One thing leads to another: Evidence on the scope and persistence of behavioral spillovers

Journal Article

Author(s):
Goetz, Alexander; Mayr, Harald; Schubert, Renate

Publication date:
2024-08

Permanent link:
https://doi.org/10.3929/ethz-b-000682607

Originally published in:
Journal of Public Economics 236, https://doi.org/10.1016/j.jpubeco.2024.105166

Journal of Public Economics 236 (2024) 105166

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

One thing leads to another: Evidence on the scope and persistence of behavioral spillovers[☆]

Alexander Goetz[a,b], Harald Mayr[c,*], Renate Schubert[a]

[a] ETH Zurich, Switzerland
[b] Swiss National Bank, Switzerland
[c] University of Zurich, Switzerland

ARTICLE INFO

JEL classification:
C93
D12
Q50

Keywords:
Behavioral spillover
Randomized controlled trial
Water consumption
Energy consumption
Welfare

ABSTRACT

Evaluations of economic interventions usually focus on one target behavior. This study extends the evaluation scope to multiple untargeted behaviors. We evaluate a hot water saving intervention in a natural field experiment. Despite an exclusive focus on hot water, the intervention changes multiple behaviors. Notably, we find a 5.6 percent reduction in room heating energy consumption that persists one year after the intervention. We show that the room heating spillover has important welfare implications.

1. Introduction

Economists have long recognized that individual behaviors are interrelated. Hicks and Allen (1934) formalize the idea of substitutes and complements in consumption, Tinbergen (1952) studies multiple targets for economic policy, and Heckman and Smith (1998) emphasize the importance of considering multiple margins in policy evaluation. Influential empirical work studies multiple outcomes. Leading examples are evaluations of the Perry Preschool Program (Rolnick and Grunewald, 2003; Belfield et al., 2006; Heckman et al., 2010) and Moving to Opportunity (Kling et al., 2007; Ludwig et al., 2008; Fryer and Katz, 2013; Ludwig et al., 2013), investigating effects on education, employment, earnings, crime, tax revenues, and the use of the welfare system. Non-targeted behaviors, frequently called "behavioral spillovers", are particularly relevant in the environmental domain.

An extensive literature evaluates interventions promoting natural resource conservation (Allcott, 2011b; Ferraro et al., 2011; Harding and Hsiaw, 2014; Brent et al., 2015; Bradley et al., 2016; Hahn et al., 2016; Wichman et al., 2016; Jessoe et al., 2021a). These studies typically focus on a target behavior, such as water or electricity consumption, but we know little about potential spillovers on other behaviors. Policy makers might be afraid that promoting one pro-environmental behavior crowds out other pro-environmental behaviors, akin to retirement saving policies that crowd out other forms of saving (Chetty et al., 2014; Blau, 2016; Choukhmane, 2024). Behavioral spillovers can also reinforce the effectiveness of an intervention, like a water saving intervention that reduces electricity consumption (Jessoe et al., 2021b). In either case, comprehensive evaluations are crucial to understand the full welfare implications of interventions in the environmental domain.

In this paper, we evaluate a hot water conservation intervention in a natural field experiment. We randomized 782 apartment buildings in Switzerland, with a total of 4,775 tenant households. To limit experimenter demand effects, all households remained unaware of

☆ We thank our partner companies for the productive collaboration. We thank the editor Robert Metcalfe and three anonymous referees for valuable comments that significantly improved the paper. We also thank Lisa Abraham, Erik Ansink, Sandro Ambühl, Björn Bartling, Petyo Bonev, Dylan Brewer, Federica Coelli, Francisco Costa, Ximeng Fang, Ernst Fehr, Hanna Fuhrmann-Riebel, Manuel Grieder, David Hémous, Michel Marechal, Ed Rubin, Wolfram Schlenker, Anna Schulze Tilling, Raisa Sherif, Roberto Weber, Collin Weigel, and participants at the 2022 AERE Summer Conference, Advances with Field Experiments, EEA-ESEM Congress, SSES Annual Congress, the 2021 Summer Workshop on Empirical Methods in Energy Economics and the Online Summer Workshop in Environment, Energy, and Transportation for helpful comments. The field experiment reported in this paper was approved by ETH Zurich's institutional review board (reference: EK-2019-N-85) and was registered on www.socialscienceregistry.org (RCT ID: AEARCTR-0004995). The research was supported by the National Research Programme 73, Switzerland (project number: 407340_172431).

* Correspondence to: Schönberggasse 1, 8001 Zurich, Switzerland.
E-mail addresses: alexander.goetz@snb.ch (A. Goetz), harald.mayr@econ.uzh.ch (H. Mayr), schubert@econ.gess.ethz.ch (R. Schubert).

https://doi.org/10.1016/j.jpubeco.2024.105166
Received 31 May 2023; Received in revised form 27 April 2024; Accepted 8 June 2024
Available online 5 July 2024

the intervention's experimental nature. The intervention applies established behavioral instruments to promote hot water conservation: consumption information, social comparison, conservation tips, a saving goal, and a lottery incentive. We study how this intervention, which focuses exclusively on hot water, impacts the consumption of cold water, electricity, and room heating energy.

We propose a taxonomy with three behavioral spillover mechanisms: complementarities, direct spillovers, and self-image spillovers. A spillover results from complementarities if the target and spillover behavior are either complementary or substitutable. Direct spillovers occur if the intervention makes not only the environmental consequences of the target behavior more salient but also those of similar behaviors. Such direct spillovers can explain positive spillovers that attenuate over time. Finally, a change in the target behavior may influence the household's environmental self-image, potentially changing behavior in the long term. Depending on whether environmental self-image and moral utility from environmental behavior are substitutes or complements, the self-image mechanism accommodates transient negative spillovers in the spirit of moral licensing (Monin and Miller, 2001) or persistent positive spillovers in the spirit of cognitive dissonance theory, where households strive to be consistent with their environmental self-image (Festinger, 1962). We provide evidence on multiple spillover behaviors (including behaviors not subject to complementarities) and on the persistence of spillovers to shed light on the mechanisms at play.

We find that the effects of the hot water intervention extend beyond the target behavior. The intervention decreases the target behavior by 4.96% ($p < 0.001$) during the four intervention months, an effect that attenuates somewhat during the 24 observed post-intervention months. The intervention does not change cold water consumption at the household level, and we find no evidence for a spillover on electricity consumption. Our most consequential finding is that the intervention reduces room heating energy consumption by 5.63% ($p = 0.021$). This effect corresponds to less than a 1 °C reduction in room temperature. Estimated elasticities of demand suggest that a 24%–33% price increase would have similar effects on room heating (Auffhammer and Rubin, 2018). The effect on room heating is persistent, at 5.92% ($p = 0.074$), one year later.

Unique tap-level data, available for 30% of the buildings in our sample, reveal additional insights. With these data, we study complementarities in mixer taps (shower, kitchen sink, and vanity basin) and investigate the persistence of spillovers where complementarities with hot water are less of a concern (dishwasher use and toilet flushing). Our mixer tap results show that households save hot water predominantly in kitchen sink use (−9.32%, $p = 0.067$). In terms of cold water spillovers, we find large reductions in dishwasher use (−21.28%, $p = 0.020$, during the intervention) that persist eight months post-intervention. Persistent behavioral spillovers may arise from self-image spillovers in the spirit of cognitive dissonance theory.

Our findings have considerable welfare implications because room heating requires substantially more energy than hot water heating. Ignoring spillovers, we estimate a marginal value of public funds (MVPF) of −0.34. This value becomes even more negative, at −1.06, when we take the room heating spillover into account. Disregarding profit losses by utility companies, we find MVPF estimates of 0.26 (without spillovers) and 1.60 (with spillovers). Spillovers are also important in a cost effectiveness analysis, where we find costs of CHF 698 per tonne of CO_2 abatement when we ignore spillovers; CHF 57 per tonne of CO_2 if we include the room heating spillover.

This paper contributes to a large literature on behavioral spillovers that originates in psychology. The idea that a behavioral change in one domain can cause a behavioral change in related domains has been studied extensively. The literature has a rich array of empirical studies and psychological explanations for spillovers (Truelove et al., 2014; Dolan and Galizzi, 2015; Nilsson et al., 2017; Maki et al., 2019), but they are subject to important methodological concerns, including "a reliance on self-reported behavior, which is known to be only weakly correlated with actual behavior" and "a reliance on correlational or longitudinal designs which are unable to shed light on causal processes". In addition, "few studies also conduct follow-up measurements, so the durability of any immediate spillover effects is unknown" (Galizzi and Whitmarsh, 2019, p. 3). Our study contributes to this literature by studying a natural field experiment that allows causal conclusions and sheds light on the persistence of behavioral spillovers.

Our paper also contributes to a growing economics literature on behavioral spillovers in the residential sector. Prior research primarily examines spillovers from water-saving interventions on electricity consumption (Tiefenbeck et al., 2013; Carlsson et al., 2020; Jessoe et al., 2021b) and spillovers in the context of waste collection (Ek and Miliute-Plepiene, 2018; Xu et al., 2018; Alacevich et al., 2021; Sherif, 2021). Our study advances this literature with multiple spillover behaviors and a focus on the persistence of spillovers. To the best of our knowledge, our study is the first to measure multiple spillover behaviors. Counting the different water taps separately, we identify seven spillover behaviors. This detailed account of behaviors, over the long time periods under consideration, facilitates new insights into the mechanisms behind behavioral spillovers. In addition, our study is the first to investigate spillovers on water consumption and room heating. These environmentally consequential behaviors are, as it turns out, particularly susceptible to spillovers.

Finally, our results touch upon several related topics in the economics literature. An emerging literature studies spillovers caused by limited attention, also called "cognitive spillovers" (Nafziger, 2020; Altmann et al., 2021; Hall and Madsen, 2021; Medina, 2021; Trachtman, 2021). In the context of fundraising, a strand of literature investigates how donation appeals for one charity affect subsequent donations (Donkers et al., 2017; Meer, 2017; Adena and Huck, 2019; Deryugina and Marx, 2021; Grieder et al., 2021). A recent study assesses unintended consequences of financial vaccination incentives (Schneider et al., 2023). The main implication of our results—that spillovers can have important welfare implications—is relevant in the context of the flourishing literature on the welfare effects of economic interventions (DellaVigna et al., 2012; Damgaard and Gravert, 2018; Jimenez-Gomez, 2018; Taubinsky and Rees-Jones, 2018; Allcott and Kessler, 2019; Taylor, 2020; Butera et al., 2022).

The rest of the paper proceeds as follows. Section 2 describes our theoretical framework. Section 3 describes our field experiment's setting and design. Section 4 presents the data, and Section 5 presents the results. Section 6 discusses potential spillover mechanisms and the welfare implications of our findings. Section 7 concludes.

2. A taxonomy of spillovers

In this section we build on Dolan and Galizzi (2015) and propose a simple theoretical framework with three spillover mechanisms: complementarities, direct spillovers, and self-image. The framework yields empirical predictions for the persistence of behavioral spillovers.[1]

2.1. Household problem

We consider T time periods $t \in \{1, 2, \ldots, T\}$. A household with income y_t consumes $x_t = (x_t^a, x_t^b)$ at prices $p_t = (p_t^a, p_t^b)$ and a numeraire good. We call x_t^a the intervention's *target behavior* and x_t^b the *spillover behavior*. The household enjoys consumption utility $u(x_t)$.

We incorporate environmental preferences similar to Dolan and Galizzi (2015), in the spirit of the "beliefs as assets" model by Bénabou and Tirole (2011). In addition to consumption utility $u(x_t)$, the

[1] In the interest of full disclosure, these are not ex ante predictions as we did not specify a theoretical framework in our pre-analysis plan.

household cares about moral utility and environmental self-image I_t.[2] We model moral utility as $m - x_t \cdot \mu_t$, where $\mu_t = (\mu_t^a, \mu_t^b)$ are moral prices that reflect the psychological costs of immoral behavior and m is the moral utility from $x_t^a = x_t^b = 0$. Environmental self-image is malleable and follows a dynamic process, where choices serve as signals for the household's environmental self-image and $I_t = I(x_{t-1}, I_{t-1})$. Moral behavior and environmental self-image may be complements or substitutes, depending on the substitution parameter ρ of a constant elasticity of substitution function. At time t, the household chooses target behavior and spillover behavior to maximize the following utility function:

$$\max_{x_t} U_t = y_t - x_t \cdot p_t + u(x_t) + \left[(m - x_t \cdot \mu_t)^\rho + I(x_{t-1}, I_{t-1})^\rho\right]^{\frac{1}{\rho}}. \tag{1}$$

We then consider an intervention in period $t = 1$. This intervention increases the target behavior's moral price, μ_1^a, leading the household to reduce x_1^a.

2.2. Spillover mechanisms

Spillovers may result from three distinct mechanisms. In what follows, we discuss these mechanisms in detail.

Complementarities. A spillover results from complementarities if the target and spillover behavior are either complementary or substitutable ($\frac{\partial x^b}{\partial p^a} \neq 0$). We expect positive spillovers where target and spillover behavior are complements, whereas in the case of substitutes, we expect negative spillovers. The intervention's effect on the spillover behavior is more similar to the effect on the target behavior the more complementary the two behaviors are. Complementarities can be due to preferences or so-called mechanical linkages. An example for the latter would be the washing machine, which uses both water and electricity.

Direct spillovers. The intervention may make not only the environmental consequences of the target behavior more salient but also the consequences of other behaviors. We call a spillover resulting from an increase in μ_1^b a direct spillover. More generally, a direct spillover may result from any mechanism that drives the target behavior, like rational inattention (Gabaix, 2014; Sallee, 2014; Bronchetti et al., 2020; Costa and Gerard, 2021) or biased beliefs (Allcott, 2011a; Werthschulte and Löschel, 2021).

Direct spillovers are transient and mimic the effect on the target behavior. If the intervention leads to higher moral prices μ_1^a and μ_1^b, we expect x_1^a and x_1^b to decrease. As both effects go in the same direction, this mechanism can only explain positive spillovers. The effects of behavioral interventions for resource use typically attenuate over time (Allcott and Rogers, 2014; Bernedo et al., 2014; Brandon et al., 2017). Accordingly, direct spillovers are expected to attenuate over time as well.

Self-image. A change in the target behavior influences self-image, which may in turn influence the spillover behavior in the subsequent period. We use a loose definition of self-image, with the idea that a change in the target behavior influences a stock variable (e.g., self-image, identity, or habit) that eventually has an effect on the spillover behavior. As described by Dolan and Galizzi (2015), a broad definition of self-image captures a wide range of psychological spillover mechanisms.

A positive spillover via self-image relies on two conditions. The first condition is that the intervention changes environmental self-image. The psychological literature offers instruments to measure this concept (Sparks and Shepherd, 1992; Dunlap et al., 2000; Martin and Czellar, 2016) and indicates that past environmental behavior indeed influences environmental self-identity (Van der Werff et al., 2013, 2014). The second condition is that environmental self-image translates into pro-environmental behavior. Recent evidence from psychology suggests that this is indeed the case (Carfora et al., 2017).

In our theoretical framework, self-image causes a spillover if moral utility depends on the household's self-image. If $\rho > 0$, self-image is a substitute for the moral utility derived from the spillover behavior. The resulting negative spillover is akin to moral licensing—a moral initial behavior leads to an undesired behavior later on (Monin and Miller, 2001; Effron et al., 2009; Merritt et al., 2010; Blanken et al., 2015). Conversely, if $\rho < 0$, self-image and moral utility from the spillover behavior are complements. This case accommodates the theory of cognitive dissonance, which stipulates that individuals strive to be consistent in their beliefs and actions (Festinger, 1962; Cialdini et al., 1995; Gawronski, 2012).[3]

The persistence of self-image spillovers depends on the sign of ρ. If $\rho > 0$, an increase in self-image causes environmentally problematic spillover behavior. Such a behavior change reverts self-image toward its original level so that the negative spillover would be transient. If $\rho < 0$, an increase in self-image causes more environmentally friendly spillover behavior, further increasing self-image. The positive spillover would build up over time and be highly persistent. To summarize, $\rho > 0$ implies a transient negative spillover, while $\rho < 0$ implies a positive spillover that builds up over time.

In the next section, we describe the experimental design. We return to the theoretical framework for the discussion in Section 6.

3. Experimental design

This section describes the setting of our field experiment and the details of the intervention, the study sample, and the randomization procedure.

3.1. General setting

We collaborated with a large real estate owner in Switzerland (henceforth the Owning Company). The Owning Company is one of the largest real estate owners in Switzerland with a portfolio of business and private properties. Our study focuses on apartment buildings. Most apartment buildings in our sample are of moderate size, with a median of 5 apartments (the interquartile range spans from 4 to 7 apartments). The buildings are spread across the three main language regions of Switzerland (German, French, and Italian), and cover urban, suburban, and rural settings. A subsidiary of the Owning Company (henceforth the Managing Company) provides real estate management services for the Owning Company. The Managing Company oversees tenant relations, including advertising vacant apartments, managing property, and overseeing all rent-related communications. It contacts tenants on limited occasions, including sending utility bills. Communication in the name of the Managing Company is familiar to and trusted by the tenants.

The intervention was implemented as a natural field experiment in the spirit of Harrison and List (2004). Households in the intervention group received the intervention, and a separate control group was not contacted until the end of the intervention. The intervention was implemented as an email campaign, called the "Hot Water Challenge", from

[2] Bénabou and Tirole (2011) consider multidimensional identities, where households care about different domains or "life-satisfaction accounts" (e.g., health, wealth, morality, environment). Our specification includes a standard consumption domain (where identity does not play a role) and an environmental domain, where both moral utility and identity matter. We refer the interested reader to Dolan and Galizzi (2015), who present a model of spillovers with multidimensional identities.

[3] The case where $\rho < 0$ is conceptually similar to habit formation (Becker and Murphy, 1988; Byrne et al., 2022). Loosely speaking, I_t may represent a habit stock at time t, which is determined by previous consumption choices.

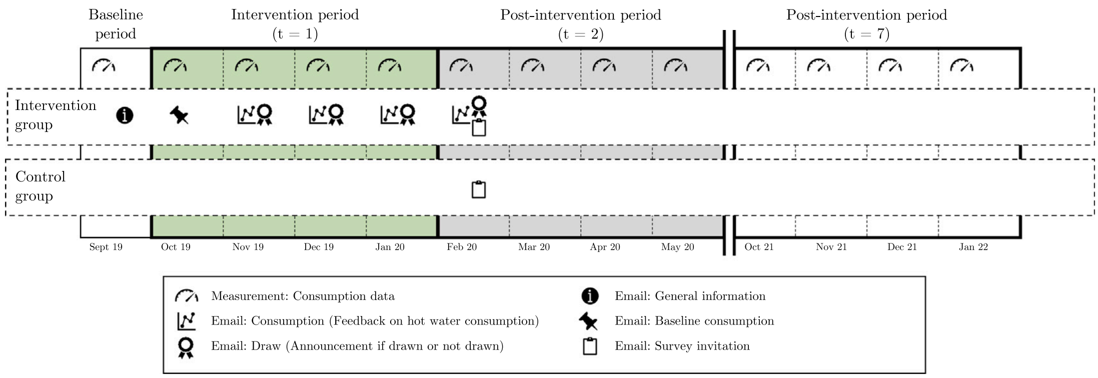

Fig. 1. Timeline of the intervention.

the Managing Company, which had not run comparable interventions before. The research team's involvement was only disclosed after the end of the intervention, as required by ETH Zurich's institutional review board (reference: EK-2019-N-85). The intervention applied to every household in the intervention group by default. Households could opt out at any time, in which case they would not receive any further intervention emails. Because the research team only gained access to fully anonymized data, opting out precluded neither the collection nor the analysis of the household's data.

3.2. Intervention

The intervention combined five behavioral instruments: information on the household's hot water consumption, social comparison, hot water conservation tips, a 5% saving goal, and a lottery tied to the attainment of the saving goal. Households received the following information about all households that had the same number of rooms: the average hot water consumption and the average hot water consumption among the 20% with the lowest consumption values. Each household's personal saving goal was set to 5% of its hot water consumption in September 2019, the baseline month of the intervention.

The saving goal was kept constant throughout the intervention. In each of the four intervention months, households that attained the saving goal could win a month's paid rent in the lottery. Online Appendix A provides English transcriptions of the intervention emails and the hot water conservation tips, which were provided in German, French, or Italian, according to the household's preferred language. Online Appendix A also depicts the graphical illustrations used in the intervention emails.

Fig. 1 shows the intervention's timeline. The email *Basic Information*, sent on September 23, 2019, informed households about the aim and content of the planned intervention, addressed privacy and legal issues, and offered an opt-out option at a mouse's click. The intervention started with the email *Baseline Consumption*, sent on October 11, 2019, which informed households about their hot water consumption during the preceding month, September 2019, and their personal saving goal.

Households received two emails at the beginning of each of the subsequent four months, November 2019 to February 2020. In early November, for instance, the email *Consumption 1* informed households about their consumption in the previous month and indicated whether the saving goal was reached in that month or not. The email *Draw 1*, sent the next day, informed them about the lottery draw. The very last email (*Draw 4*), sent on February 12, 2020, disclosed that the research team would analyze the effects of the intervention based on anonymized data.

The intervention was concluded in February 2020, before the onset of the COVID-19 pandemic in Switzerland. Appendix E describes how COVID-19 affected Switzerland and discusses how the pandemic may influence the interpretation of our long-term findings.

Two features of the intervention were varied in a cross-randomized design, as depicted in Online Appendix A. First, 50% of the households received social comparison information. The other 50% received only information about their individual consumption. Second, 50% of the households were subject to a regret lottery. These households would learn whether they were drawn in the lottery irrespective of their goal attainment.[4] The other 50% were subject to a standard lottery, and they would only learn about having been drawn if they had attained the saving goal. Figures A.3 and A.4 in the Online Appendix show that the provision of social comparison decreases hot water consumption (estimated effect of -2.9%, $p < 0.01$), while the type of lottery does not influence hot water consumption (estimated effect of 0.22%, $p = 0.749$). In terms of spillovers, Figures A.5 and A.6 show that neither social comparison (estimated effect of -0.45%, $p = 0.227$) nor the lottery type (estimated effect of 0.52%, $p = 0.508$) have a significant effect on cold water consumption. The cross-randomized treatment variations do not appear to influence the spillovers investigated in this paper. We abstract from the cross-randomized nature of our experiment in the rest of this paper and use the term "intervention" to refer to all treatment variations.

3.3. Study sample and randomization

The study sample includes all households that fulfilled the following technical requirements. Buildings had to have hot water meters at the household level, which had to be remotely read every month. We restricted the sample to households with valid hot water readings for August 2019. Eligible households had to have a valid email address in the Managing Company's database before the intervention started. In addition, tenants were only included in the sample if they rented exactly one apartment.[5] Finally, households that terminated their rental agreement before the intervention started were excluded. These criteria left us with a sample of 4,775 households in 782 buildings.

The intervention was randomly allocated at the building level. All households in a building were either assigned to the intervention group or to the control group. This form of randomization prevented control group households from learning about the intervention from other tenants in the same building. Furthermore, it avoided the possibility that a household could win in a lottery while a neighbor in the same building could not, a scenario that could be perceived as unfair. Twenty

[4] In a regret lottery, households learn about being drawn, even if they do not meet the goal to be eligible for the prize. The idea is that the anticipated feeling of regret motivates behavior change. This approach contrasts with a standard lottery, where households are only informed of the lottery draw if they have achieved the goal. The Dutch postcode lottery is a well-studied example of a regret lottery (Zeelenberg and Pieters, 2004).

[5] Very few tenants rent multiple apartments. Including them would have unduly complicated the intervention's implementation.

percent of our sample were allocated to the control group, and the remaining 80% were subject to the intervention. The randomization was performed within strata, as described in detail in Online Appendix A, on September 20, 2019.

4. Data

We use three main types of data. The first type is monthly data for hot and cold water consumption on the household level, which were delivered by a specialized company that is contracted to maintain sub-metering data for the Managing Company. The second type is electricity consumption data on the building level, and the third is data on heating energy consumption on the cost center level. A cost center represents the expenditures of a physical heating system, which may comprise one or multiple buildings. In the following subsection, we describe the three types of data in detail.

4.1. Hot and cold water consumption

Our outcome variables for hot and cold water are based on household-level data. We measure the consumed volumes (in m^3) of hot and cold water separately and denote hot water consumption of a household i in time period t as $HW_{i,t}$ and the respective cold water consumption as $CW_{i,t}$. The following equations refer to hot water, but equivalent equations apply for cold water. We average the household-level data on the level of building b (the unit of randomization with N_b households) and normalize to monthly values (with N_t months in time period t):

$$HW_{b,t} = \frac{\sum_{i \in b} HW_{i,t}}{N_b \cdot N_t}. \tag{2}$$

The intervention period $t = 1$ covers the four intervention months from October 2019 to January 2020. We consider post-intervention periods of equal duration. $t = 2$ covers the four months from February to May 2020, $t = 3$ covers June to September 2020, and so on. The last time period in our data ($t = 7$) covers October 2021 to January 2022. For each time period t, we denote the percentage change of consumption (compared to the baseline month September 2019) as $dHW_{b,t}$, where

$$dHW_{b,t} = \frac{HW_{b,t} - HW_{b,sep}}{HW_{b,sep}} \cdot 100. \tag{3}$$

See Figures B.1 and B.2 in the Online Appendix for histograms of the outcome variables $dHW_{b,t=1}$ and $dCW_{b,t=1}$.

In addition to household-level hot and cold water data, tap-level data are available for a sub-sample. This sub-sample, comprising 231 of the 782 buildings in the full sample, has hot and cold water meters for the mixer taps in the shower, sink, and vanity basin as well as a cold water meter for toilet flushing. 167 buildings have positive readings for a meter at the dishwasher.

4.2. Electricity consumption

Local electric utility companies provide us with annual electricity data on the building level. The buildings in our sample are served by 55 companies; 12 responded to a request for data, providing data for 324 of the 782 buildings in our sample (41%).

Since the intervention starts in October 2019, we use the year 2018 as our baseline period. We use the percentage change in electricity consumption from 2018 to 2019 and 2020 as our outcome variables.

$$dE_{b,2019} = \frac{E_{b,2019} - E_{b,2018}}{E_{b,2018}} \cdot 100, \tag{4}$$

$$dE_{b,2020} = \frac{E_{b,2020} - E_{b,2018}}{E_{b,2018}} \cdot 100. \tag{5}$$

See Figures B.3 and B.4 in the Online Appendix for histograms of the outcome variables $dE_{c,2019}$ and $dE_{c,2020}$.

4.3. Heating energy consumption

Our analysis of heating energy consumption is based on cost center data, which are regularly used to bill households. A cost center represents the expenditures of a physical heating system for natural gas, district heating, oil, electricity, or a mixture of those. The energy consumption of a given cost center is calculated by converting physical quantities (e.g., liters of oil) into energy content (kWh). These values include not only energy for room heating but also energy for hot water. Our estimation strategy takes this caveat into account (see Section 5.3).

Some buildings share a heating system and, consequently, a cost center. The 782 buildings in our sample share 333 cost centers. As we randomized on the building level, an individual cost center may include buildings from the intervention group and the control group. For a cost center c, we calculate the intervention share $Intervention_c$ as the number of intervention households over the total number of households. A cost center with only control group buildings has an intervention share of zero, whereas a cost center that includes only intervention buildings has an intervention share of one. A cost center with both control and intervention buildings has an intervention share between zero and one.

Cost center data refer to yearly billing periods. We calculate the relative change from the 2019 billing period to the 2020 billing period, which is potentially influenced by the intervention.[6] We then exclude 13 cost centers that changed their energy sources from the 2019 billing period to the 2020 billing period. Moreover, we exclude 20 cost centers with missing data. Our estimation sample hence comprises 300 of the relevant 333 cost centers. We calculate the outcome variable $dHE_{c,2020}$ as the percentage change from the 2019 billing period to the 2020 billing period:

$$dHE_{c,2020} = \frac{HE_{c,2020} - HE_{c,2019}}{HE_{c,2019}} \cdot 100. \tag{6}$$

We also use data on the 2021 billing period, by starting with the 300 cost centers from the 2020 billing period and excluding 3 cost centers that changed their energy sources from the 2020 billing period to the 2021 billing period. Hence, the estimation sample for the 2021 billing period comprises 297 cost centers. Similar to above, we calculate the outcome variable $dHE_{c,2021}$ as the percentage change from the 2019 billing period to the 2021 billing period:

$$dHE_{c,2021} = \frac{HE_{c,2021} - HE_{c,2019}}{HE_{c,2019}} \cdot 100. \tag{7}$$

Figures B.5 and B.6 in the Online Appendix show histograms of the outcome variables $dHE_{c,2020}$ and $dHE_{c,2021}$. Figure B.7 depicts the intervention share $Intervention_c$.

5. Empirical approach and results

This section describes our empirical approach and results for water consumption, electricity consumption, and heating energy consumption.

5.1. Water consumption

Estimation. Before using the data, we exclude outliers in two steps. First, we exclude households with the lowest 5% consumption values in the baseline month, September. Low baseline values (caused, e.g., by a vacation in September 2019) would artificially inflate our outcome variables. Second, for the remaining sample, we exclude the bottom

[6] The billing date is March 31 in the majority of cost centers. Other billing dates are April 30, May 31, June 30, and August 31. The different billing dates are not a significant concern for our analysis because all billing dates cover the entire intervention period (October 2019 to January 2020) in the 2020 billing period.

and top 1% households based on the respective outcome variable to account for errors (e.g., due to defective meters). The exclusion criteria in both steps (5% and 1%) are pre-specified.

We then estimate intention-to-treat effects, which consider the difference in outcomes between those who were initially assigned to the intervention and the control group irrespective of whether they complied with their treatment assignment (Heckman, 2010). Consequently, our estimation sample includes households whose emails bounced, who opted out, or who did not open the intervention emails. We include these households to ensure that our estimates apply to the full targeted population rather than just to the households that actually took part in the intervention. This approach allows us to avoid selection bias and measure effects that are directly policy relevant.

The randomized nature of our data allows for a straightforward empirical analysis. Following the recommendation of Athey and Imbens (2017), we conduct our analysis on the building level, i.e., the level of treatment assignment. We regress the outcome variables of interest on a constant and $Intervention$, an indicator that equals one if the building was assigned to the intervention group and zero if it was assigned to the control group. The following equation refers to hot water, but equivalent equations apply for cold water and tap-level consumption:

$$dHW_{b,t} = \alpha_{hw} + \beta_{hw} Intervention_b + \epsilon_b. \quad (8)$$

We estimate Eq. (8) with ordinary least squares (OLS) and calculate heteroscedasticity robust standard errors. α_{hw} yields the mean in the control group, while β_{hw} is the intervention effect.

Quality of randomization. Table C.1 in the Online Appendix presents hot water consumption values before the intervention started. The table shows data from June, July, August, and September 2019 for the estimation sample in the intervention period ($t = 1$). Columns (1) and (2) report mean and median values for all buildings in the control group. The average building in the control group uses 2.5 m^3 of hot water per household in June 2019, 2.1 m^3 in July, 2.3 m^3 in August, and 2.5 m^3 in the baseline month of September. Among the households with available tap-level data, hot water is used in the shower (1.4 m^3 in September), kitchen sink (0.7 m^3), and vanity basin (0.4 m^3). Columns (3) and (4) report mean and median values for all buildings in the intervention group. To assess the quality of randomization, columns (5) and (6) show differences between the control and intervention group, with p-values in column (7). All differences are statistically insignificant.

Table C.2 in the Online Appendix presents cold water consumption values. Cold water consumption in the control group is 5.7 m^3 in June 2019, 5.5 m^3 in July, 5.6 m^3 in August, and 5.4 m^3 in September. Among the households with available tap-level data, cold water is used in the shower (1.3 m^3 in September), kitchen sink (0.8 m^3), vanity basin (0.6 m^3), dishwasher (0.3 m^3), and during toilet flushing (2.9 m^3). These values are similar in the intervention group, and all differences are statistically insignificant, suggesting that the randomization achieved a balance of pre-intervention cold water consumption values.

Results for household-level water consumption. Fig. 2 summarizes the results for our target behavior, hot water consumption, during the intervention period ($t = 1$) and the post-intervention periods ($t = 2$ to $t = 7$) (see Table C.3 in the Online Appendix for detailed results).[7]

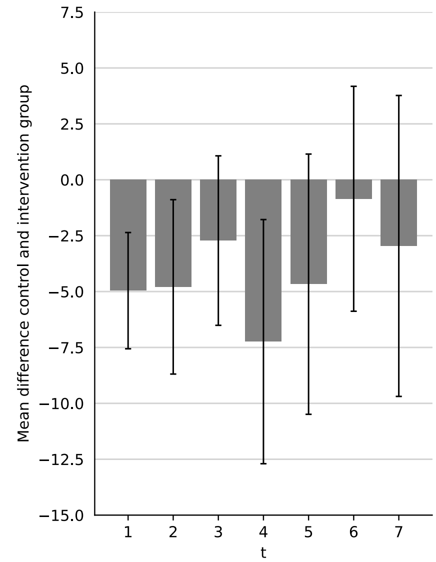

Fig. 2. Effects on hot water.
Note. The figure shows the effects on hot water for the intervention period ($t = 1$) and the post-intervention periods ($t = 2$ to $t = 7$), where each period t represents an average over four months. Detailed results are available in Table C.3 in the Online Appendix. Bars correspond to the estimated intervention effects. The error bars represent 95% confidence intervals.

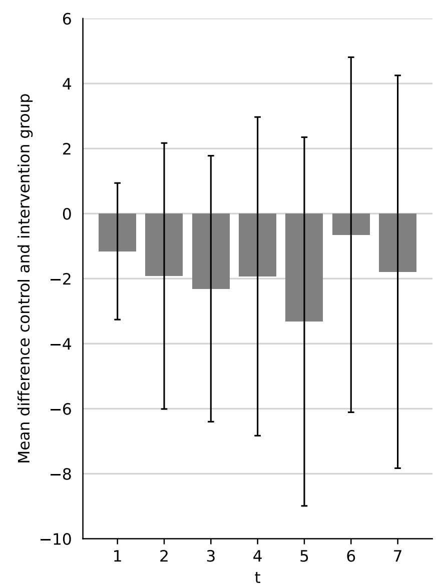

Fig. 3. Effects on cold water.
Note. The figure shows the effects on cold water for the intervention period ($t = 1$) and the post-intervention periods ($t = 2$ to $t = 7$), where each period t represents an average over four months. Detailed results are available in Table C.4 in the Online Appendix. Bars correspond to the estimated intervention effects. The error bars represent 95% confidence intervals.

We see a strong effect of the intervention. During period $t = 1$, the intervention has an effect of −4.96% ($p < 0.01$). The impact of the intervention persists in the corresponding post-intervention period $t = 2$, with a −4.79% effect ($p = 0.016$). In $t = 3$, the effect of the intervention attenuates at −2.72% ($p = 0.159$), to recur in $t = 4$, one year after the intervention, at −7.24% ($p = 0.009$). After that, the effect attenuates to −4.67% in $t = 5$ ($p = 0.116$) and is not statistically significant in $t = 6$ and $t = 7$.

Fig. 3 displays the results for the cold water consumption spillover (see Table C.4 in the Online Appendix for detailed results). The first

[7] Table C.3 in the Online Appendix shows results for all time periods $t = 1$ to $t = 7$ as well as monthly results. The first and second column report the constant terms and intervention effects respectively, with standard errors shown in parentheses. The third column lists the p-values for the intervention effects. The fourth column indicates the number of buildings in the main estimation sample for each time period. Fig. 2 summarizes the effect estimates in the second column and the p-values in the third column for the time periods $t = 1$ to $t = 7$.

bar depicts the intervention period $t = 1$. The effect of −1.16% is not statistically significant at conventional levels ($p = 0.279$). The bars for $t = 2$ to $t = 7$ show that the intervention effect hovers between −0.65% and −3.32%, but neither of these effects reaches statistical significance. In summary, we see a large reduction in hot water consumption but no significant evidence for cold water spillovers at the household level.[8] In the following subsection, we dig deeper and investigate water effects on the tap level.

Results for tap-level water consumption. We further investigate spillovers on cold water in the sub-sample with tap-level data. We discuss selection into this sub-sample in Section 6.2.

Fig. 4 summarizes the effects in the intervention period ($t = 1$) and the post-intervention periods ($t = 2$ to $t = 7$), for hot and cold water (see Tables C.6–C.13 in the Online Appendix for detailed results). The bars correspond to the estimated differences between the control and intervention group means. The three leftmost columns in the figure show results for the mixer taps in the shower, kitchen sink, and vanity basin. The intervention seems to reduce hot and cold water consumption in the post-intervention period, but these effects are mostly statistically insignificant. The kitchen sink appears pivotal in the intervention's effect on the target behavior: households strongly reduce their hot water usage when using the kitchen sink, with effect sizes between −9.32% (in $t = 1$) and −29.39% (in $t = 4$). The hot water reductions for kitchen sink use are statistically significant with $p < 0.067$ in $t = 1$ to $t = 3$ and with $p < 0.10$ in $t = 4$. Cold water consumption in kitchen sink use, however, sees no statistically significant effects. We find no statistically significant intervention effects on hot or cold water for the shower and the vanity basin.

The remaining two graphs in Fig. 4 show results for the dishwasher and toilet flushing. These taps are particularly interesting because they use only cold water and are hence less affected by complementarities with hot water. We find a large reduction in dishwasher use during the intervention period (−21.28%, $p = 0.02$) that appears persistent in the post-intervention periods (effect sizes between −10.93% and −21.06%, with $p < 0.051$ in $t = 2$ and $t = 3$). This effect is strong enough to generate substantial electricity spillovers. A back-of-the-envelope calculation indicates that the change in dishwasher use decreases electricity consumption by approximately 2%.[9] For toilet flushing, we see no statistically significant effect during the intervention period or the post-intervention period $t = 2$ but do see a significant −10.82% effect ($p = 0.038$) in the post-intervention period $t = 3$.

The tap-level results in Fig. 4, in combination with baseline consumption by tap in Table C.1, allow us to decompose the hot water effect in Fig. 2 into individual taps. Focusing on intervention effects in $t = 1$, with baseline consumption for the control group in September 2019, we find that the kitchen sink is the main driver of the hot water reduction, with 57% (2.66 percentage points) of the total effect. The vanity basin and shower contribute 25% and 18% respectively. A similar exercise would be possible for cold water. Since the effects in Fig. 3 are not statistically significant, we refrain from such a decomposition for cold water.

In Online Appendix D we assess the robustness of our water consumption results in two ways. We change the statistical test to the non-parametric Mann–Whitney U test and vary our outlier exclusion criteria. The results are generally robust to these estimation choices.

Fig. 4 includes a large number of hypothesis tests. The results presented above are not corrected for multiple testing. We investigate multiple testing in Section 6.1 and find that the tap-level results have to be taken with caution, as they are not robust to multiple testing corrections.

[8] Table C.5 in the Online Appendix shows the effects on total water consumption.

[9] A Swiss tenant household with two people uses approximately 9.6% of its electricity consumption for the dishwasher (BFE Bundesamt für Energie, 2021b). A 20% reduction in dishwasher use thus reduces total electricity consumption by 1.92%.

Table 1
Effects on electricity.

Year	Constant	Intervention	p-value	Observations
2019	−1.37 (1.26)	−0.00 (1.36)	0.997	299
2020	−12.64 (2.99)	1.50 (3.34)	0.654	299

Note. The table displays regression estimates for the effects on electricity consumption. The first and second column report the constant terms and intervention effects respectively, with standard errors shown in parentheses. The third column lists the p-values for the intervention effects. The fourth column indicates the number of buildings in the main estimation sample for each time period.

5.2. Electricity consumption

Estimation. To estimate the effects on electricity consumption, we follow the same estimation procedure as for water consumption. We hence exclude households with the lowest 5% consumption values in the baseline year 2018. For the remaining sample, we exclude the bottom and top 1% of households based on the respective outcome variable. Our results report means and medians for the control group and the intervention group. The intervention effects are calculated as the differences between the control group and the intervention group, and we use heteroscedasticity-robust standard errors to assess the statistical significance of these estimates.

Results. Table 1 shows the results for electricity consumption in the years 2019 and 2020. The regression estimates, including the constant and the coefficient for the intervention, are detailed for each year. The main estimates in the second column represent intervention effects on electricity consumption, with p-values in the third column. We find small differences in the years 2019 (mean difference −0.00%) and 2020 (mean difference 1.50%) that are not statistically significant.

Our analysis is powered to detect a 3.4% effect on electricity consumption.[10] We do not find large electricity spillovers like the −9% in Carlsson et al. (2020) or the 5.6% in Tiefenbeck et al. (2013). Conversely, our null result is in line with Jessoe et al. (2021b), who find electricity spillovers during summer months and indirect evidence for this effect being due to decreased cooling. Our null result is consistent with this explanation as cooling accounts for only 0.1% of energy consumption in Swiss households. The bulk of energy consumption in Swiss households (67% in 2019) is due to room heating (BFE Bundesamt für Energie, 2021a). We investigate potential room heating spillovers in the next subsection.

In Online Appendix D, we conduct robustness checks using the Mann–Whitney U test and for different outlier exclusion criteria. Again, we find no evidence for electricity spillovers.

5.3. Heating energy consumption

Estimation. We use a simple regression framework on the level of cost center c to regress the change in heating energy consumption ($dHE_{c,2020}$ or $dHE_{c,2021}$) on the intervention share ($Intervention_c$):

$$dHE_{c,2020} = \alpha_h + \beta_h Intervention_c + \epsilon_c. \tag{9}$$

We estimate Eq. (9) with OLS and calculate heteroscedasticity robust standard errors. Our preferred specification does not include additional control variables because $Intervention_c$ is exogenous by randomization. However, we also provide estimates from two specifications with additional control variables.

[10] We use the following numbers to calculate the minimum detectable effect. The outcome variable $dE_{b,2019}$ has a standard deviation of 8.3 after excluding outliers. Our sample includes 58 control buildings and 141 intervention buildings. We use standard values for power (0.8) and significance level (0.05). Employing the standard formula, we obtain a minimum detectable effect of 3.4. Hence, we lack adequate power to detect a potential −1.92% effect on electricity use that would result from reduced dishwasher use (see footnote 9).

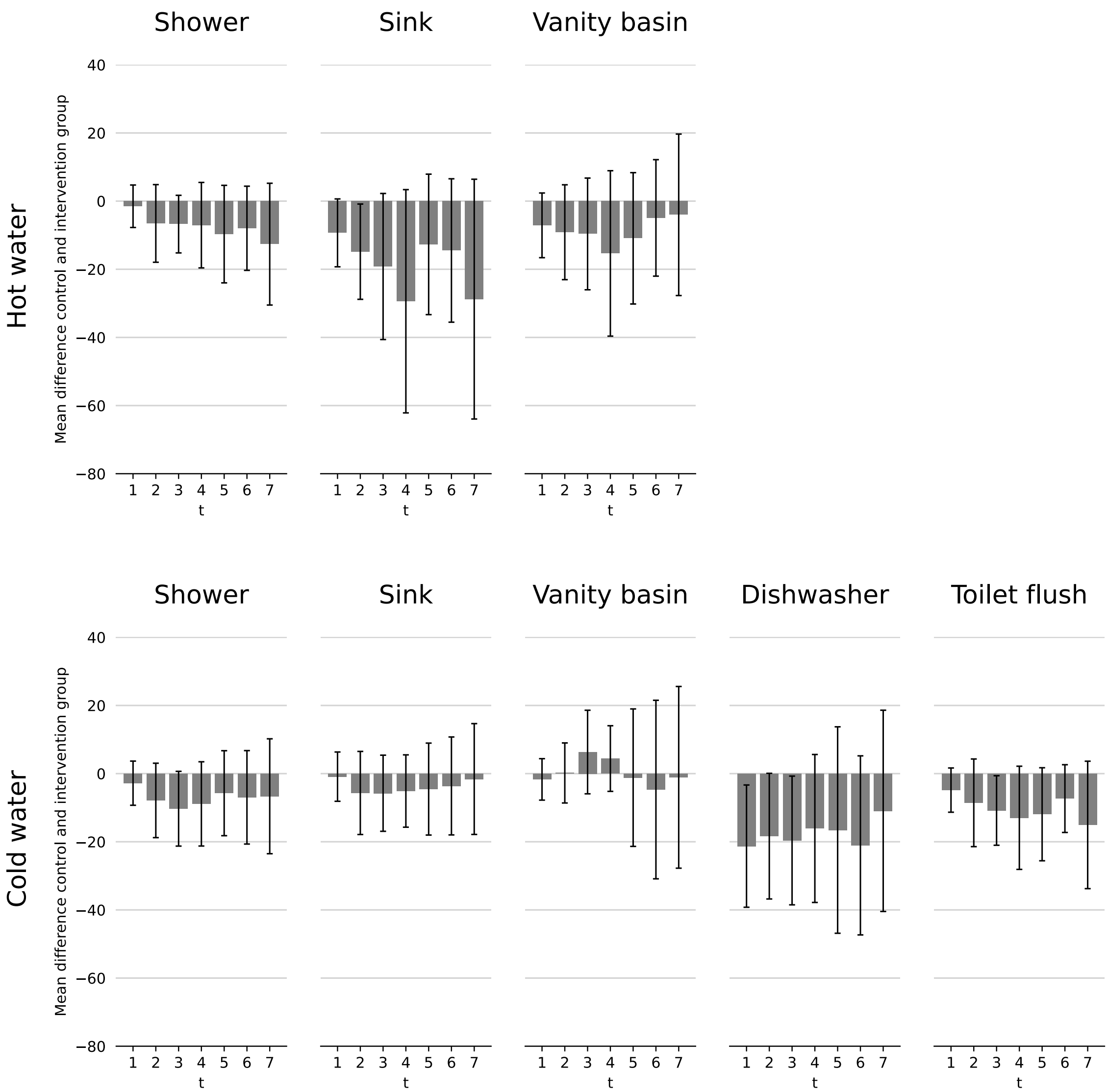

Fig. 4. Tap-level results.
Note. The figure summarizes intervention effects in the sub-sample with tap-level data. It shows the effects on hot and cold water taps for the intervention period ($t = 1$) and the post-intervention periods ($t = 2$ to $t = 7$), where each period t represents an average over four months. Detailed results are available in Tables C.6–C.13 in the Online Appendix. Bars correspond to the estimated intervention effects. The error bars represent 95% confidence intervals.

As discussed in Section 4, the cost center data include both heating energy for room heating and the preparation of hot water. Even if households do not change their use of heating energy for room heating, the effect on hot water implies $\beta_h < 0$. We denote the coefficient we would expect in the absence of an effect on room heating as β_0. Our null hypothesis is $H_0 : \beta_h - \beta_0 = 0$.

We quantify β_0 using the effect of the intervention on hot water consumption. As a first step, we calculate the hot water effect for the part of the billing period 2020 that is potentially influenced by the intervention. Using the approach from Section 5.1, we find a −5.45% effect (standard error: 1.35) on hot water consumption from October 2019 to March 2020. For the billing period 2021, we find a −6.71% effect (standard error: 2.06). These effects are relative to the baseline month September 2019.

At baseline, households in the analyzed cost centers use 10,092 m^3 of hot water in aggregate. According to the Metering Company, one m^3 of hot water requires 71 kWh of energy.[11] Hence, households use 716,530 kWh of energy for hot water in the baseline month. The hot water energy savings then add up to 234,305 kWh for the 2020 billing period and 567,950 kWh for the 2021 billing period. Putting the estimated hot water energy savings in relation to the baseline energy consumption in the sample of cost centers (71,817 MWh), we obtain $\beta_0 = -0.33\%$ for the 2020 billing period and $\beta_0 = -0.80\%$ for the 2021 billing period.

Importantly, β_0 is measured with error σ_0. The standard error of the hot water effect translates into a standard error σ_0 of 0.08 for the billing period 2020. For the billing period 2021, we obtain $\sigma_0 = 0.25$. To test our null hypothesis, we calculate the t-statistic $t = \frac{\beta_h - \beta_0}{\sqrt{\sigma_h^2 + \sigma_0^2}}$ and report p-values.

Results. Table 2 shows the results of our heating energy analysis. Column (1) shows OLS coefficients for our preferred specification that

[11] This value is based on a circulation loss factor of 1.25 and a temperature difference between hot and cold water of 49° C. It is somewhat outdated as the Metering Company uses a lower value (65.4 kWh per m^3) for modern heating systems that tend to provide lower hot water temperatures. We use the higher value as a conservative choice.

Table 2
Effects on heating energy.

	Billing period 2020			Billing period 2021		
	(1)	(2)	(3)	(4)	(5)	(6)
Intervention	−5.96	−6.45	−9.61	−6.72	−6.84	−9.56
	(2.43)	(2.34)	(3.44)	(3.29)	(3.25)	(4.88)
	[0.021]	[0.009]	[0.007]	[0.074]	[0.065]	[0.074]
Billing period April		−2.45	−0.82		7.06	22.77
		(3.00)	(18.55)		(3.41)	(10.10)
Billing period May		−5.76	−11.30		0.89	−5.59
		(1.86)	(3.89)		(1.80)	(4.33)
Billing period June		−11.49	−12.61		2.02	−7.59
		(2.17)	(3.51)		(3.02)	(6.97)
Renewable heating		−3.17	−10.24		−7.58	−8.81
		(1.75)	(3.57)		(2.02)	(5.18)
April x Intervention			−1.48			−17.49
			(19.17)			(12.01)
May x Intervention			7.13			8.26
			(4.53)			(5.09)
June x Intervention			1.63			13.07
			(4.78)			(7.97)
Ren. heating x Intervention			8.97			1.72
			(3.93)			(5.93)
Constant	7.81	10.82	13.28	14.47	15.46	17.58
	(2.38)	(2.63)	(3.46)	(2.97)	(3.19)	(4.47)
Observations	300	300	300	297	297	297

Note. The table reports coefficients from OLS regressions. Column (1) reports regression coefficients of $dHE_{c,2020}$ (the change in heating energy consumption from the 2019 billing period to the 2020 billing period) on $Intervention_c$ (the share of intervention households in the cost center). The specification in column (2) adds indicators for different billing periods (reference category March). The specification in column (3) adds interactions between these billing periods and $Intervention_c$. Column (4) reports regression coefficients of $dHE_{c,2021}$ (the change in heating energy consumption from the 2019 billing period to the 2021 billing period) on $Intervention_c$. The specification in column (5) adds indicators for different billing periods (reference category March). The specification in column (6) adds interactions between these billing periods and $Intervention_c$. Standard errors are reported in parentheses. The values in brackets represent p-values from tests of the main coefficient against β_0. All results are evaluated on the cost center level.

regresses $dHE_{c,2020}$ on $Intervention_c$. The values in parentheses represent standard errors, and the values in brackets represent p-values from tests of the main coefficient against β_0. The intervention reduces heating energy consumption by 5.96%. As discussed above, we do not test the coefficient on $Intervention_c$ against zero but against −0.33. The effect on room heating is significant with a p-value of 0.021.

The specification in column (2) of Table 2 includes control variables for billing periods, with March as the reference category, and an indicator for renewable heating.[12] As compared to our preferred specification, the coefficient of interest in column (2) is slightly larger at −6.45; this estimate is different from −0.33 with $p < 0.01$. The specification in column (3) adds interaction terms of the control variables in column (2) and $Intervention_c$. The coefficient of interest (−9.61) is larger than the coefficients in columns (1) and (2) and is statistically significant against −0.33 ($p < 0.01$). This effect refers to cost centers with fossil energy sources and a March billing period. All billing period interactions are statistically insignificant, but the interaction term for renewable heating is notable. With a coefficient of 8.97, it suggests that the intervention hardly changes heating energy consumption in cost centers with renewable heating systems. Conversely, the effect on heating energy seems to be driven by households that rely on fossil fuels.

Columns (4) to (6) in Table 2 reports results for the 2021 billing period. The room heating spillover appears to be remarkably persistent. The coefficient in column (4) indicates that the intervention reduces heating energy consumption in the post-intervention 2021 billing period by 6.72%. Again, we take the intervention's effect on hot water into account, with $\beta_0 = -0.80\%$. The test against this null hypothesis yields a p-value of 0.074. Columns (5) and (6) add the same control variables used in columns (2) and (3). The results are quantitatively similar to those columns and marginally significant, with p-values of 0.065 and 0.074, respectively.

To summarize, we find strong positive spillovers of our hot water intervention on room heating. Our preferred specification turns out to be the conservative choice, with a coefficient of −5.96. Correcting for the −0.33% influence of hot water, we find a −5.63% spillover effect on energy consumption for room heating. The room heating spillover appears to persist one year later. Correcting the coefficient of −6.72% for the −0.80% influence of hot water, the effect of the intervention on room heating is −5.92%.

The −5.63% effect on room heating energy is large yet not implausible. The timing of the hot water intervention at the start of the heating period may have been opportune for a spillover on room heating, and a behavioral change at this time of the year may be persistent if households do not re-adjust their thermostat. Even a small decrease in room temperature causes substantial energy savings. In the Swiss context, decreasing the indoor temperature by 1 °C saves between 6% and 10% of room heating energy (BFE Bundesamt für Energie, 2014).[13] The 5.63% reduction of room heating energy hence corresponds to a reduction in indoor temperatures of less than 1 °C.

What level of price increase would be necessary to achieve the −5.63% effect on room heating energy? The demand for room heating is price inelastic (Auffhammer and Rubin, 2018; Brewer, 2021). Auffhammer and Rubin (2018) estimate the elasticity of demand for natural gas in California between −0.23 and −0.17. Using these estimates, we find that the −5.63% room heating spillover could be roughly equivalent to a 24%–33% price increase.

In Online Appendix D, we assess the robustness of the room heating spillover in three ways. First, we exclude outliers according to the procedure we use for our analysis of water data. Second, we weight cost centers by the number of households in our sample. Third, we weight cost centers by their energy consumption in the baseline billing period 2019. These robustness checks corroborate our results.

6. Discussion

6.1. Pre-registration and multiple hypothesis testing

Our pre-analysis plan (RCT ID AEARCTR-0004995) was uploaded on November 7, 2019, before the research team got access to data on the intervention period. It is publicly available on www.socialscienceregistry.org. The pre-analysis plan includes a power simulation and pre-specifies all estimation choices for our water consumption analysis. Note that our pre-analysis plan specifies the Mann–Whitney U test. We opt for a regression framework because it provides us with confidence intervals and allows us to account for multiple hypothesis testing.

The analysis of heating energy consumption was pre-specified, but heating energy data were only obtainable on the level of the cost center and not, as originally foreseen, on the household level. The analysis of electricity consumption was not pre-specified because we did not foresee data availability.

6.2. Sample composition by outcome

We observe hot and cold water consumption for all households in our sample, but tap-level data, electricity data, and heating energy

[12] Of the 300 cost centers in our sample, 238 use fossil energy sources. The remaining 62 cost centers use renewable heating systems, predominantly district heating. In Switzerland, district heating comes mostly from waste incineration, renewables, and waste heat (Nussbaumer et al., 2021).

[13] Brown et al. (2013) use an estimate of 7% to evaluate the energy savings of an intervention in France.

data are only available for sub-samples. In this section, we discuss selection into the respective sub-samples and potential implications for the interpretation of our results.

As explained in Section 6.1, these data sources were not anticipated in the experimental design and our pre-analysis plan. The availability of these data sources allows us to investigate spillovers on additional outcomes. We make use of these opportunities, acknowledging their exploratory nature and the selective nature of the respective sub-samples.

Tap-level consumption. Tap-level meters are small devices with a display that shows the current reading. They are installed at each tap and can be read remotely. Tap-level meters are used where household-level submeters are not feasible. This is typically the case in older buildings, where the plumbing system is not apartment-specific. 231 of the 782 buildings in our sample have tap-level meters.

To assess how this sub-sample compares to our full sample, Tables C.1 and C.2 report pre-intervention household-level hot and cold water consumption values for the full sample and the sub-sample with tap-level data. The two samples appear almost identical in terms of pre-intervention hot water consumption. Cold water consumption is somewhat higher in the sub-sample with tap-level data, with the largest difference in the control group in June 2019 (5.68 m^3 in the full sample and 6.19 in the sub-sample, $p = 0.38$). The differences are not statistically significant, suggesting that the sub-sample with tap-level data is representative of the full sample.

We can also assess the intervention effects in the sub-sample with tap-level data. Tables C.14 and C.15 in the Online Appendix show estimates of the effects on hot and cold water consumption. The effects in the sub-sample are somewhat larger than the effects in the full sample, particularly in later time periods. Larger intervention effects in the sub-sample with tap-level data could be plausible due to the visibility of tap-level meters, potentially reminding households of the intervention. In any case, the differences in effect sizes between the full sample and the sub-sample are not statistically significant.

Electricity consumption. Our sample is served by 55 different electricity providers. These electricity providers are local monopolies and households cannot choose their provider. We requested data from all 55 companies, with 12 responding to our request, a response rate of 22%. These 12 companies are relatively large, as they serve 41% of the buildings in our sample (324 of 782). Unfortunately, we cannot assess water consumption in this sub-sample. The reason is that the water consumption data are fully anonymized and cannot be merged to the electricity data.

Heating energy consumption. We observe heating energy consumption for 300 of 333 cost centers in our sample. Of the remaining 33 cost centers, 13 changed their heating system to a different energy source and 20 have missing data. Again, we do not know water consumption in this sub-sample because the water consumption data are anonymized. However, with heating energy data for more than 90% of cost centers, selection into this sub-sample is not a major concern.

Main hypotheses. The pre-analysis plan documents our a priori interest in spillovers on two specific behaviors: cold water consumption and heating energy consumption (see p. 20 in the pre-analysis plan). After the end of the intervention, we learned about the availability of post-intervention data, tap-level data and electricity data. In the spirit of Banerjee et al. (2020), we make use of these opportunities to maximize knowledge gain from the field experiment. Abstracting from these unforeseen and exploratory analyses, we test two main spillover hypotheses: does the intervention change cold water consumption and heating energy consumption in the intervention period? The corresponding estimates are in Table C.4 (the first row, with a p-value of 0.218) and Table 2 (column 1, with a p-value of 0.021). We employ the Bonferroni correction (Bonferroni, 1936; Dunn, 1961) and multiply these p-values by two, i.e., the number of hypotheses. The Bonferroni correction is simple, but particularly conservative, as it ignores the correlation between test statistics. The resulting p-values 0.436 (cold water) and 0.042 (room heating) do not change our conclusions.

A critical reader may argue that the effect on hot water consumption, the target behavior, should count as a third hypothesis. We chose not to, because this study is primarily interested in behavioral spillovers. The adjusted p-value for the room heating spillover is 0.063 if we apply the Bonferroni correction with three hypotheses.

Water consumption outcomes. The exploratory water consumption analysis in Section 5.1 tests a large number of hypotheses. Figs. 2, 3, and 4 assesses ten behaviors over seven time periods, for a total of 70 hypothesis tests. Since all these results come from one dataset on the building level, we can use the method proposed by List et al. (2019) and the implementation provided by Steinmayr (2020) to account for multiple hypothesis testing.

We find that the intervention effect on the target outcome hot water is robust to multiple hypothesis testing, while the tap-level results are generally not. If we focus only on the intervention period $t = 1$, we find that the hot water reduction remains significant with $p < 0.01$, while cold water and all tap-level results do not. Looking at individual outcomes over time, we find hot water consumption to be significantly reduced in $t = 1$ ($p < 0.01$) and $t = 4$ $p = 0.048$, but not in other periods. No other outcome is significant in any period, once multiple hypothesis tests over time are accounted for. Finally, we account for multiple hypothesis tests across outcomes and time periods. The result for hot water consumption in $t = 1$ remains highly statistically significant ($p < 0.01$), while the remaining 69 hypothesis tests are not significant at the 10% level.

6.3. Spillover mechanisms

In this section, we discuss which mechanisms may give rise to the results reported in Section 5. A better understanding of spillover mechanisms offers practical insights that can be leveraged to design more effective interventions in the future. This is particularly important for researchers, policy makers, and practitioners aiming to replicate the success of the intervention in different contexts. In light of our theoretical framework outlined in Section 2, spillovers may result from three distinct mechanisms.[14] We discuss each mechanism in turn.

Complementarities. Mixer taps use both hot and cold water, allowing us to empirically investigate the role of complementarities. Hot and cold water consumption are tightly linked in shower, kitchen sink, and vanity basin use. Swiss authorities prescribe a minimum hot water temperature of 50 °C to prevent legionellosis, an infection caused by legionella bacteria (BAG Bundesamt für Gesundheit and BLV Bundesamt für Lebensmittelsicherheit und Veterinärwesen, 2018). Modern mixer taps conveniently mix hot and cold water to regulate the water temperature; we refer to this mixture as warm water.

Hot water consumption can be reduced in two ways. First, households can use warm water at a lower temperature. They may, for instance, wash their hands at 34 °C instead of 40 °C. However, this behavior would also increase cold water consumption—a negative spillover.[15] Second, households may reduce the quantity of warm water

[14] Behavioral spillovers can be explained by mechanisms that are not included in our theoretical framework. In Online Appendix F, we discuss imperfect procedural knowledge, physical investments, and cognitive spillovers in light of our results.

[15] Assuming 10 °C for cold water and 55 °C for hot water, using water at 34 °C instead of 40 °C decreases the hot water share from 67% to 53%. The use of cold water increases accordingly. The hot water share is the share of hot water in the mix of hot water (at temperature T_{hot}) and cold water (at temperature T_{cold}). For water at temperature T, the hot water share is calculated as $\frac{T-T_{cold}}{T_{hot}-T_{cold}}$.

used rather than its temperature. This can be achieved, for instance, by showering for four minutes instead of five while keeping the temperature constant. Doing so decreases both hot water and cold water consumption by exactly 20%, implying a positive spillover from the hot water reduction on cold water consumption.

Hot and cold water are likely complementary in the shower. Tiefenbeck et al. (2018) provide households with real-time feedback during showering and measure their shower time and temperature. They find a 21% decrease in shower time but only a 0.3 °C reduction in temperature. These findings indicate that hot and cold water are strong complements—households appear to resist deviations from their preferred shower temperature. Our results confirm this assessment as we find negative intervention effect estimates for cold water in the shower.

As compared to the shower, hot and cold water may be more substitutable in kitchen sink and vanity basin use. Tap producers even sell taps with a "cold start" feature, which provides cold water in the mixer lever's default position (Nording and Bennich, 2021). The popularity of this cold water default indicates that hot and cold water are more substitutable in kitchen sink and vanity basin use than in shower use. We find no evidence for substitutability. If anything, hot and cold water may be substitutable for vanity basin use, where we see occasional and statistically insignificant increases in cold water consumption.

Direct spillovers or self-image? As elaborated in Section 2, the persistence of behavioral spillovers informs the distinction between direct spillovers and self-image spillovers. Direct spillovers are expected to be transient and positive, mirroring the effect on the target behavior. The persistence of self-image spillovers depends on the sign of ρ, where $\rho > 0$ implies a transient negative spillover and $\rho < 0$ implies a positive spillover that builds up over time. We focus the discussion of direct spillovers and self-image on dishwasher use, toilet flushing, and room heating, where complementarities hardly apply.[16]

Our findings suggest persistent spillovers. The sizable reduction in dishwasher use is statistically significant until the post-intervention period $t = 3$, but we cannot rule out similarly large effects in the subsequent periods. The spillover effect in toilet flushing is not statistically significant in most periods, but the point estimates are consistent with a persistent effect. Finally, the room heating spillover appears fully persistent one year after the intervention. While our study does not provide a rigorous test of spillover mechanisms, the time patterns in our results are in line with self-image spillovers with $\rho < 0$, where households strive to be consistent with their environmental self-image.

6.4. Welfare implications

The spillover on room heating may have considerable welfare implications. In this section, we provide welfare estimates using the marginal value of public funds (MVPF) approach (Hendren and Sprung-Keyser, 2020, 2022) and Cost Effectiveness Analysis (CEA).

We define the MVPF as:

$$\text{MVPF} = \frac{\Delta W}{\Delta E - \Delta C} = \frac{\eta_l \Delta W_l + \eta_g \Delta W_g + \eta_h \Delta W_h + \eta_p \Delta W_p}{\Delta E - \Delta C_{\text{CO}_2} - \Delta C_{VAT}} \quad , \tag{10}$$

where ΔW are the benefits provided to the population, ΔE is the government's expenditure on the intervention, and ΔC is the long-run reduction in government costs that is caused by the intervention. In our context, ΔW includes changes in local (ΔW_l) and global (ΔW_g) emission reductions, household utility (ΔW_h), and profits of utility companies (ΔW_p). η_l, η_g, η_h, and η_p represent the social marginal utility of income of the respective population groups. ΔC includes effects on CO_2 levy revenues (ΔC_{CO_2}) and VAT revenues (ΔC_{VAT}).

Our welfare calculations cover a period of 12 months; we acknowledge that this simplification ignores longer-term effects of the intervention. We account for *local* environmental externalities (ΔW_l) resulting from NO_x, SO_2, and NMVOCs using cost factors from the integrated impact assessment model EcoSense (Schmid and Im, 2019). We further account for *global* environmental externalities (ΔW_g) from CO_2 emissions, using a social cost of carbon of USD 190 per tonne of CO_2 (EPA, 2023). Households' utility does not change as long as households are indifferent between their consumption choice without the intervention and the lower consumption choice with the intervention.[17] We follow Jessoe et al. (2021b) and use wholesale prices to approximate profits of utility companies and estimate ΔW_p. ΔW_{CO_2} includes one third of the CO_2 levy (CHF 96 per tCO_2 at the time), as two thirds of the revenue are redistributed. ΔW_{VAT} accounts for VAT (2.6% on water, 7.7% on heating fuels). All monetary values are in CHF. USD/CHF traded close to parity during the intervention period. Online Appendix G describes the assumptions behind our welfare estimates in detail.

Table 3 presents our MVPF estimates. Column (1) focuses on the target behavior hot water. During the 12 months after the start of the intervention, the average household's reduction of local externalities is valued at CHF 0.20. The reduction of global externalities is valued at CHF 3.44. The hot water savings reduce profits of utility companies by an estimated CHF 8.41. We estimate that repeating the intervention in a similar sample would cost CHF 12.63 per household. The hot water effect reduces CO_2 levy revenues (excluding lump-sum refunds) by 0.58 and VAT revenues by CHF 0.75. The resulting MVPF, ignoring the spillover on room heating, is negative at -0.34. The 95% confidence interval ranges from -0.50 to -0.17.[18]

The spillover effect is depicted in column (2) of Table 3. Room heating needs more energy than hot water heating, especially in buildings with low energy efficiency. The average household's reduction in room heating implies large benefits from reduced local (CHF 2.25) and global (CHF 38.91) externalities, but also sizable losses in profits (CHF 66.19), CO_2 levies (6.55) and VAT revenue (CHF 7.49). Column (3) shows that overall, including effects on the target behavior and spillover effects, we obtain an MVPF of CHF -1.06 (with a confidence interval between -1.28 and -0.53).

We find a negative MVPF, especially when we account for the spillover on room heating. In our setting, per-unit profits are larger

[16] Toilet flushing does not use hot water. Also, the dishwasher uses cold water (heated with electricity), but this may not be obvious. Households may mistakenly believe that the dishwasher uses hot water, a case we discuss in Online Appendix F. Room heating is independent of hot water consumption. Although both hot water and room heating may be provided by the same energy source, they are distributed via distinct systems. Hot water is heated in a boiler and is distributed to the apartment's faucets. Once used, it leaves the apartment through the sewer system. Room heating is delivered through an entirely separate system. A typical heating system heats up water to send heat to the building's radiators, floor heating, or wall heating. The medium cools down as it warms up the housing space, returns in a circulation system, and is heated up again.

[17] This assumption is motivated by our theoretical framework in Section 2, where the intervention increases the moral price of energy consumption. In this case, the intervention causes psychological costs, that may be (partly or fully) offset by monetary savings. The pure information aspect of our intervention is not modeled in our theoretical framework, but may increase welfare for misinformed households. The cross-randomized features of our experimental design, described in Section 3.2, are relevant in this discussion. Depending on the experimental group, households receive social comparison information or not; a standard lottery or a regret lottery. Social comparison and the regret lottery may induce moral or psychological costs. This is important to consider, given that different types of nudges, working through different behavioral mechanisms, induce heterogeneity in welfare effects (Sánchez et al., 2022). The assumption $\Delta W_h = 0$ is an approximation that may be reasonable on average, i.e., across experimental conditions.

[18] To calculate confidence intervals, we simulate 10,000,000 independent draws from the estimated effects and their standard errors for hot water (Table C.3) and room heating (Table 2). We calculate the MVPF (or CEA) for each sample and report the 5% and 95% quantiles.

Table 3
Marginal value of public funds.

	(1) Hot water	(2) Heating	(3) Total
Population			
Local externalities ΔW_l	0.20	2.25	2.45
Global externalities ΔW_g	3.44	38.91	42.35
Profits ΔW_p	−8.41	−66.19	−74.60
Government			
Expenditures ΔE	12.63		12.63
CO_2 levy revenue ΔC_{CO_2}	−0.58	−6.55	−7.13
VAT revenue ΔC_{VAT}	−0.75	−7.49	−8.24
Marginal value of public funds			
$\eta_l = \eta_g = \eta_p = 1$	−0.34		−1.06
	[−0.50, −0.17]		[−1.28, −0.53]
$\eta_l = \eta_g = 1; \eta_p = 0$	0.26		1.60
	[0.13, 0.38]		[0.62, 2.00]

Note. The table reports marginal value of public funds (MVPF) estimates. The values in brackets represent 95% confidence intervals.

than per-unit externalities. This means that the MVPF is negative for any consumption reduction and any (positive) intervention cost. Taken to the extreme, an intervention that decreases energy consumption at infinitesimal costs to government has a MVPF of minus infinity. This observation raises questions about the consideration of profits in our welfare evaluation, and the social marginal utilities of income η_l, η_g, and η_p.

ΔW comprises effects on different population groups. Local externalities affect the population living in the vicinity of the buildings; global externalities affect the world population; profits affect the shareholders of natural gas and heating oil suppliers. Social marginal utility of income may differ across these groups, for two reasons. First, the marginal utility of income may be lower for high-income households. Second, a policy maker may apply different social welfare weights to the different population groups. Social welfare weights are ultimately a political question.

A policy maker may propose $\eta_p = 0$, as decarbonization implies that the profits of the fossil fuel industry go to zero. Table 3 also shows the corresponding MVPF estimates. The last row in column (1) does not account for profit losses. Ignoring the room heating spillover, the MVPF is 0.26. This value is positive but clearly smaller than 1. The MVPF in column (3) accounts for the room heating spillover and finds an MVPF of 1.60, with a confidence interval between 0.62 and 2.00. This finding suggests that the intervention improves welfare when policy makers put a low weight on profit losses *and* take the room heating spillover into account.

We also consider a cost effectiveness analysis (CEA) to compare the costs of the intervention to the abatement of CO_2, NO_x, SO_2, and NMVOC. This approach may be useful for private decision makers like the Owning Company, who do not necessarily consider effects on the government's budget or profits of utility companies. Table 4 shows the CEA estimates. We focus on CO_2, the primary externality in our setting. The intervention costs CHF 12.63 per household. If we only consider the effect on the target behavior hot water in column (1), the intervention saves 0.018 tonnes of CO_2 for the average household. Hence, if we ignore the room heating spillover, the intervention costs CHF 698 per tonne of CO_2 abatement. The room heating spillover in column (2) saves 0.205 tonnes of CO_2. The total effect in column (3) is 0.223 tonnes of CO_2 abatement, with a cost of CHF 57 per tonne. The confidence interval is relatively wide, but this value is arguably lower than the social cost of carbon (EPA, 2023; Rennert et al., 2022).

7. Conclusions

This paper presents a large-scale field experiment to measure spillover effects of a behavioral intervention in the environmental domain.

Table 4
Cost effectiveness analysis.

	(1) Hot water	(2) Heating	(3) Total
Costs			
Expenditures ΔE	12.63		12.63
Benefits			
CO_2 abatement [t]	0.018	0.205	0.223
NO_x abatement [kg]	0.006	0.067	0.073
SO_2 abatement [kg]	0.001	0.010	0.011
NMVOC abatement [kg]	0.001	0.009	0.009
Cost effectiveness			
CO_2 [CHF/t]	698		57
	[452, 1,518]		[31, 219]
NO_x [CHF/kg]	2,131		173
	[1,381, 4,637]		[95, 760]
SO_2 [CHF/kg]	14,139		1,148
	[9,161, 30,768]		[630, 4,447]
NMVOC [CHF/kg]	16,372		1,330
	[10,605, 35,620]		[729, 5,149]

Note. The table reports cost effectiveness analysis (CEA) estimates. The values in brackets represent 95% confidence intervals.

The intervention exclusively focuses on hot water consumption, but its impact extends beyond this target behavior. Unique tap-level data suggest potential spillovers on cold water consumption in dishwasher use and toilet flushing. We find no electricity spillover but find a large positive spillover on energy consumption for room heating, which was reduced by 5.63%. The room heating spillover implies substantial utility bill savings and reductions in environmental externalities. Taking it into account has important consequences for welfare evaluation.

A growing literature investigates behavioral spillovers. The emerging evidence indicates that behavioral interventions in the environmental domain may have positive side effects on other resource-related behaviors. Concerns about negative side effects of energy-efficiency policies (Gillingham et al., 2013) may indeed be unfounded, but more evidence is needed. What are the spillover effects of the most widely used economic interventions? Where should we be concerned about negative spillovers, and where can we count on positive spillovers to reinforce interventions? Answering the broad question of external validity may allow us to reveal the conditions under which positive or negative spillovers prevail.

Further insights into the mechanisms behind behavioral spillovers may facilitate a deeper understanding of economic interventions. Our theoretical framework distinguishes three broad spillover mechanisms. Empirical tests require rich data, such as the multiple spillover behaviors over long time periods in this study. Future research may evaluate spillovers in different settings, perhaps with high frequency data, and study entirely new target and spillover behaviors. In many ways, we are just starting to understand the side effects of economic interventions.

Declaration of competing interest

The research was supported by the National Research Programme 73 (project number: 407340_172431).

The research received in-kind support by the company that implemented the field experiment and provided the data, as well as by the company that owns the properties that were subject to the study. As clearly defined in a collaboration contract, both companies had the right to review the manuscript for confidential information, but had no role in the analysis of the data and the preparation of the manuscript.

The authors declare no conflict of interest.

Data availability

Data available upon request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jpubeco.2024.105166.

References

Adena, M., Huck, S., 2019. Giving once, giving twice: A two-period field experiment on intertemporal crowding in charitable giving. J. Pub. Econ. 172, 127–134.
Alacevich, C., Bonev, P., Söderberg, M., 2021. Pro-environmental interventions and behavioral spillovers: Evidence from organic waste sorting in Sweden. J. Environ. Econ. Manag. 108.
Allcott, H., 2011a. Consumers' perceptions and misperceptions of energy costs. Amer. Econ. Rev. 101 (3), 98–104.
Allcott, H., 2011b. Social norms and energy conservation. J. Pub. Econ. 95 (9–10), 1082–1095.
Allcott, H., Kessler, J.B., 2019. The welfare effects of nudges: A case study of energy use social comparisons. Am. Econ. J.: Appl. Econ. 11 (1), 236–276.
Allcott, H., Rogers, T., 2014. The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. Amer. Econ. Rev. 104 (10), 3003–3037.
Altmann, S., Grunewald, A., Radbruch, J., 2021. Interventions and cognitive spillovers. Rev. Econ. Stud. rdab087.
Athey, S., Imbens, G.W., 2017. The econometrics of randomized experiments. In: Handbook of Economic Field Experiments, vol. 1, Elsevier, pp. 73–140.
Auffhammer, M., Rubin, E., 2018. Natural Gas Price Elasticities and Optimal Cost Recovery Under Consumer Heterogeneity: Evidence from 300 Million Natural Gas Bills. Working Paper, National Bureau of Economic Research.
BAG Bundesamt für Gesundheit, BLV Bundesamt für Lebensmittelsicherheit und Veterinärwesen, 2018. Legionellen und Legionellose. BAG-/BLV-Empfehlungen.
Banerjee, A., Duflo, E., Finkelstein, A., Katz, L.F., Olken, B.A., Sautmann, A., 2020. In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for Rcts in Economics. Technical Report, National Bureau of Economic Research.
Becker, G.S., Murphy, K.M., 1988. A theory of rational addiction. J. Polit. Econ. 96 (4), 675–700.
Belfield, C.R., Nores, M., Barnett, S., Schweinhart, L., 2006. The High/Scope Perry Preschool Program: Cost–benefit analysis using data from the age-40 followup. J. Hum. Resour. 41 (1), 162–190.
Bénabou, R., Tirole, J., 2011. Identity, morals, and taboos: Beliefs as assets. Q. J. Econ. 126 (2), 805–855.
Bernedo, M., Ferraro, P.J., Price, M., 2014. The persistent impacts of norm-based messaging and their implications for water conservation. J. Consum. Policy 37 (3), 437–452.
BFE Bundesamt für Energie, 2014. Heizen mit Köpfchen. Bundespublikationen.
BFE Bundesamt für Energie, 2021a. Der Energieverbrauch der Privaten Haushalte 2000–2020.
BFE Bundesamt für Energie, 2021b. Stromverbrauch eines typischen Haushalts.
Blanken, I., van de Ven, N., Zeelenberg, M., 2015. A meta-analytic review of moral licensing. Pers. Soc. Psychol. Bull. 41 (4), 540–558.
Blau, D.M., 2016. Pensions, household saving, and welfare: A dynamic analysis of crowd out. Quant. Econ. 7 (1), 193–224.
Bonferroni, C.E., 1936. Teoria statistica delle classi e calcolo delle probabilita. Pubbl. R Ist. Super. Sci. Econ. Commericiali Firenze 8, 3–62.
Bradley, P., Coke, A., Leach, M., 2016. Financial incentive approaches for reducing peak electricity demand, experience from pilot trials with a UK energy provider. Energy Policy 98, 108–120.
Brandon, A., Ferraro, P.J., List, J.A., Metcalfe, R.D., Price, M.K., Rundhammer, F., 2017. Do the Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments. Working Paper 23277, National Bureau of Economic Research.
Brent, D.A., Cook, J.H., Olsen, S., 2015. Social comparisons, household water use, and participation in utility conservation programs: Evidence from three randomized trials. J. Assoc. Environ. Resour. Econ. 2 (4), 597–627.
Brewer, D., 2021. Household Responses to Winter Heating Costs: The Remarkably Inelastic Demand for Space Heating. Working Paper.
Bronchetti, E.T., Kessler, J.B., Magenheim, E.B., Taubinsky, D., Zwick, E., 2020. Is Attention Produced Rationally? Working Paper 27443, National Bureau of Economic Research.
Brown, Z., Johnstone, N., Haščič, I., Vong, L., Barascud, F., 2013. Testing the effect of defaults on the thermostat settings of OECD employees. Energy Econ. 39, 128–134.
Butera, L., Metcalfe, R., Morrison, W., Taubinsky, D., 2022. Measuring the welfare effects of shame and pride. Amer. Econ. Rev. 112 (1), 122–168.
Byrne, D.P., Goette, L., Martin, L.A., Miles, A., Jones, A., Schob, S., Staake, T., Tiefenbeck, V., 2022. The Habit Forming Effects of Feedback: Evidence from a Large-Scale Field Experiment. Working Paper 3974371, SSRN.
Carfora, V., Caso, D., Sparks, P., Conner, M., 2017. Moderating effects of pro-environmental self-identity on pro-environmental intentions and behaviour: A multi-behaviour study. J. Environ. Psychol. 53, 92–99.
Carlsson, F., Jaime, M., Villegas, C., 2020. Behavioral spillover effects from a social information campaign. J. Environ. Econ. Manag.
Chetty, R., Friedman, J.N., Leth-Petersen, S., Nielsen, T.H., Olsen, T., 2014. Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark. Q. J. Econ. 129 (3), 1141–1219.
Choukhmane, T., 2024. Default options and retirement saving dynamics. Working Paper.
Cialdini, R.B., Trost, M.R., Newsom, J.T., 1995. Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. J. Pers. Soc. Psychol. 69 (2), 318.
Costa, F., Gerard, F., 2021. Hysteresis and the welfare effect of corrective policies: Theory and evidence from an energy-saving program. J. Polit. Econ. 129 (6), 1705–1743.
Damgaard, M.T., Gravert, C., 2018. The hidden costs of nudging: Experimental evidence from reminders in fundraising. J. Pub. Econ. 157, 15–26.
DellaVigna, S., List, J.A., Malmendier, U., 2012. Testing for altruism and social pressure in charitable giving. Q. J. Econ. 127 (1), 1–56.
Deryugina, T., Marx, B.M., 2021. Is the supply of charitable donations fixed? Evidence from deadly tornadoes. Am. Econ. Rev.: Insights 3 (3), 383–398.
Dolan, P., Galizzi, M.M., 2015. Like ripples on a pond: Behavioral spillovers and their implications for research and policy. J. Econ. Psychol. 47, 1–16.
Donkers, B., van Diepen, M., Franses, P.H., 2017. Do charities get more when they ask more often? Evidence from a unique field experiment. J. Behav. Exp. Econ. 66, 58–65.
Dunlap, R.E., Van Liere, K.D., Mertig, A.G., Jones, R.E., 2000. New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale. J. Soc. Issues 56 (3), 425–442.
Dunn, O.J., 1961. Multiple comparisons among means. J. Amer. Statist. Assoc. 56 (293), 52–64.
Effron, D.A., Cameron, J.S., Monin, B., 2009. Endorsing Obama licenses favoring Whites. J. Exp. Soc. Psychol. 45 (3), 590–593.
Ek, C., Miliute-Plepiene, J., 2018. Behavioral spillovers from food-waste collection in Swedish municipalities. J. Environ. Econ. Manag. 89, 168–186.
EPA, 2023. Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances. Technical Report, U.S. Environmental Protection Agency.
Ferraro, P.J., Miranda, J.J., Price, M.K., 2011. The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. Am. Econ. Rev.ew: Pap. Proc. 101 (3), 318–322.
Festinger, L., 1962. A Theory of Cognitive Dissonance, vol. 2, Stanford University Press.
Fryer, Jr., R.G., Katz, L.F., 2013. Achieving escape velocity: Neighborhood and school interventions to reduce persistent inequality. Amer. Econ. Rev. 103 (3), 232–237.
Gabaix, X., 2014. A sparsity-based model of bounded rationality. Q. J. Econ. 129 (4), 1661–1710.
Galizzi, M.M., Whitmarsh, L., 2019. How to measure behavioral spillovers: A methodological review and checklist. Front. Psychol. 10 (342).
Gawronski, B., 2012. Back to the future of dissonance theory: Cognitive consistency as a core motive. Soc. Cogn. 30 (6), 652–668.
Gillingham, K., Kotchen, M.J., Rapson, D.S., Wagner, G., 2013. The rebound effect is overplayed. Nature 493 (7433), 475–476.
Grieder, M., Schmitz, J., Schubert, R., 2021. Asking to Give: Moral Licensing and Pro-Social Behavior in the Aggregate. Working Paper 3920355, SSRN.
Hahn, R., Metcalfe, R.D., Novgorodsky, D., Price, M.K., 2016. The Behavioralist as Policy Designer: The Need to Test Multiple Treatments to Meet Multiple Targets. Technical Report, National Bureau of Economic Research.
Hall, J.D., Madsen, J., 2021. Can Behavioral Interventions Be Too Salient? Evidence from Traffic Safety Messages. Working Paper 3633014, SSRN.
Harding, M., Hsiaw, A., 2014. Goal setting and energy conservation. J. Econ. Behav. Organ. 107, 209–227.
Harrison, G.W., List, J.A., 2004. Field experiments. J. Econ. Lit. 42 (4), 1009–1055.
Heckman, J.J., 2010. Building bridges between structural and program evaluation approaches to evaluating policy. J. Econ. Lit. 48 (2), 356–398.
Heckman, J.J., Moon, S.H., Pinto, R., Savelyev, P.A., Yavitz, A., 2010. The rate of return to the HighScope Perry Preschool Program. J. Pub. Econ. 94 (1–2), 114–128.
Heckman, J.J., Smith, J.A., 1998. Evaluating the welfare state.
Hendren, N., Sprung-Keyser, B., 2020. A unified welfare analysis of government policies. Q. J. Econ. 135 (3), 1209–1318.
Hendren, N., Sprung-Keyser, B., 2022. The Case for Using the MVPF in Empirical Welfare Analysis. Technical Report, National Bureau of Economic Research.
Hicks, J.R., Allen, R.G., 1934. A reconsideration of the theory of value. Part I. Economica 1 (1), 52–76.
Jessoe, K., Lade, G.E., Loge, F., Spang, E., 2021a. Residential water conservation during drought: Experimental evidence from three behavioral interventions. J. Environ. Econ. Manag. 110, 102519.
Jessoe, K., Lade, G., Loge, F., Spang, E., 2021b. Spillovers from behavioral interventions: Experimental evidence from water and energy use. J. Assoc. Environ. Resour. Econ. 8 (2), 315–346.
Jimenez-Gomez, D., 2018. Nudging and Phishing: A Theory of Behavioral Welfare Economics. Working Paper 3248503, SSRN.
Kling, J.R., Liebman, J.B., Katz, L.F., 2007. Experimental analysis of neighborhood effects. Econometrica 75 (1), 83–119.

List, J.A., Shaikh, A.M., Xu, Y., 2019. Multiple hypothesis testing in experimental economics. Exp. Econ. 22, 773–793.
Ludwig, J., Duncan, G.J., Gennetian, L.A., Katz, L.F., Kessler, R.C., Kling, J.R., Sanbonmatsu, L., 2013. Long-term neighborhood effects on low-income families: Evidence from Moving to Opportunity. Amer. Econ. Rev. 103 (3), 226–231.
Ludwig, J., Liebman, J.B., Kling, J.R., Duncan, G.J., Katz, L.F., Kessler, R.C., Sanbonmatsu, L., 2008. What can we learn about neighborhood effects from the Moving to Opportunity experiment? Am. J. Sociol. 114 (1), 144–188.
Maki, A., Carrico, A.R., Raimi, K.T., Truelove, H.B., Araujo, B., Yeung, K.L., 2019. Meta-analysis of pro-environmental behaviour spillover. Nat. Sustain. 2 (4), 307–315.
Martin, C., Czellar, S., 2016. The extended inclusion of nature in self scale. J. Environ. Psychol. 47, 181–194.
Medina, P.C., 2021. Side effects of nudging: Evidence from a randomized intervention in the credit card market. Rev. Financ. Stud. 34 (5), 2580–2607.
Meer, J., 2017. Does fundraising create new giving? J. Pub. Econ. 145, 82–93.
Merritt, A.C., Effron, D.A., Monin, B.t., 2010. Moral self-licensing: When being good frees us to be bad. Soc. Pers. Psychol. Compass 4 (5), 344–357.
Monin, B., Miller, D.T., 2001. Moral credentials and the expression of prejudice. J. Pers. Soc. Psychol. 81 (1), 33.
Nafziger, J., 2020. Spillover effects of nudges. Econom. Lett. 190, 109086.
Nilsson, A., Bergquist, M., Schultz, W.P., 2017. Spillover effects in environmental behaviors, across time and context: A review and research agenda. Environ. Educ. Res. 23 (4), 573–589.
Nording, S., Bennich, P., 2021. A Comparative Study on Water Tap Labels Within the Nordsyn Cooperation. Technical Report 2021:524, Nordisk Ministerråd.
Nussbaumer, T., Thalmann, S., Hurni, A., Mennel, S., 2021. Faktenblatt thermische Netze. Technical Report, EnergieSchweiz.
Rennert, K., Errickson, F., Prest, B.C., Rennels, L., Newell, R.G., Pizer, W., Kingdon, C., Wingenroth, J., Cooke, R., Parthum, B., et al., 2022. Comprehensive evidence implies a higher social cost of CO2. Nature 610 (7933), 687–692.
Rolnick, A., Grunewald, R., 2003. Early Childhood Development: Economic Development with a High Public Return. Technical Report.
Sallee, J.M., 2014. Rational inattention and energy efficiency. J. Law Econ. 57 (3), 781–820.
Sánchez, G.E., Pellerano, J.A., Price, M.K., Puller, S.L., 2022. Not all nudges are created equal: Comparing social and pure informational nudges. Available at SSRN 4053370.
Schmid, D., Im, U., 2019. Focus Report on Health Impacts and External Costs due to Air Pollution. Technical Report 2019:1668007.
Schneider, F.H., Campos-Mercade, P., Meier, S., Pope, D., Wengström, E., Meier, A.N., 2023. Financial incentives for vaccination do not have negative unintended consequences. Nature 1–8.
Sherif, R., 2021. Are Pro-environment Behaviours Substitutes or Complements? Evidence from the Field. Working Paper 2021-03, Max Planck Institute for Tax Law and Public Finance.
Sparks, P., Shepherd, R., 1992. Self-identity and the theory of planned behavior: Assesing the role of identification with "green consumerism". Soc. Psychol. Q. 388–399.
Steinmayr, A., 2020. MHTREG: Stata Module for Multiple Hypothesis Testing Controlling for FWER.
Taubinsky, D., Rees-Jones, A., 2018. Attention variation and welfare: Theory and evidence from a tax salience experiment. Rev. Econ. Stud. 85 (4), 2462–2496.
Taylor, R.L., 2020. A mixed bag: The hidden time costs of regulating consumer behavior. J. Assoc. Environ. Resour. Econ. 7 (2), 345–378.
Tiefenbeck, V., Goette, L., Degen, K., Tasic, V., Fleisch, E., Lalive, R., Staake, T., 2018. Overcoming salience bias: How real-time feedback fosters resource conservation. Manage. Sci. 64 (3), 1458–1476.
Tiefenbeck, V., Staake, T., Roth, K., Sachs, O., 2013. For better or for worse? Empirical evidence of moral licensing in a behavioral energy conservation campaign. Energy Policy 57, 160–171.
Tinbergen, J., 1952. On the Theory of Economic Policy. North-Holland Publishing Company, Amsterdam.
Trachtman, H., 2021. Does Promoting One Behavior Distract from Others? Evidence from a Field Experiment. Working Paper 3941884, SSRN.
Truelove, H.B., Carrico, A.R., Weber, E.U., Raimi, K.T., Vandenbergh, M.P., 2014. Positive and negative spillover of pro-environmental behavior: An integrative review and theoretical framework. Global Environ. Change 29, 127–138.
Van der Werff, E., Steg, L., Keizer, K., 2013. The value of environmental self-identity: The relationship between biospheric values, environmental self-identity and environmental preferences, intentions and behaviour. J. Environ. Psychol. 34, 55–63.
Van der Werff, E., Steg, L., Keizer, K., 2014. Follow the signal: When past pro-environmental actions signal who you are. J. Environ. Psychol. 40, 273–282.
Werthschulte, M., Löschel, A., 2021. On the role of present bias and biased price beliefs in household energy consumption. J. Environ. Econ. Manag. 109, 102500.
Wichman, C.J., Taylor, L.O., Von Haefen, R.H., 2016. Conservation policies: Who responds to price and who responds to prescription? J. Environ. Econ. Manag. 79, 114–134.
Xu, L., Zhang, X., Ling, M., 2018. Pro-environmental spillover under environmental appeals and monetary incentives: Evidence from an intervention study on household waste separation. J. Environ. Psychol. 60, 27–33.
Zeelenberg, M., Pieters, R., 2004. Consequences of regret aversion in real life: The case of the Dutch postcode lottery. Organ. Behav. Hum. Decis. Process. 93 (2), 155–168.