significantly interact with the treatment any more. Column (2) shows the results for the same set of variables of interest, but this time allowing for group specific time trends. In column (3), the estimation is limited to the interaction terms that were significant in column (1); in column (4), their respective influence on the time trend is included in the model. None of the results chang: The results are robust to specifics of the model and also do not depend on a particular constellation of predictor variables.

Table 3.9 shows that environmental attitudes *do* positively interact with the treatment when both baseline consumption and environmental attitudes are included in one model. This means that after controlling for baseline consumption, people with stronger environmental attitudes respond more strongly to the intervention. In the separate analysis of the predictors, this correlation was hidden. This may be because of its strong negative correlation with baseline consumption. Section 3.4.2.2 will return to this point. In line with the results from the separate analyses, the analysis shows a positive interaction of the treatment effect with the tendency to monitor progress towards goals and a negative interaction with conscientiousness. Again, the latter is somewhat surprising, as one could have expected that more conscientious individuals would exhibit a stronger response to the treatment. Yet the results show that the device was more effective among users who in general are a less conscientious. Section 3.4.2.4 will discuss this point. As in the separate interaction model, no evidence was found that emotionality would affect the outcome of the treatment. This means that the device did not have a stronger effect on users who are more susceptible to pressure. This result will

	(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 3.9: Joint analysis of interaction effects

be discussed in section 3.4.2.6. In contrast to the separate interaction analysis, neither age, nor self-estimated savings potential significantly interacted with the treatment any more. The analysis suggests that the interaction effects identified in the separate analysis were in fact mainly due to the correlation with baseline consumption. Section 3.4.2.7 will discuss this finding.

Figure 3.16 illustrates the results for the 10th and 90th percentile, again highlighting the prominent role of baseline consumption regarding its interaction with the treatment effect. In line with the joint regression table 3.9, evidence for a significant interaction of the variables age and self-estimated savings potential is not supported. Once more, both consistent with the joint regression table and also with the separate analysis, the personality trait emotionality did not affect the outcome. Also in line with all previous analyses, the device had a stronger impact on participants scoring low on conscientiousness. Likewise, the positive interaction of the tendency to monitor progress towards goals is confirmed once more in the illustration. In contrast to the separate analysis, yet in line with table 3.9, the figure highlights the significant positive interaction of environmental attitudes.

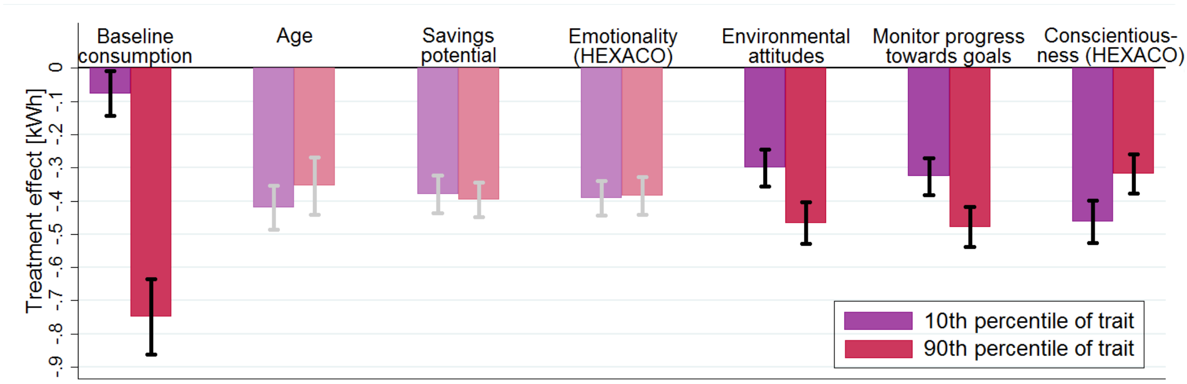


Figure 3.16: Interaction of the key variables of interest with the treatment effect, now evaluated jointly in a single regression model. Again, bars depict the mean of the 10th and 90th percentile of each variable. Black error bars indicate 95% confidence intervals.

Triple Interactions The question arises whether the level of baseline consumption also affected how environmental attitudes or the tendency to monitor progress towards goals interacted with the treatment. To investigate this, the model was augmented by triple interactions between a) treatment effect, b) baseline consumption, and c) either tendency to monitor progress towards goals, environmental attitudes, or conscientiousness. The results are displayed in column (5) and (6) (without resp. with time trend interactions) of table 3.9. They show that indeed, the tendency to monitor progress leads to a stronger reaction to the treatment when baseline consumption is high. Equally, environmental attitudes induce a stronger reaction to the treatment when baseline consumption is high. The effect of conscientiousness, however, is not magnified by the baseline consumption: While individuals with a lower level of conscientiousness do respond more strongly to the intervention, this effect was found to be stable across all levels of baseline consumption. The results are illustrated in figure 3.17, showing the magnifying effect of baseline consumption on the interacting variables environmental attitudes and individuals' tendency to monitor progress towards goals.

The outcomes of the separate and the joint regression of the variables of interest were fundamentally different, in particular for environmental attitudes, age, and self-estimated savings potential. This can probably mainly be attributed to the strong correlation of these variables with baseline consumption. Section 3.4.2.7 will further discuss this point.

Based on the results of the joint regression and on the correlation table 3.4 for the baseline consumption, figure 3.18 provides a *qualitative* illustration of the interplay of the main variables of interest. The illustration contains all variables that had interacted with the treatment effect either in the separate, or in the joint regression. It shows connectors for all the variables that ultimately moderated the treatment effect (baseline consumption, environmental attitudes, conscientiousness, and the tendency to monitor progress towards goals), as well as significant correlations between those variables before the onset of the treatment. The goal of this simplified representation is to provide a qualitative visualization of the interplay of these variables. Red connectors indicate a positive correlation, while blue, dashed connectors represent negative correlations. Connectors were left on purpose without arrowheads: While the causal direction of the correlation is clear in most cases (e.g., higher baseline use leads to higher savings effects, not the other way around), arrowheads were left out on purpose in this representation, to avoid speculations on causality for cases that are not entirely clear (for instance, the correlation between conscientiousness and tendency to monitor progress towards goals).

Overall, the results of the analysis of the psychological variables and household characteristics show a rather complex picture, but also that a careful analysis can disentangle the contribution of the different factors. The results of both the separate and the joint regression are of interest: In the case of age, for instance, it is important to understand that it is not this attribute itself which moderates the outcome; at the same time, for practitioners it is also very valuable to know that age can serve as a good proxy to identify households with a high savings potential (at least in the case of showering). While the separate regressions reveal factors that are meaningful from a net outcome point of view, the joint model shows which variables genuinely moderate the treatment effect.

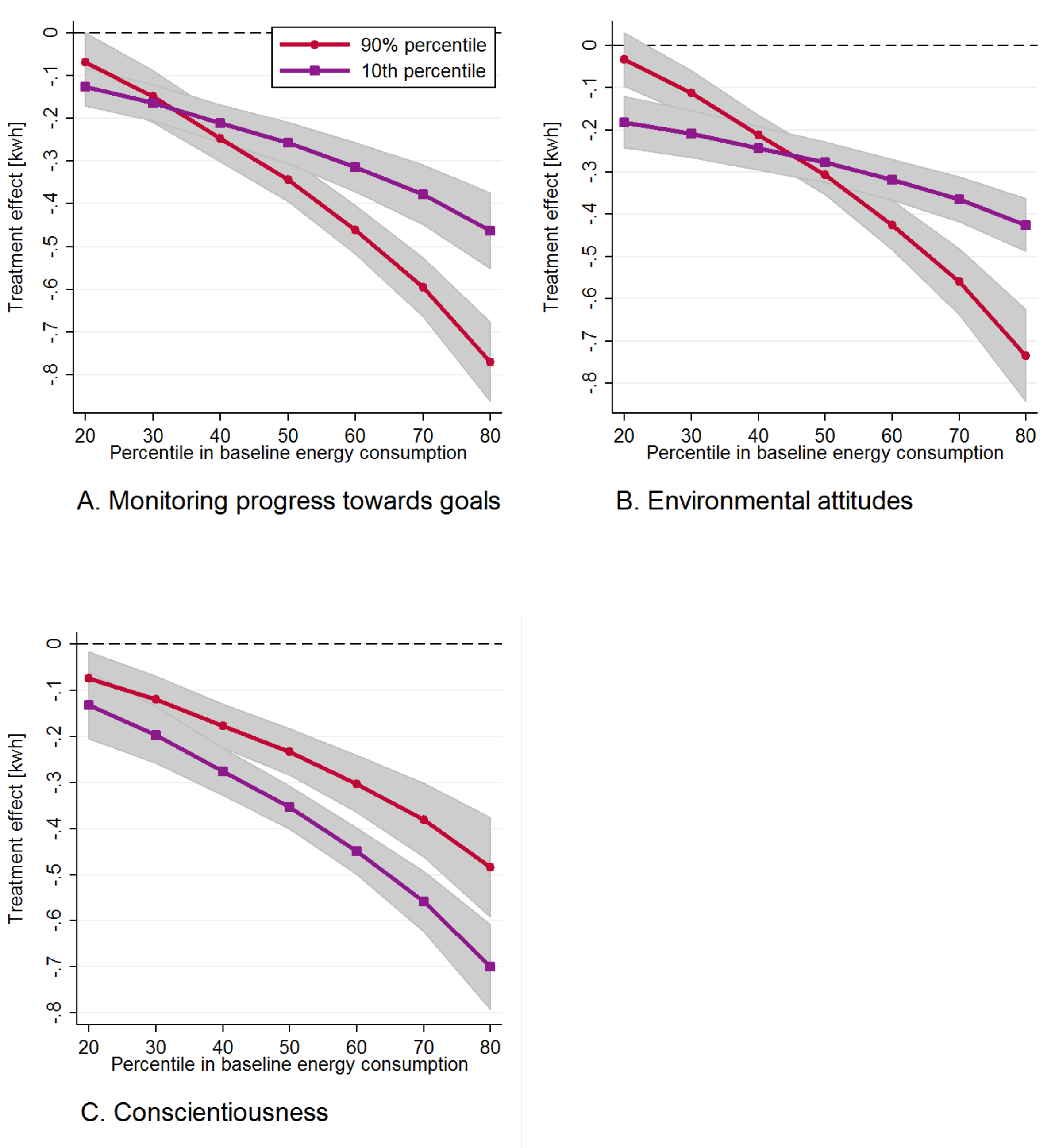


Figure 3.17: Visualizations of triple interactions showing the magnifying effect of baseline consumption on the interaction of tendency to quantify and environmental attitudes with the treatment effect

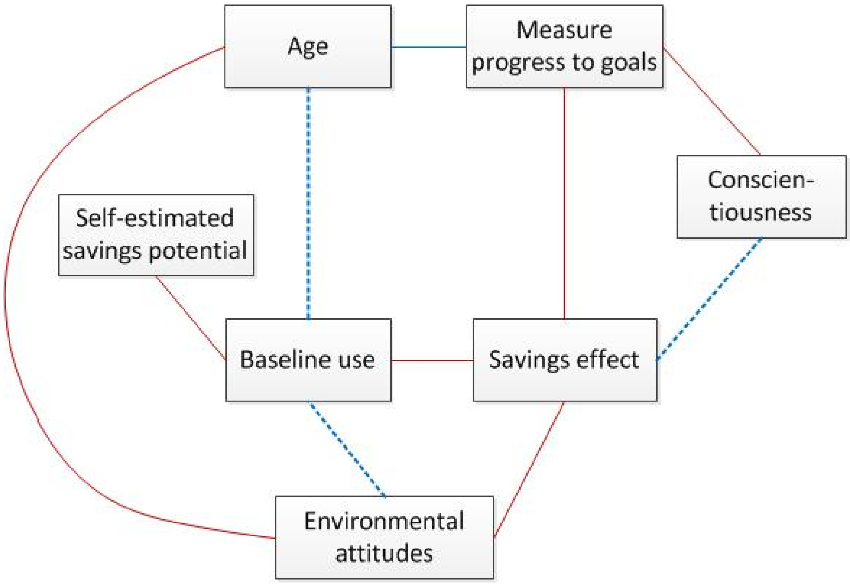


Figure 3.18: Qualitative visualization of the interplay of the key variables of interest, evaluated jointly in a single regression model.

3.4 Discussion

So far, the analysis has mainly focused on the outcomes themselves and provided only a very limited interpretation of the findings and their implications. This section will first quantitatively assess the main treatment effect (section 3.4.1). This is followed by an interpretation of the findings with respect to the underlying psychological mechanisms and the influence of household characteristics (section 3.4.2); the interpretation covers the key variables of interest (ex-ante consumption, environmental attitudes, conservation intent, self-efficacy, emotionality, conscientiousness, tendency to monitor progress towards goal, and age), highlights the importance of their joint evaluation, addresses the topic of external validity, and discusses the correlation of age with baseline consumption. The third part of this discussion, section 3.4.3, examines the main implications of this study in a broader research and policy perspective: It evaluates the intervention in the context of behavioral nudging, highlights the relevance of the study as a demonstration of scalability and cost-effectiveness of behavior-specific real-time feedback and its importance to overcome market barriers for innovative technologies. Furthermore, effects of market segmentation (profiling) are discussed as well as the relevance of the correlation of age with baseline consumption for energy and water demand forecasts.

3.4.1 Quantification of the Treatment Effect

This section provides a quantitative evaluation of the main treatment effects: per household, for a large-scale deployment, compared to electricity smart metering, and in the light of potential Hawthorne effects.

3.4.1.1 Quantification of Direct Savings per Household

As outlined in section 3.3.5, the use of the shower meter induces an average reduction of 23% or 0.38 kWh of thermal energy consumption per shower. This number, however, assumes 100% boiler efficiency and zero generation and storage losses. Actual boiler efficiency depends on boiler size, fuel type, and age; it averages 65% efficiency in Swiss households (Prognos AG (2013b)). Also depending on infrastructure and heating system properties, average distribution losses amount to 24 to 36% (Tschui and Stadelmann (2006)). Taking a rather conservative estimate for distribution losses of 20%, considering the Swiss average for boiler efficiency of 65%, and assuming a cold water temperature of 12 degrees Celsius (Geberit (2011)), the actual reduction is 0.56 kWh per shower.

Given the Swiss average household size of 2.2 persons (SFSO (2012)) and assuming a use frequency of one shower per person per day, this results in a reduction of 74 kWh of heat energy per household over the two-month study period alone. Assuming persistence of the effects (see section 3.4.4.4), this amounts to energy savings of 443 kWh and an emissions reduction of 94 kg of CO_2 per household and year (cf. section 1.1 for Swiss carbon intensity of water heating).

In addition to heat energy savings and emissions reduction, the reduced water consumption per shower results in a yearly per-household reduction of 8,500 liters. This may be a less critical issue in a water-rich country like Switzerland than in an international context: 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 3.3 and current utility prices, the reduced energy and water consumption results in average savings of CHF 110 per household and year[2]. As a consequence, the average payback period ranges between six months (bulk purchase) and eight months (individual purchase).

3.4.1.2 Quantification of the Impact of a Large-Scale Deployment

The cost per kWh saved is CHF 0.041 (dividing the initial cost through the kilowatt-hours saved over the course of a 3-year lifetime period and assuming persistent saving effects – an additional long-term study to evaluate this assumption is still ongoing). 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 roughly half of the marginal cost

[2] Underlying assumptions: Fuel mix for water heating shown in figure 3.3 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/m^3, waste water 2.35 CHF/m^3 (Statistisches Amt des Kantons Basel Landschaft (2013); Zanzi (2011)); oil 0.105 CHF/kWh (Hauseigentümerverband Schweiz (2012); Hauseigentuümerverband Schweiz (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))

of current electricity production of 0.074 CHF/kWh (Kost et al. (2012)). At least 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 the first study worldwide, these findings demonstrate 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.

Technically, given the standardized threads of shower fittings, the device can 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. A larger roll-out with a 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, cf. previous section). For comparison, the total production of all Swiss wind power plants in 2012 was 85 GWh of electricity. Moreover, for the installation cost of a single onshore 2 MW-wind turbine (Rehfeldt and Wallasch (2013)), 60,000 households - roughly one third of the households the city of Zurich - could be equipped with such a shower meter.

3.4.1.3 Comparison with Electricity Smart Metering Pilots

As outlined in section 2.1.2, randomized controlled trials with smart meters that evaluate the effectiveness of feedback on residential electricity consumption typically report reductions between 2 and 6%. Based on sample size, recruitment strategy, involvement of the study administration and research design, this study falls into "class B" of the classification scheme by McKerracher and Torriti (2013) (see section 2.1.2). According to that meta-review, class B studies have a weighted mean conservation effect (weighted by sample size) of 4.5%. The relative treatment effect observed in this study is five times as high.

But also in absolute terms, the device compares favorably to the energy reduction of other smart metering studies which 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, this discussion will not go into the details of converting one form of energy into another, yet at least for the 25% of Swiss households with 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 1.1, carbon intensity of water heating in Switzerland is 212 g/kWh, compared to 122 g/kWh for household-level electricity. In the two-month study period alone, the average shower monitor reduced heat energy consumption by 74 kWh and carbon emissions by 16 kg of CO_2, compared to 2 kg of CO_2 abated in two months by the smart metering program described in Degen et al. (2013) with the same (!) pool of participating households. From a Swiss carbon abatement perspective, the shower monitor thus has the 8-fold impact of an electricity smart meter.

As a bottom line, the study shows a) that a simple device applied to a very specific domain can achieve energy savings that are quantitatively important, even for 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 targets such as overall household electricity usage. Further research is needed to determine to what extent the magnitude of the savings are proprietary to the specific shower context, and to what extent the findings can also be applied to other domains.

3.4.1.4 Hawthorne Effect

As outlined in section 3.3.5, the per-shower consumption of the control group slightly increased over time. This increase cannot be explained by seasonal effects, as temperatures did not show any particular trend during the study period (see section 3.2.6.5). Instead, this trend might be attributed to what is referred to as *Hawthorne effect* in the literature: It describes the phenomenon that participants modify their behavior because they are aware that they are being monitored (McKerracher and Torriti (2013)). The effect has been inconsistently observed and may confound the outcome of studies (Schwartz et al. (2013)). McKerracher and Torriti (2013) reports 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%.

The present study shows a modest, yet mostly insignificant increase in the energy consumption of the control group over the study period. In line with Schwartz et al. (2013), the conjecture is that participants might have reduced their resource consumption in response to the feeling of being observed, in particular at the beginning of the study. Prior research suggests that after a while, participants get used to being part of an experiment or even forget that they are being monitored; as a result, the influence of the Hawthorne effect diminishes over time (Martinussen and Hunter (2009)). The finding that the control group has slightly improved their sense for their water consumption per shower over the course of the study (see section 3.3.6.2) confirms this conjecture: Even without the feedback information, the installation of the device has increased their awareness for their consumption. Baseline consumption was subject to the biggest influence of the Hawthorne effect, resulting in a reduced resource consumption in particular at the beginning; over time, participants habituated to the presence of the shower meter and the control group relapsed to their normal shower habits. This highlights the importance of using a difference-in-differences strategy to capture the causal effects unconfounded with these kind of time trends.

3.4.2 Interpretation of the Psychological Mechanisms and Role of Household Characteristics

This section takes the evaluation of the results of section 3.3.6 a step further: It interprets the findings with respect to the question of what drives individuals' response to behavioral interventions and outlines implications of these findings. First, it discusses key findings for the most important variables of interest (sections 3.4.2.1-3.4.2.5). Section 3.4.2.6 pulls these aspects together and evaluates whether the saving effects in this study are driven by positive or negative mechanisms (section 3.4.2.6). This is followed by a discussion on the importance of a comprehensive analysis (section 3.4.2.7), on external validity (section 3.4.2.8), and on evolving norms in the context of showering (section 3.4.2.9).

3.4.2.1 Ex-Ante Usage Level (Baseline Consumption)

As section 3.3.6.4 shows, the treatment effect is much higher for households with a high baseline consumption. This is in line with previous research (Davis (2011); Allcott (2011b)). The analysis also shows that baseline consumption is by far the best predictor of the treatment effect. Yet baseline consumption itself is not a given, but a mixture of many factors and accordingly correlated with many other variables (see table 3.4 for a first, though incomplete impression). This makes it particularly difficult to disentangle which variables ultimately drive the savings effect. Apart from confounding factors, there are several possible interpretations for why households with a high baseline consumption respond more to the treatment: First of all, it may simply be easier for households with a high ex-ante usage to reduce it: An individual who already takes short showers, mainly to fulfill the practical purpose of cleaning, will find it harder to shorten her shower by a minute than an individual who tends to take prolonged showers for various reasons including pleasure, awakening, invigoration, and warming. An alternative explanation is based on Festinger's cognitive dissonance theory (Festinger (1957)): As figure 3.12 shows, high users in particular tend to fundamentally underestimate their water consumption. As the discrepancy between expected use and actual use is particularly high for them, they may perceive a stronger desire to either update their beliefs or to align their consumption with what they thought they would use. In any case, although baseline consumption may be more of an outcome of many factors and processes itself, the results show the importance of incorporating it as a predictor of savings effects to uncover masked correlations (as in the case of environmental attitudes) and prevent spurious correlations (as in the case of age).

3.4.2.2 Environmental Attitudes

A broad body of literature has covered the relationship of environmental attitudes and behavioral outcomes (section 2.1.3.3), yet with mixed results. The results of this study are also ambiguous: While environmental attitudes, when analyzed separately, do not affect the net treatment effect (section 3.3.6.3), they do significantly interact with the treatment effect after controlling for baseline consumption (section 3.3.6.3).

On the one hand, the results show that the treatment effect is not driven by individuals with an extremely strong motivation to conserve energy: Participants with less strong environmental attitudes equally reduce their consumption. Thus, from a practitioner's perspective, the treatment appears to work equally well independent of environmental attitudes; the savings are not driven by a small subset of individuals with a particularly green mindset. This is relevant for the external validity of the study and for scaling up the intervention to a larger number of households (section 3.4.3.4).

On the other hand, 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 care more about the feedback and actually pay more attention to it. In the final survey of this study, users with stronger pro-environmental attitudes report paying more attention to the device and discussing their consumption more frequently within the household (in the case of the two-person households). However, it is very likely that they had already paid more attention to their resource consumption in the shower upfront, which would explain their lower baseline usage. As users with weaker pro-environmental attitudes tend to start out from a higher baseline use, they may find it easier to reduce their consumption. At the same time, individuals with stronger pro-environmental attitudes seem to compensate for this with more efforts. In the end, their higher efforts appear to be balanced out by a higher difficulty to further reduce their consumption. This is in line with previous literature which states that people who report stronger pro-environmental behaviors do not necessarily use less energy: While people may report substantial changes in behavior, these outcomes may often not be measurable (Heberlein (2012); Gatersleben et al. (2002)). Given the strong correlation of environmental attitudes with baseline usage, it may well be the case that people with stronger pro-environmental attitudes actually try harder, yet this does not necessarily manifest itself in a higher impact on resource consumption.

To summarize, the results show that individuals' pro-environmental attitudes are strongly correlated with lower baseline consumption. Only after controlling for the level of ex-ante consumption, it becomes apparent that environmental attitudes do in fact interact with the treatment effect. Thus, although environmental attitudes do contribute to the saving effects in the background, this can be masked in the net treatment effect. As a result, environmental attitudes do not significantly affect the *net* treatment effect and households with less strong environmental attitudes equally reduce their resource consumption.

3.4.2.3 Conservation Intent and Belief in Conservation Ability

As outlined in section 2.1.3.3, intent and self-efficacy - the belief in one's own ability to complete tasks and reach goals - are generally considered as good predictors of behavior change. Yet the findings of this study do not provide support for the importance of either one. The interaction between individuals' intent to reduce their consumption and the treatment effect is not significant (section 3.3.6.3). The results of this study 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. With respect to self-efficacy, while the results from section 3.3.6.3 suggest that participants' self-estimated resource conservation potential (i.e., their belief in their ability to reduce their consumption) moderates the treatment effect, table 3.4 shows a strong positive correlation of this variable with baseline consumption. As a result, the analysis in section 3.3.6.4 reveals that participants' self-estimated conservation potential does not significantly interact with the treatment effect. Rather, individuals with a high baseline consumption seem to be aware of the fact that they could reduce their water and energy consumption quite a bit and assess their conservation potential accordingly. Thus the findings do not provide any evidence that individuals' belief in their ability to reduce their consumption would change the outcome. The study results suggest that it is the de-facto state of affairs itself (i.e., participant's shower behavior before the onset of the intervention) rather than individuals' perception thereof that predicts their subsequent response. These findings conflict with Bandura's work (Bandura (1997); Bandura et al. (1980)), which qualified self-efficacy as the most important precondition for behavioral change (see section 2.1.3.3). It is possible that the results are specific to showering and that they may not generally apply to other behaviors. Yet further research is necessary to determine whether some of the influence that has previously been attributed to self-efficacy might simply be an artifact of the de-facto state of affairs and individuals' - rather realistic - perception thereof.

3.4.2.4 Influence of Emotionality and Conscientiousness

As outlined in section 2.4, one hypothesis regarding the mechanisms driving feedback was that it operates through psychological pressure: Feedback may induce negative sensations in individuals about their resource consumption, create the feeling that their behavior is being observed and that resource conservation is expected from them. This sensation would reduce the utility of the behavior (showering), as its associated moral cost (Levitt and List (2007)) is now made visible. This might generate negative affect, which individuals seek to partially offset by changing their behavior. While behaviorally effective, the conservation effect would thus come at the price of individuals' utility, making the device less desirable. Moreover, this would probably make the effect less persistent, as people might simply remove the device to avoid these negative sensations.

If feedback operates through these negative channels, then individuals who are more susceptible to pressure should be particularly responsive to this kind of intervention. In particular, one would expect a stronger reaction from individuals who score high on the traits emotionality and conscientiousness of the HEXACO personality scale: Emotionality measures anxiety and susceptibility to pressure, while conscientiousness measures individuals' propensity to adhere socially prescribed norms (Roberts et al. (2009)).

As both section 3.3.6.3 and 3.3.6.4 show, the results do not support any evidence for the influence of the emotionality trait. As for conscientiousness, the findings even provide evidence for the opposite: The device is particularly useful for *less* conscientious users. This is in line with the inattentive choice model presented by Taubinsky (2013), which states that limited attention prevents individuals from acting on their intentions and preferences. The shower meter may thus serve as a reminder that elicits attention and reflective thinking by making behavioral outcomes on resource consumption salient. Section 3.4.3.1 will pick up on this point.

Overall, these findings strongly indicate that the device does not operate through psychological pressure. The following sections 3.4.2.6 will further discuss the role of the personality trait conscientiousness.

3.4.2.5 Tendency to monitor progress towards Goals

Previous research has established that feedback is essential for goal pursuit (Fishbach and Finkelstein (2012)) and individuals' goal orientation shapes feedback-seeking behavior (Ashford (1986); VandeWalle (2003)). As section 3.3.6.4 shows, individuals' general propensity to monitor progress towards goals positively interacts with the treatment effect. Protection of the environment is a goal that the vast majority of the Swiss people and most study participants support (see section 3.4.2.8). Yet humans do not always act in line with their goals, and achieving pro-environmental goals can be particularly challenging (Gutsell et al. (2012)). Feedback can help individuals monitor how they fare relative to their personal goals. 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.

In the final survey of this study, 50% of the treatment group participants indicated that they had set themselves a specific goal for their shower behavior. While they do reduce their consumption significantly more than participants who had not set themselves a goal, this may be confounded with other variables: Setting oneself a goal or not is probably also highly correlated with environmental attitudes, for instance. Nevertheless, the fact that the tendency to monitor progress towards goals interacts with the treatment effect implies that the device resonates with this preference. Among the general population, propensity to monitor goals is quite a prevalent preference (see also section 2.1.3.3). Likewise, 74% of the study participants indicated upfront that they frequently compare their performance against self-set goals; by contrast, only 28% reported that they frequently compare themselves with their peers' performance.

As outlined in section 2.1.3.3, technological progress (e.g., the ubiquity of smart phones) makes it increasingly easy to collect and track metrics about one's life. These technologies facilitate individuals' goal pursuit by providing her with relevant information about the *present* state of affairs. Moreover, they can serve as behavioral cues or nudges that remind the individual of the her *preferred* state of affairs (as described in Thaler and Sunstein (2009), see section 3.4.3.1). As a result, such self-tracking technologies and applications have already found widespread adoption. The feedback system used in this study thus resonates with the "quantified self" movement, which a series of newspaper articles (e.g., Bradley (2013); Hay (2013); Snyder (2013)) have recently described as a major trend that has reached mainstream popularity (see section 2.1.3.3).

3.4.2.6 Positive vs. Negative Mechanisms

Overall, the findings do not support evidence for any of the negative mechanisms that might have explained the response to the intervention. Individuals who are more susceptible to pressure do not respond stronger to the treatment; conscientious people respond even less to it. The results rather suggest a mechanism that makes resource consumption more salient and facilitates the pursuit of conservation goals. First of all, table 3.2 shows that the treatment group perceives the device as more effective, helpful, pleasant, interesting, and entertaining than the control group; the device is also rated as slightly more relaxing rather than stressful. Moreover, individuals' propensity to monitor progress towards goals and environmental attitudes both significantly and positively moderate the treatment effect, while conscientiousness negatively interacts with it (section 3.3.6.4). In line with Woodside (2011) and Taubinsky (2013), the findings indicate that many individuals have a strong desire to conserve energy (and water), but forget about it in the rush of their daily lives; or they don't have the necessary information to act in line with their preferences. When provided, especially individuals with a strong tendency to monitor progress towards goals, individuals with strong pro-environmental attitudes, and inattentive (less conscientious) individuals respond strongly to the treatment. The role of conscientiousness is rather difficult to interpret, but might indicate that the device is particularly useful for individuals who on a daily basis are not mindful of their good intentions, while conscientious people might not need frequent reminders of their goals: They manage to act in line with them on their own.

Section 3.3.6.2 shows that the information is indeed actively being processed by the participants. This contrasts with previous research on social norm-based feedback which - despite effectively reducing water consumption - did not find evidence for learning (Mitchell et al.

(2013)). Several reasons might explain why learning effects occur in the present study, but not in the program analyzed by Mitchell et al. (2013): Compared to feedback on household's monthly water consumption, it could be the real-time component, the frequency of the feedback, or the relatively small numbers, which individuals can remember and relate to more easily. One could also imagine that in programs with home energy / water reports, people's attention focuses on the social comparisons and on relative numbers (their resource usage compared to the neighborhood), instead of absolute consumption metrics.

In any case, the present study shows that feedback interventions can effectively operate without emphasizing social norms and that simple salience of consumption metrics effectively induces both learning and substantial behavior change. Instead of putting individuals under negative pressure to reduce their consumption, these findings are suggestive of a mechanism that operates through making resource consumption more salient and the pursuit of conservation goals easier.

3.4.2.7 Importance of a Joint Evaluation of Key Variables of Interest

As sections 3.3.6.3 and 3.3.6.4 show, when comparing the results of the separate and joint regressions, the outcome is fundamentally different for several key variables of interest, in particular for age, environmental attitudes, and self-estimated savings potential. While this may seem trivial to the experienced statistician, this puts the results of many previous studies into question. As outlined in section 2.4, most studies that evaluated the impact of behavioral interventions only had access to a limited set of predictor variables. As a result, those studies might have reported biased outcomes: They may either have failed to uncover mechanisms that were concealed by other correlations (as the separate regression did for environmental attitudes, see section 3.4.2.2); or, on the other hand, some may falsely have attributed outcomes to variables that were cross-correlated with unobserved variables (as in the separate regression, which identified age as moderator of the savings effects). This discrepancy shows the importance of a comprehensive analysis as carried out in section 3.3.6.4. This approach makes it not only possible to determine which variables genuinely moderate the treatment effect, but also to uncover hidden interactions (like the effect of environmental attitudes after controlling for baseline consumption). At the same time, the results of the separate regressions are also of interest from a net-impact point of view. For instance, while age may not be the driver of higher savings effects, in the absence of better predictors it can still serve as a good proxy to identify households with higher savings potential. On the other hand, while environmental attitudes may make a difference "behind the scenes", it is also a fundamental finding that ultimately, they did not affect the net savings outcome in the study. Ultimately, both the separate and the joint analysis contribute to complementary insights, but the discrepancy of their outcomes shows once more that caution is warranted in the interpretation of correlations.

3.4.2.8 Discussion of the External Validity of the Study

One of the key questions in the evaluation of the study results is to what extent the savings effects might have been affected by self-selection of participants. This section will examine what the findings on the behavioral mechanisms imply with respect to self-selection biases. This is not only meaningful for the generalization of the results of this study, but also for the external validity of other trials with opt-in design.

First of all, one should keep in mind that randomized controlled trials have been identified as the most reliable method to evaluate the impact of energy conservation programs: As Allcott (2011b); Allcott and Mullainathan (2012) showed, non-experimental estimators perform dramatically worse than experimental estimators. Nevertheless, opt-in studies generally face criticism regarding a self-selection bias towards "green" consumers: It is argued that participants of these kinds of studies might be more conscientious and to have a stronger pro-environmental mindset than the general population. As outlined in section 2.1.2 for real-time feedback 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 an opt-in design. While these results are comparable in magnitude, the difference is not negligible.

The results of this study directly address these concerns. With respect to environmental attitudes, as discussed in section 3.4.2.2, the findings indicate that from a *net* impact point of view, the treatment works equally well independent of environmental attitudes. This suggests that the effectiveness of these interventions is not restricted to a small subgroup with particularly strong environmental attitudes. Moreover, the sample recruited for this study even scored slightly lower on pro-environmental attitudes ($M_A = 3.48, N = 643, SD = 0.9$) than the nationally representative Swiss sample presented in the most recent Swiss Environment survey ($M_D = 3.80, N = 3,352, SD = 1.0$, Diekmann et al. (2008)). Therefore, a selection bias regarding environmental attitudes can practically be excluded.

Regarding conscientiousness, the results indicate that the device is actually more effective for less conscientious users (see section 3.4.2.5). In the case that the recruitment strategy should favor a self-selection of more conscientious individuals, the findings imply that such a bias would rather reduce than increase the treatment effect.

Apart from environmental attitudes and conscientiousness, the study results show that two more variables moderate the treatment effect, namely baseline consumption and the tendency to monitor progress towards goals. Therefore, these two will also be discussed briefly in the light of self-selection concerns. Regarding baseline consumption, given the scarcity of reference data, one cannot say with certainty whether the baseline consumption of the study sample is representative for the general population: This study probably collected the world's largest dataset on shower behavior so far. However, the baseline values measured are in line with previous (smaller) studies (section 2.3.3). Furthermore, given the strong negative correlation of baseline consumption and environmental attitudes, concerns regarding self-selection would primarily predict an over-representation of ex-ante low consumers. Yet a low baseline consumption is the strongest predictor for a smaller effect size. Thus, also with respect to this dimension, a self-selection bias would rather under- than overestimate the effect size.

Finally, the forth significant predictor for savings effects identified is individuals' tendency to monitor progress towards goals. As outlined in section 3.4.2.5, there is evidence that a high share (roughly 70%) of the general population tracks metrics on their daily life on a regular basis. These numbers are comparable with the share of study participants who stated their general preference for monitoring progress towards goals (74%). While this may not be a perfect indicator to assess the representativeness of the sample regarding this dimension, the numbers indicate that the prevalence of this preference among the study sample is at least roughly in line with the share of goal-tracking adherents among the general population.

Another relevant aspect with respect to a potential self-selection bias is that the 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 on electricity consumption induced by in-home displays was 3.2%. This magnitude is in line with other large-scale electricity smart metering studies (Schleich et al. (2013)), but far below the 23% reduction achieved in the present study. 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.

Notwithstanding the above, caution is still warranted in generalizing the results to other contexts, populations, and behaviors. More research is needed to evaluate in particular whether the findings on an equal net treatment effect regardless of environmental attitudes or the evidence of a higher impact for less conscientious users is valid for the general population or for other countries. This point will be picked up again in the discussion of the study limitations 3.4.4.1, which also elaborates on the question to what extent the findings (effect size and underlying mechanisms) may be due to the specifics of the application context "shower behavior" or of the feedback system *amphiro a1*.

3.4.2.9 Influence of Age on Shower Behavior and Treatment Effect

This study revealed another interesting aspect: A strong correlation between age and the resource consumption per shower in the baseline period. As section 3.3.3 and in particular figure 3.9 show, there is a strong and continuous trend in the dataset between age and energy (resp. water) consumption: 20-29 year-old participants end up using 2.3 times as much energy and water per shower as participants over 64. This is likely to be the first study that provides data from a larger number of households on the higher resource consumption per shower of younger people. While it is impossible with the current - if any - dataset to determine with certainty whether this is due to an age effect or a cohort effect, none of the literature reviewed for this study supports the explanation of age effects. Still, at least for elderly people, one could argue that they might find it difficult to remain in a standing position over extended periods of time, which might shorten the time they spend in the shower. Yet this does not explain why per-shower consumption continuously increases from participants in their 60ies compared to those in their 50ies, 40ies, 30ies, and 20ies. By contrast, there is already a rather broad and solid body of literature that strongly supports the explanation of cohort effects - yet with a mainly qualitative focus and without providing detailed quantitative data, as this study does.

The study results contribute to this body of literature by providing new evidence for what sociologists have described as *socio-technical transformation of conventions and behaviors*: A substantial change of norms and conventions for perceptions of comfort, cleanliness, and convenience, leading to increasingly resource-intensive consumption patterns (Shove (2004, 2003)). While these creeping transformations are described as dramatic in magnitude from a mid- to long-term perspective, they often gone unnoticed for their lack of sudden, disruptive shifts. As a result, these transformations are practically absent in policy analysis and in the public debate.

Aside from showering, similar dramatic behavioral shifts have also been observed for the use of space heating, air conditioning, and laundry quantities: U.S. studies report an increase in the amount of laundry per person to the threefold within 50 years (Biermeyer (2001)); similarly, Shove (2003) observes a five-fold increase in the frequency of bathing, showering, and washing clothes over the last century. Whereas most of the related literature describes an increasing shower *frequency* over the past decades, the present data suggest that there is an even bigger transformation going on in the resource consumption *per shower* than in shower frequency. Also, while the existing body of literature that describes similar changes of patterns mainly focuses on qualitative aspects to understand this shift of paradigm in the society, this study analyzes field data to quantify the magnitude of this trend. The findings are also interesting in the light of the widespread belief that young people are particularly concerned about the environment. This will be discussed more in detail in section 3.4.3.6.

An alternative explanation to the strong negative correlation between age and baseline consumption could be monetary aspects. Yet neither in the existing literature examined, nor in the dataset at hand, shower behavior is correlated with income (neither the baseline consumption, nor the treatment effect). Moreover, the dataset does not indicate any difference in the baseline consumption or in the response to the treatment between households who pay a fixed price for water / water heating ("Warmmiete"), and households who are charged in function of their utility consumption. This suggests that the alternative explanation of income effects can be excluded.

On the other hand, in terms of response to the intervention, section 3.3.6.3 suggests the device might have a bigger conservation effect on young people. Section 3.3.6.4 clarifies that this is merely an artifact of a spurious correlation, confounded with the higher baseline use of younger users: The apparent stronger reaction of younger people is actually driven by their high baseline usage. Yet even though age may not drive the treatment effect, from a practitioner's point of view, the discrepancy in the reaction of different age groups could be used to close the generation gap and also be useful for profiling, as section 3.4.3.5 will describe.

3.4.3 Research and Policy Implications

This section discusses the implications of the findings of this study in a broader context. It highlights the relevance of the main results from a research and policy point of view. More specifically, this section discusses the intervention in the context of behavioral nudging and libertarian paternalism and underlines the importance of the study as a demonstration of the cost-effectiveness of real-time feedback on a larger scale. It further highlights the relevance of the findings with respect to market barriers for innovative technologies and points out the implications of the findings for market segmentation (profiling). The section closes with a discussion of the findings in the light of changing norms and conventions and the relevance thereof for energy and water demand forecasts.

3.4.3.1 Amphiro as Example of Nudging

Human behavior is not always the outcome of controlled, conscious processes. Many decisions are based on automatic, unconscious processes and heuristics - simple and efficient rules that serve as mental shortcuts - instead of logic and rational choice (Tversky and Kahneman (1974); Kahneman (2013)). As a consequence, individuals often do not act in line with their preferences and behave in ways they would not if they were perfectly informed and rational decision makers ((Thaler and Sunstein (2009)). Most people care about certain ideals and have adopted goals for their behavior that are in line with them: Exercising on a regular basis, eating healthy food, reading more books, limiting their gin tonic consumption, reducing their carbon footprint, to name just a few examples. Yet when it comes down to acting on them, only a few manage to follow through with their good intentions and to bring their behavior in line with their goals. This is particularly true for the pursuit pro-environmental goals (Gutsell et al. (2012)): While many people care about the environment and want to live a sustainable lifestyle, in their daily lives, they fall short of their pro-environmental goals (Gutsell et al. (2012)). Or, as an article published in Nature phrased it, *"people do want to use less energy, but forget or put it aside in the rush of routine, or they don't know how"* (Woodside (2011)).

Over the past few years, behavioral economists have identified a number of cognitive biases which systematically prevent individuals from acting in line with their (long-term) interests; at the same time, they have developed strategies to overcome the negative outcomes associated with them. In this context, the widely read book on behavioral nudges by Thaler and Sunstein

(Thaler and Sunstein (2009)) has influenced the policy making of many governments and organizations (Hansen and Jespersen (2013)). The book describes a nudge as a small change in the decision making context that *"alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives"* (Thaler and Sunstein (2009), p.6).

The shower meter used in this study serves as such a nudge: It makes behavioral outcomes - individuals' resource consumption while showering - salient at the very moment when the behavior takes place. That way, the device draws attention to the consequences of the behavior. In line with the inattentive choice model by Taubinsky (2013), the device may thus serve as an *attention-based foundation for the formation of "good" habits*. Furthermore, by providing continuous feedback on the current behavior, the device can help individuals to monitor how they fare relative to their self-set goals. Moreover, its allows them to update misconceptions about their daily resource use (Attari et al. (2010); Allcott (2011b)).

Nevertheless, the device does not prescribe individuals how they should act on this information. They can choose to act on the information displayed, but they don't have to. This approach was coined as "*libertarian paternalism*" by Thaler and Sunstein (2009). This term conveys both the gentle and freedom-preserving character of these approaches and the goal of steering decision-making in directions that maximize both individuals' and society's welfare. The shower meter does not prescribe in any way how and to what extent users should reduce their energy consumption in the shower: In theory, they could reduce water temperature, shorten shower duration, reduce the flow rate, temporarily turn off the water, e.g., while they are soaping - or they could also decide to not act on it or to completely ignore it.

The data show a clear user preference for shortening shower duration. Flow rate, by contrast, is hardly manipulated. This is interesting from a policy point of view: Many states and countries 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 consumption (Power (2011)); the phenomenon of direct rebound is also discussed in this context, as people might respond by extending their shower time. By contrast, 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. Policymakers often privilege technical solutions, automation, and regulation over solutions involving the user-in-the-loop. While those approaches reduce the need for active user decisions and collaboration, they also curtail individuals' freedom. Conversely, this study serves an example that behavioral interventions which actively involve the user can yield a high conservation impact *and* preserve individuals' freedom to decide for herself.

In general, the nudging approach has received both widespread recognition and harsh criticism. The main line of criticism is that nudging works by *"manipulating people's choices"*

(Bovens (2009); Vallgårda (2012)). In response to this, Hansen and Jespersen (2013) distinguish between four broad types of nudges: The authors draw a line between transparent and non-transparent nudges, as well as between nudges that influence *"the attention and premises of reflective thinking" ("type 2")* vs. nudges that influence the automatic system and *"do not involve deliberation, judgment, and choice" ("type 1")*. In this conceptual framework, the shower meter qualifies as a *"transparent type 2 nudge"*, hence as a non-manipulative behavioral instrument: The devices makes consequences salient and facilitates consistent choices in a transparent way. This is corroborated by the result that inattentive, less conscientious users are more responsive to the treatment. At the same time, individuals who do not agree with the ends or means do not need to comply with the prompt for behavior change.

To summarize, the shower meter embodies a transparent nudge that elicits attention and reflective thinking about resource consumption. As such, it serves as an example for an *"effective way to influence citizens' behaviour without restricting freedom of choice"* (Hansen and Jespersen (2013)): By making behavioral outcomes salient, the device prompts decisions that are in line with individuals' goals and long-term preferences; yet at the same time, the user preserves the freedom to act on the information or not.

Overall, this suggests that the device operates through positive mechanisms, thus maximizing individuals' and society's welfare: It makes inattentive individuals better off by making resource use more salient; at the same time, nobody is made worse off, as the intervention does not coerce other individuals into a behavior change; it is thus particularly desirable from a normative point of view.

3.4.3.2 Real-time Feedback to Correct Misconceptions about Resource Consumption

The results of section 3.3.6.2 show that prior to the intervention, the majority of participants considerably underestimated their water consumption per shower. This is in line with previous research that shows a poor public understanding for their energy (Attari et al. (2010)) and water consumption (Attari (2014)). In particular - and also in line with previous research (Beal et al. (2013)) - people seem pretty unaware what "high" or "low" consumption is for a particular behavior like showering: While many high-users are convinced that their per-shower consumption is average or even at the low end of the spectrum, actual low-users tend to rank their consumption higher than it actually is. The conjecture is that *"self-nominated high users may be setting themselves a higher benchmark on what is low or personally acceptable consumption"* (Beal et al. (2013)).

At the end of the study, the treatment group had developed a good sense of their average water consumption (in absolute terms). By contrast, prior research shows that home water reports with social normative feedback on household's monthly water consumption do not improve knowledge and quantification abilities of program participants (Mitchell et al. (2013)). While the social normative component of these reports might possibly have given participants of that program a better sense for how they fare relative to peer households - an assumption that yet needs to be validated - participants' ability to quantify their consumption does apparently not improve. By contrast, real-time feedback in the shower substantially improved participants' ability to accurately estimate their average per-shower consumption. A better public understanding of the environmental contribution of different actions is considered as key to reduce resource consumption: *"When people are faced with a laundry list of advice, they may feel confused and overwhelmed, and consequently take no action, or they may carry out one or two actions - probably the easiest to remember and perform"* (Gardner and Stern (2008)). As a consequence, an intervention like the shower meter described in this chapter helps correct misconceptions and direct people's resource conservation efforts to high-impact behaviors. A better public understanding of water and *"energy use and savings could pay large dividends"* (Attari et al. (2010), p.1) and help to realize the *"substantial potential of behavioral interventions"* (Attari et al. (2010), p.6).

3.4.3.3 Evidence for the Cost-Effectiveness of Behavior-Specific Real-Time Feedback

This is probably one of the first studies that investigated the impact of real-time feedback on a particular behavior on a larger scale. Although several small pilot studies have shown very promising results for such systems, most of these systems have never overcome prototype status. As a result, a validation of the effect size on a larger scale had been missing so far. This study closes this gap by showing that it is feasible and cost-effective to scale up behavior-specific real-time feedback. First of all, the findings provide evidence that this kind of feedback can prompt substantial behavior change and yield immense conservation effects. Second, the study shows that conservation effects of this order of magnitude are possible even on a larger scale.

Altogether, the findings demonstrate that such technologies and interventions can be implemented cost-effectively, with substantial conservation impact, and with a short payback period.

3.4.3.4 External Validity Issues as a Market Entry Barrier for Innovative Technologies

A central question in the evaluation of most studies is what extent the results can be generalized to other systems, contexts, or larger populations. Given the magnitude of the effect size in this study, this question is of particular interest: Was the large effect size observed due to specific characteristics of the implementation strategy or of the study sample? Or is it possible to replicate these high conservation effects also on a large scale with the same or another real-time feedback system? Is real-time feedback on specific behaviors really the panacea to successfully promote resource conservation? Are the mechanisms identified proprietary to this specific device and sample of households, or would they be the same among the general public or for a different behavioral intervention? These questions are crucial for evaluations of the cost-effectiveness, carbon abatement, and competitiveness of future projects. Section 3.4.2.8 discussed what the findings on the psychological mechanisms signify for the external validity of this study and of related projects. The conjecture is that based on the study results regarding the psychological mechanisms, selection bias is less of a problem: In case the recruitment strategy favors a self-selection of individuals with particularly strong *pro-environmental attitudes*, the results of this study suggest that this does not affect the net treatment effect. Similar, if these biases advantage the self-selection of particularly *conscientious* individuals or individuals with a *lower ex-ante consumption* than the general population, the results of the present study indicate that this would rather lead to an underestimation of the effect size than to an overestimate. Unless the extensive survey data have failed to cover additional key aspects relevant to the treatment effect, the intervention should work equally well also among a larger and more representative population.

These findings could have far-reaching consequences on existing market entry barriers for innovative feedback technologies: Many pilot studies have yielded quite encouraging results that showed higher treatment effects for behavior-specific real-time feedback applications compared to the currently prevalent forms of feedback (Ehrhardt Martinez et al. (2010)). Yet this superior performance in small-scale field tests is generally questioned with respect to effects of self-selection. It is argued that these results cannot be extrapolated to large-scale roll-outs where the conservation effects would be dramatically pushed downwards (McKerracher and Torriti (2013)). As a result, even in the case of very encouraging pilot tests, new feedback technologies face credibility issues regarding their performance in large-scale deployments. Compared to traditional products, this generates an enormous additional barrier to the development and market entry for these technologies: Not only do they have to fulfill all the technical tests, standards requirements that new products typically need to go through for a proof of concept: Even if a novel feedback technology can prove its reliability and superior performance in a small-scale pilot study, this is generally not accepted as a reliable validation of performance, as conservation effects superior to the state of the art meet credibility concerns regarding the external validity of these outcomes. Unless a company manages to roll out its product on a really large scale, the conservation effects will be put into question and discounted for concerns of external validity. In combination with the considerable investment that is necessary to develop such hardware-based systems, the difficulty to overcome the credibility issues concerning the external validity of pilot studies is clearly a market entry barrier for companies before they can profit from economies of scale. This is a particularly challenging endeavor for small companies.

In the case of the research partner of this study, the company Amphiro AG, the sheer size of the treatment effect and the sample size have already contributed to validating the cost-effectiveness of the device in the field. Yet the findings on behavioral mechanisms may be similarly relevant from a practical point of view, as they suggest that the intervention can work equally well for a broad audience.

3.4.3.5 Implications for Profiling

Section 3.3.6.3 and 3.3.6.4 identified several variables which 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 particularly tends 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 the 50% of the households with a baseline consumption above the median would raise the treatment effect by 74%. 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% when focusing solely on households with an above-average baseline consumption. These numbers are in line with Allcott (2011b) who found that restricting the Opower program to half the eligible population (those with a baseline consumption above the median) would increase the treatment effect by 74% and the cost-effectiveness by 43% (using the same cost calculation as in section 3.4.1.2).

From a practical point of view, however, in the case of showering, it may not be as easy to identify households with a high baseline consumption. Not only for policymakers and program designers it is difficult to know which households are on the high use end of the spectrum of energy and water consumption in the shower (at least as of today, water and energy end uses are not disaggregated, thus e.g., a high water bill is not necessarily a good predictor for resource consumption in the shower): User self-assessment prior to the provision of feedback is also poor, as section 3.3.6.2 and previous research show (Beal et al. (2013)). Given that baseline consumption is by far the best predictor for subsequent savings, age - which is highly correlated with baseline consumption - could be used as a good proxy to identify high users and, by extension, households with a high conservation potential. In the current dataset, restricting the intervention to participants below 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 study sample 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 validate this.

To summarize, profiling could considerably further increase 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 also be useful as a more accessible, yet effective proxy to identify households with high expected savings.

3.4.3.6 Implications of Age-Dependent Baseline Usage for Projections on Future Energy Demand

As section 3.3.3 shows, the results point out a strong correlation between age and per-shower baseline consumption. The discussion in section 3.4.2.9 attributes this to a cohort effect: A dramatic increase in resource consumption per shower due to socio-technical transformations. These findings contribute to the body of research that describes a dramatic increase in resource consumption, in particular in the domain of *"comfort, cleanliness, and convenience"* (Shove (2004, 2003)): This increase is explained with creeping changes in conventions on lifestyle, social norms, and paradigms in society. Similar phenomena have previously been described for the use of space heating, air conditioning, laundry quantities, and shower frequency.

The implications of these phenomena extend far beyond the domain of showering. First of all, changing lifestyle conventions, norms, and paradigms almost certainly also affect behavior in other domains, for instance mobility: Swiss transportation statistics show that younger people travel the threefold distance on a daily basis compared to older people (FDHA (2013), p.11). In that case, however, cohort effects are more difficult to disentangle from age effects (e.g., no more commute to work after retirement) and factors like increased wealth and income, rebound effects, and other trends. In the case of showering (or at least for the consumption per shower), these alternative explanations can be excluded to a large extent (see section 3.4.2.9).

To understand to relevance of these issues, one must keep in mind the magnitude of the changes observed: For showering, this dataset suggests an increase in resource consumption within a single generation to the 2.7-fold (the 2.3-fold per-shower consumption along with an increased frequency). Prior studies on other behaviors in the domain of comfort, cleanliness, and convenience even report increases to the three- resp. five-fold over the past decades. The behaviors described affect the most energy- and water-intense areas of residential resource consumption, in particular space and water heating. Yet none of this is taken into account in projections on future resource demand. Those scenarios typically account for technological progress and substitution, and - as of recently - increasingly include rebound effects. Yet other than those financially and technologically-driven changes, behavior is assumed to be stable.

The conjecture is that behavioral shifts due to slowly changing norms and conventions tend to go unnoticed due to their lack of observable, disruptive shifts. Nevertheless, they can have profound repercussions on future demand of energy and water. Over the past decades, a growing and increasingly solid 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)). These now widely debated rebound effects may not be the only mechanisms that potentially thwart the resource efficiency gains achieved through technological progress and through policy interventions: It may be the case that profound changes of lifestyle, norms, and paradigms additionally undermine the outcomes of these efforts in a considerable way. Given the magnitude of these changes within a few decades, it would be wise to further investigate them and to consider them in energy and water demand projections.

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 do not live up to their ideals, using by far more resources for daily actions than older generations, as in the case of showering.

More quantitative research in particular is needed to investigate these issues 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, as they allegedly use the most energy and water in the shower. At the same time, the data also suggest that as high users, young people are also likely to be more responsive to feedback. Therefore feedback might help to address these issues and to reduce the generation gap.

3.4.4 Limitations

Although every effort has been made to ensure the internal and external validity of the study, a couple of aspects should be kept in mind - or might be a good starting point for future research. In addition to various questions regarding the generalizability of the findings, this includes in particular the following five aspects: potential heterogeneity of attitudes and personality within two-person households, mechanisms of social normative feedback, persistence of the effects, and the role of goals.

3.4.4.1 Further Considerations on External Validity

While section 3.4.2.8 already addressed the topic of external validity and section 3.4.3.4 discussed practical implications of the behavioral mechanisms identified with regard to this topic, some facets have not been covered so far. Remaining key questions in this context concern the validity for other behaviors and presentation media: To what extent are the results of this study proprietary to the specifics of the showering context? To what extent are they specific to the feedback device used, and to what extent can the findings be transferred to other feedback systems, to other behaviors, and to other domains of resource consumption (or beyond)?

Validity for other Behaviors Based on the findings of this study, it is impossible to predict to what extent the magnitidue of the treatment effect in particular can be generalized to real-time feedback on other behaviors and to what extent it is proprietary to the showering domain. Compared to other areas of resource consumption, showering and feedback on resource consumption in the shower may be particular in several aspects, as pointed out in table 3.1: First of all, there are few distractions available in the shower: No smart phones, only a very limited set of other tasks that could be carried while in the shower, and - in most cases - no presence of other individuals. Consequently, shower meters are much more likely to receive user attention than traditional in-home displays or web portals. Yet these particular set of characteristics may be harder to encounter or to implement in other areas of resource consumption. In addition to that, water consumption in the shower is visible and tangible to the user at the point of consumption. While a few other behaviors share these characteristics (e.g., tap water use), most processes of electricity, oil, and gas consumption in the household operate out of sight and behind the scenes (e.g., space heating). Third, in the case of showering, the metrics are intelligible: People are familiar with liters, which is not the case for e.g., kWh, therms, or BTU. Furthermore, strategies to reduce resource consumption (in particular, by reducing the shower duration) are quite straightforward to understand and to implement. Moreover, users know that the resource consumption displayed is not the result of many people's contribution, but solely the outcome of their individual current activity. While this aspect could also be generalized to other areas of resource consumption (e.g., tap water usage, bathing, car usage, food intake, etc.), for most processes, it is more complicated to attribute responsibility for resource consumption: End uses like space heating, air conditioning, lighting, or refrigeration would have to be split between household members according to some reasonable allocation scheme. Yet that would dilute both the "accountability" and the savings achieved from a single person to a group level. Finally, showering is a frequent, roughly daily behavior for most people. As frequent rehearsal leads to habit formation (Taubinsky (2013)), it is likely that users start to adopt new behavioral routines (e.g., taking shorter showers) after a while. By contrast, the adoption of behaviors that occur on a less frequent basis (e.g., maintenance of household equipment) might be a more lengthy process.

Validity for other Presentation Media The feedback medium and form of presentation should also be considered in the context of relevance for and transfer to other domains. One should keep in mind that the shower meter *amphiro a1* differs from traditional IHDs for electricity consumption in several dimensions, including level of data aggregation, immediacy of the feedback, prominent location of the display, direct link to a precise behavior, and visual format of the display, to name the most prominent ones (see table 3.1). It is not clear to what extent some of these characteristics may have influenced the outcome of the intervention. While a detailed investigation, e.g., of the visual display characteristics, is clearly beyond the scope of this study (and dissertation), it could be of interest to investigate these questions from a computer-human interaction (CHI) research point of view. Some aspects like immediacy of the feedback might be part of a follow-up research study (see section 5.3.2).

One key aspect regarding the presentation medium is clearly that the information is automatically being provided to the user (data push), lowering the barrier for user access to the feedback information. By contrast, other real-time feedback applications require repeated user actions (data pull). Many IHD displays need to be activated by the user and most require regular battery replacements. The *amphiro a1*, once installed, powers itself and displays its information automatically and in a prominent place (information push). In that sense, it is more comparable with paper-based feedback which is mailed to the households without them initiating the mailing. Data push approaches may be more effective than data pull strategies, in particular in the long run (3.4.4.4).

3.4.4.2 Potential Heterogeneity of Attitudes and Personality within Households

A potential shortcoming of the dataset collect may be the fact that one person per household filled out the survey. For simple demographics (e.g., number of household members), this is irrelevant, as the answers would be consistent. As far as attitudes, environmental orientation, etc. are concerned, the data are based on the survey respondent's answers, which means that in two-person households, no information on attitudes and the personality of the second household member was collected (or rather, was assumed to be consistent with the survey respondent's profile). While this assumption does probably not hold for personality traits, this approach is unlikely to fundamentally have affected the outcome of the study. First of all, 92% of the participating two-person households were 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 good proxy for both partner's attitudes.

Second, if this issue should affect the dataset in any meaningful way, it will mainly add noise to the dataset and reduce the strength of correlations, not increase them. A couple of sensitivity analyses could be run to quantify the impact of a random household member vs. a household member who entirely shares the attitudes and personality of the survey respondent. While this should not fundamentally change to the results, follow-up studies could collect data on attitudes and personality for all shower users in order to provide a more complete picture.

3.4.4.3 Social Normative Feedback and Positive Mechanisms

The results indicate that the treatment effect has been generated mainly by positive mechanisms, not by peer pressure. The question arises whether this is also the case for studies where peer comparison information is more salient, as it is the case e.g., in the Opower or BEN Energy home energy reports. The device used in the study does deliberately not communicate any descriptive norm (e.g., the typical resource consumption per shower for the average household), nor does it provide an explicit injunctive evaluation (e.g., "good" or "below average") of individuals' behavior. The only indicator that conveys some sort of normative feedback is the energy efficiency class (and the polar bear animation that is associated with it); yet without specifying which class the average user would fall into. However, people seem to be interested in normative feedback: Comparisons with other households and mean consumption values was a common theme among user questions and suggestions.

In this study, the emotionality trait - which captures sensitivity to pressure - as well as individuals' self-reported tendency to compare themselves with other people did not affect the treatment effect. The question is if this also holds true for interventions which emphasize the social comparisons feature more and which give an explicit injunctive rating. In a similar vein, the additional display of information on other household member's consumption data - or the presence of a second household member - did not increase the treatment effect in this study. However, one cannot exclude the possibility that this feature might have failed to produce any significant difference in the outcome due to shortcomings in its implementation. Maybe participants simply did not understand the feature or pay enough attention to the device to observe it. Given the broad evidence from previous feedback studies that shows how social normative feedback can affect subsequent resource consumption (e.g., Goldstein et al. (2008); Nolan et al. (2008)), more research is needed to evaluate whether a wisely chosen social normative feedback component might a) further increase the treatment effect and b) whether this shifts the psychological mechanisms away from positive mechanisms, towards negative processes that operate through pressure and are more effective on individuals who are susceptible to pressure.

3.4.4.4 Persistence of the Effects

One of the key questions of this study is the persistence of the effects. While the results do not support evidence for a decay of the conservation effect during the study, they do not provide certainty on how people respond to the device in the long run. This question is crucial for evaluations of impact and cost-effectiveness of the device. As outlined in section

2.1.2, the literature so far has found mixed evidence for the long-term effect persistence. Recent literature on habit formation suggests that it takes on average two months for modified daily actions to become an automated process that no longer requires self-control (Lally et al. (2010)). This is in line with Taubinsky (2013), who also stresses the importance of rehearsal of behaviors for habit formation. Thus, for the majority of the participants of this study, the new, less resource intense showering process should already have become a habit. While many recent studies find a high degree of persistence of the savings (e.g., Raw and Ross (2011); Allcott and Rogers (2014); Foster and Mazur Stommen (2012)), a number of feedback studies report that the effect eventually dissipate and households return to pre-intervention consumption levels (e.g., Fielding et al. (2013); van Dam et al. (2010)).

Recent meta-studies have reached the following three conclusions: a) the vast majority of savings can be attributed to behavior change, not to the adoption of new, energy-efficient technologies (Ehrhardt Martinez et al. (2010)), b) savings are much more persistent than previously generally assumed (Allcott and Rogers (2014)), and c) effects tend to persist for feedback systems with data push, while they diminish for systems with data pull (i.e., which require repeated and active user action to receive feedback). In line with these findings, the effects of the *amphiro a1* shower meter can be expected to be quite persistent as long as participants keep the device installed in their shower. This topic is currently being investigated by another team member, Vojkan Tasic, who analyzes long-term data both of web portal users and of devices that were collected after 12 months. His work will provide a more complete answer that is built on an evaluation of datasets of actual long-term measurements (14 months).

3.4.4.5 Side Effects

The evaluation of the device was limited to the impact on shower behavior. However, it might be the case that behavior change in one domain also spills over into other areas. This could create both positive side effects (positive spillover) and negative side effects (moral licensing). While the measurement data in this study are limited to shower behavior, chapter 4 will evaluate this question in detail in a different setting. Yet as the chapter will point out, such a system-level perspective would allow for a more comprehensive cost-benefit evaluation of these kinds of interventions. In anticipation of the results of the following chapter, one of the major implications of the study on side effects is that policymakers should refrain from interventions on behaviors with a small impact (e.g., unplugging cell phone chargers) or from interventions with low impact. Instead, they should a) focus even more on environmentally significant behaviors such as space heating, water heating, mobility or food consumption and b) use interventions that prompt substantial behavior change. Given the relatively large environmental footprint of showering (see section 3.1) and the magnitude of the treatment effect, both in relative and in absolute terms (section 3.4.1.3), both applies to real-time feedback in the shower.

3.4.4.6 Role of Goals

The results indicate that goals or individuals' goal orientation (tendency to monitor progress towards goals) significantly moderate the treatment effect. Furthermore, the data suggest that participants who report having set themselves a savings goal reduce 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 might be worthwhile to further explore the role of goals - self-set and imposed externally.

To conclude, while some of these aspects should be kept in mind and be considered with caution, the device highlights the potential of behavior-specific real-time feedback. One can easily imagine similar forms of feedback directly integrated into building infrastructure and appliances. Yet further research is needed to determine to what extent the findings can be generalized to other behaviors, appliances, forms of feedback, the general population, and to other cultures. The results also raise a series of related questions, in particular on the long-term stability of the effects, on the value of real-time feedback, and whether the mechanisms observed are also applicable for other forms of feedback (for instance, deferred feedback, social normative feedback, or feedback relative to imposed goals). These questions are excellent candidates for follow-up research studies.

Chapter 4

Field Study on Side Effects

4.1 Motivation

It is a common, well-known practice among dieters to treat themselves to a snack or richer meal after having completed an exhausting, demanding, or unpleasant physical task (Fishbach and Dhar, 2005). This is a typical example of moral licensing: feeling entitled to a self-indulgent behavior that one would not permit oneself without first having done a positive action. Recent contributions in consumer research and policy, marketing, and social psychology journals provide evidence of moral licensing in various behavioral domains including purchasing decisions, nutrition, racism, and sexism. A recent online article in Science Norton (2012) reports that drivers of hybrid cars violate crosswalk laws more often than drivers of conventional cars, attributing the observed difference to moral licensing. The same pattern may apply to environmental behavior: resource conservation in one area may make people more wasteful elsewhere. And just as the rewarding food treat might contain by far more calories than those consumed during the activity that licensed it, some environmental campaigns might do more harm than good to the environment overall when the licensing effect is taken into account. On the other hand, just as people dieting for weight loss also tend to exercise more, behavior change in one environmental domain might also open a window of opportunity for positive spillover into other domains through increased awareness or motivation Lawson and Flocke (2009). This study sheds light on cross-domain effects of conservation campaigns in an energy-intensive and frequently targeted area: residential energy consumption. This study thereby responds to the call made by Stern (2011) for research on the effect of taking one pro-environmental action on subsequent actions. He points out the contradictory predictions for subsequent actions made by behavioral scientists as one of the fundamental research questions for future psychological research: "Which of these mechanisms predominates with high-impact behaviors, and under what conditions, are fundamental research questions of obvious importance to limiting climate change."

The residential sector accounts for 21% of the CO_2 emissions from fossil fuel combustion EPA (2011a) and for approximately 22% of total primary energy consumption in the U.S. (U.S. Department of Energy (2012)). U.S. primary energy consumption in the residential sector has more than doubled since the 1960s eia (2014), while per capita residential electricity consumption more than tripled between 1960 and 2008 IEA (2011). Consequently, residential energy demand has received considerable attention in programs that aim at reducing energy consumption. In the past several years, particular attention has been paid to nonmonetary incentives, such as neighborhood comparisons of the consumption of electricity Schultz et al. (2007); Ayres et al. (2009); Allcott (2011b) or water Ferraro and Price (2013). These programs typically yield savings on the order of 2-5% across the population in the targeted area of utility consumption, equivalent to the effect of a price increase of 11-20% Allcott (2011b). A large number of such isolated environmental campaigns have been researched, and many more have been undertaken – yet typically only the singular effects of the target behavior in isolation are analyzed. For a full cost-benefit analysis of environmental campaigns, however, the complete change in energy consumption must be taken into account. A better understanding of these mechanisms and the quantification of their impact is crucial for well-informed policy decisions.

This paper explores whether a behavior change campaign in one domain (water consumption and the directly associated energy for water heating) also has a measurable impact on the consumption of other utilities – in this case, electricity. Water and electricity consumption are chosen as the dependent variables for four reasons: First of all, they are the outcome of everyday behaviors and relevant for every household (unlike airplane travel). Second, both account for a large share of a household's carbon footprint: Water heating is the second largest energy end use after space heating in residential buildings, accounting for 18% of the site energy use. Water heating accounts for 13%, and electricity accounts for 71% of residential primary energy consumption (U.S. Department of Energy (2012)). Third, water and electricity consumption reflect the aggregated real-world impact of multiple behavioral decisions instead of a single action that may or may not be relevant for a household. And fourth, thanks to existing infrastructure and technology, they are easier to measure than the quantity of waste produced or recycled.

This study investigates the impact of a water conservation campaign in a multifamily building complex on residents' electricity consumption. By providing weekly water conservation tips and individual feedback on water usage to half of the study participants, the study examines whether evidence for the dominance of positive spillover or moral licensing can be detected in the residents' electricity consumption. If the effect on water consumption is viewed in isolation, the campaign can be considered another example of a successful non-price-based behavioral intervention. Yet by taking the analysis one step further, this study shows that apartments exposed to the water conservation campaign do indeed increase their electricity consumption relative to the control group, which can probably be attributed to the dominance of moral licensing.

The following section provides an overview of related work including the key findings. Section 3 describes the study setting, the intervention, and the data analysis methods used. Section 4 outlines the data collection procedure and summarizes the impact of the water feedback intervention on residential water and electricity use. Finally, section 5 concludes with implications for policy and further research.

4.2 Methodology

The study was conducted at a multifamily building complex with 200 apartments to investigate the impact of an environmental campaign on water and electricity consumption in the field. Apartment water consumption was measured daily and electricity consumption weekly. After two weeks of baseline data collection, half of the apartments received weekly feedback on their per capita water consumption along with water conservation tips for seven weeks.

4.2.1 Partners and Collaborators in the Spillover Study

This study was developed, implemented, and analyzed in collaboration between the Fraunhofer Center for Sustainable Energy Systems (Cambridge, MA, USA) and ETH Zurich (Bits to Energy Lab). The implementation was carried out in collaboration and with the support of Corcoran Properties, the managing company of the study site. Corcoran Properties granted the researchers permission to carry out the intervention on their premises and to collect the utility data for research purposes from the participating apartments. They also helped with the collection of the paper surveys and as a first-level support for residents.

4.2.2 Timeframe, Site Description, and Recruitment

The study was carried out from May to July 2011 at a multifamily property in Lynnfield, Massachusetts, a town in the Greater Boston area. The property consists of 200 apartments in three neighboring five-floor buildings constructed in 2009 with identical floor plans and a similar building orientation (Figure 1), managed and rented out by a single property management company. Apartment size varies from 74 m^2 (smallest one-bedroom units) to 113 m^2 (largest two-bedroom units) with a mean of 91 m^2, compared to a U.S. average of 129 m^2 for new multi-family building units built in 2010 (U.S. Census Bureau (2011)). According to the property management, residents are a mix of all age groups, with an upper-medium level of income and education. At the beginning of the study, 14 units were vacant, with the majority of the units occupied by one (48%) or two persons (38%) (M=1.72, SD=0.84, N=186). According to the property management, demographics and rental policy are the same across all three buildings.

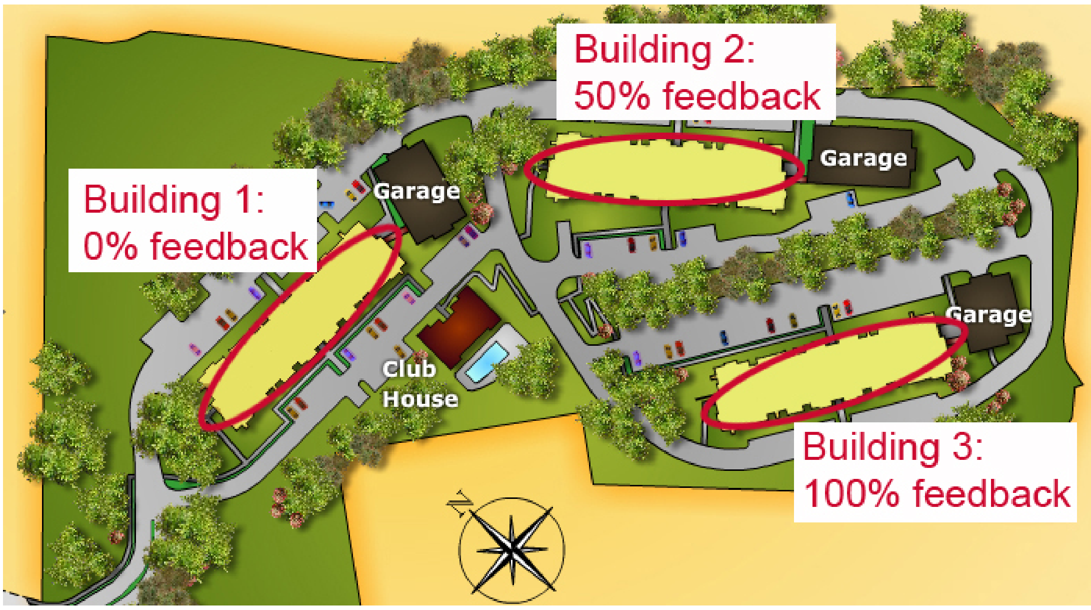


Figure 4.1: Property layout with the triplet buildings and their assigned percentage of apartments in the treatment condition.

In contrast to most multifamily buildings in the U.S., all utilities (electricity, gas, and water) are submetered at the apartment level; tenants pay for electricity and gas, but not for water usage. This implies that one can exclude direct microeconomic income effects that would ascribe the increased electricity consumption to the additional disposable income generated by reduced expenses for water. All units are equipped with the same space and hot water heating (gas) and cooling (electric) system built into each apartment, the same water fixtures (faucets, toilets), and the same or very similar major appliances. The only exception is that 25% of the apartments – evenly distributed among the three buildings – have a gas instead of an electric oven). This excludes equipment-specific and building-structural aspects (e.g., level of insulation) as major factors influencing usage and attribute most of the variance in the utility consumption among apartments to behavioral factors and observed factors, such as number of occupants and floor space. The property management company emphasizes the "green living" aspects of the community, e.g., energy-efficient appliances, low-flow water fixtures, and dual flush toilets in all apartments.

4.2.3 Feedback Flyer Description

The intervention consisted of a series of seven double-sided water consumption feedback flyers that were slipped under the door of treatment group households on a weekly basis (see Figure 4.2 for an example of a front and back side). The flyers were placed in such a way as to be barely visible from outside the apartments for privacy reasons and to avoid drawing attention from control group households, yet allowing the researchers to check whether they had been picked up from the floor. With the exception of residents who were absent over extended periods, flyers of the previous week had always been picked up by the residents when the next flyer was distributed.

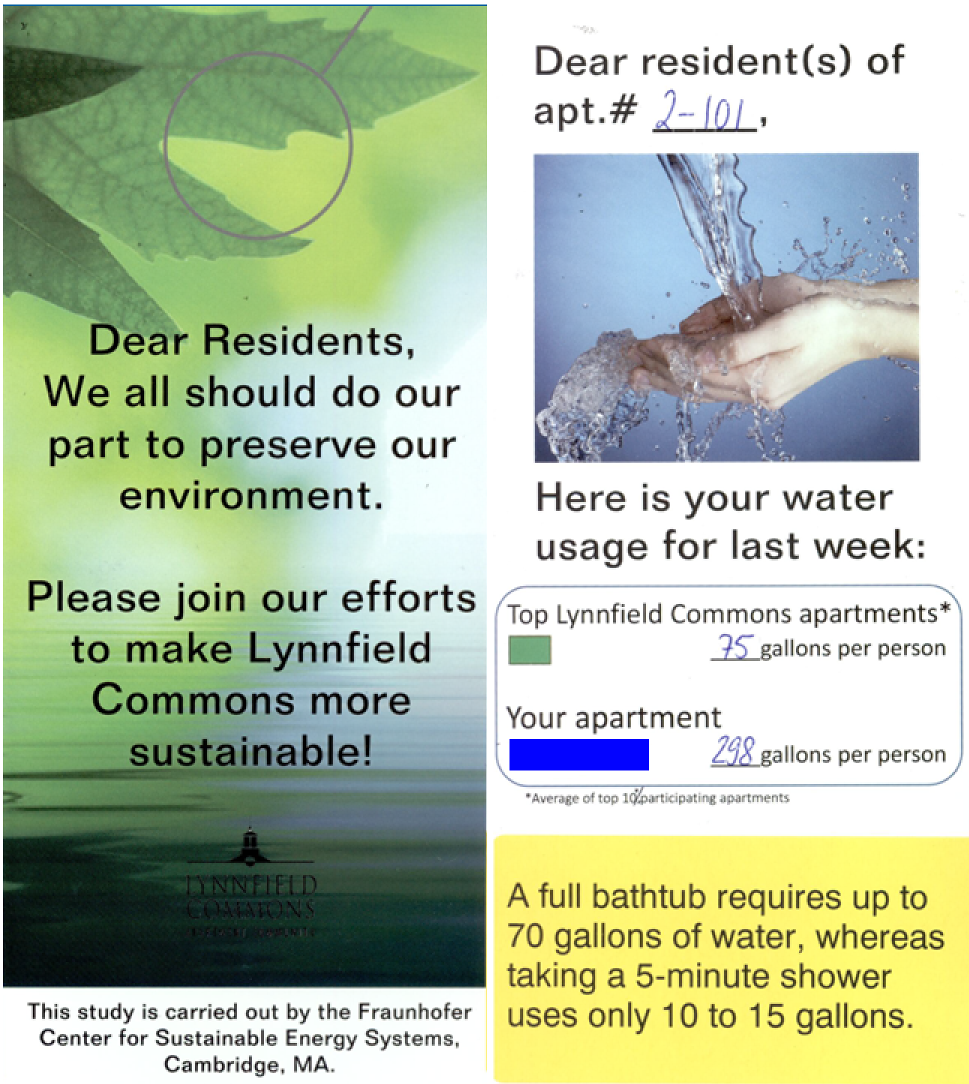


Figure 4.2: Feedback flyer distributed to treatment group apartments. On the left: front side of weekly flyer - on the right: example for the personalized back side.

To facilitate recognition of subsequent flyers as part of the same campaign, all seven feedback flyers came with the same front side (Figure 4.2 on the left) with an appeal to environmental social norms ("Dear Residents, We should all do our part to preserve the environment. Please join our efforts to make Lynnfield Commons more sustainable!"), the property logo to underline community identity and the fact that the campaign was backed by the property management. The backside (Figure 4.2 on the right) contained that week's water conservation tip and a personalized section with the apartment number and its per capita water usage of the last week compared to the "Top Lynnfield Commons apartments" (average of top 10% participating apartments). The top 10% apartments were chosen instead of the mean of all apartments to avoid having to distinguish between apartments above and below the mean, respectively. Otherwise, below-average consumers might be encouraged to adjust their water consumption upwards towards the social norm (Schultz et al. (2007)). By taking the mean of the apartments below the 1st weekly decile, only two to six apartments per week received a message that their apartment's consumption was below the mean of the 10% reference group. The remaining apartments were in the control condition or above the reference group mean; those below the reference mean were typically units that had not been occupied for several days that week, so their inhabitants would probably not interpret their low consumption value as a consequence of being excessively "green." On the other hand, when asked for the usefulness of that comparison in a follow-up survey, none of the respondents indicated the suspicion that the relatively low consumption value of the reference group might be due to a higher absence rate for that group. All water conservation tips stated a concrete action and its associated weekly water savings potential based on the water flow rate of the fixtures installed at the property (e.g., "Shorten your shower by a minute or two and you'll save up to 20-35 gallons per person per week. Turn off the water while soaping or shampooing"). The flyers did not contain advice that would have simultaneously affected electricity consumption, such as behavior concerning the dishwasher or washing machine.

4.2.4 Research Design

Before information on the planned study was distributed to the residents, apartments were assigned to two experimental conditions, one that would receive weekly water consumption feedback (treatment group) for seven weeks and one that would not (control group). To facilitate the feedback distribution process and reduce the likelihood of information spillover via discussions between participants in the two experimental conditions, a quasi-experimental design was implemented, using the following group assignment: Building 1 was entirely assigned to the control group, building 3 entirely to the treatment group, and apartments in building 2 were randomly assigned to the treatment and control groups. The two groups did not reveal a significant difference in any of the observed variables.

Two weeks before the study began, every apartment received a one-page information sheet describing the organization conducting it including a contact address, the utility data that would be collected anonymously from each apartment for research purposes over the next eleven weeks, and the possibility to opt out. The information page for the two experimental conditions differed in two respects: while apartments assigned to the control group were told that they would receive feedback on their utility consumption at the end of the study, treatment group apartments were informed that they would "receive energy-saving tips as well as feedback on your household's consumption in the form of a paper card that a researcher will slip under your door once a week." Their version also included a small image of the feedback flyer to facilitate recognition in the future.

4.2.5 Participation Rate and Filtering

Three apartments opted out before the study began, and another three did so during the study. They were excluded them from further data collection and analysis. In contrast to experimental studies requiring participants to actively opt into a study, this opt-out recruitment model allows for collection of data from a more representative sample (only about 3% opted out), while respecting the choice of households who do not wish to participate.

The raw dataset collected consisted of daily water and weekly electricity meter readings from May 4, 2011 through July 19, 2011 (eleven weeks) of all 200 apartments in the complex. In addition, the property management provided data on the number of occupants per apartment, floor space, location of each apartment number in the building, the major appliances installed in the apartments, and a list of move-ins and move-outs during the study. Figure 4.3 depicts the study timeline with the seven feedback distribution events after two weeks of the baseline period as well as the measurement events (daily for water, weekly for electricity).

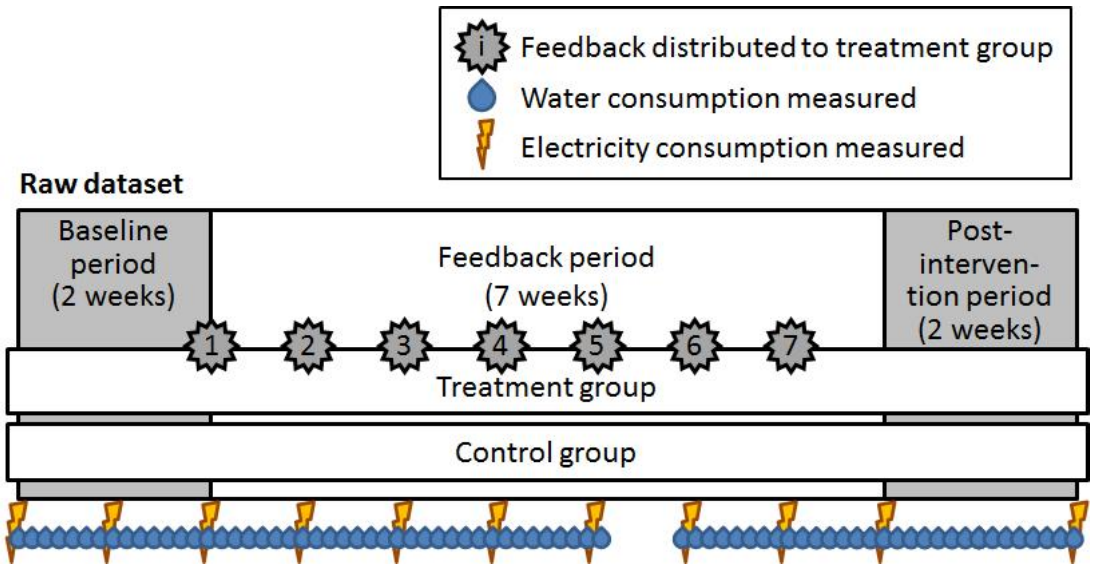


Figure 4.3: Illustration of the study (raw dataset) with its three phases baseline period, feedback period (flyer distribution to the treatment group), and post-intervention period. Water consumption data collected daily, electricity readings weekly.

Figure 4.4 gives a schematic overview of the steps taken to analyze the data. After several apartments were filtered out (described just below), water and electricity consumption data were controlled for observed variables (e.g., number of occupants per apartment). Thereafter, treatment and control group datasets were separated; both were normalized to the control group mean of each measurement interval to adjust for time-dependent factors (e.g., weather). Then the normalized data of each group were pooled by study periods (figure 4.4).

To ensure that differences between the groups were not simply due to different occupancy patterns, water consumption values served as proxy to infer apartment vacancy. Unlike elec-

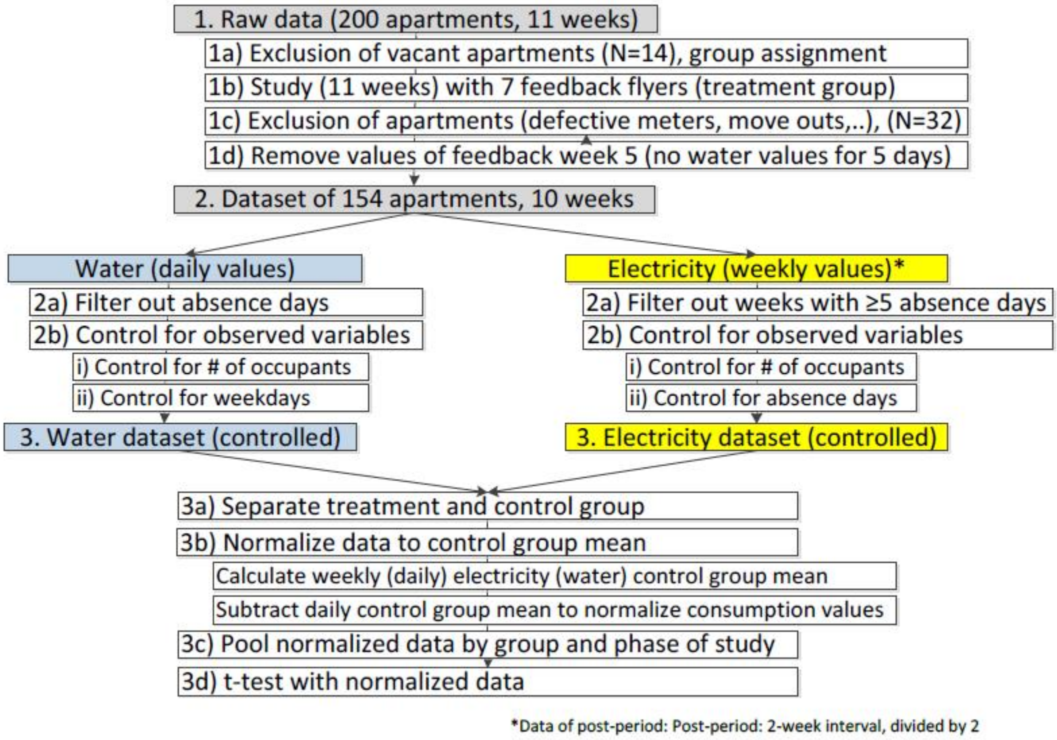


Figure 4.4: Steps undertaken to analyze water and electricity consumption data from raw meter readings to the final dataset analyzed.

tricity, water is usually only consumed when someone is at home. This allows an accurate inference of the vacancy of an apartment over several days using the daily water consumption data. An exploratory analysis of the dataset revealed that water meters reported water consumption values up to two gallons (eight liters) per day for vacant apartments (measuring uncertainty). Days with water consumption up to eight liters were therefore considered as "absence days." Apartments with longer periods of absence were entirely excluded from the study (a total of 16 units; see the following paragraph); short periods of absence and single absence days were excluded from the remaining water consumption dataset and controlled for in the weekly electricity data to reduce the variance. Both experimental groups showed no difference in the number of absence days in all phases of the study (see the analysis of electricity data for details).

Based on the same exclusion criteria for all apartments, 14 apartments were excluded due to vacancy / late move in at the beginning of the study; nine apartments were excluded for technical reasons (water meters reporting zero/constant consumption every single day); six due to opt-out of the residents; and another 16 were excluded due to move-outs, change of tenants, or extended periods of absence (absences of 15 consecutive days or more or more than eight days during baseline period). One more apartment was excluded as an extreme outlier (leakage or a defective meter assumed), as its weekly per capita water consumption was up to 10 times the average consumption of the other participants. In the end, data from

154 apartments or a total of 275 individuals were analyzed for the study, with 77 apartments in the treatment and 77 in the control group. The two groups did not show a significant difference in floor space (M_T=90.7 m^2 vs. M_C = 90.4 m^2, p = 0.89), the number of occupants (M_T = 1.71 vs. M_C = 1.86, p = 0.31), utility consumption (see following sections), percentage of apartments with a gas oven (22% and 23%, respectively), or any other observable factors. Due to a failure of the water meters (five days without daily updates in week 5 of the feedback period, June 17-21), that week was entirely excluded from the analyses in order to analyze water and electricity data of identical time periods. The final study period therefore covered two weeks of the baseline period, six weeks of the intervention, and two weeks of the post-intervention period.

The variability in the water consumption data was high, both within and between households. An exploratory analysis showed a strong correlation of household water consumption with the number of occupants and weekdays; therefore, these factors were controlled for after excluding absence days. For that purpose, the values of apartments with more than one occupant were adjusted with a correction factor based on the ratio of means: the mean water consumption of all apartments with i occupants (i=1, 2, 3, 4) was calculated for each day. Then the ratios of these means were taken on a daily basis; the means of these daily ratios were used as correction factors (for other examples and more details on this ratio correction factor method see e.g., Cundiff et al. (1966); Breslow and Day (1975); Gfroerer (1998); Ruijter et al. (2006)). The same approach was then followed to control for weekdays (for instance, the mean water consumption on Fridays and Saturdays was on average twice as high as on Sundays and Mondays).

4.2.6 Logistics and Implementation

Water meter readings for each apartment were collected by the Inovonics' TapWatch submetering system and updated every afternoon. Every apartment's utility meter is connected to a pulse counter/wireless transmitter unit that sends its meter reading once a day to a central data concentrator and communicator unit, from which daily meter readings are retrieved and stored by the system provider. Feedback flyers on the previous week's water consumption were distributed on Wednesdays by members of the research team. During these visits, they also read the electricity meters of all participating apartments.

4.3 Results

4.3.1 Impact on Water Consumption (Primary Effect)

Figure 4.5 shows the variation of the daily median water consumption before and after excluding absence days and controlling for the number of occupants and weekdays. This procedure reduced the absolute value of the standard deviation of the daily means from 148 to 64 liters, and the ratio of the standard deviation/mean of daily means from 0.42 to 0.27. Taking into account all 10,780 observations (70 days, 154 apartments), the mean daily water consumption was 356 liters/apartment/day, with a median of 265 liters/apartment/day and a standard deviation of 350 liters/apartment/day (98% and 132% of the mean and median value, respectively).

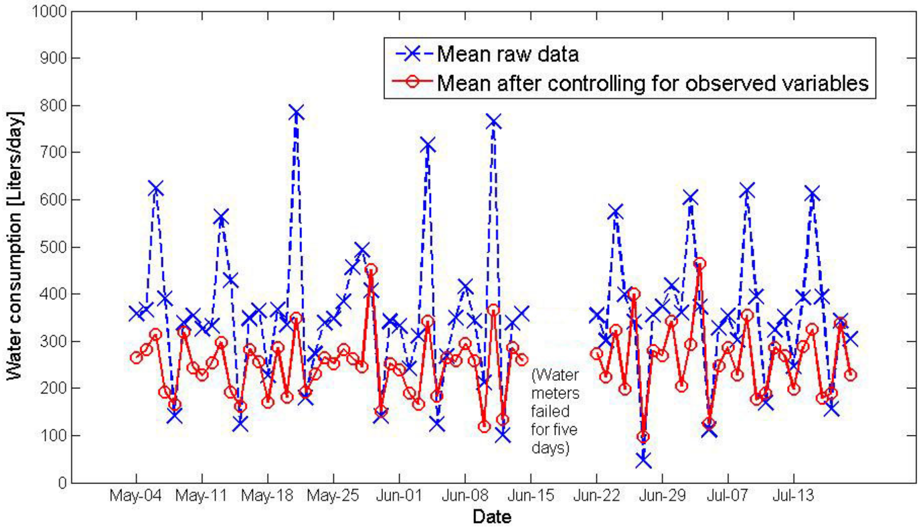


Figure 4.5: Means of daily water consumption before and after controlling for absence, number of occupants, and weekdays.

Hereafter (step 3a in Figure 4.4), the dataset was separated into treatment and control groups. The daily control group mean was subtracted from every apartment's daily water consumption to normalize for unobserved time-dependent effects before pooling the data of the two experimental groups into the three periods of the study baseline (two weeks), feedback (six weeks) and post-intervention period (two weeks).

During the baseline period, the treatment and the control group used a similar amount of water (M_C = 238 liters/person/day, M_T = 242 liters/person/day); the daily treatment group mean was on average 1.9% above the control group's. By contrast, during both the feedback and the post-intervention period, the mean daily treatment group consumption was 4.1% below the control group mean on average. Figure 4.6 illustrates the mean water consumption (aggregated on a weekly basis) of the two experimental groups.

Table 4.1 shows the test statistics of the daily water consumption for baseline, feedback, and post-intervention period after normalizing both experimental groups to the daily control group mean. Whereas there was no significant difference in the water consumption between the two groups during the baseline period (p_{baseline}=0.53), the treatment group used significantly less water during the feedback period (p_{feedback}=0.0036); consumption in the post-intervention period did not show a significant difference (p_{post}=.27). Normalizing to median values instead of means (medians being more robust to outliers) yielded similar values (p_{baseline}=.55, p_{feedback}=.033, p_{post}=.26). One can thus assume that the campaign did have a measurable impact on the target behavior (water consumption).

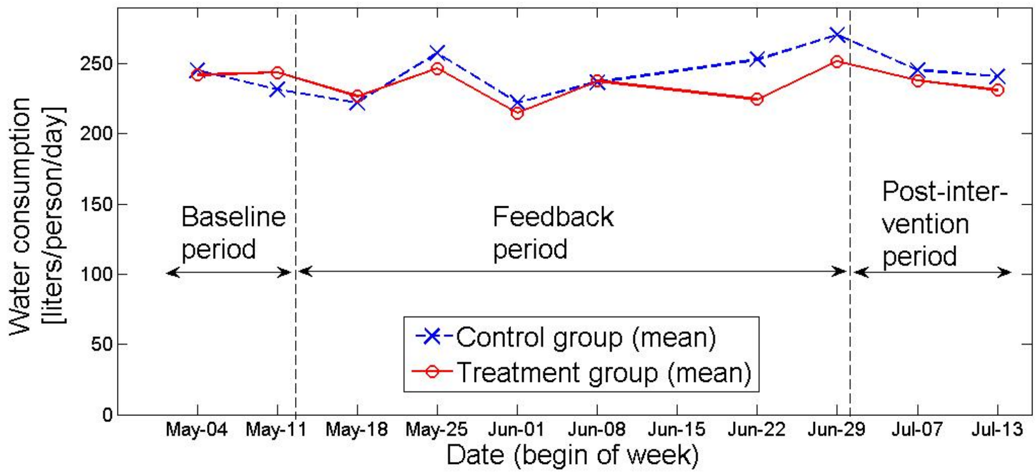


Figure 4.6: Daily water consumption means (averaged on weekly basis) after controlling for the number of occupants, weekdays, and filtering out absence days.

Study phase	N treatment (no. of observ.)	N control (no. of observ.)	Effect size vs. baseline (%)[a]	t-statistic	p-value
Baseline	1008	995	/	-0.63	0.529
Feedback	2961	2926	6.0	2.10	0.036*
Post	997	997	5.5	1.09	0.275

[a] Calculation of the effect size with DiD-method; * $p < 0.05$

Table 4.1: t-test of pooled water consumption values, normalized to daily control group mean for baseline, feedback, and post-intervention period.

4.3.2 Impact on Electricity Consumption (Side Effect)

Electricity consumption data in this study differ from water consumption data in three major respects. First, they were only collected on a weekly basis. Second, they were less subject to absence than water usage, and third, they were highly dependent on outdoor conditions, as air conditioning represents a large fraction of electricity consumption in the summer. As electricity consumption data are aggregated over one week, it is not possible to filter out "absence days" from the analysis in the case of water. Instead, after controlling for the number of occupants (by following the same procedure as described for water), the data were controlled for the number of absence days per week. Absence days per week were inferred from the water consumption data. Correction factors were calculated by following the same procedure as for the number of occupants. The weekly mean electricity consumption was calculated over the number of absence days per week, then the values were adjusted with this correction factor.

Weeks with five or more absence days were excluded due to the small number of data points. As an additional check, the number of absence days per week was compared between the two groups, as this might explain a difference in electricity consumption. However, the two conditions did not show a significant difference of absence days per week during any phase of the study, neither in the baseline period (treatment: M_T=0.33, SDT=0.77; control: M_C=0.31, SD_C=0.73, $t(300)$=-0.18, p=0.85), nor the feedback period (treatment: M_T=0.45, SD_T=1.07; control: M_C=0.45, SD_C=1.12, $t(905)$=0.01, p=0.99), or the post-period (treatment: M_T=0.25, SD_T=0.68; control: M_C=0.36, SD_C=0.31, $t(291)$=0.68, p=0.50). As electricity data for the post-intervention period had only been collected after two weeks' time, this value was divided by two and half of the sum of absence days of these two weeks was used as control factor, thus creating a comparable metric for the post-intervention period. In the end, the analysis included 1,362 valid observations (154 apartments times 9 (2+6+1) measurement intervals for baseline, feedback, and post-intervention period, reduced by 24 data points for absence over five or more absence days of the week). The mean daily electricity consumption was 111 kWh/person/week, with a median of 101 kWh/person/week and a standard deviation of 54 kWh/person/week (48% and 53% of the mean and median value, respectively).

Figure 4.7 shows the weekly electricity consumption means. As one can see, electricity consumption between the first and last week of the study increases by approximately 75% for both groups. This is due to increased outdoor temperatures and resulting higher electric consumption due to air conditioning: while the first weeks of the study took place in moderate conditions (14 °C] daytime average), the last weeks of the study coincided with the warmest days of the year (28 °C] daytime average). During the feedback period, the treatment group mean is on average 6.9% above the control group mean, compared to 1.3% during the baseline period and 1.7% during the post-intervention period.

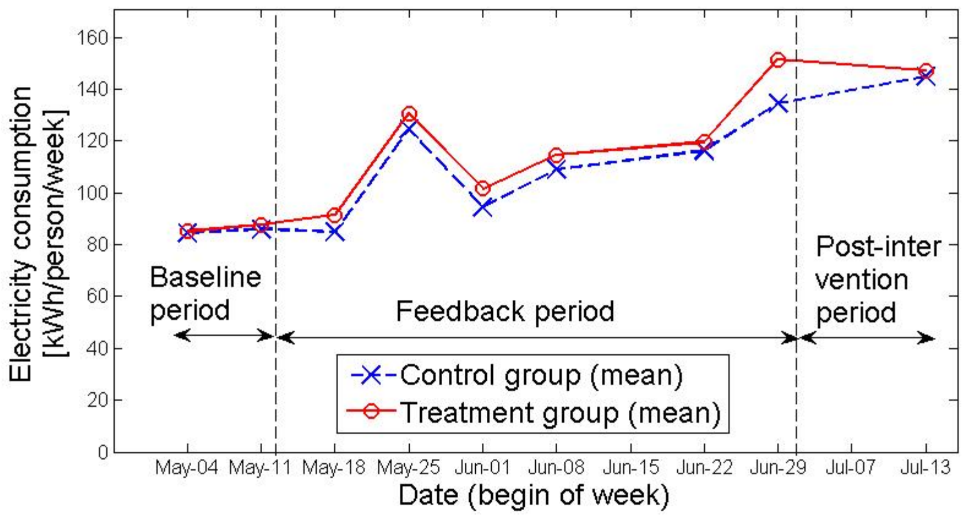


Figure 4.7: Weekly electricity usage means of the two experimental groups after controlling for the number of occupants.

Table 4.2 shows the test statistics of electricity usage for the baseline, feedback, and post-intervention periods after normalizing both experimental groups to the weekly control group mean. Whereas the difference in the electricity consumption between the two groups was not significant during the baseline period (p=.757), a significant difference between the two groups was found during the feedback period (p=.035). Additionally, the values were normalized to the control group median (instead of the mean) as a value that is more robust to outliers; the results obtained were very similar (baseline period: p=.760, feedback period: p=.037).

Study phase	N treatment (no. of observ.)	N control (no. of observ.)	Effect size vs. baseline (%)[a]	t-statistic	p-value
Baseline	154	149	/	0.03	0.757
Feedback	456	451	5.6	-1.88	0.035*
Post	76	76	0.3	-0.34	0.660

[a] Calculation of the effect size with DiD-method; * $p < 0.05$

Table 4.2: t-Test of pooled electricity consumption data, normalized to qeekly control group median, for baseline, feedback, and post-intervention period.

4.4 Discussion of the Study on Side Effects

This study represents one of the first quantitative field trials on cross-domain effects in residential utility consumption behavior. It investigates the side effects of a water conservation campaign on residents' electricity consumption and finds evidence that people exposed to the water campaign did reduce their water consumption as expected. Yet at the same time, they increased their electricity consumption relative to the control group. While the data cannot provide a proof for the precise mechanism of the psychological process at work, the findings are consistent with what the majority of articles that analyze incongruous actions describe as moral licensing. In the energy economics field, the concern for increased consumption following the adoption of an environmental action is typically subsumed as "boomerang effect" (Goeschl and Perino (2009); Harding and Rapson (2013)), yet this also includes phenomena from neoclassical economic theory, e.g., income effects. Some studies that report incongruous actions frame the mechanism as "guilt reduction" (Gneezy and Rustichini (2000)), "moral cleansing" (Sachdeva et al. (2009)) or a "warm glow effect" similar to charitable giving (DellaVigna et al. (2012); Kotchen and Moore (2007)). Ultimately, the boundaries between the terms are blurry and the topic is currently subject to a rapidly growing body of research across disciplines. After all, the implications of such observations are highly relevant across domains, not only for environmental campaigns and energy policy.

4.4.1 Net Energy Balance of the Study on Side Effects

The following paragraph quantifies the impact of behavioral spillover in this study using a net energy balance and then proceeds to a discussion of its potential limitations. The chapter concludes with direct implications of the findings for the design and evaluation of environmental campaigns and policy. These findings call for the adoption of a more comprehensive system-level perspective in the evaluation of environmental programs.

To quantify the environmental impact of the observed cross-domain effects, one can compare the energy saved through water conservation with the increased electricity consumption. In this study, the treatment group reduced its water consumption by 6.0% or 15 liters/person/day relative to the control group. Typically, approximately 40% of domestic (excluding yard) water consumption is hot water (U.S. Department of Energy, 2012); assuming that residents saved a similar ratio of hot and cold water, the energy conserved by reduced water usage in the study is 0.5 kWh/person/day[1] On the other hand, electricity consumption increased by 5.6%, resulting in an additional electricity use of 0.89 kWh/person/day (111 kWh/7 days · 5.6%). Thus, in terms of on-site energy balance, the energy saved by reduced (hot) water consumption was offset by the increased electricity consumption by nearly a factor of two. Extending the lens to source energy, one would have to take into account source-site factors for losses that are incurred in the production, transmission, and delivery to the site. Using EPA's national average values for electricity (3.34) and natural gas (1.047) (EPA,2011c), the net energy balance from a source energy perspective is even more negative, by a ratio of about 1:6 (energy savings from hot water conservation vs. additional energy from increased electricity consumption). By comparison, the energy conserved from reduced water consumption in terms of water treatment is much smaller, approximately 0.2 kWh/person/day [2]. Thus, both in terms of on-site usage and even more in source-energy balance, this campaign had a clearly net negative energy outcome despite its success with respect to water conservation. Although a growing body of literature has found broad evidence for moral licensing, most of these studies have been carried out in a laboratory setting or are based on self-reported behaviors in surveys. A limited number of very recent studies have looked into implications of moral licensing for green electricity tariffs, but negative side effects through moral licensing are still a blind spot in program design and evaluation. After all, these kinds of psychological mechanisms are not limited to patterns of household electricity consumption, but can also extend to energy consumption and supply in general, food, transportation, and overall consumer choice.

[1] Underlying assumptions: Energy E [kWh] required: E= m $\cdot c_p \cdot \Delta T/EF$, where m is the mass [kg], c_p the specific heat capacity [kJ/kg/K], ΔT the temperature difference between cold and hot water [K], typically 45 K (from 10 to 55°C), and EF the water heater energy factor, typically 0.61 for gas heaters (U.S. Department of Energy, 2011). This results in energy savings of 15 · 0.4 · 4.179 J/liter/K · 45K / 3600 J/kWh/.61 = 0.5 kWh per person per day.

[2] EPRI (1996) reports a total electricity use of 1.4 kWh/1000 gallons of water produced at a typical water treatment plant and 1.8 kWh/1000 gallons for production at a typical groundwater utility, resulting in approximately 15 liters/(3.785 liters/gallon) * 1.6 kWh/gallon * 3.34 = 0.2 kWh source energy conservation per person per day from the reduced amount of water treated (including energy for pumping).

4.4.2 Implications of Study on Side Effects

Our findings raise many questions about the net outcome of energy efficiency information campaigns and policymaking. First, environmental campaigns that are motivated by a sense of "every bit helps," or the hope that they might create a window of opportunity to more meaningful environmental behavior, can potentially be harmful and should be evaluated carefully. Not only can they waste individuals' time and effort on low-impact activities, they might also generate a warm glow effect of "already doing something," both among individuals in their daily lives and among policymakers in programs they support. This might be amplified by the general public's poor understanding of the energy consumption impact associated with different behaviors (Attari (2011)): people might invoke low-impact behaviors of such campaigns as confirmation of their environmental engagement. As Gardner and Stern (2008) put it in their "Short List" of household actions to curb climate change: "When people are faced with a laundry list of advice, [...] they may carry out one or two actions - probably the easiest to remember and perform. However, the behaviors that are easiest to remember and perform, for example, turning out lights when leaving rooms, tend to have minimal impact on climate change. Thus, long and unranked lists of behaviors are likely to be ineffective at best and may even be counterproductive, if they lead people to feel satisfied that they have done their part after accomplishing very little." In combination with the human tendency to choose the easier alternative of environmental actions for oneself (Attari et al. (2011)), these campaigns might actually crowd out environmental actions that would result in higher energy savings or CO_2-abatement, or license negative behaviors that people might otherwise abstain from, such as increased electricity consumption or airplane travel. Second, it might be the case that a considerable amount of the current environmental program efforts and funds actually generate a much smaller – or even a negative – net impact on CO_2 emissions than the current program evaluations suggest. These considerations should be taken into account in the evaluation of future environmental campaigns and policy. In particular, the long-term implications will be of interest here: just as the positive effects of many behavior-based efficiency campaigns fade over time, it is important to understand whether the negative side effects caused fade away even more quickly or are more persistent than the positive outcomes. Policymakers need to know whether they have to account for some short-lived side effects, or whether a program might actually create side effects that reduce, negate, or even exceed its benefits not only in magnitude, but also in persistence over time. The findings are also relevant for the ongoing environmental policy debate on policy strategy and individual responsibility. Energy efficiency can be achieved through individual behavioral change (curtailment) or through better technology and structural changes. The latter is often costly and resulting energy efficiency benefits may be affected by the rebound effect. Consequently, many interventions encourage individuals to change their attitude, values and behavior (the 'ABC' paradigm of attitude, behavior, and choice). However, that kind of policy faces increased criticism for yielding only marginal, incremental improvements, while reinforcing the status quo of the current system, and deflecting attention away from the many institutions involved in structuring possible courses of action (Shove (2010); Stern (2000)). If individual curtailment programs license other negative behaviors and crowd out investments in better technology, curtailment programs should be analyzed with even greater caution for their potential to solve environmental problems. Third, it is important to acquire a better understanding of these mechanisms that can influence behavior across domains and of their magnitude in order to develop programs that minimize their risk or impact. Therefore, it might be necessary to develop environmental messaging that prevents people from overestimating the positive impact of their pro-environmental actions or that focuses on actions with the greatest impact. It will be a challenge to find the right balance between communicating that an individual's behavior is important, without providing people with a license for less pro- environmental choices in other domains. Theory on moral licensing that has been developed in other behavioral domains could help to address this issue, for example by making environmental behaviors more important to individuals' identity, framing campaigns with respect to goal commitment instead of progress, or making hypocrisy visible (Dickerson et al. (1992); Miller and Effron (2010)).

If these energy policy implications are not taken seriously, the current environmental campaign focus on a single behavior might result in missing out crucial parts of the whole picture and could lead to repercussions on other energy-related behaviors that may be greater than the outcome of the targeted behavior. This calls for the adoption of a more comprehensive view in the evaluation of programs and future scenarios that incorporates the potential impact of licensing effect. This would help to develop more accurate economic models and make predictions about CO_2 emissions more realistic and reliable.

4.4.3 Limitations of the Study on Side Effects

The results of this study are subject to a number of limitations. The study focuses on the analysis of directly measurable short-to-mid-term behavior change, not long-term predictions. Certain behavior change processes might require more than a couple of weeks to take effect and to stabilize into long-term habits. For instance, in the study by Harding and Rapson (2013), the crowding-out of conservation diminished after 3-6 months. In that regard, research on mid- and long-term processes will certainly contribute important complementary aspects to this analysis of directly measurable short-to-mid-term impacts, but it will also be subject to an even greater variety of other unobservable factors. This study makes a valuable contribution to a better understanding of the direct and measurable side effects of behavior change programs; further research is necessary to investigate the long-term effects of such interventions. Caution is also warranted concerning the interdependence of water and electricity consumption. To a certain extent, household water and electricity usage is coupled.

For some behaviors, such as clothes washing or running the dishwasher, conserving water also conserves energy. On the other hand, no typical household behavior could be identified in this context that substitute electricity consumption for water consumption. From this perspective, savings or increases in water consumption would result in changes in the same direction for electricity consumption. To reduce the interdependence, behaviors with a known interrelation were not included in the water conservation tips. Nevertheless, as the feedback campaign addressed apartment water usage in general, it is likely that at least some individuals did not only engage in the water-saving actions that were explicitly pointed out as water conservation tips on the flyers. In addition, they might have themselves identified some commonsense means of conserving water (e.g., reducing the number of laundry loads) that would also have affected electricity savings – however, in a positive correlation with water conservation, whereas the results do not support evidence for increased electricity consumption. This implies that the real impact of the licensing effect might be even greater than the data suggest, as the additional electricity consumption of some licensed electricity consumption behaviors would be offset in the data by the electricity savings from a reduced number of laundry loads. Another limitation of the study results concerns the seasonal nature of water and electricity consumption. As parts of the study coincided with some of the warmest weeks of the year, the use of air conditioning highly influenced electricity consumption: the median electricity consumption in the warmer weeks was more than 50% higher than in the more moderate weeks. Both air conditioning in summer and space heating in winter account for a large portion of a household's energy consumption. Furthermore, both are more subject to regular user adjustments than other end uses, making the occurrence and measurability of licensing more likely than in moderate climate conditions. Finally, by using water and electricity consumption data that are aggregated on a household level, it is impossible to specify which behaviors were influenced by the intervention and to what extent. Also, these measurement data do not reveal participants' perceived efforts, attitudes towards the campaign and towards their own behavior. Follow-up research is necessary to better understand and the specific psychological mechanisms that lead to the observed difference in electricity usage between the two groups.

Despite these caveats, the results of this study underscore the need for more research to better understand the underlying psychological mechanisms, to verify and quantify the environmental impact of moral licensing in similar contexts and other areas of environmental consumer behavior, as well as to reliably quantify the magnitude and persistence of such cross-domain effects on a larger scale.

Chapter 5

General Discussion and Implications

Chapter 5 summarizes the key findings of this thesis and outlines their implications for research and policymaking. After briefly recapitulating the context and motivation of the work, results of the field studies on behavior-specific feedback interventions and moral licensing are taken up. Effect sizes of the interventions under study, their suitability for large target groups, and the underlying mechanisms driving behavior change are given due attention in the context of a practical application of such measures. The chapter closes with a discussion of the limitations of this work, an outlook on future research, and a conclusion.

5.1 Context and Motivation

Humans are goal-oriented and adaptive in their decision-making; yet due to imperfect information, bounded rationality, and cognitive biases, they sometimes fail to act in line with their preferred outcomes (Jones (1999)). In many situations, human behavior is a product of automatic and unconscious processes, not the result of conscious and deliberate choices (Kahneman (2013); Cialdini (2009)). As a consequence, people often act in ways they would not if they had paid attention, had access to better information, and had higher self-control (Thaler and Sunstein (2009)). For instance, most people care about the environment and want to conserve resources; but even though they would derive financial and moral utility from living up to their attitudes and ideals (Levitt and List (2007)), in their daily lives they act contrary to their preferred outcomes (Woodside (2011); Hansen and Jespersen (2013)).

Over the past few years, behavioral economists have identified a number of cognitive biases that cause systematic deviations from rational decision-making and prevent individuals from acting in line with their long-term preferences. Based on these insights, a variety of behavioral strategies has been developed to counteract these systematic patterns and the negative outcomes associated with them. The consumption of natural resources is one of the domains where these strategies have already been implemented with success in large scale interventions to address global and regional challenges such as water scarcity, energy security, or air pollution. Behavioral interventions are often viewed as a politically feasible instrument to reduce inefficient consumption of natural resources quickly and at scale (Gardner and Stern (2008)). In particular, feedback - providing information about one's own or other people's behavior - has been identified as a cost-effective and relatively persistent tool to affect consumer choices (Allcott and Mullainathan (2010); Allcott and Rogers (2014)).

While behavioral interventions promoting resource conservation have been rolled out to millions of households, the underlying psychological mechanisms are not well understood (Ferraro and Price (2013)). A better comprehension of these mechanisms would pave the way for more effective strategies, while minimizing adverse effects. In particular, it is still unclear how feedback interventions on resource consumption affect individuals' utility. In the case that they help people act in line with their preferences, they increase individuals' utility. On the other hand, if they coerce individuals into conservation behaviors through psychological pressure, this would induce a loss of individual utility, potentially undoing all of the welfare gains of reduced resource consumption. Another critical issue in this context are spillover processes: So far, it is still unclear whether these programs trigger cross-domain adoption of additional environment-friendly behaviors (positive spillover) or reduced engagement elsewhere (moral licensing). Thus, a thorough evaluation of the real net performance of behavioral programs is lacking. Moreover, feedback interventions do not reap their full potential yet: The prevalent forms of feedback deployed today are not very well suited to make processes of resource consumption salient, transparent, and controllable for the user (Faruqui et al. (2010); Fischer (2008)): Enhanced billing, the currently most prevalent form of feedback, is limited to quarterly or, at best, monthly bills aggregated to the household level. As a result, consumers are not able to make the link between the individual action and its outcome on resource consumption. While highly cost-effective, these programs generally yield a rather modest conservation impact between 1 and 3% for electricity and between 4% and 7% for water. Electricity smart metering pilots, which typically provide feedback in 15-minute intervals or even in real time on in-home displays or web portals, still require the user to actively seek out this information ("data pull system") and only provide data aggregated to the household level. Recent pilot studies report treatment effects between 1 and 6%. While *behavior-specific real-time* feedback systems have repeatedly been described as more effective (Ehrhardt Martinez et al. (2010); Fischer (2008)), they still need to demonstrate their feasibility and performance in the field. In particular, it is unclear whether the promising potential suggested by small pilot studies is limited to a specific subset of individuals, for instance those with particularly strong pro-environmental attitudes or exceptionally high technology affinity, or whether the deployment of such systems is also beneficial and cost-effective on a large scale.

In a nutshell, behavioral interventions have already been identified as powerful, scalable, and cost-effective tools to promote resource conservation. In order to maximize their impact on resource consumption and to minimize adverse outcomes, it is crucial to a) understand the underlying psychological mechanisms, b) investigate potential side effects, and c) demonstrate the scalability and cost-effectiveness of behavior-specific real-time feedback.

5.2 Key Findings and Implications

In order to better understand the issues outlined in section 5.1, two field studies have been designed and conducted in the context of this thesis. Both studies sought to overcome limitations and gaps of previous research. The datasets collected are unique in their scope and quality, providing a solid basis to investigate mechanisms, moderating factors, and side effects in depth. This section will briefly describe the properties of the datasets collected before presenting the seven key findings and implications of this thesis, discussing each of them in a separate paragraph.

Large and Unique High-Quality Datasets Most existing research on resource conservation can be broadly grouped into two categories: On the one hand, studies analyzing large utility consumption datasets along with a handful of household characteristics; on the other hand, survey-based studies collecting an extensive set of psychological variables (attitudes, beliefs, preferences, etc.), but typically only self-reported behaviors. Hardly any study unites the best of both worlds: A combination of granular real-world measurements of resource consumption *and* detailed survey data. The study in chapter 3 in particular offers this rare set-up by providing fine-grained behavior-specific resource consumption measurements with extensive survey information.

Both studies measured the outcomes of participants' behavior in naturalistic and highly relevant settings. The study on side effects of behavioral interventions described in chapter 4 probably presents the first quantitative field study on psychological cross-domain effects in residential utility consumption behavior. While there is ample evidence from various field for negative psychological side effects (moral licensing), the vast majority of those studies had been carried out in a laboratory setting. The specific setup of the study in one housing complex made it possible to control for confounding non-behavioral variables and to rule out alternative explanations like income effects.

By the same token, the study presented in chapter 3 collected granular pre- and post-intervention data on the resource consumption related to a specific behavior (showering) over two months in 697 households, resulting in the world's largest dataset on shower behavior. This dataset is complemented with high-quality survey data on household characteristics, attitudes, preferences, beliefs, and personality traits. This combination of psychological variables and measurement data is unique. While previous research only had access to a subset of these variables (or was exclusively relying on self-reported data), the dataset presented here allows for a joint analysis of variables that previous research had identified as key determinants of residential water and energy consumption. This makes it possible to evaluate which factors genuinely interact with the treatment and which correlations are spurious.

In this regard, the results of chapter 3 put into question the findings of a number of previous studies: While it may seem trivial that variables that are relevant to the model should be analyzed jointly, in practice, most studies using actual measurement data only had access to very limited complementary data - typically, a set of basic demographics. While separate regressions are useful to establish correlations between predictors and net outcomes, they do not allow to disentangle underlying mechanisms and the role of household characteristics. Furthermore, separate regressions can produce biased estimates, attributing effects to wrong predictors or failing to uncover mechanisms that are concealed by cross-correlations, as the discrepancy of the results of the separate and the joint models in chapter 3 demonstrates. Nevertheless, several studies in the past have attempted to establish causal links between the set of variables collected, although essential predictors were missing in their model. In contrast, the combination in chapter 3 of granular measurements, detailed survey data, a focus on a specific behavior (showering), and of a huge effect size make it possible to reveal links between variables that are typically masked in the presence of the confounding factors that characterize datasets collected in the field. These properties make the combination of shower meters and survey data an excellent research platform to empirically investigate relationships that otherwise might be undetectable with a sample size of this order of magnitude. The insights collected as part of this thesis point out the potential of this platform for future research on behavioral interventions, resource consumption, or even broader fields of application.

Based on these datasets, the following seven findings have been identified as particularly relevant:

Real-time feedback on concrete behaviors can cost-effectively prompt substantial behavior changes. First, while previous research indicated that real-time feedback on specific behaviors could yield higher savings than deferred feedback that is based on aggregated data, recent meta-studies share a rather pessimistic view, questioning both the external validity of previous trials and the cost-effectiveness of large-scale deployments of systems providing this kind of feedback: Existing studies providing behavior-specific real-time feedback were limited to a handful of households due to high equipment costs or due to the fact that they were experimenting with prototypes; in other cases, these studies used feedback systems that (would have) required active user collaboration on a regular basis ("data pull systems", e.g., logging into a web portal, or charging and activating an in-home display). In contrast to that, the study in chapter 3 used the relatively inexpensive feedback system *amphiro a1* that - once installed - automatically provides real-time feedback on shower behavior. The results suggest that feedback on resource consumption, when implemented correctly, can yield higher savings effects than the currently established literature has settled for. As suspected by several meta-studies, but so far not validated on a larger scale, real-time feedback on concrete behaviors can cost-effectively prompt substantial behavior changes. The average treatment household reduced the energy and water consumption per shower by 23%. This is a substantially larger reduction than the currently most prevalent form of feedback on residential energy consumption, paper-based consumption feedback mailed to households by the U.S. company Opower or by the Swiss company BEN Energy, typically yielding savings between 1 and 3%. In particular, the reduction is considerably higher than the treatment effect achieved by recent smart metering trials: Those programs provided real-time feedback information on household electricity consumption using in-home displays or on web portals, yielding reductions ranging between 1% and 6% (Schleich et al. (2011); Darby (2012); Schleich et al. (2013); Degen et al. (2013); Carabias-Huetter (2013)).

One might argue that in contrast to programs with mailed reports or with electricity smart meters, the current study was "just about showering", yet the results speak for themselves: Projected to one year, the 23% treatment effect in the shower amounts to energy savings of 443 kWh for the average Swiss household, as well as a yearly conservation of 8'500 liters of drinking water. This reduces the carbon footprint by 94 kg per year for the average household. During the two-month study period, no decay of the effects was observed (see 3.4.4.4 for a discussion of long-term effects). Assuming a 3-year lifetime period, the cost per kWh saved is CHF 0.041 - not accounting for the additional benefits of water conservation. This number compares favorably with marginal generation costs for most energy carriers: It is roughly half of the marginal cost of current electricity production (0.074 CHF/kWh); 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 the first study 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.

Technically, given the standardized threads for shower fittings, the device can be deployed in 97% of Swiss (and European) showers. The program could be scaled up cost-effectively and quickly to large scale. Based on the numbers above, a 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). For comparison, the total electric production of all Swiss wind power plants in 2012 was 85 GWh. Moreover, for the installation cost of a single onshore 2 MW-wind turbine, 60,000 households- roughly one third of the households in the city of Zurich - could be equipped with such a shower meter.

The study further shows that real-time feedback on a specific behavior can even yield considerably higher *absolute* energy savings than electricity smart metering programs: The shower meters yielded a kWh-reduction that was between three and five times as high as savings reported in recent smart metering studies in Switzerland and Austria using in-home displays or web portals[1] - at a small fraction (roughly 10%) of the equipment cost. As a result and contrary to rather pessimistic conclusions in the established literature, the study shows that behavior-specific real-time feedback can be very cost-effective. The results further indicate that profiling - targeting households with particular characteristics in future interventions - can raise the conditional treatment effect and increase the cost-effectiveness of future deployments even further: Focusing on high consumers - for instance, households with an above-average baseline consumption - could literally double the treatment effect and reduce the cost per kWh conserved by nearly 50%. Similarly, targeting younger individuals - a strategy that might be easier to implement than identifying high per-shower consumers - would also considerably reduce the cost per kWh conserved.

Moreover, one should not forget that most regions in the world do not have access to abundant water resources as Switzerland does. As of today, already one fifth of the world's population experience absolute water scarcity; by 2030, almost half the world's population are expected to live in regions of high water stress (United Nations (2013)). National intelligence units are already warning of major regional and global conflicts over water, in particular in North Africa, the Middle East, and South Asia, contributing to political instability, posing risks to food markets and harming economic performance in many regions of the world (U.S. Intelligence Community (2012)). Many U.S. states also expect increasingly severe water shortages in the future; 2013 was the driest year on record for several of them. Begin of 2014, the U.S. southwest experienced unprecedented drought conditions (Weaver (2014)); after an official drought declaration, the governor of California urged citizens to reduce their water use by 20 percent (Dearen and Williams (2014)). This is just one example to illustrate that in other parts of the world, the shower meter's "additional benefit" of a substantial water conservation effect, as it is framed in Switzerland, may be even more crucial than the large impact of the device on energy consumption.

The savings effects are driven by positive mechanisms, not by negative psychological pressure. Second, a series of existing studies on behavioral interventions have expressed the need to better understand the underlying mechanisms and their welfare implications. It is important to understand if an intervention operates through positive or negative channels: If an intervention achieves positive effects for the society like carbon abatement and water conservation mainly through psychological pressure, these societal benefits come at the expense of individuals' utility. As a consequence, the intervention would be normatively less desirable as a policy instrument. Similarly, from a practical point of view, only a small group

[1] Note that a 100 kWh-reduction in the case of smart metering refers to electric energy, whereas in this study the number relates to the Swiss fuel mix for water heating (25% electric, 40% oil, 21% natural gas). For the 25% using electric heating, the numbers can be directly compared; for the remaining 75%, more complex calculations on energy conversion processes would be necessary and beyond the scope of this dissertation; in this context, one should keep in mind that the carbon intensity of water heating in Switzerland is 212 g/kWh, compared to 122 g/kWh for the generation of household-level electricity)

of users would probably be willing to install and to keep a device whose main effect it is to make them feel bad on a daily basis. While behaviorally effective, previous studies on feedback did not explain through which mechanisms feedback operates: They only had very limited or no access to data on user preferences, attitudes, and personality traits.

The combination of a large effect size and detailed survey data allows the study in chapter 3 to disentangle the underlying psychological mechanisms. The results suggest that the large savings effects are driven by positive mechanisms and do not support any evidence for the alternative explanation of psychological pressure. In particular, people who are more anxious or susceptible to pressure (measured by the HEXACO trait emotionality) do not respond stronger to the treatment. This is important, as it is politically - and commercially - much easier to promote the adoption of an instrument that unambiguously enhances individual and social welfare, than an instrument that makes (at least certain) individuals worse off for the sake of society. Yet the results do not provide any evidence that this might be the case. Furthermore, if the device operates through pressure, one would expect that it would be particularly effective for more conscientious users. Yet the findings suggest the opposite: More conscientious individuals respond less to the feedback information than the less conscientious. This indicates that the device is a particularly effective tool for people who tend to be unmindful of their long-term preferences in their daily actions: The device brings these preferences to people's attention at the very moment where they can incorporate those preferences in the decision-making context, helping them to act accordingly. In line with this, the results show that pro-environmental attitudes and individuals' propensity to monitor progress towards goals both interact positively with the treatment effect. Prior research shows that individuals can derive utility from resource conservation and pro-social behavior in general (De Young (1986); Bamberg (2003); DellaVigna et al. (2012)). Likewise, the ability to track one's behavior is one of the key ingredients to self-control, which in turn enables the individual to align her behavior with her preferred state (Baumeister (2002); Gutsell et al. (2012)). The results thus corroborate the explanation of positive mechanisms, that the shower meter resonates with individuals' preferences: Users who generally enjoy monitoring progress towards goals can use it as a tool to quantify their behavior and to compare the current state with their preferred ends. Likewise, individuals with strong pro-environmental attitudes (measured by their willingness to incur costs and efforts for the sake of the environment) are better able to live up to the benchmarks they have created for themselves. Altogether, the results suggest that the device does not exert pressure on individuals; it rather helps them to conserve resources by bringing information to their attention that is relevant to their preferences, yet which they would normally not be able to incorporate in the decision-making context.

These insights are fundamental for behavioral interventions: Energy conservation, in particular behavior change like taking shorter showers, are sometimes equated with personal sacrifice, unpleasant curtailment, and self-imposed constraints - things only a few particularly motivated environmentalists would be willing to inflict upon themselves - incarnated in Jimmy Carter's famous sweater (to make up for turning down the heat in the White House). By contrast, the results of the study in chapter 3 point out that feedback does not need to resort to negative mechanisms: It can prompt substantial behavior change by appealing to individuals' preferences and help users to follow through with their long-term goals. Individuals thus derive utility from using the device. This study thus puts feedback instruments into a new light: As mainstream tools that enable individuals to act in line with their preferences, instead of serving as remorse-inflicting gadgets for a small group of environmental zealots.

External validity and scalability: The net conservation effect is independent of environmental attitudes. Third and related to the previous paragraph, the question arises whether the large treatment effect in chapter 3 was driven by a small subset of households with particularly strong pro-environmental attitudes. This is crucial for studies with opt-in recruitment, as they might potentially be subject to self-selection biases: If the effects are driven mainly by a few environmentalists, and if this particular subset of people tends to self-select into these kind of studies, the external validity of those trials would be compromised and the findings could not be extrapolated to the general population. The study in chapter 4 followed an opt-out recruitment strategy, for which by design self-selection is rarely an issue (only three households opted out of the intervention). In contrast, the study in chapter 3 used an opt-in recruitment strategy. However, the findings show that real-time feedback can work equally well for a broad audience. In particular, they suggest that environmental attitudes did not affect the net outcome of the intervention: Thus, the savings were not driven by a few environmentalists. The study results indicate that people with very "green" environmental attitudes may well have tried harder to reduce their consumption (e.g., by paying more attention to the device); yet people with less pro-environmental attitudes appear to have found it easier to reduce their consumption, as they tended to start out from a higher ex-ante level. In the end, the "facilitating" conditions (of a higher ex-ante level) for users with weaker pro-environmental attitudes seemed to compensate the higher efforts of the "green" users. At this point, one might argue that these observations only hold true for the sample recruited, that only individuals above a certain minimum level of pro-environmental attitudes had opted into this study in the first place. However, the sample recruited for this study scored similar - even slightly weaker - on pro-environmental attitudes than the nationally representative sample in the most recent Swiss Environment Survey. To the extent that self-selection mainly occurs on environmental attitudes - and typically, this dimension is the main concern that is brought forward regarding self-selection - the findings can be extrapolated to the Swiss population. One might further argue that self-selection bias might still be present, as the recruitment strategy might favor participation of a certain subgroup of individuals along other dimensions. In particular the other three variables moderating the treatment effect are of interest here (baseline usage, tendency to monitor progress towards goals, and conscientiousness). However, as outlined in section 3.4.2.8, individual's tendency to monitor progress towards goals seems to be similarly prevalent among the general population as among the study sample; with respect to to conscientiousness and baseline use, one could even rather expect a negative than a positive self-selection bias with respect.

The finding that the device can be equally effective for a broad audience can have far-reaching consequences: In line with the findings of the previous paragraph, this implies that this kind of interventions can be much more encompassing than previously thought. This is crucial for cost-benefit projections when scaling up the deployment of feedback interventions to a larger number of households.

Moreover, these insights can also be vital for other innovative technologies and companies to overcome existing barriers to market entry: Although a number of other real-time feedback applications have yielded encouraging results in small-scale pilot studies, their superior performance compared to traditional feedback instruments is generally attributed to self-selection biases. This external validity issue creates a chicken-and-egg problem: As the scalability of the conservation effect is questioned, it is difficult to identify customers who are willing to deploy such devices on a larger scale - at least large enough for an initial investment in tooling and manufacturing. But without a deployment on a larger scale, it is impossible to demonstrate the scalability of initial pilot studies. The result that the device works equally well on a broad audience helps break this circle. These insights are not only crucial for the particular product used in this trial, they may also help other companies to overcome credibility issues regarding the scalability of their pilot studies: The study in chapter 3 serves as the first example worldwide that real-time feedback on the resource consumption of a specific behavior can yield a substantial impact and work equally well on a broad audience. It shows that the intervention can be scaled up cost-effectively (at least in Switzerland). That way, the program can make a valuable contribution to corporate, regional and federal resource conservation and carbon abatement goals.

Both feedback interventions are examples of non-manipulative behavioral nudges: They help many individuals act in line with their preferred outcomes, yet are freedom-preserving. Forth, both studies (chapter 3 and 4) present feedback interventions that are examples of behavioral nudges and of a policy approach coined as *"libertarian paternalism"* (Sunstein and Thaler (2003); Thaler and Sunstein (2003)). While a few fervent environmentalists or particularly conscientious people may achieve to generally act on their well-informed intentions, the vast majority of people fails to align their behavior with their preferred outcomes on a regular basis as a result of bounded rationality and cognitive biases. Following through on environmental goals can be particularly challenging (Gutsell et al. (2012)). Due to cognitive biases, heuristics, and fallacies, individuals often behave in ways that are not in line with their well-informed intentions and preferences (Hansen and Jespersen (2013)). Behavioral nudges have been identified as effective strategies to overcome these biases in order to maximize both individuals' and society's welfare. Nudges are small changes in the decision making context that alter *"people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives"* (Thaler and Sunstein (2009)). Both the water consumption feedback flyer with the social comparison element used in the study in chapter 4 and the shower meter described in chapter 3 constitute nudges, more precisely transparent nudges that elicit attention and reflective thinking about resource consumption. As such, they don't qualify as manipulative - a frequent criticism which applies to some other forms of nudges (Hansen and Jespersen (2013)). Behavior-specific real-time feedback, as provided by the shower meters, makes behavioral outcomes salient while the behavior is taking place. That way, these systems draw attention to aspects that are relevant for people's preferences, which they might otherwise not consider in the decision-making context. This attempt to "*influence choices in a way that will make choosers better off, as judged by themselves*" (Thaler and Sunstein (2009), p. 5) preserves nevertheless individuals' freedom: Both in the case of the shower meter and of the feedback flyers used in chapter 4, individuals are free not to act on the feedback if they do not agree with the ends or means of the intervention (Hansen and Jespersen (2013)). Moreover, at least in the case of the shower meters, the conservation effect is not based on a negative mechanism like social pressure. As a consequence, the feedback intervention increases the utility of many individuals, in particular for the less conscientious (inattentive) and for those with a tendency to monitor progress towards goals, while making nobody worse off. This makes the intervention a particularly desirable policy instrument from a normative point of view.

Behavior-specifical real-time feedback can improve knowledge about resource consumption and correct public misconceptions about the environmental contribution of daily behaviors. Fifth, the results of chapter 3 corroborate previous research that shows that people are generally poorly informed about their resource consumption (Attari et al. (2010); Attari (2014)); at the same time, this thesis also indicates a pathway to correct fundamental public misconceptions about the environmental contribution of daily behaviors. This is particularly true for their perceptions of comparative energy uses of different behaviors: While people tend to slightly overestimate the contribution of low-energy activities, they fundamentally underestimate the energy consumption of high-energy activities (Attari et al. (2010)).

In contrast to previous studies using deferred feedback on an aggregated level (Mitchell et al. (2013)), the study in chapter 3 indicates that behavior-specific real-time feedback can substantially improve individuals' understanding of their resource consumption and correct their misconceptions about the environmental contribution of different actions. This allows people to focus their efforts on high-impact behaviors that actually make a difference and to act in line with their good intentions to protect the environment.

Behavioral interventions can entail measurable negative side effects (moral licensing), which can more than offset the benefits they generate. Sixth, chapter 4 investigated side effects of a behavioral intervention in the field. So far, it is still unclear whether such programs trigger cross-domain adoption of additional socially approved behaviors (positive spillover, built on individuals' desire for consistency in their actions) or reduced engagement elsewhere (moral licensing - also referred to as negative spillover effects, moral cleansing, guilt reduction, or discussed in the context of boomerang effects and warm glow effects). While various laboratory studies have provided broad evidence for moral licensing for different behavioral domains, including purchasing decisions, nutrition, racism, and sexism, the topic is still a blind spot in the design and evaluation of behavioral interventions, at least in the domain of residential resource consumption. The study presented in chapter 4 is probably the first quantitative field study that investigated cross-domain effects in the domain of residential resource consumption with actual utility data. It measured the side effects of a water conservation campaign on residents' electricity consumption. While individuals exposed to a water conservation campaign did reduce their water consumption as expected, they also increased their electricity consumption relative to the control group. This provides new evidence that the side effects of behavioral interventions can be measurable and that negative side effects (moral licensing) seem to prevail over potential positive repercussions on subsequent actions. The study further shows that these side effects can more than offset the benefits on the primarily targeted behavior, at least in the short-term: Both with respect to on-site usage and even more in source-energy balance, the campaign had a clearly net negative energy outcome despite its success with respect to water conservation. Regarding on-site energy balance, the energy saved by reduced (hot) water consumption was offset by the increased electricity consumption by nearly a factor of two. From a source energy perspective, the net energy balance is even more negative, by a ratio of about 1:6. Thus, from an energy consumption point of view, the side effects did not only reduce or cancel out the benefits of the campaign, but the overall campaign produced a clear negative impact.

These results show the importance of adopting a more comprehensive perspective in the assessment of program costs and benefits. In particular, campaigns that root in the idea that "every little bits helps" should be evaluated carefully: They may not only waste individuals' time and efforts and deflect their attention away from more environmentally significant behaviors: Given public misconceptions of comparative resource consumption and people's tendency to choose easier alternatives, such campaigns can also give people the impression that they are already "doing their part". As a result, these low-direct-impact programs may even crowd out actions that would result in higher energy savings or CO_2-abatement, or license negative behaviors that people might otherwise abstain from. As the study presented in chapter 4 shows, if the program does not focus on the environmentally most significant behaviors, or if the intervention used is not very powerful, the negative impact may even exceed the primary benefits. This may not only be true regarding magnitude, but could potentially also be the case regarding persistence over time. This implies that a substantial amount of efforts, time, and money may currently be spent on interventions that actually have very small or even negative net outcomes.

The findings entail important implications for policymakers: First, these results advocate the adoption of a more comprehensive system-level perspective in the evaluation of environmental programs. Second, behavioral interventions should a) focus on environmentally significant behaviors and b) apply powerful tools. That way, individuals' efforts and attention are directed to the domains that really matter - and the positive margin from which adverse effects are discounted from is larger. If pro-environmental behavior in high-impact domains then licenses more wasteful behavior for environmentally less significant actions, the net balance of the program is still clearly positive. Third, the study in chapter 4 also points out the importance of correcting fundamental public misconceptions about the environmental contribution of daily behaviors. Real-time feedback on a specific behavior can achieve this, as the study in chapter 3 demonstrated and as discussed in the previous paragraph.

The study in chapter 4 demonstrates the importance of taking side effects into account in the evaluation of environmental programs. Thus, it responds to the call made by Paul C. Stern who points out the contradictory predictions for subsequent actions made by behavioral scientists as key questions for future psychological research: *"Which of these mechanisms predominates with high-impact behaviors, and under what conditions, are fundamental research questions of obvious importance to limiting climate change"* (Stern (2011)). Moreover, the question of behavioral spillover is not only relevant in the domain of resource consumption: There is also an ongoing debate in consumer research and policy, marketing, and social psychology on how actions that are perceived as socially desirable (like pro-environmental behavior) influence subsequent actions in a positive or in a negative way. Yet most of the research so far has been carried out in the lab. The study thus also contributes to research on other behaviors in other disciplines.

The results provide quantitative evidence for a growing intensity of resource use of daily behaviors due to changing norms and conventions. Seventh, the previous section highlighted the importance of extending the lens in the evaluation of behavioral programs. The adoption of a more comprehensive view is also important for energy and water demand projections in the light of changing norms and conventions, as the study in chapter 3 pointed out. The data reveal a strong correlation of age and per-shower resource consumption before the onset of the intervention: The results show that younger people (20-29 years) use 2.3

times as much energy and water per shower as persons over 64 (with a continuous trend for the age cohorts in between). The resource intensity *per shower thus increased to the 2.3-fold within a single generation*. While age effects (i.e., individuals taking shorter showers as they get older) cannot be ruled out entirely and with absolute certainty (cf. discussion in section 3.4.2.9), the increased consumption per shower is probably due to cohort effects (i.e., individuals who are born more recently tend to take longer showers than older generations). This is probably the first study with a larger number of households that points out this trend in the resource consumption per shower. In combination with a higher observed shower frequency, this amounts to the 2.7-fold resource intensity for showering.

The implication of these findings extend far beyond the domain of showering: They expand on a body of literature that describes a socio-technical transformation of conventions and behaviors: A substantial change of norms and conventions in particular for perceptions of comfort, cleanliness, and convenience, leading to increasingly resource-intensive consumption patterns (Shove (2004, 2003)). Several areas in that domain experienced similar growth rates over the past decades: laundry quantities, shower frequency, the use of space heating, and air conditioning. Yet in contrast to the existing literature, which mainly focused on qualitative aspects, the study in chapter 3 provided quantitative evidence for the phenomenon. The behaviors described affect the most energy- and water-intense areas of residential resource consumption, in particular space and water heating. Changing lifestyle conventions, norms, and paradigms almost certainly also affect other sectors like mobility: Younger people in Switzerland travel three times the distance on a daily basis that elder people do; yet in that case, cohort effects are much more difficult to disentangle from age effects (e.g., no more commute to work after retirement), financial factors, and rebound effects. In the case of showering (or at least the resource consumption per shower), these alternative explanations can be excluded to the largest extent.

Although the creeping transformations of norms and conventions lead to dramatic growth of resource consumption from a mid-to long-term perspective, they often go unnoticed for their lack of sudden, disruptive shifts. As a result, they are practically absent in policy analysis and in the public debate. In particular, these aspects are ignored in energy and water demand projections. Those scenarios typically account for technological progress and substitution. As of recently, they increasingly include rebound effects, which have been found capable of eroding substantial parts of the savings achieved by technological progress, or of entirely negating them. Yet other than those financially and technologically-driven changes, demand projections assume behavior to be stable. Yet as the results of this study show, it may be the case that profound changes of lifestyle, norms, and paradigms also substantially undermine the benefits of technological progress. Given the magnitude of these changes within a few decades, it would be wise to further investigate these transformations and to consider them in energy and water demand projections.

The results are also interesting from another angle: Younger people are generally considered as more aware of and concerned about environmental issues. Among other indicators, this is reflected in attitudinal polls, in their relatively strong support for green parties, and in souring numbers of enrollment in environmental studies. Nevertheless, the results indicate that the so-called "Generation Green" does not live up to their standards in their daily actions, leading a much more resource-intensive lifestyle than older generations. At the same time, the results suggest that given their high baseline consumption and their relatively strong pro-environmental attitudes, younger people are also more responsive to feedback. As a consequence, feedback interventions might help to address these issues and to close or at least reduce the generation gap.

5.3 Limitations of this Thesis and Research Outlook

As discussed in sections 3.4.4 and 4.4.3, the two studies at the core of this thesis have several limitations. Some of these limitations open new paths for further research, identified through insights generated in the context of this thesis. This section discusses the limitations, followed by the perspectives perspectives for future research.

5.3.1 Limitations

Despite all efforts made to ensure the internal and external validity of the studies conducted, the results are subject to a number of limitations. These can be classified into three main categories: limitations due to the characteristics of field studies, the question of effect persistence, and questions on generalizability.

First of all, while conducting studies in real-world settings has many benefits, in particular with respect to external validity (see e.g., Levitt and List (2007)), this environment entails a number of issues that make this kind of research more challenging than research under clean and controlled laboratory conditions. In general, *"choosing between lab experiments and field data usually requires a tradeoff between the pursuit of internal and external validity"* (Roe and Just (2009); also see Jimenez-Buedo and Miller (2010)). In field studies, the data collected are subject to a series of exogenous influences, unobserved factors and erratic patterns that add noise to the dataset and are potential sources of biases. In both studies, this concerns questions of absence, visitors, seasonal influences, or changes in the household infrastructure. Furthermore, it is typically more challenging to collect survey data from participants in a naturalistic setting (even more so while keeping the level of interference and intrusion at a minimum). Moreover, field studies tend to require more preparation upfront, making it extremely difficult or costly to repeat trials or to carry out additional experiments to cover missing aspects ex post. Moreover, access to data can be challenging: Identifying a building complex with two submetered utilities was a clear challenge in the study on the side effects; in the study with the shower meters as well, smart metering data were only available from a small subset of study participants, a sample too small to investigate cross-licensing effects. In the study on side effects, despite the measurement data and the strong evidence from prior studies in lab conditions, a lack of survey data made it difficult to pin down the effect to a specific behavior and mechanism.

This was a lesson from which the study in chapter 3 (conducted later in time than the study in chapter 4) clearly profited: The later of the two studies collected very detailed data on attitudes, perceived efforts, specific behaviors, changes in household structure, etc. Despite the extensive sets of questions asked in the study, the dataset still has one shortcoming worth pointing out: The answers of the survey respondent were extrapolated to the second household member (if applicable). As discussed in section 3.4.4.2, however, this should be a minor concern given homogamy (high concordance on attitudes among couples); moreover, this would rather add noise to the dataset and weaken correlations than create spurious ones. And finally, one should not forget that compared to the existing literature, this dataset is among the best that has ever been combined with granular utility consumption data. As a point of reference, Allcott and Rogers (2014) had access mainly to some basic demographics and overall census-track statistics (e.g., income median in the census tract) and described that dataset as "exceptionally good household-level data". One should keep this in mind to realize that the dataset at hand is clearly not a source of limitation, but one of the strengths of the study.

The second major limitation of both studies is that they collected data only for two resp. three months, thus on short-to mid-term effects, not long-term data. In the case of the study on side effects, for instance, it is unclear whether side effects are more short-lived than primary effects of interventions. In the study with the shower meters, the effect persistence is crucial for projections on savings and cost-effectiveness. Previous research indicates that for data push feedback systems, as employed in both studies, effects tend to be more persistent than feedback systems that require user interaction to access feedback information (e.g., logging into a portal). The topic is currently being explored by a fellow Ph.D. student at ETH Zurich (see section 5.3.2).

Finally, the question of external validity or generalizability is an important issue in several respects. First of all, the potential for selection bias is only a minor concern for both studies: The study in chapter 4 followed an opt-out recruitment strategy, which by design is quite robust to this issue. Turning to the study in chapter 3, in general, the by far biggest source of concern regarding selection bias for opt-in studies regards environmental attitudes: The conjecture is that participants who opt into these kinds of studies have stronger pro-environmental attitudes than the general population and therefore exhibit a treatment response which may not be representative. However, as discussed in section 3.4.2.8, the sample recruited scores comparably - or even slightly lower - on this dimension than the national representative sample of the most recent Swiss Environment Survey. As a result, the sample recruited can be considered as representative for Switzerland, at least with respect to the major concern for selection bias. Beyond selection bias, the question arises to what extent the findings are specific to the behaviors, feedback systems, season, or participants, and to what extent they can be extrapolated to other settings. In the study on side effects, for instance, the effects might have been particularly pronounced due to the very hot weather. Thus the effect might be more difficult to observe in less extreme weather conditions. Regarding the study in chapter 3, it would be interesting to find out to what extent the large savings are specific to the showering context, and to what extent real-time feedback can make a difference on other behaviors. The characteristics of showering (see table 3.1) make a particularly good candidate for real-time feedback. To what extent similar large treatment effects can be replicated on other behaviors is another interesting path for future research. Moreover, there is the question on the specific format of the feedback: To what extent can the findings be extrapolated to social normative feedback, to deferred instead of real-time feedback, or to setting where a certain goal is communicated? How do these parameters affect the impact of the intervention, the underlying psychological mechanisms, and the persistence of the effect? These are all questions that are clearly of interest for future research projects.

Overall, despite the efforts made to ensure internal and external validity, the aspects outlined here should be taken into account when extrapolating the findings to other settings, longer durations, or other users. At the same time, these points present interesting opportunities for future research: Based on these insights of these studies, certain aspects have been identified as particularly promising avenues of research to explore in follow-up projects. The following section 5.3.2 will provide more perspectives on this.

5.3.2 Outlook

As outlined in the previous section 5.2, both studies described in this dissertation have not only contributed novel insights to theory, but the findings are very relevant from a practical point of view for policymakers, companies in the utility sector, technology startups, and many other stakeholders. Several insights generated by the two studies described in this thesis have already started to create a first real-world impact and provided a valuable base for future projects.

This is particularly the case for the study with the shower meters described in chapter 3. First of all, the results of the study have already served the startup company Amphiro AG as a useful proof for the real-world impact of their product: In the dialog with prospective partners and customers, the study serves as a quantitative proof of the cost-effectiveness of the devices and of their sizable impact household's resource consumption. The collaboration with the utility company ewz in the course of this project has also paved the way for similar corporate social responsibility programs, be it in the form of employee engagement campaigns or as a service to their customers. Building on the promising results of the study described in chapter 3, two major follow-up projects in Switzerland and Singapore are currently in an advanced stage of planning and proposal submission. The large effect size makes the devices not only an effective means to realize meaningful energy and water savings, but also qualifies them as an excellent tool to empirically investigate related or broader research questions which would otherwise require a much larger sample size. In combination with detailed survey data, the granular measurement data collected make the devices well-suited instruments to investigate other hypotheses. Some of those hypotheses are also a direct outcome out of the study described in chapter 3.

For instance, the study showed that real-time feedback on a specific behavior can successfully prompt substantial behavior change and achieve this through positive mechanisms. The study further revealed that large conservation effects are possible without social normative feedback. The majority of literature on feedback interventions, however, uses social normative feedback, in most of the cases provided for aggregated data and with substantial time lag (see chapter 2). As a result, several questions impose themselves: What is the value of real-time feedback in comparison to deferred feedback, and how does it compare with the impact of social normative feedback when both are applied to the same setting of showering? Could the combination of real-time feedback on a specific behavior and social normative feedback raise the treatment effect even further? Does social normative feedback also operate through positive channels? Does it rather enhance positive mechanism by acknowledging socially desirable behavior, or does psychological pressure, maybe the fear of losing social approval, coerce people into conservation behaviors?

Another recurring motive in the study in chapter 3 was the prominent role of goals. This raises the question whether the effect could be enhanced by providing ambitious, yet realistic goals, instead of hoping that people will end up formulating a goal on their own (as half of the study participants did). Self-efficacy and the definition of an effective goal are interesting aspects in this context. Again, the question arises if this operates through psychological pressure and how more goal-oriented strategies affect the long-term persistence of the savings effects.

The question of effect persistence is in general a pressing issue for future research. Vojkan Tasic, a fellow Ph.D. student at the Bits to Energy Lab, is currently investigating this question with two datasets that cover 14 months each. The results will answer whether the conservation effects induced by the shower meters is rather short-lived or whether it results in long-lasting sustainable behavioral practices.

In the context of behavioral spillover and moral licensing, the study presented in chapter 4 has brought up the question to what extent savings achieved by campaigns need to be discounted for by negative behavioral spillover to other domains. Exploring this would also be of interest in future research projects with the shower meter, by investigating whether resource conservation in the shower licenses more wasteful behaviors elsewhere. Or, as other researchers have already done it in the very context of showering (and in other domains): They had people remember their pro- environmental attitudes, then asked them to call into their memory instances of them not living up to these preferences; as a result, they tried to reduce this cognitive dissonance by making up for it in subsequent environmentally significant behaviors, by taking shorter showers (Stone and Fernandez (2008)).

Finally, the ongoing technology development of the shower meters towards wireless two-way communication capabilities will open the door to a whole set of new possibilities for user engagement and research projects and for the integration with other systems and services. This could allow for much more tailored feedback and inspire various ideas in the domain of interactive user interaction or gamification, for instance.

To summarize, the research presented in this thesis contributes to laying the groundwork for diverse future projects and developments, both for follow-up research studies and for large-scale carbon abatement or resource conservation programs. On a practical side, the insights can facilitate the implementation of environmentally significant large-scale field deployments (while paying attention to minimizing negative primary or side effects). On the research and development side, the insights of both studies provide a good starting point for promising follow-up studies. These in turn can guide and support future product developments. Moreover, the shower devices have proven themselves as an excellent research platform to explore additional questions and theories.

5.4 Conclusion

This thesis concludes by picking up the quote which already served as its epigraph:

"Nuanced research into human behavior and energy-use decisions is not new, nor is the idea that energy efficiency may be generally cost-effective. 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." - Hunt Allcott & Sendhil Mullainathan, Science Magazine (327), 2010

The research efforts made in the context of this thesis sought to realize this concerted "engineering" work in the two studies carried out as part of this endeavor. Both studies were deliberately not designed in an isolated ivory tower of academic research and carried out in a clean laboratory setting: The conception, design, implementation, and analysis of these field trials actively involved a number of stakeholders from different continents, ranging from researchers from several universities, utility company employees, founders and employees of a technology start-up, members of the Swiss Federal Office of Energy, researchers at a non-profit research organization, consultants for community-based sustainable development, employees of a property management company, realtors, and, last but not least, approx. 900 participating households. As a result of these concerted efforts, this thesis has generated a series of novel insights into behavioral interventions that are highly relevant for theory and practice alike.

While energy efficiency has repeatedly been described as the low hanging fruit and as the cheapest "fuel" available, picking that fruit has turned out not to be not as straightforward as many might have hoped: Human decision-making is subject to bounded rationality and cognitive biases, not to mention that humans have different preferences, personalities, and living environments. As a result, the "engineering" work in this context cannot rely on a set of natural laws and equations to build a model that provides a universal and accurate description of inputs and outcomes. Yet insights from behavioral economics, social psychology, information systems research, and other disciplines are valuable sources to develop strategies that actually make a difference. In combination with the broadening ubiquity sensors, smart phones, communication networks, and other technologies, it is increasingly possible to apply these insights in large-scale deployments in order to harvest a growing portion of the (maybe not so) low hanging fruit of efficient resource usage.

The goal of this thesis was to develop and to investigate some of these strategies, to understand underlying psychological mechanisms, and to explore potential side effects. The results show that resource consumption feedback can be more powerful, more positive and more mass-compatible than previously assumed; at the same time, the results of both studies emphasize the importance of adopting a comprehensive view in the evaluation of studies, behavioral programs, and resource demand projections. The thesis is probably the first work which demonstrated that real-time feedback on the resource consumption of a specific behavior can both yield a large impact and work equally well for a broad audience. This implies that such interventions - when implemented properly - can be very cost-effective and scalable. In this context, the findings show that feedback interventions can operate through positive mechanisms, helping people to act in line with their preferences, and not through psychological pressure. The results make these instruments much more amenable to widespread adoption. However, the results of this thesis also warrant caution regarding perverse side effects and regarding changing norms and conventions, both of which could potentially undermine the benefits generated by technological progress and behavioral instruments.

Overall, the insights show that the exploration of behavioral interventions holds great potential to prompt behavior change and to promote resource conservation quickly and at scale in order to effectively reduce the negative externalities associated with the consumption of resources.

Chapter 6

Bibliography

Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, 25(3):273–291.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2):179–211.

Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32(4):665–683.

Ajzen, I. and Fishbein, M. (1969). The prediction of behavioral intentions in a choice situation. *Journal of Experimental Social Psychology*, 5(4):400–416.

Akerlof, G. A. and Kranton, R. E. (2010). *Identity economics: how our identities shape our work, wages, and well-being*. Princeton University Press, Princeton.

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. (2011a). Rethinking real-time electricity pricing. *Resource and Energy Economics*, 33(4):820–842.

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

Allcott, H. and Greenstone, M. (2012). Is there an energy efficiency gap? Technical report, National Bureau of Economic Research.

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.

Allen, D. and Janda, K. (2006). The effects of household characteristics and energy use consciousness on the effectiveness of real-time energy use feedback: a pilot study. In *Proceedings of the ACEEE summer study on energy efficiency in buildings*, page 7–1.

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? Image motivation and monetary incentives in behaving prosocially. *The American Economic Review*, 99(1):544–555.

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>.

Ashford, S. J. (1986). Feedback-seeking in individual adaptation: a resource perspective. *Academy of Management Journal*, 29(3):465–487.

Attari, S. Z. (2011). I'll do the easy thing, you do the hard thing. Behavior, Energy and Climate Change Conference (2011). Washington, DC, USA. <http://cred.columbia.edu/research/all-projects/saving-energy-ill-do-the-easy-thing-you-do- the-hard-thing/>.

Attari, S. Z. (2014). Perceptions of water use. *Proceedings of the National Academy of Sciences*, 111(14):5129–5134.

Attari, S. Z., DeKay, M. L., Davidson, C. I., and Bruine de Bruin, W. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences*, 107(37):16054–16059.

Attari, S. Z., DeKay, M. L., Davidson, C. I., and de Bruin, W. B. (2011). Changing household behaviors to curb climate change: how hard can it be? *Sustainability: The Journal of Record*, 4(1):9–11.

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.

Baca-Motes, K., Brown, A., Gneezy, A., Keenan, E. A., and Nelson, L. D. (2013). Commitment and behavior change: evidence from the field. *Journal of Consumer Research*, 39(5):1070–1084.

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>.

Bamberg, S. (2003). How does environmental concern influence specific environmentally related behaviors? A new answer to an old question. *Journal of environmental psychology*, 23(1):21–32.

Bamberg, S. and Moeser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: a new meta-analysis of psycho-social determinants of pro-environmental behaviour. *Journal of Environmental Psychology*, 27(1):14–25.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2):191.

Bandura, A. (1997). *Self-Efficacy: the Exercise of Control.* Worth Publishers, New York, 1st edition edition.

Bandura, A., Adams, N. E., Hardy, A. B., and Howells, G. N. (1980). Tests of the generality of self-efficacy theory. *Cognitive Therapy and Research*, 4(1):39–66.

Barroso, J. M. D. (2008). Europe's climate change opportunity. Speech to the European Parliament (2008/01/23). Brussels. <http://europa.eu/rapid/press-release_SPEECH-08-34_en.htm>.

Baumeister, R. F. (2002). Yielding to temptation: self-control failure, impulsive purchasing, and consumer behavior. *Journal of Consumer Research*, 28(4):670–676.

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.

Beckel, C., Sadamori, L., and Santini, S. (2013). Automatic socio-economic classification of households using electricity consumption data. In *Proceedings of the fourth international conference on Future energy systems*, page 75–86. ACM.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Bento, A. M., Ho, B., Poe, G. L., and Taber, J. T. (2010). Culpability and willingness to pay to reduce negative externalities: a contingent valuation and experimental economics study. *Selected Paper prepared for presentation at the Agricultural & Applied Economics Association.*

BergInsight (2012). Smart Metering in Europe – 9th edition. <www.berginsight.com/ReportPDF/ProductSheet/bi-sm10-ps.pdf>.

Bicchieri, C. and Muldoon, R. (2014). Social norms. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, spring 2014 edition.

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

Boccaletti, G., Löffler, M., and Oppenheim, J. M. (2008). How IT can cut carbon emissions. *The McKinsey Quarterly*. <http://storage.globalcitizen.net/data/topic/knowledge/uploads/20111102956347O5.pdf>.

Bollinger, B. and Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science*, 31(6):900–912.

Bovens, L. (2009). The ethics of nudge. In Grüne-Yanoff, T. and Hansson, S. O., editors, *Preference Change: Approaches From Philosophy, Economics and Psychology*, pages 207–219. Springer.

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>.

Bradley-Geist, J. C., King, E. B., Skorinko, J., Hebl, M. R., and McKenna, C. (2010). Moral credentialing by association: the importance of choice and relationship closeness. *Personality & social psychology bulletin*, 36(11):1564–1575.

Brandon, G. and Lewis, A. (1999). Reducing household energy consumption: a qualitative and quantative field study. *Journal of Environmental Psychology*, 19(1):75–85.

Breslow, N. E. and Day, N. E. (1975). Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases*, 28(5–6):289–303.

Cain, D. M., Loewenstein, G., and Moore, D. A. (2005a). Coming clean but playing dirtier: the shortcomings of disclosure as a solution to conflicts of interest. In *Conflicts of Interest*. Cambridge University Press.

Cain, D. M., Loewenstein, G., and Moore, D. A. (2005b). The dirt on coming clean: perverse effects of disclosing conflicts of interest. *The Journal of Legal Studies*, 34(1):1–25.

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.

Chiou, W.-B., Wan, C.-S., Wu, W.-H., and Lee, K.-T. (2011). A randomized experiment to examine unintended consequences of dietary supplement use among daily smokers: taking supplements reduces self-regulation of smoking. *Addiction*, 106(12):2221–2228.

Choi Granade, H., Creyts, J., Derkach, A., Farese, P., Nyquist, S., and Ostrowski, K. (2009). Unlocking energy efficiency in the U.S. economy. Technical report, McKinsey Global Energy and Materials.

Chong, A., Karlan, D., Shapiro, J., and Zinman, J. (2013). (Ineffective) messages to encourage recycling: evidence from a randomized evaluation in Peru. Technical report.

Cialdini, R. B. (2009). *Influence: science and practice*. Pearson Education, Boston.

Cisco (2008). The sustainable business practice study. Technical report. <http://www.cisco.com/ cisco/web/UK/Cisco-SBPS.pdf>.

Clot, S., Grolleau, G., and Ibanez, L. (2011). Do good deeds make bad people? Document de Recherche 2011-21, Laboratoire Montpelliérain d'Economie Théorique et Appliquée, Montpellier. http://www.lameta.univ-montp1.fr/Documents/DR2011-21.pdf.

Conway, P. and Peetz, J. (2012). When does feeling moral actually make you a better person? Conceptual abstraction moderates whether past moral deeds motivate consistency or compensatory behavior. *Personality and Social Psychology Bulletin*, 38(7):907–919.

Cornelissen, G., Pandelaere, M., Warlop, L., and Dewitte, S. (2008). Positive cueing: promoting sustainable consumer behavior by cueing common environmental behaviors as environmental. *International Journal of Research in Marketing*, 25(1):46–55.

Crompton, T. and Kasser, T. (2009). Meeting environmental challenges: the role of human identity. Technical report, WWF, London, UK.

Cundiff, L. V., Willham, R. L., and Pratt, C. A. (1966). Additive versus multiplicative correction factors for weaning weight in beef cattle. *Journal of Animal Science*, 25(4):983–987.

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. (2010). Smart metering - what potential for householder engagement? *Building Research & Information*, 38(5):442–457.

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.

De Young, R. (1986). Encouraging environmentally appropriate behavior: the role of intrinsic motivation. *J. Environmental Systems*, 15(4).

Dearen, J. and Williams, J. (2014). California drought prompts unprecedented water conservation efforts. Huffington Post. 2014-01-31. http://www.huffingtonpost.com/2014/01/31/california-drought-reserv_n_4706344.html.

DEFRA (2008). A framework for proenvironmental behaviours. Technical report. <http://www.defra. gov.uk/publications/files/pb13574-behaviours-report-080110.pdf>.

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.

DellaVigna, S. (2009). Psychology and economics: evidence from the field. *Journal of Economic Literature*, 47(2):315–372.

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

Dickerson, C. A., Thibodeau, R., Aronson, E., and Miller, D. (1992). Using cognitive dissonance to encourage water conservation. *Journal of Applied Social Psychology*, 22(11):841–854.

Dickson, J. and Albaum, G. (1977). A method for developing tailormade semantic differentials for specific marketing content areas. *Journal of Marketing Research*, 14(1):87.

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.

Dobson, J. K. and Griffin, J. A. (1992). Conservation effect of immediate electricity cost. feedback on residential consumption behavior. *Proceedings of the 7th ACEEE summer study on energy efficiency in buildings*, 2.

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

Eccles, M. P., Hrisos, S., Francis, J., Kaner, E. F., Dickinson, H. O., Beyer, F., and Johnston, M. (2006). Do self- reported intentions predict clinicians' behaviour: a systematic review. *Implementation Science*, 1(1):28.

Effron, D. A., Cameron, J. S., and Monin, B. (2009). Endorsing Obama licenses favoring Whites. *Journal of Experimental Social Psychology*, 45(3):590–593.

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

eia (2013b). What are the energy-related carbon dioxide (CO2) emissions by source and sector for the United States? <http://www.eia.gov/tools/faqs/faq.cfm?id=75&t=11>.

eia (2014). Monthly energy review February 2014 of the U.S. Energy Information Administration. <http://www.eia.gov/totalenergy/data/monthly/pdf/sec2_5.pdf>.

Ek, K. and Söderholm, P. (2010). The devil is in the details: household electricity saving behavior and the role of information. *Energy Policy*, 38(3):1578–1587.

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>.

EPA (2011a). ENERGY STAR performance ratings methodology for incorporating source energy use. <http://www.energystar.gov/ia/business/evaluate_performance/site_source.pdf>.

EPA (2011b). Pick 5 for the environment. <http://www.epa.gov/pick5/>.

Epley, N. and Dunning, D. (2000). Feeling "holier than thou": are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology*, 79(6):861–875.

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.

European Environment Agency (2011). % change in household final energy consumption per person, 1990-2008. Technical report.

European Environment Agency (2012). Households and industry responsible for half of EU greenhouse gas emissions from fossil fuels.

European Environment Agency (2013). Progress on energy efficiency in Europe (ENER 037) - assessment published Mar 2013. Technical report, European Environment Agency.

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.

FDHA (2013). Mobility and transport - pocket statistics 2013. Technical Report 839-1300-05, Swiss Federal Department of Home Affairs FDHA, Federal Statistical Office FSO, Neuchatel.

Ferraro, P. J., Miranda, J. J., and Price, M. K. (2011). The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *American Economic Review*, 101(3):318–322.

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.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford University Press, Stanford, Calif.

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/>.

Fishbach, A. and Finkelstein, S. R. (2012). How feedback influences persistence, disengagement, and change in goal pursuit. *Frontiers of social psychology: Goal-directed behavior*, page 203–230.

Fishbein, M. and Ajzen, I. (1975). *Belief, attitude, intention, and behavior: an introduction to theory and research.* Addison-Wesley Pub. Co., Reading, Mass.

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.

Froehlich, J., Findlater, L., Ostergren, M., Ramanathan, S., Peterson, J., Wragg, I., Larson, E., Fu, F., Bai, M., and Patel, S. (2012). The design and evaluation of prototype eco-feedback displays for fixture-level water usage data. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, page 2367–2376. 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/>.

Garcia, J. H. and Wei, J. (2013). On social sanctions and beliefs - a pollution norm example. *Environment for Development Discussion Paper Series.*

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.

Gatersleben, B., Steg, L., and Vlek, C. (2002). Measurement and determinants of environmentally significant consumer behavior. *Environment and Behavior*, 34(3):335–362.

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

Gfroerer, J. (1998). *Household survey on drug abuse: preliminary estimates, 1995*. DIANE Publishing.

Gilbert, B. and Zivin, J. S. G. (2013). Dynamic salience with intermittent billing: evidence from smart electricity meters. Working Paper 19510, National Bureau of Economic Research.

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

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4):191–210.

Gneezy, U. and Rustichini, A. (2000). A fine is a price. *J. Legal Stud.*, 29:1.

Goelz, S. (2011). Smart Energy und der Endkunde - Ergebnisse und Schlussfolgerungen aus der Smart Energy Forschung. Energietouren 2011 "Smart Metering – Smart Home" Intelligent mit Strom umgehen. 2011-11-29. Freiburg. <www.hwk-freiburg.de/web/adb/output/asset/19616>.

Goeschl, T. and Perino, G. (2009). On backstops and boomerangs: environmental R&D under technological uncertainty. *Energy Economics*, 31(5):800–809.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482.

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

Greening, L. A., Greene, D. L., and Difiglio, C. (2000). Energy efficiency and consumption — the rebound effect — a survey. *Energy Policy*, 28(6-7):389–401.

Gregory, G. D. and Leo, M. D. (2003). Repeated behavior and environmental psychology: the role of personal involvement and habit formation in explaining water consumption. *Journal of Applied Social Psychology*, 33(6):1261–1296.

Griskevicius, V., Tybur, J. M., and Van den Bergh, B. (2010). Going green to be seen: status, reputation, and conspicuous conservation. *Journal of Personality and Social Psychology*, 98(3):392–404.

Gutsell, J. N., Inzlicht, M., and Gutsell, J. (2012). A neuroaffective perspective on why people fail to live a sustainable lifestyle.

Haas, R., Auer, H., and Biermayr, P. (1998). The impact of consumer behavior on residential energy demand for space heating. *Energy and Buildings*, 27(2):195–205.

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

Harding, M. and Hsiaw, A. (2014). Goal setting and energy conservation.

Harding, M. and Rapson, D. (2013). Does absolution promote sin? The conservationist's dilemma. Technical report, Working paper. Available at< http://www.econ.ucdavis.edu/faculty/dsrapson/HR_Absolution.pdf.

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

Hauseigentuümerverband 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/>.

Heberlein, T. A. (2012). *Navigating Environmental Attitudes*. Oxford University Press.

Heise, D. R. (1969). Some methodological issues in semantic differential research. *Psychological Bulletin*, 72(6):406–422.

Heslop, L. A., Moran, L., and Cousineau, A. (1981). Consciousness in energy conservation behavior: an exploratory study. *Journal of Consumer Research*, 8(3).

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>.

Heywood, J. L. (2002). The cognitive and emotional components of behavior norms in outdoor recreation. *Leisure Sciences*, 24(3-4):271–281.

Hines, J. M., Hungerford, H. R., and Tomera, A. N. (1987). Analysis and synthesis of research on responsible environmental behavior: a meta-analysis. *The Journal of Environmental Education*, 18(2).

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.

IEA (2011). Energy statistics and balances of OECD countries. <http://www.iea.org/stats/index.asp>.

IEA (2012). *Energy Technology Perspectives 2012 - Pathways to a Clean Energy System*. ISBN 978-92-64-17488-7. <http://www.iea.org/w/bookshop/add.aspx?id=425>.

IEA (2013). Key world energy statistics 2013. Technical report, International Energy Agency. <http://www.iea.org/publications/freepublications/publication/name,31287,en.html>.

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.

intelliekon (2011). Intelliekon - achieving sustainable energy consumption with smart metering, communication and tariff systems. Technical report, Freiburg.

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/>.

Jacobsen, G. D., Kotchen, M. J., and Vandenbergh, M. P. (2012). The behavioral response to voluntary provision of an environmental public good: evidence from residential electricity demand. *European Economic Review*, 56(5):946–960.

Jaffe, A. B. and Stavins, R. N. (1994). The energy-efficiency gap - what does it mean? *Energy policy*, 22(10):804–810.

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

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

Jimenez-Buedo, M. and Miller, L. M. (2010). Why a trade-off? The relationship between the external and internal validity of experiments. *THEORIA. An International Journal for Theory, History and Foundations of Science*, 25(3):301–321.

Jochelson, K. (2007). Paying the patient. *London: Kings Fund*. <http://www.kingsfund.org.uk/sites/files/kf/field/field_document/paying-the-patient-kicking-bad-habits-supporting-paper-karen-jochelson.pdf>.

Jones, B. D. (1999). Bounded rationality. *Annual review of political science*, 2(1):297–321.

Jordan, J., Mullen, E., and Murnighan, J. K. (2011). Striving for the moral self: the effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*, 37(5):701–713.

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

Judge, T. A., Erez, A., Bono, J. E., and Thoresen, C. J. (2002). Are measures of self-esteem, neuroticism, locus of control, and generalized self-efficacy indicators of a common core construct? *Journal of Personality and Social Psychology*, 83(3):693–710.

Kahneman, D. (2013). *Thinking, fast and slow*. Farrar, Straus and Giroux, New York, reprint edition.

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.

Khan, U. and Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research*, 43(2):259–266.

Khan, U. and Dhar, R. (2007). Where there is a way, is there a will? The effect of future choices on self-control. *Journal of Experimental Psychology: General*, 136(2):277–288.

Kinzig, A. P., Ehrlich, P. R., Alston, L. J., Arrow, K., Barrett, S., Buchman, T. G., Daily, G. C., Levin, B., Levin, S., Oppenheimer, M., Ostrom, E., and Saari, D. (2013). Social norms and global environmental challenges: the complex interaction of behaviors, values, and policy. *BioScience*, 63(3):164–175.

Kivetz, R. and Simonson, I. (2002). Self-control for the righteous: toward a theory of precommitment to indulgence. *Journal of Consumer Research*, 29(2):199–217.

Kolstad, C. D. (2010). *Environmental Economics*. Oxford University Press, USA, New York, 2 edition edition.

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.

Kotchen, M. J. and Moore, M. R. (2007). Private provision of environmental public goods: household participation in green-electricity programs. *Journal of Environmental Economics and Management*, 53(1):1–16.

Krampf, R. F., Burns, D. J., and Rayman, D. M. (1993). Consumer decision making and the nature of the product: a comparison of husband and wife adoption process location. *Psychology and Marketing*, 10(2):95–109.

Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6):1121–1134.

Kruger, J. and Gilovich, T. (2004). Actions, intentions, and self-assessment: the road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin*, 30(3):328–339.

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.

Lawson, P. J. and Flocke, S. A. (2009). Teachable moments for health behavior change: a concept analysis. *Patient education and counseling*, 76(1):25–30.

Lee, K. and Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2):329–358.

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.

Lester, R. and Hart, D. (2012). *Unlocking energy innovation: how America can build a low-cost low-carbon energy system*. MIT Press.

Levitt, S. D. and List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174.

Loock, C.-M., Staake, T., and Thiesse, F. (2013). Motivating energy-efficient behavior with green IS: an investigation of goal setting and the role of defaults. *Management Information Systems Quarterly*.

Madlener, R. and Alcott, B. (2009). Energy rebound and economic growth: a review of the main issues and research needs. *Energy*, 34(3):370–376.

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.

Mazar, N. and Zhong, C. B. (2010). Do green products make us better people? *Psychological Science*, 21(4):494–498.

McCoy, D. and Lyons, S. (2014). Better information on residential energy use may deter investment in efficiency: case study of a smart metering trial.

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

McKenzie-Mohr, D. (2013). *Fostering sustainable behavior: an introduction to community-based social marketing*. New Society Publishers.

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.

Melville, N. (2010). Information systems innovation for environmental sustainability. *Management Information Systems Quarterly*, 34(1):1–21.

Merritt, A. C., Effron, D. A., Fein, S., Savitsky, K. K., Tuller, D. M., and Monin, B. (2012). The strategic pursuit of moral credentials. *Journal of Experimental Social Psychology*, 48(3):774–777.

Miller, D. T. and Effron, D. A. (2010). Psychological license. In *Advances in Experimental Social Psychology*, volume 43, pages 115–155. Elsevier.

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.

Monin, B. and Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of personality and social psychology*, 81(1):33.

Naderi, I. (2011). Green behavior: cncern for the self or others. In *AMA Summer Educators' Conference Proceedings*, volume 22, page 163.

National Council and Sustainable Development Commission (2006). I will if you will. <http://www.sd-commission.org.uk/publications/downloads/I%20WillIfYouWill.pdf>.

Neuman, K. (1986). Personal values and commitment to energy conservation. *Environment and Behavior*, 18(1):53–74.

Nolan, J. M., Schultz, P. W., Cialdini, R. B., Goldstein, N. J., and Griskevicius, V. (2008). Normative social influence is underdetected. *Personality and Social Psychology Bulletin*, 34(7):913–923.

Norton, E. (2012). Shame on the rich. *Science NOW*. 2012-02-27. <http://news.sciencemag.org/2012/02/shame-rich>.

Olmstead, S. M. and Stavins, R. N. (2009). Comparing price and nonprice approaches to urban water conservation. *Water Resources Research*, 45(4):W04301.

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

Pew Research Center for the People and the Press (2009). Section 9: The environment and the economy.

Poortinga, W., Steg, L., and Vlek, C. (2004). Values, environmental concern, and environmental behavior: a study into household energy use. *Environment & Behavior*, 36(1):70–93.

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 (2013a). Der Energieverbrauch der Privaten Haushalte 2000 - 2012. Ex-Post-Analyse nach Verwendungszwecken und Ursachen der Veränderungen. Technical report. <http://www.bfe.admin.ch/themen/00526/00541/00542/02721/ index.html?lang=de&dossier_id=02722>.

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

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

Rehfeldt, K. and Wallasch, A.-K. (2013). Kostensituation der Windenergie an Land in Deutschland. Technical Report SP13008A0, Deutsche WindGuard GmbH.

Roberts, B. W., Jackson, J. J., Fayard, J. V., Edmonds, G., and Meints, J. (2009). Chapter 25. Conscientiousness. In *Handbook of Individual Differences in Social Behavior*. The Guilford Press, New York, 1 edition edition.

Roe, B. E. and Just, D. R. (2009). Internal and external validity in economics research: tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics*, 91(5):1266–1271.

Ruijter, J. M., Thygesen, H. H., Schoneveld, O. J. L. M., Das, A. T., Berkhout, B., and Lamers, W. H. (2006). Factor correction as a tool to eliminate between-session variation in replicate experiments: application to molecular biology and retrovirology. *Retrovirology*, 3:2.

Ryan, R. M. and Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*, 25(1):54–67.

Sachdeva, S., Iliev, R., and Medin, D. L. (2009). Sinning saints and saintly sinners the paradox of moral self-regulation. *Psychological science*, 20(4):523–528.

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.

Samuelson, C. D. (1990). Energy conservation: a social dilemma approach. *Social Behaviour*, 5(4):207–230.

Schleich, J., Klobasa, M., Brunner, M., Gölz, S., and Götz, K. (2011). Smart metering in Germany and Austria: results of providing feedback information in a field trial. Technical report, Working paper sustainability and innovation.

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>.

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

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.

Schwartz, S. H. (1977). Normative influences on altruism. In *Advances in Experimental Social Psychology*, volume 10, pages 221—279.

Schwarzer, R. (2008). Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology*, 57(1):1–29.

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

SFOE (2013b). 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>.

Sheeran, P. (2002). Intention—behavior relations: a conceptual and empirical review. *European Review of Social Psychology*, 12(1):1–36.

Shiklomanov, I. A. (1999). International hydrological programme - UNESCO's inergovernmental scientific programme in water resources - dynamics of water use in the world by continents (km3/year). Technical report.

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

Shove, E. (2004). *Comfort, cleanliness and convenience: the social organization of normality*. Bloomsbury Academic, Oxford, England; New York.

Shove, E. (2010). Beyond the ABC: climate change policy and theories of social change. *Environment and Planning A*, 42(6):1273–1285.

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

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.

Steg, L. and Vlek, C. (2009). Encouraging pro-environmental behaviour: an integrative review and research agenda. *Journal of Environmental Psychology*, 29(3):309–317.

Stern, P. C. (2000). New environmental theories: toward a coherent theory of environmentally significant behavior. *Journal of social issues*, 56(3):407–424.

Stern, P. C. (2011). Contributions of psychology to limiting climate change. *American Psychologist*, 66(4):303–314.

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.

Stone, J. and Fernandez, N. C. (2008). To practice what we preach: the use of hypocrisy and cognitive dissonance to motivate behavior change. *Social and Personality Psychology Compass*, 2(2):1024–1051.

Strahilevitz, M. A. and Myers, J. (1998). Donations to charity as purchase incentives: how well they work may depend on what you are trying to sell. SSRN Scholarly Paper ID 1728172, Social Science Research Network, Rochester, NY.

Sunstein, C. R. and Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, page 1159–1202.

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

Thaler, R. H. and Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, page 175–179.

Thaler, R. H. and Sunstein, C. R. (2009). *Nudge: improving decisions about health, wealth, and happiness*. Penguin Books, New York.

Thøgersen, J. (1999a). The ethical consumer. moral norms and packaging choice. *Journal of Consumer Policy*, 22(4):439–460.

Thøgersen, J. (1999b). Spillover processes in the development of a sustainable consumption pattern. *Journal of Economic Psychology*, 20(1):53–81.

Thøgersen, J. and Crompton, T. (2009). Simple and painless? The limitations of spillover in environmental campaigning. *Journal of Consumer Policy*, 32(2):141–163.

Thøgersen, J. and Ölander, F. (2003). Spillover of environment-friendly consumer behaviour. *Journal of Environmental Psychology*, 23(3):225–236.

Tiefenbeck, V., Götte, L., Degen, K., Tasic, V., and Staake, T. (2014). Investigating the effectiveness of real-time feedback: the influence of demographics, attitudes, and personality traits. ewz-amphiro study final report to the swiss federal office of energy.

Tiefenbeck, V., Staake, T., Roth, K., and Sachs, O. (2013a). For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. *Energy Policy*, 57:160–171.

Tiefenbeck, V., Tasic, V., Schoeb, S., and Staake, T. (2013b). 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.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157):1124–1131.

Ueno, T., Inada, R., Saeki, O., and Tsuji, K. (2006). Effectiveness of an energy-consumption information system for residential buildings. *Applied Energy*, 83(8):868–883.

Ueno, T., Tsuji, K., and Nakano, Y. (2003). Effectiveness of displaying energy consumption data in residential buildings: to know is to change.

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

Unit, E. I. (2008). Managing the company's carbon footprint - the emerging role of ICT. Technical report.

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

U.S. Census Bureau (2011). Characteristics of new housing - highlights of annual 2010 characteristics of new housing.

U.S. Department of Energy (2012). Buildings energy data book. Technical report. <http://buildingsdatabook.eren.doe.gov/ChapterIntro2.aspx>.

U.S. Intelligence Community (2012). Global water security. Technical report. <http://www.dni.gov/files/documents/Newsroom/Press%20Releases/ICA_Global%20Water%20Security.pdf>.

Vallgårda, S. (2012). Nudge - a new and better way to improve health? *Health Policy*, 104(2):200–203.

van Dam, S. S., Bakker, C. A., and van Hal, J. D. M. (2010). Home energy monitors: impact over the medium-term. *Building Research & Information*, 38(5):458–469.

VandeWalle, D. (2003). A goal orientation model of feedback-seeking behavior. *Human Resource Management Review*, 13(4):581–604.

Viscusi, W. K., Huber, J., and Bell, J. (2011). Promoting recycling: private values, social norms, and economic incentives. *American Economic Review*, 101(3):65–70.

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.

Watson, R., Boudreau, M.-C., and Chen, A. (2010). Information systems and environmentally sustainable development: energy informatics and new directions for the IS community. *Management Information Systems Quarterly*, 34(1):23–38.

Weaver, F. (2014). The unprecedented water crisis of the American Southwest. The Week. 2014-02-01. <http://theweek.com/article/index/255814/the-unprecedented-water-crisis-of-the-american-southwest>.

Webb, D. A., Boyd, N. R., Messina, D., and Windsor, R. A. (2003). The discrepancy between self-reported smoking status and urine cotinine levels among women enrolled in prenatal care at four publicly funded clinical sites. *Journal of Public Health Management & Practice*, 9(4):322–325.

Wilcox, K., Vallen, B., Block, L., and Fitzsimons, G. J. (2009). Vicarious goal fulfillment: when the mere presence of a healthy option leads to an ironically indulgent decision. *Journal of Consumer Research*, 36(3):380–393.

Williams, E. (2011). Environmental effects of information and communications technologies. *Nature*, 479(7373):354–358.

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.

Wilson, C. and Dowlatabadi, H. (2007). Models of decision making and residential energy use. *Annual Review of Environment and Resources*, 32(1):169–203.

Wood, G. and Newborough, M. (2003). Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. *Energy and Buildings*, 35(8):821–841.

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

WTO (2010). World trade report 2010 - trade in natural resources. <http://www.wto.org/english/res_e/booksp_e/anrep_e/world_trade_report10_e.pdf>.

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

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 HEXACO Personality Trait Emotionality

Table A.2: 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 × 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 × 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.3 Treatment Interaction Effect with HEXACO Personality Trait Conscientiousness

Table A.3: 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

A.4 Treatment Interaction Effect with Tendency to Measure Progress Towards Goals

Table A.4: 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.5 Treatment Interaction Effect with Tendency to Compare Oneself with the Performance of Others

Table A.5: 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.6 Treatment Interaction Effect with Self-estimated Savings Potential

Table A.6: 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 × 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 × 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.7 Treatment Interaction Effect with Age

Table A.7: 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$\times$ 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 $\times$ 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.8 Treatment Interaction Effect with Gender

Table A.8: 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.9 Treatment Interaction Effect with Happiness

Table A.9: 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.10 Treatment Interaction Effect with HEXACO Personlity Trait Honesty

Table A.10: 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× 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 × 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.11 Treatment Interaction Effect with HEXACO Personlity Trait Extraversion

Table A.11: 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.12 Treatment Interaction Effect with HEXACO Personlity Trait Agreeableness

Table A.12: 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

List of Tables

List of Figures