European Economic Review 160 (2023) 104612

Contents lists available at ScienceDirect

European Economic Review

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

Can self-set goals encourage resource conservation? Field experimental evidence from a smartphone app[☆]

Andreas Löschel[a,b,c], Matthias Rodemeier[d,e,f], Madeline Werthschulte[g,h,*]

[a] Ruhr-University Bochum, Germany
[b] RWI – Leibniz Institute for Economic Research, Germany
[c] Fraunhofer Center for International Management and Knowledge Economy IMW, Germany
[d] Bocconi University, Italy
[e] University of Chicago, United States of America
[f] BAFFI – Centre on Economics, Finance and Regulation, Italy
[g] Heidelberg University, Germany
[h] ZEW – Leibniz Centre for European Economic Research, Germany

ARTICLE INFO

JEL classification:
C93
D91
Q49

Keywords:
Nudging
Goal setting
Field experiments
Scalability
Energy
Behavioral welfare economics
Smartphones

ABSTRACT

This study leverages a large RCT to examine the potential of goal-setting nudges to encourage resource conservation at scale. We randomize a feature that allows subjects to set themselves energy consumption targets in a popular smartphone app. We document negative effects of the nudge on app utilization and estimate null effects on energy consumption with confidence intervals that rule out estimates from observational studies. A complementary survey identifies the mechanisms underlying these behavioral responses. Using a structural model and random variation of the app's price, we estimate that the average user is willing to pay 7.41 EUR to *avoid* the nudge.

1. Introduction

Policymakers are increasingly relying on tools that leverage insights from psychology and behavioral economics to affect individual choices. These mostly non-pecuniary incentives, referred to as "nudges", simplify choice environments and help individuals to implement privately or socially desirable actions. One of the most important fields of policy intervention is resource conservation: Nudges are frequently used as complementary tools to more classic interventions, such as Pigouvian taxes or legal mandates.

☆ We would like to thank an anonymous Associate Editor and two anonymous reviewers for their very helpful comments and suggestions. We are grateful to Eszter Czibor, Lorenz Götte, Ken Gillingham, Sébastien Houde, Jörg Lingens, John List, Martin Kesternich, Mike Price, Matthew Rabin, Julia Nafziger, and Jeroen van de Ven for insightful discussions about this project. We also thank seminar participants at ANU, Chicago, ETH Zürich, Hamburg, Mannheim, Neuchâtel, NHH Bergen, AERE 2020, and EAERE 2021 for helpful comments. Anna Beikert, Lukas Reinhardt and especially Malte Lüttmann provided excellent research assistance. The experiment was pre-registered at the AEA RCT Registry as trial AEARCTR-0003003. IRB approval was obtained by the University of Münster. Financial support for this project by the European Union's Horizon 2020 research and innovation program under grant agreement 723791 is gratefully acknowledged.

* Corresponding author at: Heidelberg University, Germany.
E-mail addresses: andreas.loeschel@ruhr-uni-bochum.de (A. Löschel), matthias.rodemeier@unibocconi.it (M. Rodemeier), madeline.werthschulte@awi.uni-heidelberg.de (M. Werthschulte).

https://doi.org/10.1016/j.euroecorev.2023.104612
Received 21 July 2022; Received in revised form 5 October 2023; Accepted 7 October 2023
Available online 12 October 2023

Our study is the first to test in a field experiment the effectiveness of a goal setting prompt – an ex-ante promising nudge – in encouraging resource conservation. The design of the nudge is motivated by a literature showing that prompting people to set goals helps them to follow through with desired behavior in areas other than environmental policy. It is therefore a natural step to analyze the role of these promising interventions in encouraging resource conservation and combating climate change.

To take the intervention to scale, we cooperate with a large public utility and a specialized IT company to develop and launch an energy savings app that can be accessed by the majority of the German population. Within this app, we randomize a goal-setting feature that prompts users to set themselves energy consumption targets for the upcoming month. In addition, we randomize a financial incentive that allows us to monetize the effect of the goal-setting prompt on consumer welfare. We then promote the rollout of the app through a mass-marketing campaign that includes considerable financial incentives.

Over a total period of seven months, we find a precisely estimated null effect of the goal-setting prompt on electricity consumption. The nudge does not affect behavior, although users set meaningful goals that are highly predictive of future consumption. Poor targeting properties of the app might explain the null effect of the nudge. As a complementary survey shows, app users are characterized by an already low baseline energy consumption and high levels of energy-related knowledge. Further, the average consumer is neither present-focused nor loss averse, two features commonly used in the theoretical literature to explain how goal setting affects behavior (Koch and Nafziger, 2011; Hsiaw, 2013).

Second, we find that the goal-setting nudge causes disutility to consumers, as it significantly reduces app utilization. Using exogenous price variation in our financial incentive, we structurally estimate that consumers are willing to give up 7.41 EUR to avoid the nudge. These results cast doubt on the prospects for mobile technologies to act as cost-effective scaling devices for behaviorally-motivated energy policies.

Our study makes three main contributions. It is the first study to use randomly assigned goal-setting prompts to evaluate their causal effect on energy conservation. Goal-setting is an important behavioral phenomena that has been modeled by both psychologists and economists (Kahneman and Tversky, 1979; Locke et al., 1981; Heath et al., 1999; Kőszegi and Rabin, 2006). Recent empirical evidence shows that prompting people to make plans and set goals helps them to reduce smoking (Armitage and Arden, 2008), eat healthier (Achtziger et al., 2008), get vaccinated (Milkman et al., 2011), and vote during elections (Nickerson and Rogers, 2010).[1] An overview article of the effectiveness of plan-making by Rogers et al. (2013) concludes that goal-setting prompts should play a larger role in public policy.

Motivated by this literature, we investigate the potential of goal-setting prompts in the context of resource conservation policies. Following the established literature, we prompt users to *self-set* a goal to test for endogenous reference point formation (Kőszegi and Rabin, 2006). Further, our design of the goal-setting feature follows best-practice in the literature (e.g., Locke and Latham (1990)) as it (i) allows users to set explicit and simplified goals, (ii) draws attention to the procedures to save energy, (iii) provides reinforcing feedback, and (iv) allows for learning.

Previous behavioral interventions have proven to be effective tools for reducing resource consumption, such as the widely used social norm comparisons (e.g., Allcott (2011), Ferraro et al. (2011), Ferraro and Price (2013), Dolan and Metcalfe (2015), Pellerano et al. (2017), Andor et al. (2020), and many others). In the area of goal-setting prompts, we are aware of only one related study on self-set energy consumption goals. Harding and Hsiaw (2014) use an event study design to evaluate the effects of self-set electricity consumption targets in the United States. The study finds that self-set goals reduced consumption by, on average, 8 percent in the short term, but identification hinges on the assumption that the timing of program take-up is quasi-random. Our estimated confidence intervals rule out these optimistic treatment effects.

Second, we add to the emerging literature on the scalability of policy interventions by investigating the extensive margin decision of adopting an intervention that promotes resource conservation. Policymakers often need to base their decisions on studies from small-scale and highly selected samples, which may yield disappointing results when the intervention is brought to scale (Al-Ubaydli et al., 2017b, 2019; Czibor et al., 2019; DellaVigna and Linos, 2022; Andor et al., 2022). Our intervention is rolled out using an easily accessible smartphone application that has the potential to be adopted by the majority of the population.

We also provide first evidence on how smartphones perform in scaling up energy policies. Recent studies have examined the ability of other devices, such as smart thermostats, to encourage energy savings, but these studies have yielded conflicting results (Blonz et al., 2021; Brandon et al., 2022). Smartphone interventions have received greater attention in the medical literature, where they have shown to increase patient compliance with medication intake (for an overview, see Vervloet et al. (2012) and Sarabi et al. (2016)).[2] Our understanding of how these digital technologies can help people to follow through with plans is very limited, despite the substantial demand for goal-setting apps. Two examples are "Goal Meter" and "Goalmap", which have over 1.1 million downloads combined at the Google Play Store.

Third, our experiment expands upon a small set of studies estimating the welfare effects of nudges, providing the first estimate of the welfare effects of goal-setting prompts. Understanding how nudges affect consumer surplus and social welfare is fundamental for the identification of optimal policy. In our setting, a complete cost-benefit calculation must not only consider the energy cost savings for consumers but also take into account how the nudge changes consumer utility. It is therefore crucial to obtain an estimate of consumers' valuation of the nudge.

[1] There are also few studies that find no effect of planning prompts. For instance, a recent study by Carrera et al. (2018) asks gym visitors to make plans and estimates a precise null effect of plan-making on gym visits.

[2] The success of these medical interventions has led economists to advocate for more studies that use smartphones for behavioral policy interventions (Al-Ubaydli et al., 2017a).

We find that the average consumer is willing to give up the considerable amount of 7.41 EUR to *avoid* the nudge. This estimate compares negatively to the prominent social comparison nudges that show households their peers' energy consumption (Allcott, 2011). Allcott and Kessler (2019) estimate an average willingness-to-pay of up to USD 4.36 (approx. 3.93 EUR) for a bundle of four comparison letters. The stark difference in willingness-to-pay in our study highlights the importance of estimating the structural parameters of behavioral models to advance our understanding of how different nudges affect utility. Only a handful of other studies have taken a structural approach to behavioral economics to natural field experiments: DellaVigna et al. (2012, 2016) and Butera et al. (2022), all of whom estimate social preferences; Rodemeier and Löschel (2020), who measure informational biases about energy efficiency; and Rodemeier (2023), who quantifies willingness-to-pay for carbon mitigation.[3]

We structure the presentation of our study as follows. Section 2 describes the experimental design, the energy app, and the strategy to promote app adoption. Section 3 presents reduced-form results and the underlying mechanisms. In Section 4, we estimate the welfare effects of the goal-setting nudge based on a simple theoretical model. Section 5 discusses our results in relation to previous evidence. Section 6 concludes.

2. Experimental design

The experiment was conducted in cooperation with the municipal utility provider in Münster, a German city with some 300,000 inhabitants. The utility is owned by the municipality and is the universal service provider in the area, servicing about 80 percent of the households in the municipality. The experiment spans a period of seven months and was implemented from May 2018 to November 2018. In the following, we first lay out the design of the mobile app and our treatment. We then elaborate on how the app was promoted. The experimental design was pre-registered at the AEA RCT Registry.[4]

2.1. The energy app

Our experiment intends to use a promising behavioral intervention and take it to a large scale. Mobile devices are often considered suitable scaling devices, as they are easily accessible by the majority of the population. Accordingly, we developed a mobile app with a goal-setting feature that prompts users to set themselves energy consumption targets. To causally identify the effect of a goal-setting feature on energy consumption, the availability of the feature was randomly assigned among app users. The randomization of the treatment also implies that the app needs to include other desirable features such that users in the control group find it worthwhile to use. The app we developed therefore provided two useful features to every user, irrespective of treatment assignment.

First, the app allowed households to scan their meter with their phone and to submit the meter reading to the utility electronically. This is a useful feature in the German context because the vast majority of German households do not have smart electricity meters. For meter readings, they are required to schedule an annual meeting with a representative of the utility who then reads the electricity meter manually. The app circumvents the hassle of manual meter readings by allowing electronic submission. As depicted in Fig. 1(a), the developed feature automatically recognizes and reads the electricity meter if the user points her phone's camera at the meter. The user then only needs to confirm the scan by clicking a button to upload the data to the utility provider's server. In the event the user experiences technical issues with the scanning process, she can also manually enter the meter value. However, the app always takes a picture of the meter so that the self-reported meter value can be verified. Both the digital scans and the actual pictures are available to us.

Second, the app provides simple information on the electricity usage of various household appliances. Fig. 1(b) shows a screenshot of the information translated into English. This information is provided to all subjects irrespective of the experimental group assignment. Thus, control group users also received information on the electricity consumption of various appliances. Providing consumers with information becomes especially important for the treatment group because savings goals are likely to only affect consumption if subjects can set meaningful targets and know how to save energy in the first place. Of course, the information provided may also affect consumption without the goal-setting feature, thus encouraging subjects in the control group to alter their behavior. Our study therefore identifies the causal effect of a goal-setting nudge when consumers are informed about the energy consumption of their appliances. This also means that our design avoids the interpretation that goals may not affect behavior because people do not understand how choices translate into outcomes.

Fig. 2 gives an overview of the experimental timeline. For each app user, the experiment lasts for four experimental periods, where each period corresponds to 30 days. Upon signing up for the app, users are randomly assigned to either the control group or a treatment group with equal probability. Users need to enter their meter number, postal code, and meter type to exclude the possibility of the same household participating with multiple devices.[5] Participants then conduct a first digital meter reading with

[3] DellaVigna (2018) provides an overview of studies estimating structural behavioral parameters using lab, field, and observational data. List et al. (2022) estimate behavioral parameters in various markets based on prior studies. The policy implications of behavioral models are discussed in Bernheim and Taubinsky (2019).

[4] The trail number is AEARCTR-0003003. When we pre-registered the experiment, we intended to have one additional treatment arm, in which subjects could re-adjust their pre-set goal at any point in time. Since the first month of our experiment was the same for all treatments, we were able to adjust our design based on the surprisingly low sign-up rate. We decided to drop this additional treatment arm to have enough statistical power to identify the causal effect of goals on energy conservation.

[5] The combination of meter number and postal code uniquely identifies a household. The meter type is either a regular meter or an HT-NT meter. An HT-NT meter records peak electricity consumption separately from off-peak consumption. This information is needed for the scanner functions, as with an HT-NT meter, two scans have to be made.

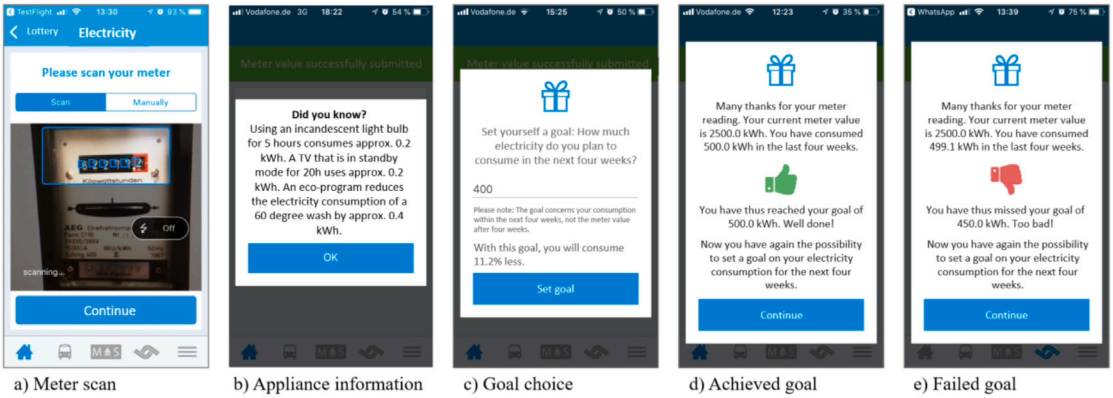

Fig. 1. Screenshots of energy app. *Note:* The figure shows English translations of screenshots of the energy app. For the original version in German see Figure A.1.

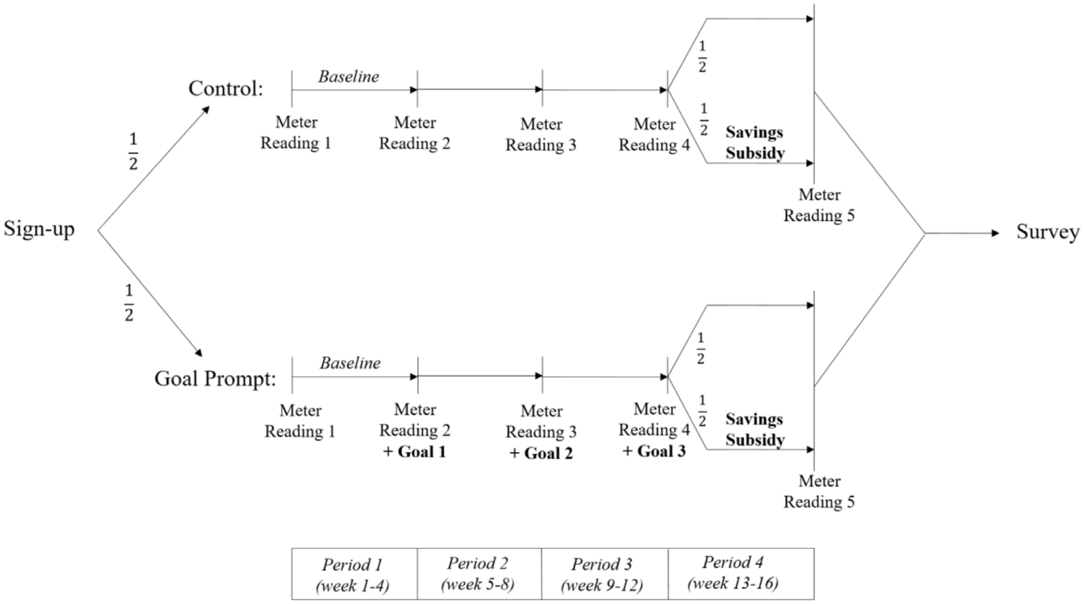

Fig. 2. Experimental design and timeline.

their phone. After subjects have done the first scan, they are informed that the next scan is due in 30 days. They are given the option to automatically save the due dates of all upcoming meter scans in their phone calendar. As we will discuss in more detail in the next section, subjects received a lottery ticket for every regular meter scan they submitted. The lottery assured that meter scans were incentivized for all subjects. Throughout the experiment, all participants received reminders to scan their meter one day before, one day after, and exactly on the due date. We offered participants a grace period to scan their meter from two days before the due date to two days after the due date. If participants failed to scan the meter within this grace period, the scanning function was deactivated until another 26 days had passed. After this deactivation period, they could continue to use the scanning function.

We call the first experimental period our baseline period since these 30 days are identical in the treatment and control groups. Upon completing the second scan after 30 days, participants in the treatment and control groups receive the aforementioned information on the energy consumption of different appliances. Afterwards, participants in the treatment group are asked to set themselves an energy consumption goal for the upcoming 30 days. Fig. 1(c) shows the goal-setting screen. Participants enter their desired consumption in kilowatt-hours for the next 30 days. The app also tells the subject how the consumption goal translates into percentage savings relative to the baseline period, allowing them to try out different consumption figures and get a feeling for a realistic consumption goal.

When designing the intervention, we tried to closely follow established best-practice in the long-standing psychology literature on goal setting. First, a number of studies suggest that self-set goals are more effective at increasing performance compared to exogenously assigned goals (e.g., Latham et al. (1978), Locke et al. (1981), Latham (2004) and Latham and Locke (2006)), which is why we ask participants to set themselves a goal as opposed to exogenously imposing goals.

Second, the literature suggests that making goals specific and reducing the complexity of the task increases treatment effects (e.g., Latham et al. (1978), Locke and Latham (1990) and Latham and Locke (2006)). Consequently, we asked participants to enter a specific goal for planned kilowatt-hour consumption. Further, we simplify the task by (i) informing participants by how much their absolute goal in kilowatt-hours translates into a relative goal in percent, and (ii) by giving participants practical tips on how to save energy. The provision of energy savings tips is motivated in part by the finding that goals are more effective when cognitive attention is drawn towards the *procedures* necessary to achieve the goal (e.g., Seijts and Latham (2005) and Latham et al. (2008))

Third, performance feedback should be accurate rather than ambiguous or manipulated (Ilgen et al., 1979; Podsakoff and Farh, 1989). Therefore, after the third scan has been submitted, participants in the treatment group receive precise feedback how many kilowatt-hours they consumed and how this compares to their set goal. If they consumed less than or equal to the planned amount, they are congratulated and are shown a "thumbs up", as depicted in panel (d) of Fig. 1. If they fall short of their goal by consuming more than intended, they are shown a "thumbs down" (see panel (e)). We chose to provide reinforcing feedback about the subject's performance, which has shown to increase effectiveness in other contexts (e.g., Locke et al. (1981), Latham (2004), Ilies and Judge (2005) and Fishbach et al. (2014)).[6]

Fourth, our design intentionally allows for learning over time by giving subjects the chance to set goals regularly, which has shown to boost effectiveness (e.g., Carver and Scheier (1990), Stock and Cervone (1990), Latham and Locke (2006) and Amabile et al. (2011)). As illustrated in Fig. 2, after the first performance feedback, participants set a new goal for the next 30 days.

The control group does not have access to the goal-setting feature and just completes the scan.

In experimental period 4 we randomize an additional treatment among all subjects that provides a financial incentive to save energy (see Figure A.5 for a screenshot). This "energy saving subsidy" treatment appears immediately after the fourth scan has been submitted and after a subject has potentially set a goal for the final period. Specifically, with a probability of one-half, the user is informed that she is participating in a lottery. If she wins the lottery, she receives 1 EUR per kilowatt-hour saved in month 4 relative to her electricity consumption in month 3. The total amount a subject may receive from saving energy is limited to 100 EUR. Prizes are paid out in the form of vouchers for the online shop Amazon. The lottery draws 15 users with equal probability (and no replacement) from the pool of eligible users, which results in a winning probability of approximately 1.85 percent. The winning probability is the same for every eligible user and is communicated to the subject. With an average electricity price of 0.30 EUR per kWh, our savings subsidy corresponds to an increase in the expected electricity price of approximately 6.1 percent.

Thus, there are a total of four experimental groups in the last period: the control group with and without the savings subsidy and the goal-setting group with and without the savings subsidy. As we will show in Section 4, the savings subsidy enables us to estimate users' willingness-to-pay for the nudge.

Finally, participants are reminded of the fifth and last scan. After completing this last scan, all of them are invited to an online survey. The survey elicits individual characteristics, qualitative statements about goal-setting behavior, and measures of electricity price beliefs, loss aversion, and present-focus. We use this survey to investigate the behavioral mechanisms underlying the treatment effects.

2.2. Rollout and promotion of the energy app

The energy app is easily accessible and designed for ease of use by anyone capable of operating a smartphone, and therefore has the potential to be used by the majority of the population. However, adoption rates may be low if the app is not properly marketed. To maximize the chances of large-scale diffusion, we worked closely together with industry experts. Specifically, the design of the app was developed by an established IT company specialized in mobile app development. The promotion of the app was conceived by both an external marketing agency and marketing experts of the utility who have experience with the successful rollout and implementation of energy apps.

Importantly, the app was promoted under the name of the local utility, which had a vested interest in the app's success. Working with the utility and subcontracted IT company, we invested in rigorous prototype testing involving joint workshops and discussions with focus groups and students. The primary goal was to generate a user-friendly process and to ensure that the text content was easily comprehensible.

We view our close cooperation with industry experts as a realistic simulation of how governments would ideally implement and scale up behavioral interventions: they partner with large players in the industry and leave much of the app's design and marketing strategy to industry experts. In our context, this approach avoids the concern that the app simply fails to deliver because its creation and promotion was entirely designed by researchers not trained for these purposes. Whenever we advertised the app, we were careful to only mention the features that were available to all users and not the randomized treatments.

The rollout of the app consists of three main steps. First, the energy app is integrated into a larger and widely used mobile application. The "münster:app" has been downloaded by over 122,000 smartphone users and involves functions such as real-time

[6] In order to tightly connect to established models of goal-setting and reference points in psychology and behavioral economics (e.g., Heath et al. (1999) and Kőszegi and Rabin (2006)), we decided against providing a monetary incentive for goal achievement. This literature understands a goal as a purely non-monetary tool that alters behavior by forming a reference point to which outcomes are compared. In addition, literature in psychology points towards potential pitfalls of implementing bonus schemes for goal achievement, such as setting strategically low goals (e.g., Locke (2004) and Latham and Locke (2006)).

information on changes in bus schedules, notification of free parking spots downtown, and local news.[7] The integration of the energy app into the larger münster:app makes its adoption particular convenient, as many households do not have to download a new application. All of the münster:app users were notified about the new features by a push message on their phone and by a large banner displayed on the app's landing page.

Second, a promotion campaign targeted the entire municipality. Over 69,000 utility customers received direct and personalized mails promoting the app (see Figure A.3), and a popular local radio station played frequent advertisements. As shown in Figure A.2, 14,000 flyers were enclosed with annual electricity bills sent by mail. The same flyers served as the basis for a print ad in an edition of the local newspaper that reached 18,000 households. Another local newspaper with a print of 48,000 copies announced the new app, and the social media outlets of the local utility advertised the app. The main student cafeteria, which has around 1600 visitors per day, displayed advertisement posters and flyers. An additional 4000 flyers were handed out by research assistants either in public spaces or by going door-to-door.

Third, we financially incentivized the use of the app (see Figure A.4 for a screenshot). *Every* app user receives a 45 EUR voucher for an online shop that sells household appliances if she completes all five meter scans and the online survey. In addition, all users (irrespective of how many scans they submit) participate in a lottery with various prizes such as holiday trips worth 1000 EUR, Apple iPads, and 100 EUR vouchers for local activities. The total amount of the lottery prizes is 6000 EUR. As previously mentioned, the chance to win in the lottery can be increased by conducting the digital meter readings: for every regular scan participants send in, they obtain one additional ticket for the lottery. Subjects who submit all regular scans therefore receive five additional lottery tickets.[8]

2.3. Sample

Table 1 presents the summary statistics for our sample. A total of 1627 participants signed up and sent in the first scan. Using information on meter type, we know whether subjects have a double tariff involving different nighttime and daytime electricity prices. Around 3 percent of the sample have such a contract, and these subjects are balanced between the treatment and control groups. Meters also vary in how they record electricity consumption and may have different decimal places and numbers of digits before the decimal point. Both of these variables do not significantly differ between treatment and control groups.

The baseline electricity consumption is calculated as the difference between the second and first meter scan. Since not every subject submits a second meter scan, baseline consumption is not available for all subjects who signed up. Around 50 percent of those who signed up submitted two or more scans. Importantly, the probability to submit a second scan does not differ between the treatment and control. In total, we have information on baseline electricity consumption for 844 subjects, who consume around 190 kWh, on average, in the baseline month. Baseline consumption also does not significantly differ between the treatment and control groups. Importantly, baseline consumption is far below the German average of 264 kWh for the respective month.[9] This suggests that the energy app is attracting consumers with an already low baseline energy consumption and might imply poor targeting—a point we turn to later in more detail.

Note also that the number of consumers adopting the app is relatively low given the substantial efforts we made in recruiting participants. Since every subject might have received a variety of advertisements, we cannot pin down the true response rate. However, we can create a very conservative upper bound on the response rate. Since at least 83,000 individual households were contacted through direct mailing or through flyers enclosed with their energy bill, the most optimistic response rate is 1.96 percent. This rate indicates a very low demand for an app related to energy consumption, especially given our mass-marketing campaign and that participation was financially incentivized through 45 EUR vouchers and a lottery with considerable prizes.

To assure that our electricity data are reliable, our research assistants verified every meter scan. In particular, they compared the value reported by the digital scan (or the manual data entry by the user) with the pictures that the app made of the meter. Of a total of 3610 m scans, the research assistants were not able to verify 297 scans (e.g., because these scans did not match the pictures). In Table A.1 in the Appendix we present results from a regression of the probability to report non-verifiable meter scans on treatment and we find no significant effect. We therefore drop these observations from our analysis. We also drop four users who have a missing first scan, which technically should not be possible and which prevents the calculation of baseline consumption.

3. Reduced-form results

3.1. Extensive margin choices: App utilization

We begin by analyzing subjects' utilization of the app. In every experimental period $e \in \{1, 2, 3, 4\}$, the subject has the choice to actively use the app and submit a scan or to not use the app. We call this utilization choice the extensive margin. We code the outcome variable $Utilization_{ie}$ such that it equals one if subject i submitted a meter scan at the beginning and at the end of period

7 The münster:app is run by the local utility provider and is available both through the Google Play Store and the Apple App Store.

8 The lottery to encourage participation is independent of our randomly assigned Pigouvian lottery in period 4 that incentivizes energy savings.

9 The annual electricity consumption of an average German household is 3111 kWh (Federal Statistical Office of Germany, 2019). To adjust for seasonal variation in consumption, we use the weights calculated for national load profiles for different months of the year in Fünfgeld and Tiedemann (2000). In the month of April, load profiles are about 8.5 percent of annual consumption, resulting in an average consumption of 264 kWh for the month of April.

Table 1
Summary table.

	Control		Treatment		Difference	
Double tariff (1 = yes)	0.036 (0.187)		0.029 (0.167)		−0.008 (0.009)	
Decimal places of meter	1.046 (0.342)		1.049 (0.398)		0.002 (0.018)	
Number of digits before decimal point	5.806 (0.605)		5.833 (0.615)		0.027 (0.030)	
Submitted at least two scans (1 = yes)	0.527 (0.500)		0.509 (0.500)		−0.017 (0.025)	
Baseline consumption (in kWh)		188.466 (109.235)		193.996 (123.006)		5.530 (8.384)
N	824	434	803	410	1627	844

Note: This table presents the mean of observable variables for the treatment and control group and the difference in means between these groups. Standard deviations for control and treatment are reported in parentheses. For the last two columns, standard errors of the differences are reported. Information on meter characteristics is available for everyone who signed up for the energy app. Information on baseline electricity consumption is available for anyone who scanned their meter at the beginning and end of period 1.

e. If the subject did not submit the scan at the beginning or at the end of the period, the outcome equals zero. To investigate how the treatments affect utilization choice, we estimate the following linear probability model:

$$Utilization_{ie} = \alpha_e + \beta_e G_{ie} + \mathbb{1}_{e=4}(\gamma S_i + \delta \times S_i \times G_{ie}) + \epsilon_{ie}, \tag{1}$$

where G_{ie} equals one if subject i was in the treatment group in period e and zero otherwise. The average baseline probability to use the app in period e is given by α_e, and the error term is denoted by ϵ_{ie}. In period 4 we also randomized the savings subsidy, for which we assign the treatment dummy S_i. The coefficient γ is the treatment effect of the savings subsidy on utilization, and δ measures the interaction effect of the subsidy with the nudge.

Table 2 reports the results. Around 52 percent of subjects in the control group who signed up also submitted the second scan and therefore count as being part of period 1. For period 1, the difference in utilization between treatment and control is small and statistically insignificant, which is unsurprising given that period 1 was identical for subjects in both groups. The probability of using the app in the following periods becomes dramatically smaller for both treatment and control subjects. In the control group, only 29 and 23 percent use the app in periods 2 and 3, respectively. For these periods, the treatment coefficient of the goal-setting nudge is negative but remains economically small and statistically insignificant. This provides evidence of no selection on the extensive margin during the first two treatment periods and suggests we can causally identify the effect of the nudge on electricity consumption among users.

Columns 4 and 5 report the effects of the nudge and subsidy on utilization in period 4. While column 4 shows pooled effects, column 5 interacts the subsidy with the goal-setting nudge. In the pooled model, the goal-setting nudge decreased the utilization probability by 4.2 percentage points. The effect is statistically significant at the 5 percent level. By contrast, the savings subsidy involves an increase in the utilization probability of 2.9 percentage points. While the effect is not statistically significant at conventional levels, the p-value of 0.148 is relatively small. An independent t-test shows that the treatment effect of the nudge and the subsidy are significantly different at the 1 percent level. Column 5 reports an even larger negative treatment effect of the nudge and a slightly smaller effect of the subsidy in isolation. When the nudge is combined with the subsidy, subjects are 1.6 percentage points more likely to use the app than when they are only treated with the nudge.

Overall, the results from period 4 suggest that the goal-setting prompt sufficiently pressured or disturbed some of the subjects such that they stopped using the app.[10] Compensating consumers with the savings subsidy reduced this tendency.

3.2. Intensive margin choices: Goal setting and electricity consumption

3.2.1. Goal setting behavior

Before we turn to the treatment effect of the goal-setting nudge on energy consumption, we briefly describe how subjects set their consumption targets. Specifically, we investigate whether subjects chose meaningful goals, as this indicates whether subjects carefully engaged with the app and the goal-setting prompt. If consumption goals are meaningful, we would expect them to be correlated with actual consumption. By contrast, if subjects just set goals irrespective of their true consumption, the correlation would be zero. Fig. 3 plots the consumption goal (planned consumption) for a month against the actual consumption in that same

[10] We do not find a significant correlation between failing to achieve the goal and dropping out of the app in the next period. While this correlation does not generally have a causal interpretation, it provides suggestive evidence consistent with the idea that subjects receive negative utility from the goal-setting prompt *directly* rather than from a failure to reach the goal.

Table 2
Effect on extensive margin: Probability of using the app over time.

	(1) Period 1	(2) Period 2	(3) Period 3	(4) Period 4	(5) Period 4
Goal treatment	−0.006	−0.013	−0.014	−0.042**	−0.050*
	(0.026)	(0.023)	(0.022)	(0.020)	(0.028)
Savings subsidy				0.029	0.021
				(0.020)	(0.029)
Goal × subsidy					0.016
					(0.040)
Constant	0.517***	0.293***	0.237***	0.190***	0.194***
	(0.018)	(0.017)	(0.016)	(0.018)	(0.020)
N	1493	1493	1493	1493	1493

Note: The outcome variable is a dummy for whether a subject submitted the meter scan at the beginning and the end of the respective period. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses.

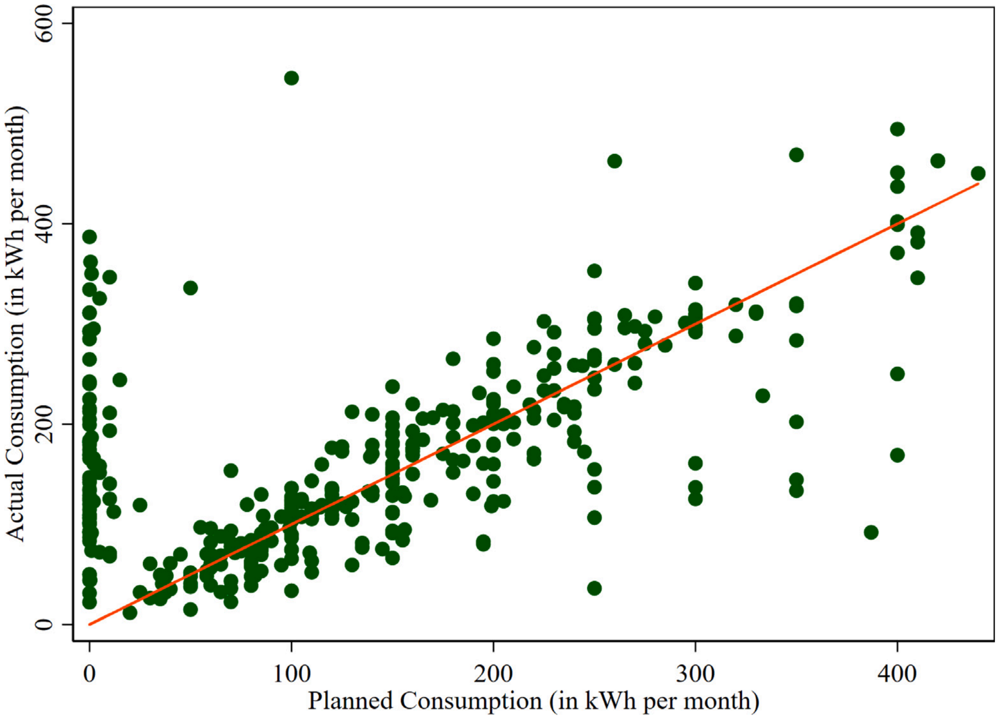

Fig. 3. Actual versus planned consumption. *Note:* The graph plots the consumption goals against the actual consumption for the treatment group. The orange line represents the 45-degree line. To account for outliers, we restrict the graph to the 95th percentile of consumption goals. Observations from all three treatment periods are included.

month. We also plot the 45-degree line indicating when planned and actual consumption are equal. To adjust for outliers, we exclude the top 5 percent of goals.

Visual inspection reveals a striking correlation between planned and actual consumption. Many subjects choose goals that are highly predictive of their consumption. The figure also shows that there is a non-negligible share of consumers who choose consumption goals equal to zero. In principle, a zero consumption goal is feasible (e.g., when subjects go on vacation) but is highly unlikely to be a realistic goal for this many households. A more likely interpretation is that these subjects did not carefully engage with the goal-setting feature and a value of zero is just a mentally convenient default. Around 14 percent of all goals fall into this category.

Among the remaining 86 percent of meaningful goals, we can draw distinctions based on goal ambition. About 28 percent of all goals are *lenient*, meaning they are equal to or greater than baseline consumption. By contrast, the majority of 59 percent are

Table 3
Effect on intensive margin: Electricity consumption.

	(1) Log (kWh)	(2) Log (kWh)	(3) Log (kWh)
First goal	0.015 (0.026)	0.008 (0.024)	
Second goal	0.047 (0.036)	0.053 (0.037)	
Third goal	−0.034 (0.046)		
Savings subsidy	0.028 (0.040)		
Goals (pooled)			0.027 (0.025)
Period 4 consumption included	Yes	No	No
N	1813	1538	1538

Note: The outcome variable is the natural logarithm of electricity consumption measured in kWh. Month and user fixed effects are included in all regressions. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered on subject level are in parentheses.

savings goals, which are smaller than baseline consumption and reflect an intention to save energy.[11] However, lenient goals may also imply an intention to save energy because the need for electricity may be larger in future periods than in the baseline period. This may be due to seasonal variation in energy demand or due to idiosyncratic consumption spikes (e.g., social gatherings, being at home more due to vacation or remote work, etc.). Since, by randomization, the control group is experiencing the same temporal variation in energy demand, even lenient goals may result in electricity savings compared to the control.

In a nutshell, there are two main groups of subjects in the treatment group. The first one, which includes the vast majority of the sample, sets meaningful consumption goals that are highly predictive of their actual future consumption. The second smaller, but non-negligible group, chooses meaningless goals of zero. Overall, there is a slight tendency for consumers to fall short of their goal, as most observations lie above the 45-degree line.[12]

3.2.2. *Effect on electricity consumption*

The electricity consumption data are a panel data set with four experimental periods. When subjects stop using the app and do not scan their meters, electricity consumption is missing. We therefore have an unbalanced panel, which does not seem problematic for identification in the first three experimental periods because attrition rates do not systematically differ between treatment and control. Asymmetric attrition becomes a concern in period 4, in which the treatment decreased the likelihood of app usage (see Section 3.1). Hence, we will use an unbalanced panel but run the analysis separately with and without the inclusion of period 4. Standard errors will be clustered at the subject level. Our empirical model is

$$log(\text{kWh}_{iet}) = A_i + E_e + T_t + \tau_e G_{ie} + \gamma_t S_{ie} + \epsilon_{iet},$$

where $log(\text{kWh}_{iet})$ is the natural logarithm of participant i's electricity consumption in experimental period e and calendar month t. We follow the literature in energy economics and logarithmize electricity consumption, but this does not change the qualitative interpretation of our results. Recall that subjects could submit their meter scan up to 2 days before and 2 days after the due date. We therefore normalize the outcome variable to resemble 30-days consumption.[13] The variables $G_{ie} \in \{0, 1\}$ and $S_{ie} \in \{0, 1\}$ indicate the nudge treatment and savings subsidy in experimental period e, respectively. Individual fixed effects and experimental period fixed effects are denoted by A_i and E_e, respectively. To control for seasonal variation, we also include months fixed effects, T_t. The consumption recorded in an experimental period e may belong to two calendar months because subjects do not necessarily start with the experiment on the first day of a calendar month. T_t therefore comprises a fixed effect for the calendar month in which consumption started and another fixed effect for the calendar month in which it ended.

The coefficient τ_e can be interpreted as the approximately average percentage change in electricity consumption in period e caused by the goal-setting prompt at the beginning of the same period. Table 3 reports the treatment effect coefficients for each period.

The model used to produce results in column 1 includes observations from period 4 and must be interpreted with caution. We do, however, see that the exclusion of period 4 in column 2 does not substantially alter results. In both columns, we observe economically small treatment effects of the first goal-setting prompt of 1.5 and 0.8 percent on electricity consumption, respectively. Both coefficients are statistically indistinguishable from zero and involve tight confidence intervals. We can exclude treatment effects

[11] We further explore the association between goal achievement/failure in one period and the goal set for the subsequent period, but do not find a significant correlation.

[12] The median distance between planned and actual consumption is 7 kWh.

[13] Specifically, we divide the outcome variable by the number of days that lie between the first and second scan and then multiply the result by 30.

smaller than –3.6 percent and –3.9 percent with 95 percent confidence in columns 1 and 2, respectively. Results are similar for the second goal-setting prompt, but the lower bounds of the confidence intervals are even closer to zero. Columns 1 and 2 rule out treatment effects smaller than –2.4 percent and –2.0 with 95 percent confidence.

The third goal prompt involves an insignificant but negative coefficient. The confidence interval is still small but is larger than for the previous goals, which may be attributable to the fact that we have few observations for this period. Of course, it may also be a result of systematic selection on the extensive margin during this period. For the same reason, the coefficient of the savings subsidy may be imprecisely estimated and/or have no causal interpretation. The coefficient is small and statistically insignificant.

In column 3 we pool the goal-setting treatments to increase statistical power even further. We again observe a precisely estimated null effect and can rule out treatment effects smaller than –2.2 percent with 95 percent confidence.

Taking all findings together, our results provide considerable evidence that the goal-setting prompt failed to reduce electricity consumption but instead caused direct disutility to consumers as it reduced app utilization.

3.2.3. Heterogeneity

We next explore heterogeneity in treatment effects by baseline energy consumption. Table A.2 in the Appendix reports intensive margin treatment effects when the goal is interacted with an indicator for above-median baseline consumption. We do not find evidence of heterogeneous treatment effects as we estimate null effects for households with both above- and below-median baseline consumption. The lack of treatment effect heterogeneity is a particularly stark result that speaks against the efficacy of the intervention as it implies that even the targeting of high consumption households would not induce energy savings.

The result complements previous findings in the literature which show that nudges that successfully manage to encourage resource conservation typically do so by inducing large treatment effects for high consumption households (see, e.g., Allcott (2011) and Andor et al. (2020)).

Finally, we investigate the relationship between the ambitiousness of the goal and energy savings. Table A.3 shows the regression results for consumer subgroups that set themselves (1) a lenient goal, i.e., a goal that is equal to or greater than baseline consumption, (2) a goal equal to zero, and (3) a savings goal, i.e., a goal that is smaller than baseline consumption. It is important to note that differences in energy consumption across these three groups do not necessarily have a causal interpretation since the goal is endogenously determined by subjects.

For the first goal, we do not find a statistically significant correlation between the goal difficulty and energy savings. The same is true for the second goal among users who set themselves either a zero-goal or a savings goal. By contrast, a lenient second goal is associated with a statistically significant *increase* in energy consumption compared to control. A potential explanation for this correlation is that users who intend to weakly increase consumption (i.e., set a lenient goal) in fact do so when prompted to set a goal. On the other hand, savings goals are not associated with significant reductions in electricity consumption in any period.

While this is only correlational evidence, we do not find evidence of substantial heterogeneity based on goal difficulty. In particular, the patterns do not suggest that savings goals and lenient goals led to opposite effects, such that these effects canceled each other out in the aggregate.

3.3. Mechanisms

To investigate the mechanisms underlying the treatment effects, we first need to understand the theoretical argument for why goal setting should affect behavior. Economic models on goal setting typically argue that people set themselves goals to reduce overconsumption resulting from self-control problems (Koch and Nafziger, 2011; Hsiaw, 2013). Overconsumption is modeled as an implication of the well-established $\beta - \delta$-model (Laibson, 1997), in which consumers focus too much on the present when making choices. Goal setting may then be used as a commitment device for present-focused agents to mitigate overconsumption. The mechanism is that goals create reference points to which agents compare their behavior (Heath et al., 1999). A crucial component of these models is loss aversion, meaning that consumers dislike falling short of a self-set goal by a certain distance more than they value achieving a goal by the same distance.

Our empirical setting features several intertemporal trade-offs that would result in overconsumption for present-focused consumers. Recall that in the German context, final accounting for energy costs only takes place once a year. While there are exceptions to this billing cycle, 96 percent of our post-experimental survey respondents state that they receive energy bills annually. Even though every household pays a fixed monthly payment that is set to approximately cover monthly energy costs, this payment amount is determined based on past usage, and has no direct relation to current consumption. This means that the costs of increasing current energy consumption is entirely delayed to the (distant) future for the majority of our sample. In addition, externalities create another intertemporal trade-off, as discussed by Harding and Hsiaw (2014): the negative environmental consequences of excessive resource consumption only accrue in the future. Thus, even if consumers have altruistic preferences, they may focus too little on the externalities that result from energy consumption.

Motivated by these theoretical arguments, we elicited several core model parameters in the post-experimental survey (see Appendix D). To measure present focus, we use the standard approach of two incentivized multiple price lists (see, e.g., Coller and Williams (1999) and Cohen et al. (2019)). The first list asks participants to choose between receiving 100 EUR within the next 24 h or an alternative amount in one month. The second list includes a trade-off of either receiving 100 EUR in one month or an alternative amount in two months. In both lists the alternative amount increases from 100 to 160 EUR across 14 decisions. Importantly, choices are incentivized since one choice from the two price lists is randomly picked as the actual payment. To ease

Table 4
Behavioral parameters: Present focus, loss aversion, and price beliefs.

	Sample average	Percentile					N	Representative average for comparison	Comparison study
	(Std. error)	10th	25th	50th	75th	90th		(Std. error)	
β	1.030 (0.007)	0.972	1	1	1.013	1.090	353	0.95 (0.02)	Imai et al. (2021) (Meta-analysis)
λ	0.826 (0.100)	−0.933	0	0.933	1.25	1.875	352	1.31 (0.11)	Walasek et al. (2018) (Meta-analysis)
$p_{max} - p_{min}$	6.658 (1.589)	0	2	4	9	15	193	12.10 (0.519)	Werthschulte and Löschel (2021) (German average)

Note: β denotes the present-focus parameter, and λ is the loss-aversion coefficient. $p_{max} - p_{min}$ gives the differences between the maximum and minimum energy price participants believe they pay, measured in cents. Standard errors of the mean are in parentheses.

research expense, we randomly chose a subset of subjects to be eligible for this actual payment. Randomizing eligibility has been shown to have no considerable effect on choices, as revealing true preferences remains optimal (Charness et al., 2016).[14]

We infer discount rates, denoted δ, by assuming that utility, $u(z, \cdot)$, is linear in experimental payments, z, and indifference between the earlier and later payment at the midpoint, $\bar{z}$, between the payments at which the participant switches from preferring the earlier amount over the later amount.[15] Under these assumptions, we can calculate discount rates by $\delta = \frac{u_t(z,\cdot)}{u_{t+1}(\bar{z},\cdot)}$ for each participant and for both price lists. Here, t refers to the earlier date and $t + 1$ to the later date of the respective list. The present focus parameter, denoted β, is then identified by the ratio of the discount rates inferred from the first and second multiple price list (Cohen et al., 2019). $\beta = 1$ implies time-consistent discounting, as both discount rates are equal, while $\beta < 1$ implies present focus.

Similarly, we measure loss aversion in our survey from multiple choices between either participating in a lottery, in which participants can win or lose 150 EUR with equal probability, or receiving a safe payment. The safe payment varied in 31 decisions from –150 to 150 EUR (Koch and Nafziger, 2019). In contrast to the time preference elicitation, choices are hypothetical. As in Falk et al. (2023), we use the staircase method to reduce survey length. The staircase method condenses the 31 decisions of the multiple price list into five consecutive choices. This means that the first decision between the lottery and safe payment determines the second choice, the second decision determines the third choice, etc. Although choices are hypothetical, Falk et al. (2018) show that preferences elicited from the staircase method correlate with a number of economic outcomes, such as education, savings, and consumption patterns.

To elicit loss aversion, we follow the standard assumption that $u(z, \cdot) = z$ if $z \geq 0$ and $u(z, \cdot) = \lambda z$ if $z < 0$. Here, λ denotes the degree of loss aversion. If $\lambda = 1$, people value gains by the same amount to which they dislike losses of equal absolute size, while $\lambda > 1$ implies loss aversion. We assume the amount at which the participant is indifferent between the lottery and safe payment, $\bar{z}$, to be the midpoint of the two safe payments at which the participant switches from preferring the lottery to preferring the safe payment.[16] With this assumption, we can solve the indifference equation between lottery and safe payment, i.e. $\mathbb{E}[u(z, \cdot)] = u(\bar{z}, \cdot)$, for λ. Specifically, if $\bar{z} \geq 0$, $\lambda = \frac{\bar{z}-0.5\cdot150}{0.5\cdot(-150)}$, and if $\bar{z} < 0$, $\lambda = \frac{0.5\cdot150}{\bar{z}-0.5\cdot(-150)}$.

Table 4 shows the distributional properties of present focus and loss aversion in our sample. The average subject features a present focus parameter of 1.03 and a loss-aversion parameter of 0.83. Both estimates are indistinguishable from the benchmarks of no present focus ($\beta = 1$) and no loss aversion ($\lambda = 1$). In the absence of measurement error, these values imply that the average participant is neither present-focused nor loss-averse. We can see that this is not only true for the average consumer but also for the majority of the survey sample, as most reported percentiles involve values close to 1. These values differ markedly from other studies. A meta-analysis by Imai et al. (2021) examines 220 estimates from the literature and estimates an average present focus parameter of 0.95 that is significantly different from one. For loss aversion, a meta-analysis by Walasek et al. (2018) estimates a median λ of 1.31 and excludes our finding of equal gain-loss weighting with 95 percent confidence. Other literature reviews regularly report higher loss-aversion parameters (e.g., the average λ of the studies summarized by Booij et al. (2010) amounts to 2.069).

Our particular parameter estimates are supported by results from two survey questions, as depicted in Figs. 4 and 5. As a measure of self-control, subjects were first asked how often they intend to save energy and then how often they fail to implement these intentions. Possible answers were "Never", "Sometimes", "Often", and "Always". We can see in Fig. 4 that the distribution of answers to the first question is oppositely skewed to the distribution of answers to the second question. While the modal subject

[14] We communicated to participants a probability of payment, which was selected based on the number of app users such that in expectation, three participants will be paid. Participants were paid with Amazon vouchers.

[15] If the participant switches multiple times between the earlier and later payments, we use the first switching point to determine the indifference amount. For participants always choosing the later payment, i.e., preferring 100 EUR later over 100 EUR earlier, we assume indifference at 99.5 EUR. Further, we assume participants always choosing the earlier payment to be indifferent at 165 EUR. However, whether imposing switching points or excluding never-switching participants, the average present focus estimate differs only slightly (1.030 when imposing versus 1.006 when excluding).

[16] If participants always preferred the lottery, we assume a switch to the safe payment when offered 160 EUR. Likewise, if they always preferred the safe payment, we assume a switch to the lottery when confronted with a safe loss of 160 EUR. Yet, when we instead exclude never-switching participants, the mean loss-aversion coefficient remains largely unchanged (0.826 when imposing switching points versus 0.868 when excluding participants).

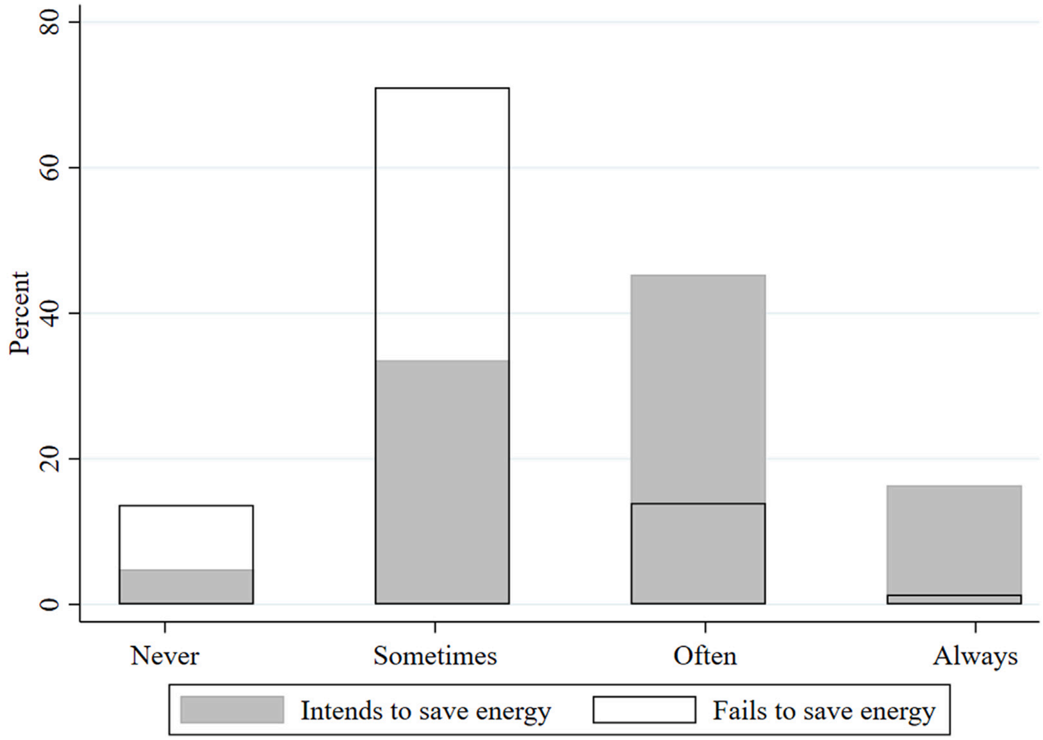

Fig. 4. Intentions and self-control.

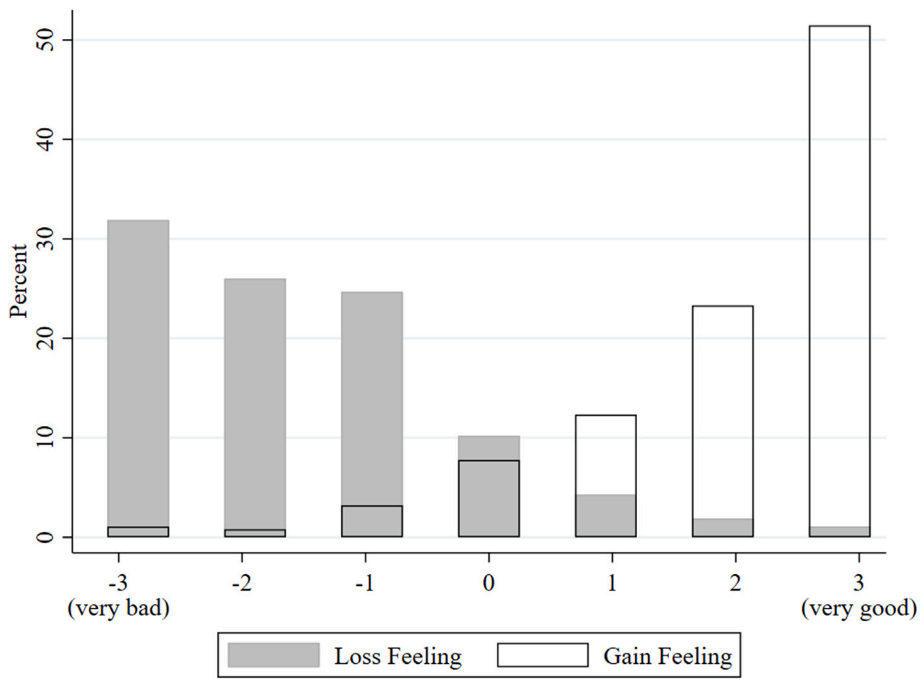

Fig. 5. Gain and loss feelings.

intends to save energy on a regular basis, she rarely fails to implement this intention. These results are also in line with the previously discussed finding that planned and actual consumption often coincide in the experiment (recall Fig. 3).

We also find additional support for the estimated loss-aversion parameter. Specifically, we asked subjects how they would feel to receive a refund of 100 EUR from the utility at the end of the year or if they had to pay an additional 100 EUR to the utility. Answers were ordered on a 7-point Likert scale from –3 (very bad) to +3 (very good). Fig. 5 illustrates the distribution of responses, and we can see that self-reported emotions are almost perfectly symmetric around zero. Subjects do not systematically report stronger feelings when losing versus gaining 100 EUR. If anything, they value gains more than they dislike equal losses.

A theoretically driven explanation for the empirical null effect is therefore that the subject pool is not characterized by the behavioral anomalies that typically explain why goal setting affects behavior. However, this does not mean that the general population fails to display these anomalies. Instead, subject features are likely to be a result of unfavorable self-selection into the pool of app users.[17] In fact, this form of systematic self-selection is evident by a number of additional results. For example, and as previously reported, the baseline consumption of app users is below the national average, which is important since previous research consistently finds larger energy savings effects among households with a higher baseline consumption (e.g., Allcott (2011) and Andor et al. (2020)).

In addition, app users appear to have higher levels of "energy literacy" than the average German household. This becomes evident by another survey question eliciting subjects' confidence about the energy price they pay. Subjects were asked to state the minimum and maximum price they think they pay for electricity. The last row in Table 4 reports the difference between the maximum and minimum as a measure of confidence. The average participant reports a relatively small interval of 0.07 EUR. To put this into perspective, we compare this estimate to the belief interval elicited in a nationally representative survey conducted by Werthschulte and Löschel (2021). In their sample the average deviation between maximum and minimum perceived electricity price is 0.12 EUR (i.e., almost twice as large as for our subject pool), suggesting consumers with an already high knowledge for energy-related topics use the app.

Additional evidence for systematic sample selection can be found in the sociodemographics, as documented in Table A.5 in the Appendix. App users are predominantly male (23 percent female versus 51 percent nationwide) and are better educated than the average German (76 percent with a high school degree versus 33 percent nationwide). The average participant is also slightly older (46 years versus 44 years) and earns a higher income (2515 EUR per month versus 1770 EUR per month). Participating households are even characterized by larger dwellings (107 square meters versus 98 square meters) and a larger household size than the national average (2.54 persons per household versus 1.98 persons per household). This means that energy consumption per capita is far below the average, since the total energy consumption per household was already relatively low. A plausible explanation is that sample participants consume energy more efficiently than a national representative household does, consistent with their high level of energy literacy.

In sum, the metrics point at very poor targeting properties of the app. Based on a theoretical argument, the ideal consumer to target would be loss averse and would have self-control problems and high levels of baseline energy consumption. Instead, we find the opposite: subjects who decide to use the app are well-informed consumers who are neither loss averse, nor present-focused, and they have low levels of baseline energy consumption. More generally, our results highlight the importance of carefully documenting selection into the pool of study participants in order to understand treatment effects, as advocated by List (2020).

4. Welfare analysis

4.1. A simple model of app utilization

We develop a simple theoretical model to quantify the efficiency effects of the nudge. In our experiment, subjects can adjust their behavior on two margins. On the extensive margin, they choose app utilization $j \in \{a, o\}$, where a denotes the energy savings app and o the outside option. On the intensive margin, they potentially choose an energy savings goal (if the feature is available) and then choose their actual energy consumption. We first derive the optimal intensive margin choices conditional on a utilization choice and then characterize the optimal utilization choice on the extensive margin. Since an empirical welfare analysis requires exogenous price variation, we restrict our theoretical model to the last experimental period, in which we also varied the expected energy price. We therefore use a static model that applies to the fourth experimental period.

A consumer gets utility $v(x)$ from x units of energy. We make the standard assumptions about the properties of $v(x)$: $\frac{\partial v}{\partial x} > 0$ and $\frac{\partial^2 v}{\partial x^2} < 0$. The cost of energy consumption is given by $c_j(x, \cdot)$ and may depend on app utilization j and other factors, which we will explain below. We assume a quasi-linear utility function such that small variations in the energy price do not cause income effects. We consider this a reasonable assumption, as our experimentally induced price variation was relatively small in expectation. "Material utility" from consumption for a consumer with income Y is therefore given by $U = v(x) + Y - c_j(x, \cdot)$.

Consumers typically cannot steer their energy consumption perfectly ex-ante due to exogenous factors such as varying weather conditions. They may also have uncertainty about how behavior maps onto energy consumption.[18] We therefore assume that x is stochastic and the consumers can affect x through effort e. Formally, effort induces a draw from the conditional cumulative distribution function $H(x|e)$. We assume that $H(x|e)$ is differentiable in e.

[17] Note that a typical issue with small-scale studies is *favorable* self-selection of subjects into the pool of participants—a phenomenon (Al-Ubaydli et al., 2017a) label "adverse heterogeneity". Favorable self-selection implies that subjects select on gains such that study participants have larger treatment effects than the overall population.

[18] While our experimental design involved information about the energy use of various activities, there may still be a remaining degree of uncertainty, and we therefore model it explicitly.

To allow for the possibility that a goal-setting nudge affects choices, the consumer also receives "psychological utility", $B(\phi, x, g) = \phi + R(x, g)$, if she is treated with a goal-setting prompt $G_j \in \{0, 1\}$. Here, $G_j = 1$ indicates treatment and $G_j = 0$ no treatment. We assume that utility from the nudge can be decomposed into two parts. First, it may cause direct utility unrelated to energy consumption, denoted ϕ. This term is positive if consumers enjoy being prompted to set a goal irrespective of its effect on energy consumption, and it is negative if they feel pressured by the prompt. The second term, $R(x, g)$, is the reference-dependence term, reflecting that the consumer receives utility from comparing her consumption to the self-set goal.

Since we use a simple static model, the consumer chooses the energy consumption goal and effort simultaneously. This is without loss of generality and can be easily translated to a dynamic model in which the consumer first chooses the goal and then chooses energy consumption in the next period. We denote the **optimal intensive margin choices** by the pair (e_j^*, g_j^*), which is simply given by

$$(e_j^*, g_j^*) = \underset{e,g}{\arg\max} \left\{ \mathbb{E}[U(x, c_j) + B(\phi, x, g)G_j | e] \right\}. \tag{2}$$

Next, we derive the extensive margin choice in which the consumer decides whether to use the app. Aside from the utility she gets on the intensive margin, she may also like the energy app for other reasons. We define ϵ_j as the app-related taste parameter. We call $\epsilon = \epsilon_a - \epsilon_o$ the relative taste parameter and let it follow an atomless distribution function $F(\epsilon)$. Furthermore, we let u denote the relative utility on the intensive margin of choosing the outside option: $u = \mathbb{E}[v(x) - c_o(x, \cdot) + B(\phi, x, g_o^*)G_o | e_o^*] - \mathbb{E}[v(x) - c_a(x, \cdot) + B(\phi, x, g_a^*)G_a | e_a^*]$. The consumer chooses the app if the sum of the utility on the intensive margin and the app-specific taste parameter is larger for the app than for the outside option. Formally, the **optimal extensive margin choice** is to choose $j = a$ if and only if

$$\epsilon \geq u. \tag{3}$$

Demand for the app is then given by $D = \int_u dF(\epsilon)$.

We now show how this model can be used to empirically identify the effect of the goal setting nudge on consumer welfare. First, note that subjects in our experiment have $c_a(x, p, s, r, \pi) = px - \mathbb{1}_{r \geq x}(r - x)s\pi$, where p is the marginal energy price and s is a savings subsidy offered by the app a on every unit saved relative to a consumption benchmark r. r corresponds to the energy consumption in the previous month and $s = 1$ EUR. Since the subjects were randomly drawn to receive the subsidy, the subsidy is multiplied by the probability of winning the lottery, π.[19] Those who are not randomized into the subsidy group simply have $c_a = px$. Without loss of generality, we also set $c_o = px$.

Relating this model to our empirical results, we find that the goal-setting prompt had no statistically significant effect on energy consumption in the first two treatment periods but significantly reduced app adoption in the third treatment period.[20] Proposition 1 establishes that when the effect of the nudge on energy consumption is negligible, we can identify the (dis)utility subjects get from the goal-setting nudge through knowledge of a small set of sufficient statistics: the treatment effect of the nudge on app adoption, denoted $\Delta_G D$; the treatment effect of the savings subsidy on app adoption, denoted $\Delta_s D$; and the first-order energy cost savings due to the subsidy. Knowledge of these statistics also enables us to approximate the effect of the nudge on consumer welfare. We denote consumer surplus by $CS(G_a, s)$.

Proposition 1. *If the effect of the goal setting prompt on energy consumption is negligible, then willingness-to-pay for the nudge is given by*

$$\phi \approx \underbrace{\frac{\Delta_G D}{\Delta_s D}}_{\substack{\text{ratio of treatment} \\ \text{effects on extensive} \\ \text{margin}}} \underbrace{\mathbb{E}[(r - x) | e_a^*, r \geq x]\pi \Delta s}_{\substack{\text{first-order cost savings} \\ \text{on intensive margin} \\ \text{due to subsidy}}}. \tag{4}$$

The effect of the nudge on consumer surplus (CS) is then

$$CS(1, s) - CS(0, s) \approx \phi \left(D + \frac{\Delta_G D}{2} \right) \tag{5}$$

$$\approx \frac{\Delta_G D}{\Delta_s D} \mathbb{E}[(r - x) | e_a^*, r \geq x]\pi \Delta s \left(D + \frac{\Delta_G D}{2} \right). \tag{6}$$

If the nudge does not affect energy consumption, the consumer's willingness-to-pay for the nudge simply equals the direct utility she gets from the nudge. As we prove in the Appendix, a first-order approximation of this term is simply the ratio of treatment effects of the nudge and the subsidy on demand for the app, multiplied by the first-order electricity cost savings due to the subsidy (see Eq. (4)). These cost savings are identified in our experiment by averaging the difference $(r - x)$ for all control group subjects who have $r \geq x$ and then multiplying this average with the expected change in the subsidy.

[19] The assumption of a quasi-linear utility function also implies risk neutrality.

[20] While the coefficient of the nudge on energy consumption in the third treatment period is also insignificant, it may not have a causal interpretation because treated subjects were more likely to opt out of the app. Given the tightly estimated null effects in the previous periods, it is, however, unlikely that the effect on consumption was large in the third period. As an alternative, we could also impose additional structure to model the selection process and then estimate the effect on the intensive margin in the last period with this additionally imposed structure.

Table 5
Structural estimates and consumer welfare.

Willingness-to-pay for nudge (in EUR)	Effect on consumer welfare (in EUR per consumer)
−7.41	−4.32

Note: This table reports structural parameters calculated as described in Section 4.1. Specifically, we use the estimated treatment effects on the probability to use the app in period 4 conditional on pre-period utilization (see column 3 in Table A.4). All numbers therefore apply to the fourth experimental period.

Intuitively, Eq. (4) is an equivalent-price metric: it gives the necessary change in the price of the app that induces a demand response equivalent to the nudge. To obtain this metric, we first approximate the slope of the app demand function by dividing the change in the price of using the app (i.e., $\mathbb{E}[(r-x)|e_a^*, r \geq x]\pi\Delta s$) by the corresponding change in app demand ($\Delta_s D$). We then approximate the equivalent-price metric by dividing the change in demand due to the nudge ($\Delta_G D$) by the demand slope. This implicit price change equals willingness-to-pay for the nudge.

Knowledge of willingness-to-pay for the nudge allows us to approximate its effect on consumer surplus. Eq. (5) is obtained by a Taylor approximation up to second order and only involves one additional statistic compared to Eq. (4)—namely, control group demand for the app. Since this is obviously observed in our experiment, we have all of the ingredients to estimate the effect of the nudge on consumer surplus.

Eq. (5) provides two insights. To first order, the welfare loss is simply the disutility from the nudge multiplied by the share of consumers using the app, i.e., ϕD. To second order, we need to take into account that some consumers avoid the loss in utility from the nudge by no longer using the app (i.e., the share $\Delta_G D$). This reduces the total surplus loss by $\phi \frac{\Delta_G D}{2}$. Since selecting out of the app also means losing access to other app features (e.g., energy savings tips and meter scanning), only those consumers select out for whom the disutility from the nudge is larger than the benefits from the other features of the app.

4.2. Structural estimates

To calculate willingness-to-pay for the nudge, we use the first part of Proposition 1. Note that we must use the treatment effects on the probability to use the app in period 4 for the subsample of consumers who were still using the app in period 3. This is because the subjects were offered the savings subsidy after submitting the scan at the end of period 3. Thus, those who dropped out earlier, e.g., at the end of period 2, were never offered the subsidy. Table A.4 shows the results of a regression of treatments on the probability to use the app conditional on having used the app in the previous period. Column 3 involves the relevant results for the identification of the structural parameters. The nudge decreased the probability to use the app in period 4 by 12.4 percentage points, while the savings subsidy increased the probability by 2 percentage points. These are the estimates of the relevant treatment effects, $\Delta_G D$ and $\Delta_s D$. We also calculate the average first-order savings of the subsidy as described in the previous section. We find that $\mathbb{E}[(r-x)|e_a^*, r \geq x] = 68.72$ kWh and multiply this result by the winning probability, $\pi = 1.85\%$, and $\Delta s = 1$ EUR. This yields expected first-order savings of 1.27 EUR per consumer.

Accordingly, column 1 in Table 5 shows that the average consumer's willingness-to-pay for the nudge is –7.41 EUR. Thus, the average consumer is willing to give up 7.41 EUR to *avoid* the nudge.[21] This amount is relatively large and compares negatively to other nudges that intend to encourage resource conservation. Allcott and Kessler (2019) estimate a positive willingness-to-pay for home energy reports that compare a household's energy consumption with that of other similar households. Willingness-to-pay estimates in their study range from USD 2.58 to USD 4.36 for a bundle of four home energy reports.

Column 2 presents the effect of the nudge on consumer surplus. The average consumer loses 4.32 EUR in utility due to the nudge. Obviously, this number is closer to zero than willingness-to-pay because consumers can avoid the full loss in utility by reducing their probability of using the app.[22]

To put the implications of the nudge for consumers into perspective, we run a back-of-the-envelope calculation to show how a nationwide rollout of the app would have affected consumer welfare. Münster has approximately 310,000 residents, of which 343 subjects still used the app at the end of the third month after the rollout (i.e., 0.1 percent of all residents). Assuming that this ratio is the same for a nationwide rollout of the app, we would expect 91,858 people out of a total of 83.02 million German residents to use the app in the fourth month. The goal-setting nudge would then reduce consumer welfare by 396,827 EUR over a period of only four weeks. Additionally, one would have to subtract non-negligible costs for promoting the app nationwide.

[21] If consumers are risk-averse, our estimate is a lower bound of (negative) willingness-to-pay. While we might therefore overestimate the disutility of the nudge, any reasonable degree of risk aversion over these small-stake lotteries can only change our results slightly. Also recall that the average consumer in our sample is *not* loss-averse according to our survey measures.

[22] The effect of the nudge on *social* welfare would be the effect on consumer surplus minus the nudge provision costs. The latter are costs of developing and promoting the energy app. The total costs resulting from programming and promoting the energy app in our case were approximately 60,000 EUR. Since these are one-time fixed costs, we do not include them in the cost–benefit analysis. In that respect, the reported loss in consumer surplus due to the nudge can be considered as lower bound of the total loss in social welfare.

5. Discussion

In this section we discuss how our results resonate with the existing literature on goal setting. Our reduced-form and structural metrics suggest that the goal-setting nudge failed to deliver and is not a cost-effective policy tool for encouraging resource conservation. We find substantial evidence that the lack of success of our intervention is at least partially driven by the unfavorable selection of highly energy-literate and seemingly rational consumers into the subject pool.

Besides the poor targeting properties of the app, other factors may explain the null effect. One difference to other studies lies in the nature of energy consumption compared to other consumption dimensions. Much of the existing literature finds goal-setting prompts to be effective in domains where the goal targets one single action, such as getting vaccinated or voting. By contrast, repetitive behavior, such as regularly going to the gym, has been shown to be less affected by goal prompts (Carrera et al., 2018). Reducing electricity consumption often requires repetitive conservation actions and a high degree of awareness while enjoying energy services. Each conservation action, such as switching off a light bulb, typically only saves small amounts of electricity. This is particularly true for European households, which typically consume lower levels of electricity compared to US households, due to lower reliance on air conditioning, electric heating, and other energy-intensive activities (Andor et al., 2020).

Alternative options for conserving energy with less repetitive effort include energy-efficient home refurbishment or the installation of energy-saving appliances. Yet several other market frictions established in the literature may impede energy efficiency investments, such as credit constraints on low-income households (Berkouwer and Dean, 2022) and false beliefs about how a product's energy efficiency maps onto actual savings (e.g., Attari et al. (2010)).

There is, however, evidence from an event study by Harding and Hsiaw (2014) that a goal-setting prompt significantly reduces energy consumption. Their study investigates an energy savings program in the US that enabled subjects to set energy consumption goals at the utility provider's website. The goal-setting prompt significantly reduced electricity consumption by 8 percent in the first two months after program take-up and by 4.4 percent over the longer term. In contrast to our randomized controlled trial, their identification strategy crucially relies on the assumption that the timing of subjects' program adoption is quasi-random.

Since our study both directly randomizes the goal-setting feature and involves variation in the timing of app adoption, it also allows us to evaluate whether an event study would have identified the treatment effect in our sample. We therefore analyze whether our study would have yielded results similar to those of Harding and Hsiaw (2014) had we implemented the same identification strategy as opposed to the randomization into treatment and control groups.

To implement their event study approach, we restrict our data to consumers who are in the goal-setting group. We then compare electricity consumption of those who are in periods 2, 3 and 4 to those who are in baseline period 1, while controlling for time fixed effects. As in the event study by Harding and Hsiaw (2014), the identifying assumption is that the timing of consumer sign-up for the Energy App is quasi-random. This allows us to compare treatment users who are already setting goals with treatment users who are in baseline and do not (yet) set goals.[23] More formally, we use the empirical specification:

$$log(\text{kWh}_{iet}) = \gamma_i + \alpha_t + \delta Event_{ie} + \kappa S_{ie} + \xi_{iet}, \tag{7}$$

where $Event_{ie} \in \{0, 1\}$ is an indicator equal to zero when consumption of user i belongs to the baseline period, and equal to one when consumption belongs to one of the treatment periods 2, 3 and 4. Calendar-month fixed effects are denoted by α_t and are coded as described in Section 3.2. Individual fixed effects are given by γ_i, and the error term is denoted ξ_{iet}. The dummy S_{ie} controls for our additional financial reward, which is not present in the study by Harding and Hsiaw (2014).

Results are presented in Table 6. Column 1 includes all treatment periods, while column 2 only includes the first two months after the treatment started. In both specifications we find statistically significant coefficients that indicate reductions in electricity consumption of 9.5 and 7.1 percent, respectively. Interestingly, these coefficients are close to the estimated treatment effect of 8 percent over the same time period (two months) in Harding and Hsiaw (2014). Recall that our estimated 95 percent confidence intervals for the identified treatment effects in the RCT exclude these values. Our results imply that the timing of app adoption is not quasi-random in our experiment and may cast doubt on event study designs in these settings. Controlling for time and individual fixed effects does not eliminate the selection bias.

We stress that this is only suggestive evidence reconciling the difference to previous results. While the event study is not identified in our setting, event studies may be identified in other settings. Our exploratory analysis simply suggests that methodological differences may explain the lack of congruence to previous findings. In a broader sense, this also relates to the external validity of empirical work. As several studies have noted, the influence of site-specific effects may impede the transferability of findings across various contexts, such that our observed null-effects may not generalize (Allcott, 2015; Al-Ubaydli et al., 2017b; Vivalt, 2020; Andor et al., 2022).

6. Conclusion

Our study mimics a large-scale policy intervention designed to leverage insights from psychology to encourage resource conservation. We build on the promising results in the literature on goal setting and plan-making nudges and examine the role that goals can play in energy consumption. We scale the intervention through a newly developed energy app for mobile phones that

[23] Since we observe only one period of baseline consumption, we are not able to test for parallel pre-trends.

Table 6
Effect on electricity consumption as estimated from an event study.

	(1) Log (kWh)	(2) Log (kWh)
Event	−0.095**	−0.071**
	(0.039)	(0.030)
Savings subsidy	−0.050	
	(0.062)	
Period 4 consumption included	Yes	No
N	872	751

Note: The outcome variable is the logarithm of electricity consumption measured in kWh. Calendar-month fixed effects are included. The variable "Event" equals one when the observation belongs to one of the treated periods: E2, E3, and E4. It equals zero if the observation belongs to the baseline period E1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered on subject level are in parentheses.

is easy to use and accessible to the majority of the population. We monetize the welfare effects of our intervention for consumers by randomly offering financial incentives that reward app utilization.

Despite substantial marketing efforts and financial incentives to participate, we find surprisingly little demand for the energy app. Those subjects adopting the energy app do not alter their actual energy consumption in response to the nudge despite setting meaningful goals that are highly predictive of future consumption. Observable subject characteristics point to suboptimal targeting properties of the app as a likely mechanism for the null effect. The average subject who selected into the pool of app users has an already low baseline level of energy consumption. A complementary survey eliciting behavioral parameters shows that the average user is neither present-focused nor loss averse—the two core features in the theoretical literature that explain why self-set goals affect behavior.

Further, the nudge significantly reduced the probability of app use over time, indicating that the goal-setting prompt caused direct disutility by pressuring subjects. Using random price variation, we estimate that the average user is willing to pay a relatively large amount of 7.41 EUR to *avoid* the nudge. Structural estimates imply that a goal-setting prompt could cause substantial welfare losses if implemented nationwide.

Our results are also relevant for current policy debates on digital consumer technologies – referred to as "smart" devices – as potential measures to reduce energy consumption. Both the low demand for the energy app and the null effects of the nudge among those selecting into the app suggest a limited role of the use of mobile applications to scale up behaviorally-motivated energy policies.

Finally, it is important to note that our results may be specific to energy conservation. Previous studies in other fields show that goal-setting nudges can help people to follow through with their intentions. Our results do not stand in contrast to these studies, but rather show the importance of identifying the fields in which goal-setting nudges can be effective. We encourage future research to isolate the particular factors that can predict the success of nudges in different areas of public policy.

Data availability

The data that has been used is confidential.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.euroecorev.2023.104612.

References

Achtziger, Anja, Gollwitzer, Peter M., Sheeran, Paschal, 2008. Implementation intentions and shielding goal striving from unwanted thoughts and feelings. Pers. Soc. Psychol. Bull. 34 (3), 381–393.

Al-Ubaydli, Omar, List, John A, LoRe, Danielle, Suskind, Dana, 2017a. Scaling for economists: Lessons from the non-adherence problem in the medical literature. J. Econ. Perspect. 31 (4), 125–144.

Al-Ubaydli, Omar, List, John A., Suskind, Dana L., 2017b. What can we learn from experiments? Understanding the threats to the scalability of experimental results. Amer. Econ. Rev. 107 (5), 282–286.

Al-Ubaydli, Omar, List, John A., Suskind, Dana, 2019. The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments. Working Paper.

Allcott, Hunt, 2011. Social norms and energy conservation. J. Public Econ. 95 (9–10), 1082–1095.

Allcott, Hunt, 2015. Site selection bias in program evaluation. Q. J. Econ. 130 (3), 1117–1165.

Allcott, Hunt, Kessler, Judd B., 2019. The welfare effects of nudges: A case study of energy use social comparisons. Am. Econ. J.: Appl. Econ. 11 (1), 236–276.

Amabile, Teresa M., Kramer, Steven J., et al., 2011. The power of small wins. Harv. Bus. Rev. 89 (5), 70–80.

Andor, Mark A., Gerster, Andreas, Peters, Jörg, 2022. Information campaigns for residential energy conservation. Eur. Econ. Rev. 144, 104094.

Andor, Mark A, Gerster, Andreas, Peters, Jörg, Schmidt, Christoph M, 2020. Social norms and energy conservation beyond the US. J. Environ. Econ. Manag. 103, 102351.

Armitage, Christopher J., Arden, Madelynne A., 2008. How useful are the stages of change for targeting interventions? Randomized test of a brief intervention to reduce smoking. Health Psychol. 27 (6), 789.

Attari, Shahzeen Z., DeKay, Michael L., Davidson, Cliff I., De Bruin, Wändi Bruine, 2010. Public perceptions of energy consumption and savings. Proc. Natl. Acad. Sci. 107 (37), 16054–16059.

Berkouwer, Susanna B., Dean, Joshua T., 2022. Credit, attention, and externalities in the adoption of energy efficient technologies by low-income households. Amer. Econ. Rev. 112 (10), 3291–3330.

Bernheim, B. Douglas, Taubinsky, Dmitry, 2019. Behavioral public economics. In: Bernheim, B. Douglas, Della Vigna, Stefano, Laibson, David I. (Eds.), Handbook of Behavioral Economics, Vol. 1. Elsevier.

Blonz, Joshua, Palmer, Karen, Wichman, Casey J, Wietelman, Derek C, et al., 2021. Smart Thermostats, Automation, and Time-Varying Prices. Working Paper.

Booij, Adam S., Van Praag, Bernard M.S., Van De Kuilen, Gijs, 2010. A parametric analysis of prospect theory's functionals for the general population. Theory and Decision 68 (1–2), 115–148.

Brandon, Alec, Clapp, Christopher M, List, John A, Metcalfe, Robert D, Price, Michael, 2022. The Human Perils of Scaling Smart Technologies: Evidence from Field Experiments. Working Paper.

Butera, Luigi, Metcalfe, Robert, Morrison, William, Taubinsky, Dmitry, 2022. Measuring the welfare effects of shame and pride. Amer. Econ. Rev. 112 (1), 122–168.

Carrera, Mariana, Royer, Heather, Stehr, Mark, Sydnor, Justin, Taubinsky, Dmitry, 2018. The limits of simple implementation intentions: Evidence from a field experiment on making plans to exercise. J. Health Econ. 62, 95–104.

Carver, Charles S., Scheier, Michael, 1990. Principles of Self-Regulation: Action and Emotion. The Guilford Press.

Charness, Gary, Gneezy, Uri, Halladay, Brianna, 2016. Experimental methods: Pay one or pay all. J. Econ. Behav. Organ. 131, 141–150.

Cohen, Jonathan D, Ericson, Keith, Laibson, David, White, John Miles, 2019. Measuring time preferences. J. Econ. Lit..

Coller, Maribeth, Williams, Melonie B., 1999. Eliciting individual discount rates. Exp. Econ. 2 (2), 107–127.

Czibor, Eszter, Jimenez-Gomez, David, List, John A., 2019. The dozen things experimental economists should do (more of). South. Econ. J. 86 (2), 371–432.

DellaVigna, Stefano, 2018. Structural behavioral economics. In: Handbook of Behavioral Economics: Applications and Foundations 1, Vol. 1. Elsevier, pp. 613–723.

DellaVigna, Stefano, Linos, Elizabeth, 2022. RCTs to scale: Comprehensive evidence from two nudge units. Econometrica 90 (1), 81–116.

DellaVigna, Stefano, List, John A., Malmendier, Ulrike, 2012. Testing for altruism and social pressure in charitable giving. Q. J. Econ. 127 (1), 1–56.

DellaVigna, Stefano, List, John A., Malmendier, Ulrike, Rao, Gautam, 2016. Voting to tell others. Rev. Econom. Stud. 84 (1), 143–181.

Dolan, Paul, Metcalfe, Robert, 2015. Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation. Working Paper.

Falk, Armin, Becker, Anke, Dohmen, Thomas, Enke, Benjamin, Huffman, David, Sunde, Uwe, 2018. Global evidence on economic preferences. Q. J. Econ. 133 (4), 1645–1692.

Falk, Armin, Becker, Anke, Dohmen, Thomas, Huffman, David, Sunde, Uwe, 2023. The preference survey module: A validated instrument for measuring risk, time, and social preferences. Manage. Sci. 69 (4), 1935–1950.

Federal Statistical Office of Germany, 2019. Federal Statistical Office of Germany: Stromverbrauch der privaten Haushalte nach Haushaltsgrößenklassen. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Umwelt/Materialfluesse-Energiefluesse/Tabellen/stromverbrauch-haushalte.html, (Retrieved: 06/29/2020).

Ferraro, Paul J., Miranda, Juan Jose, Price, Michael K., 2011. The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. Am. Econ. Rev.: Pap. Proc. 101 (3), 318–322.

Ferraro, Paul J., Price, Michael K., 2013. Using nonpecuniary strategies to influence behavior: Evidence from a large-scale field experiment. Rev. Econ. Stat. 95 (1), 64–73.

Fishbach, Ayelet, Koo, Minjung, Finkelstein, Stacey R., 2014. Motivation resulting from completed and missing actions. Adv. Exp. Soc. Psychol. 50, 257–307.

Fünfgeld, Christian, Tiedemann, Remo, 2000. Anwendung Der Repräsentativen VDEW-Lastprofile: Step-By-Step. Tech. Report, VDEW.

Harding, Matthew, Hsiaw, Alice, 2014. Goal setting and energy conservation. J. Econ. Behav. Organ. 107, 209–227.

Heath, Chip, Larrick, Richard P., Wu, George, 1999. Goals as reference points. Cogn. Psychol. 38 (1), 79–109.

Hsiaw, Alice, 2013. Goal-setting and self-control. J. Econom. Theory 148 (2), 601–626.

Ilgen, Daniel R., Fisher, Cynthia D., Taylor, M. Susan, 1979. Consequences of individual feedback on behavior in organizations. J. Appl. Psychol. 64 (4), 349.

Ilies, Remus, Judge, Timothy A., 2005. Goal regulation across time: the effects of feedback and affect. J. Appl. Psychol. 90 (3), 453.

Imai, Taisuke, Rutter, Tom A., Camerer, Colin F., 2021. Meta-analysis of present-bias estimation using convex time budgets. Econ. J. 131 (636), 1788–1814.

Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47 (2), 263–291.

Koch, Alexander K., Nafziger, Julia, 2011. Self-regulation through goal setting. Scand. J. Econ. 113 (1), 212–227.

Koch, Alexander K., Nafziger, Julia, 2019. Correlates of narrow bracketing. Scand. J. Econ. 121 (4), 1441–1472.

Kőszegi, Botond, Rabin, Matthew, 2006. A model of reference-dependent preferences. Q. J. Econ. 121 (4), 1133–1165.

Laibson, David, 1997. Golden eggs and hyperbolic discounting. Q. J. Econ. 112 (2), 443–478.

Latham, Gary P., 2004. The motivational benefits of goal-setting. Acad. Manag. Perspect. 18 (4), 126–129.

Latham, Gary P., Locke, Edwin A., 2006. Enhancing the benefits and overcoming the pitfalls of goal setting. Organ. Dyn. 35 (4), 332–340.

Latham, Gary P., Mitchell, Terence R., Dossett, Dennis L., 1978. Importance of participative goal setting and anticipated rewards on goal difficulty and job performance. J. Appl. Psychol. 63 (2), 163.

Latham, Gary P., Seijts, Gerard, Crim, Dan, 2008. The effects of learning goal difficulty level and cognitive ability on performance. Can. J. Behav. Sci./Revue Can. Sci. Comportement 40 (4), 220.

List, John A., 2020. Non est Disputandum de Generalizability? A Glimpse into The External Validity Trial. Working Paper.

List, John A, Rodemeier, Matthias, Roy, Sutanuka, Sun, Gregory, 2022. Judging Nudging: Understanding the Welfare Effects of Nudges Versus Taxes. Working Paper.

Locke, Edwin A., 2004. Linking goals to monetary incentives. Acad. Manag. Perspect. 18 (4), 130–133.

Locke, Edwin A., Latham, Gary P., 1990. A Theory of Goal Setting & Task Performance.. Prentice-Hall, Inc.

Locke, Edwin A, Shaw, Karyll N, Saari, Lise M, Latham, Gary P, 1981. Goal setting and task performance: 1969–1980. Psychol. Bull. 90 (1), 125.

Milkman, Katherine L, Beshears, John, Choi, James J, Laibson, David, Madrian, Brigitte C, 2011. Using implementation intentions prompts to enhance influenza vaccination rates. Proc. Natl. Acad. Sci. 108 (26), 10415–10420.

Nickerson, David W., Rogers, Todd, 2010. Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. Psychol. Sci. 21 (2), 194–199.

Pellerano, José A, Price, Michael K, Puller, Steven L, Sánchez, Gonzalo E, 2017. Do extrinsic incentives undermine social norms? Evidence from a field experiment in energy conservation. Environ. Resour. Econ. 67 (3), 413–428.

Podsakoff, Philip M., Farh, Jiing-Lih, 1989. Effects of feedback sign and credibility on goal setting and task performance. Organ. Behav. Hum. Decis. Process. 44 (1), 45–67.

Rodemeier, Matthias, 2023. Willingness to Pay for Carbon Mitigation: field Evidence from the Market for Carbon Offsets. Working Paper.

Rodemeier, Matthias, Löschel, Andreas, 2020. The Welfare Effects of Persuasion and Taxation: Theory and Evidence from the Field. Working Paper.

Rogers, Todd, Milkman, Katherine L, John, Leslie, Norton, Michael I, 2013. Making the Best-Laid Plans Better: How Plan Making Increases Follow-Through. Working Paper.

Sarabi, Roghayeh Ershad, Sadoughi, Farahnaz, Orak, Roohangiz Jamshidi, Bahaadinbeigy, Kambiz, 2016. The effectiveness of mobile phone text messaging in improving medication adherence for patients with chronic diseases: a systematic review. Iran. Red Crescent Med. J. 18 (5).

Seijts, Gerard H., Latham, Gary P., 2005. Learning versus performance goals: When should each be used? Acad. Manag. Perspect. 19 (1), 124–131.

Stock, Jennifer, Cervone, Daniel, 1990. Proximal goal-setting and self-regulatory processes. Cogn. Ther. Res. 14 (5), 483–498.

Vervloet, Marcia, Linn, Annemiek J, van Weert, Julia CM, De Bakker, Dinny H, Bouvy, Marcel L, Van Dijk, Liset, 2012. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J. Am. Med. Inform. Assoc. 19 (5), 696–704.

Vivalt, Eva, 2020. How much can we generalize from impact evaluations? J. Eur. Econom. Assoc. 18 (6), 3045–3089.

Walasek, Lukasz, Mullett, Timothy L., Stewart, Neil, 2018. A Meta-Analysis of Loss Aversion in Risky Contexts. Working Pape.

Werthschulte, Madeline, Löschel, Andreas, 2021. On the role of present bias and biased price beliefs in household energy consumption. J. Environ. Econ. Manag. 109, 102500.