WIK

Working Paper

No. 7

Saving behaviors of private households under varying tariff structures, price levels and incentives – Experimental evidence

Franziska Harpenau

Katrin Marques Magalhaes

Dr. Nico Steffen

Dr. Lukas Wiewiorra

Bad Honnef, December 2023

WIK
Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste

IMPRINT

WIK WORKING PAPERS

The working papers published in the series constitute work in progress circulated to stimulate discussion and critical comments. Views expressed represent exclusively the authors' own opinions and do not necessarily reflect those of the editor.

WIK Wissenschaftliches Institut für
Infrastruktur und Kommunikationsdienste GmbH
Rhöndorfer Str. 68
53604 Bad Honnef, Germany
E-Mail: info@wik.org
www.wik.org

Person authorised to sign on behalf of the organisation

General Manager	Dr Cara Schwarz-Schilling
Director	Alex Kalevi Dieke
Head of Department	
Smart Cities/Smart Regions	
Director	
Head of Department	
Networks and Costs	Dr Thomas Plückebaum
Director	
Head of Department	
Regulation and Competition	Dr Bernd Sörries
Head of Administration	Karl-Hubert Strüver
Chairperson of the Supervisory Board	Dr Thomas Solbach
Registered at	Amtsgericht Siegburg, HRB 7043
Tax No.	222/5751/0926
VAT-ID	DE 329 763 26

Date: July 2023

Saving behaviors of private households under varying tariff structures, price levels and incentives – Experimental evidence

Franziska Harpenau[1] Katrin Marques Magalhaes[1]
Nico Steffen[1] Lukas Wiewiorra[1]

[1] Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK)

December 22, 2023

Abstract

More efficient and sustainable energy consumption behaviors are crucial to mitigate the adverse effects of climate change. This paper examines how dynamic personal pricing and externality cost incentives interact and affect energy conservation behaviors. We conduct an online lab experiment in which participants complete real-effort tasks under different cost schemes. Increasing personal costs that reduce individual bonuses, significantly decreases participants' energy usage, although it requires more effort in the form of additional time. However, emphasizing increases in externality costs, representing environmental damage through reduced donations, does not impact performance. This suggests that the introduction of such prosocial incentives matters more than their magnitude. While environmental attitudes predict baseline usage, they do not affect marginal responses to price changes. Our results provide novel evidence on the motivational nuances underlying energy conservation and have key implications for policies considering a combination of incentives.

Keywords: Energy consumption, Energy conservation, Behavioral, Environment
JEL Classification: Q41, Q56, D91

[1]We are thankful to Marvin Benedikt Riemer for excellent research assistance.
E-mail addresses: F.Harpenau@wik.org, K.Marques-Magalhaes@wik.org, N.Steffen@wik.org, L.Wiewiorra@wik.org
Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste (WIK), Rhöndorfer Str. 68, 53604 Bad Honnef, Germany.

1 Introduction

The global consumption of energy and resources[1] is at the center of an urgent debate on sustainable behavior and climate protection. With growing awareness of environmental issues, the role of incentives in changing individual energy usage has become a critical focus. Financial incentives through prices, tax breaks or rewards, and intrinsic motivators stemming from personal beliefs or social factors, may impact energy conservation. Understanding the influence of these incentives is key, not just for human decision-making theories but for developing policies that effectively promote sustainability.

This paper investigates how personal pricing and externality cost incentives interact to shape energy-saving behaviors. Specifically, does heightening the personal costs of usage reduce consumption? Does conveying more pronounced externalities (e.g. of environmental damage) alter usage? And how do the two types of incentives combine - is their joint impact simply additive or does it display more complex interplays? Answering these questions provides novel evidence on the motivational nuances underlying energy conservation. Much research has examined how either personal or prosocial incentives independently affect ecological behavior. However, their interplay, especially in response to marginal changes versus initial introduction, remains less understood. Testing their combined effect is also critical for policies aiming to balance the right incentives.

We address this gap through an online lab experiment mimicking energy consumption decisions across different cost schemes. This controlled approach isolates the causal impacts of personal and externality price changes. We document three core findings: First, increased personal costs significantly reduce usage, consistent with prior evidence. However, changes in externality costs have no impact, contrary to some studies on social and intrinsic incentives. Second, combining the two price changes leads to additive effects, with no signs of synergies or interference. Third, environmental attitudes predict baseline usage but not marginal responses. Our study makes two key contributions. We incorporate experimental variation in externality pricing of energy usage, unlike past focus on norms or goals. This sheds light on the different motivations underlying general pro-environmental attitudes, which predict overall lower energy use, versus context-specific responses to price changes, which do not vary by environmental views in our study setting. We also employ a design suited to identify standalone versus combined effects of pricing incentives. Our findings carry important implications for policies considering combinations of incentives to promote energy conservation.

The remainder of the paper is structured as follows. Section 2 elaborates on the relevant strands of literature to which we contribute. Section 3 explains the laboratory experiment in detail and presents hypotheses with the resulting treatment definitions. Section 4 addresses the participant sample and subsequent analyses of the treatment effects. Fi-

[1]In this paper, the term 'energy (consumption)' is used to refer collectively to the usage of resources including electricity, transportation fuels, natural gas, and other energy sources, as well as water.

nally, in Section 5 we delve into the implications of our findings and draw conclusions based on our results.

2 Literature

Our paper is associated with three related, but distinct research domains. First, this paper relates to the research field examining the effect of higher prices on energy consumption. While a reduced energy demand when facing higher prices makes rational sense, behavioral biases, a completely inelastic demand or the like might lead to a different reaction in reality (e.g. Tversky and Kahneman, 1974; Smith, 2003; Gabaix, 2019; Zhao *et al.*, 2015). However, in the field of energy usage, the expected rational response is backed up by several studies. Higher energy prices during peak-demand hours reduced consumption during those hours by about 14-17% and even slightly during the rest of the day in the field experiment of Ito *et al.* (2018). In line with this evidence, monetary rewards for energy conservation during peak-demand hours or in general reduced energy consumption by 4-6% (Murakami *et al.*, 2022; Mizobuchi and Takeuchi, 2013). Moreover, using a regression discontinuity design, Bastos *et al.* (2015) identified a reduction of roughly 4% in gas consumption as the result of a price increase. The generality of those results is backed by the recent meta-analysis of Sloot and Scheibehenne (2022), which includes a large variety of geographic regions, that observed a 2% reduction in energy consumption when the financial incentives target the overall consumption and a 10% reduction when the focus is the peak consumption.

Second, this paper is also linked with the literature concerning the effects of non-extrinsic incentives on exerted effort and connected outcomes. Related to this, Tonin and Vlassopoulos (2015) showed in an online experiment that the presence of donations, independent of their amount, hence also independent of the ratio of the private benefit to the donation, and independent of whether they are contingent on performance or not, increases productivity by around 13%. They therefore concluded that this enhancement is neither explained by pure altruism nor an efficiency concern but rather by a new social dimension in the perception of the task. Furthermore, splitting the sample by whether individuals have volunteered in the previous year or not leads to an insignificant effect for those who have not but increases the treatment effect for the rest. Those results are supported by Charness *et al.* (2016). In their experimental design, the introduction of donations leads to a higher willingness to work but no additional increase in this willingness is achieved if the amount of donations is quadrupled. As an explanation for those findings, they quote the model of impure altruism (Andreoni, 1989) and related to that an increased utility due to a warm glow from donating per se, as opposed to additional utility derived from marginal changes. The same result holds in the experiment of Imas (2014) in which the introduction of donations increased the exerted effort, but multiplying the donations per unit of provided effort times 40 did not influence the effort. Regarding the field of energy consumption, Buckley and Llerena (2022) demonstrated in their common pool resource game that the usage of a nudge, which informs individuals whether they consumed more or less than the social optimum through an emoticon, results in a reduction of virtual energy consumption by around 19%. This effect is especially pronounced for individuals with a high environmental sensitivity. Moreover, the field experiment by Dolan and Metcalfe (2015) illustrates that using a social norm, which compares the subject's energy use to their neighbors' one, reduces energy consumption by 4.4%. Additionally, the field experiment of Ghesla *et al.* (2020) in which subjects were assigned an energy savings goal and in further treatments were additionally incentivized by the planting of a tree when meeting the savings goal led to a reduction in energy consumption. This effect was especially large when the consequences of not meeting the goal were framed as the loss of a planted tree. In this treatment, the electricity savings amounted to 5%. Furthermore, additionally to varying prices, Ito *et al.* (2018) also examined the effect of "moral suasion", namely sending messages promoting energy reduction during peak-demand hours. This intervention reduced energy consumption as well; however, the effect is only half the size of the price effect and not as persistent over time as the price effect.

Third, this paper is related to the analysis of synergies when combining interventions that target the intrinsic and the extrinsic motivation respectively. As pointed out by Drews *et al.* (2020), most studies trying to identify the effects of treatments targeting those types of motivations on environmentally friendly behavior do not regard all necessary treatments for the correct identification of the presence or absence of synergies among the treatments. For this, a control group, both interventions alone and the combination of them would need to be regarded. Following this design, the online experiment concerning the energy consumption of a virtual washing machine by Fanghella *et al.* (2021) showed that the effect of combining a self-set goal with feedback about one's performance with a financial reward when meeting this goal and the stand-alone effects of the single treatments are all indistinguishable from each other. For the subgroup with high goals, the effects of the self-set goal with feedback and of the combination are significant and indistinguishable from each other, though the former coefficient is slightly bigger. Therefore, no positive synergy effect was visible, and the combination of the treatments might even backfire. As a potential explanation for those results, they quote the crowding-out of the individuals' attention, i.e. a distraction of attention from the pecuniary benefits due to the nudge.

Examining the carbon footprint of purchases made in an online shopping experiment, a recall of past environmentally friendly behaviors coupled with displaying the amount of saved CO_2 equivalents due to this behavior and a carbon tax were introduced alone and in conjunction by Panzone *et al.* (2021); both treatments led to a reduction in the carbon footprint, but the interaction effect of the treatments in conjunction is not significant. Similarly, exploring the effects of price decreases of hypothetical train tickets as well as the introduction of a norm promoting train usage, Hilton *et al.* (2014) observed that both interventions increase the fraction of individuals choosing the train compared to the plane, but they did not find a significant interaction of those treatments. Furthermore, the meta-analysis of Alt *et al.* (2022), which deals with synergies of pro-environmental interventions, concluded that negative synergy effects are expected, meaning the effect of the combination of treatments is not as pronounced as the sum of the individual effects; but it is expected to be still bigger than the effect of any intervention itself. Specifically, Alt *et al.* (2022) find that combining interventions from different domains, i.e. one from the traditional economic and one from the behavioral domain, is more effective than combining interventions from the same domain and in the former case no negative or positive synergy effects are expected. Moreover, the environmental behavior that seems to be least affected by negative synergy effects is energy conservation for which no significant synergy effect was found. However, their definition of energy conservation seems to be very broad and the interventions from the behavioral domain differ greatly.

Our paper enriches this previous literature by two main contributions. On the one hand, by using and varying the (displayed) benefit to society (through donations) in an experimental setting of energy consumption, this paper sheds light on the effects of varying externalities of energy consumption. Previously, papers studying the effects of non-extrinsic motivators on energy consumption have instead focused on the introduction of goal setting, norms and social comparison. On the other hand, we are employing a study design that is suited to correctly identify any stand-alone as well as synergy effects in the energy context such that we are contributing to a still very scarce literature, especially when considering the different focus of previous studies regarding their employed non-extrinsic motivators.

3 Methodology

To scrutinize the impact of different cost components and changes therein on consumers' energy consumption decisions, an online laboratory experiment simulating an energy consumption scenario has been conducted. Consumers' decisions were incentivized and individuals were sourced from the Prolific (2023) academic panel. The experiment lasted on average 37 minutes. The participants' remuneration, which they received upon completion of the experiment, is based on their performance in a real-effort task (cf. Augenblick *et al.*, 2015; Lezzi *et al.*, 2015; Benndorf *et al.*, 2019). In addition to a fixed participation fee of £3, the performance-based part of the payout ranged from £0 to £4. The sessions were held online from November 23^{rd} to November 27^{th} 2023. Each participant accessed the sessions using their personal desktop device and web browser[2]. A total of 603 subjects took part in this study. The experiment's interface and features were created using the LIONESS toolkit developed by Giamattei *et al.* (2020). Further details regarding the experimental design and variations in treatments will be discussed in the subsequent sections.

[2]Due to technical reasons, the participants had to use and confirm to be using either a Microsoft Edge or Google Chrome browser.

3.1 Experimental design

The real-effort task

The energy consumption decision setting is mimicked by 16 real-effort tasks, which are split into two rounds of eight tasks each. The structure of those tasks aligns with a real-effort task as defined by Charness *et al.* (2018), providing significant external validity. The tasks consist of clicking on occurrences of the number three in a grid of digits, representing an adaption of the standard "counting-zeros" task (Abeler *et al.*, 2011). Similar to Werthschulte (2023), participants could only continue if and only if all target symbols and no other symbols were selected. The structure of the real-effort task means that no prior knowledge is required and the test subjects have no other incentives than those set. Consequently, the cost of effort can be well represented in such tasks. Compared to the original counting-zeros implementation of Abeler *et al.* (2011) we did not use zeros and ones, in order to increase the difficulty and to prevent potential cheating efforts by participants[3]. Instead, we implement threes, eights, and nines, which have a similar shape, making it harder to identify and select the right number. Below the grid, a button to switch on the light and hereby lower the task's difficulty by changing the background color from the default color of dark blue to white is provided. During each round, the difficulty of the tasks without light usage is progressively increased by a reduced contrast owing to an increasingly dark background. Figures 1a and 1b depict versions of the task without and with light usage respectively, displaying also the real-time cost meters running below. No limits concerning the number of times the light switch is pressed or the duration of light usage have been set.

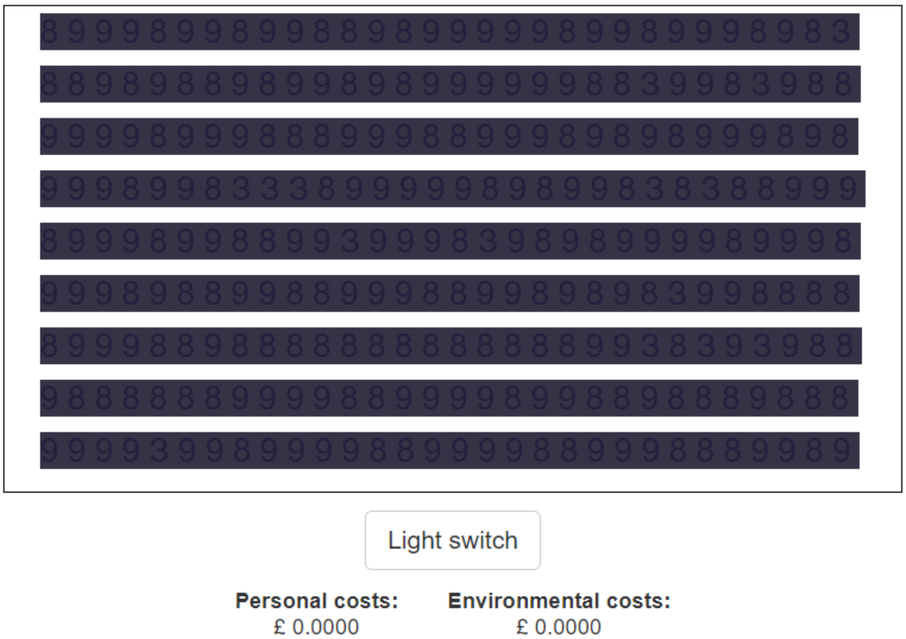
The participants have the option to turn the light on or off at any point during the work on any task. This reduces the required effort for completion but brings along two kinds of costs for energy consumption. On the one hand, their personal payment decreases with every second of light usage. On the other hand, the amount of an environmentally-friendly donation diminishes simultaneously. This donation represents the negative externalities, or more precisely, the environmental costs of energy usage.

Procedure

The online experiment starts for all treatment groups with a detailed explanation of the procedure, the payout scheme and the real-effort tasks. Participation and behavior are monetarily incentivized. The better the individual performance in the experiment, the higher the monetary payout. This provides a high internal validity of the results. In our setting, the payment consists of a base participation fee and a personal bonus they can attain in the real-effort tasks. The donation payout depends solely on the effort. At the beginning of each round, participants receive an endowment of £2 for the personal bonus and an endowment of £0.5 for the donation bonus from which the respective costs de-

[3] We also used a non-standard font to display the character 3 in order to further prevent abuse of the text search function.

(a) Real-effort task without light usage



(b) Real-effort task with light usage

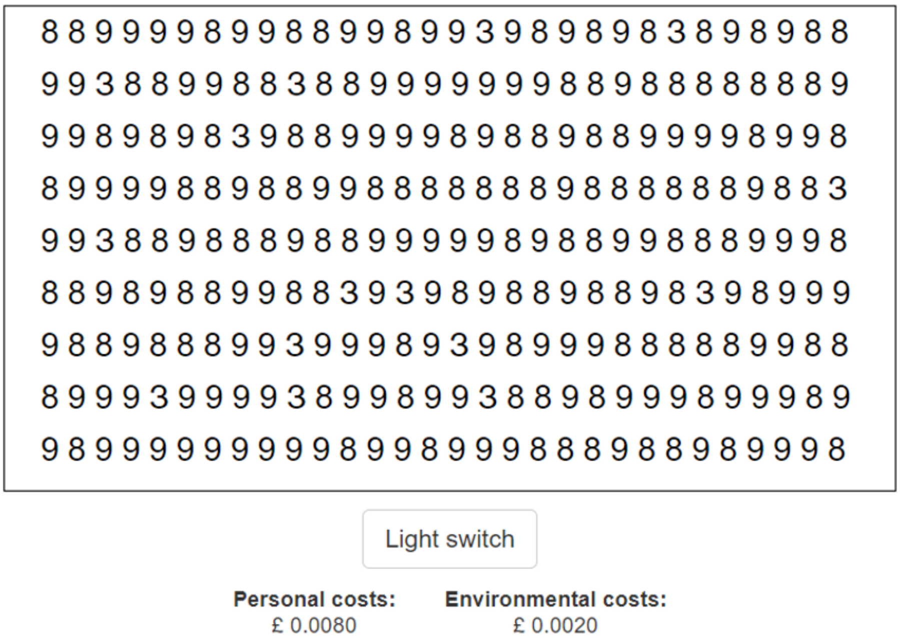


Figure 1: Real-effort task

pending on the exerted effort and the therewith related light usage are deducted. For the donation, the organization *One Tree Planted Inc.* was chosen, since they are committed to protecting the environment and plant a tree for every £0.82 (1 USD) they receive. Since up to £1 can be gained by each participant for a donation, they thereby have a real incentive to make a direct contribution to the environment, even at this relatively small stake. The donation organization is an official partner of the United Nations Decade on Ecosystem Restoration 2021-2030 and it has a platinum-level Guidestar participant, demonstrating the commitment to transparency. To establish trust and confirm the real contribution, a donation receipt was promised to the participants and was provided after the experiment had concluded.

Before a test round starts, the participants see two exemplary screenshots of the real-effort task, one with and one without light. This is followed by the test round consisting of one task and control questions to check whether the test subjects have understood the task. Note that the test round, like the actual tasks in the two rounds, can only be completed once the task has been fully solved. Afterward, the first of the two actual rounds of tasks begins. The prices for the first round are set equally for all 603 participants: For every second of light usage, 0.2 pennies are deducted from the personal bonus and 0.05 pennies from the donation bonus. The more light is used across the eight tasks, the lower both bonuses get. In each round, the occurring costs are presented in real-time below the grid and are added up along the tasks, visible at all times. While previous studies have analyzed the effect of introducing such "meters" (e.g. Gans *et al.*, 2013), this instrument and thus potential feedback effects are held constant throughout all treatments of this experiment.

After the first round, the resulting bonuses thus far are then summarized and shown to the participants. The second round follows the same logic. Here, however, the treatments differ in the amount of personal and externality costs. The next section will present the treatment groups in more detail. In the final part of the experiment, the participants are asked to complete a survey that consists of socio-demographic questions as well as questions relating to altruism, patience and environmental attitudes. The survey on altruism and patience is based on Falk *et al.* (2018) and Falk *et al.* (2016). The papers describe a validated survey module used to measure important economic preferences such as risk aversion, patience, trust, altruism, and reciprocity. It is a reliable and cost-effective tool whose survey instruments accurately predict preferences as observed in incentivized choice experiments. The ecological attitude was quantified using a revised version of the widely used New Environmental Paradigm (NEP) Scale. Dunlap *et al.* (2000) define 15 questions to measure pro-environmental orientations, which we incorporated in the survey. On the last page, the participants were informed about their total payout and the donation amount from both rounds.

3.2 Treatments and Hypotheses

There are four treatment groups which differ only in the costs they are facing in the second round. In the first round, the personal costs are 0.2 pennies and the environmental costs are 0.05 pennies per second of light usage for all participants. These costs can double depending on the treatment group. In order to accurately estimate the impacts of changing the cost components on their own as well as synergy or crowding-out effects, one baseline and three additional treatments have been implemented. Table 1 shows for each group if personal and environmental costs have doubled in the second round compared to the first round.

Table 1: Overview of cost changes (doubling of prices) in the second round

Treatment	**Baseline**	**Personal**	**Environment**	**Both**
Personal costs	No	Yes	No	Yes
Environmental costs	No	No	Yes	Yes

No change in costs occurs for the control group (Baseline), which was designed to account for potential fatigue or learning effects over the course of the experiment. Both directions have been observed in many settings and need to be controlled for (Savage and Waldman, 2008; Matthews and Desmond, 2002; Araujo *et al.*, 2016; Benndorf *et al.*, 2019), in order to not attribute any treatment effects to those changes that are already occurring in the absence of any changes in costs.

The treatment labeled Personal, in which the environmental costs remain unchanged at 0.05 pennies, but the personal energy costs double to 0.4 pennies, examines the impact of a changed financial incentive on consumption behavior. The amount of both cost components and the level of the price increases were calibrated using pretests. Various studies consistently reveal reduced energy consumption in response to heightened prices. For instance, during peak-demand hours, Ito *et al.* (2018) observed a significant decrease in energy consumption in response to higher prices; also interventions offering monetary rewards for conservation resulted in lower consumption (Murakami *et al.*, 2022; Mizobuchi and Takeuchi, 2013). A meta-analysis by Sloot and Scheibehenne (2022) across diverse regions showed a general 2% decrease in overall energy usage in response to heightened financial stimuli and a remarkable 10% drop during peak hours when the incentives targeted those specific consumption periods. Despite potential behavioral influences, these consistent findings across studies affirm the association between higher personal energy prices and increased energy conservation. The treatment labeled Personal is directly linked to our first hypothesis:

Hypothesis 1. *A higher personal energy price leads to more energy savings.*

The treatment labeled Environment allows analysis of the impact of more pronounced externalities on the energy consumption behavior as the personal costs remain unchanged but the environmental costs are doubled to 0.1 pennies compared to the first round. Research by Tonin and Vlassopoulos (2015) suggests that the presence of donations, irrespective of size or contingency, enhances productivity, hinting at a social dimension in task perception. Similarly, findings by Charness *et al.* (2016) and Imas (2014) indicate increased effort with the presence of donations but no additional impact when donation amounts per unit of effort increase. Relating these behavioral nuances to energy conservation, studies by Buckley and Llerena (2022) and Ghesla *et al.* (2020) showcase reductions in energy usage through nudges and incentives without explicit connection to heightened environmental prices. These insights underscore the complexity of human motivations in energy consumption, suggesting that while non-extrinsic motivators influence behavior, the direct impact of increased environmental prices on substantial energy savings remains uncertain. Thus, we formulate our second hypothesis as follows:

Hypothesis 2. *While an explicit environmental price can lead to energy savings, a higher price does not affect energy savings further.*

In the treatment labeled Both, the personal as well as the environmental cost double in the second round compared to the first one. Therefore, personal costs for light usage are 0.4 pennies per second and environmental costs are 0.1 pennies per second. Intuitively one would guess, that higher price levels for both price components will result in higher energy savings compared to the other treatments. However, some studies, such as the one by Fanghella *et al.* (2021), indicate that combining certain interventions may yield effects restricted to the magnitude of their individual impacts, while others, like Panzone *et al.* (2021) and Hilton *et al.* (2014), reveal an increase in impact compared to the individuals ones but limited interaction effects when interventions are combined in altering behaviors like reducing carbon footprints or influencing choices. Moreover, insights from meta-analyses such as Alt *et al.* (2022) suggest that the joint impact of heightened personal and environmental prices might not rigidly adhere to a linear aggregation of their individual effects. This notion prompts an exploration into the complex dynamics underlying the combined influence of these pricing mechanisms on behavior and outcomes. We thus propose the following exploratory research question:

Exploratory research question 1. *How does the combined impact of increased personal and environmental costs influence energy savings behavior in comparison to the cumulative effect of their individual influences?*

Our research question delves into examining how the simultaneous increase in personal and environmental costs might not merely equate to the sum of their individual impacts. We aim to explore potential synergies or crowding-out effects, aligning with previous studies exploring how interventions addressing both intrinsic and extrinsic motivations interact.

4 Data and Results

4.1 Participant sample

This section addresses the descriptive statistics that characterize our participant sample to ensure the results' validity. These statistics, which are essential to understanding subsequent analyses and interpretations, provide insight into the key demographic characteristics and variables within our sample population. In addition, all exclusion criteria of the data set are presented.

Socio-demographic characteristics

The demographics of our survey respondents are partly shaped by the selection effects of who participates in the Prolific academic panel. They still represent a broad distribution of participants from various backgrounds, which is arguably more representative of the broad population than standard university lab experiments. Based on the following, we suspect that those living in the United Kingdom, those who have a worse outside option for earning money, and those with a higher level of education are over-represented in the Prolific population. These points are also made plausible by considering how incentives and the ability to earn money with English-language online polls vary in the general population. We extended invitations to the online experiment to Prolific users across 31 European countries, encompassing all current EU member states[4] and the United Kingdom, in addition to Switzerland, Liechtenstein, and Norway. 19 of them are represented in our sample but not proportional to their population. Table 2 breaks down the distribution of participant's country of residence. The majority (32.7%) of the participants reside in the United Kingdom, further 37.3% are distributed across the Mediterranean countries Italy, Portugal, Greece and Spain, the second- to fifth-most represented countries. The remaining 30% are scattered across various EU countries.

Socio-demographic statistics are important as they provide contextual insights into the diverse characteristics of the participant sample, potentially influencing behaviors or responses observed in the subsequent analysis. Table 3 and Table 4 show further socio-demographic attributes of our sample. We used a balanced sample in which male and female participants are evenly distributed.[5] Ages range from 18 to 76, with a mean of 34 years and a standard deviation of 13 years, suggesting a robust representation across various working-age adult groups within our sample. The sample however skews towards young people, which is connected to 33.8% of respondents being students. As we use a vision-based effort task to determine the saving behavior, we checked for a potential bias regarding visual ability.

[4] Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden.

[5] The balanced distribution of gender is a selectable setting on the platform Prolific.

Table 2: Country of residence of sample

Country	**Proportion**	**Country**	**Proportion**
United Kingdom	0.327	Estonia	0.020
Italy	0.101	Germany	0.017
Portugal	0.101	Slovenia	0.013
Greece	0.091	Belgium	0.012
Spain	0.080	Austria	0.010
Poland	0.071	Ireland	0.010
Hungary	0.050	Latvia	0.007
France	0.038	Finland	0.005
Netherlands	0.023	Sweden	0.005
Czech Republic	0.020		
No. Obs. $= 603$			

Table 3: Socio-demographic statistics

Variable	**Type**	**Mean (Std. Dev.)**	**Min**	**Max**	No. Obs.
Female	Binary	0.501			603
Age	Discrete	34.441 (13.227)	18	76	603
Student	Binary	0.338			603
Visual impairment	Binary	0.506			595
Ecological	Discrete	3.684 (0.651)	1.267	4.867	585

The variable Ecological in Table 3 describes the ecological orientation in our sample based on the questions from Dunlap *et al.* (2000), resulting in a one to five NEP scale of endorsement with higher values indicating more pro-environmental orientations. Both strong endorsement and rejection, are present in our sample, although with a mean of roughly 3.7 and a standard deviation of about 0.7, most participants tend to hold moderate ecological views. Table 4 illustrates that all levels of education and employment statuses are present in our data. Similarly, participants' yearly household incomes range from below 10,000 € to above 100,000 €. However, the skew of the Prolific sample resulted in an over-representation of participants who hold a graduate (non-doctorate) degree, are unemployed, or earn a low income.

Table 4: Socio-demographic statistics II

Highest Education	Proportion	Employment status	Proportion	Income (in €)	Proportion
Undergraduate degree (BA,BSc,...)	0.322	Full-time	0.463	<10,000	0.201
Graduate degree (MA,MSc,MPhil,...)	0.245	Part-time	0.187	10,001-25,000	0.302
High school diploma /A-levels	0.237	Unemployed	0.153	25,001-40,000	0.227
Technical /Community college	0.093	Not working	0.104	40,001-55,000	0.128
Secondary education (GED,GCSE,...)	0.073	Other	0.070	55,001-70,000	0.068
Doctorate degree (PhD,...)	0.018	New job	0.023	70,001-85,000	0.025
No formal qualifications	0.007			85,001-100,000	0.018
Other	0.005			>100,000	0.020
No. Obs.= 603		No. Obs.= 603		No. Obs.= 596	

Exclusion criteria

In the course of the experiment, 19 individuals have experienced technical problems in the tracking of their data and are therefore excluded.[6] Moreover, in certain cases, the storage of survey answers encountered technical issues resulting in missing answers. When controlling for certain characteristics surveyed in the questionnaire, the sample size is therefore slightly reduced as depicted in the regression tables. In the main specification, the observations with incomplete survey data are however included since this issue does not influence the data on task performance. To ensure the integrity of our analysis, individuals exceeding 20 minutes per task completion or utilizing light for over four minutes within a single task have been excluded. This exclusion criterion aims to mitigate any impact from technical issues or interruptions on their behavioral changes across the two rounds. These criteria eliminate 14 additional observations in total, nine due to the first criterion and five because of the second one. The above-mentioned thresholds have been selected based on our pre-tests in the design phase of the experiment, which indicate that no uninterrupted fulfillment of the tasks without technical problems comes close to exceeding those thresholds. To substantiate this presumed extremeness, the z-score method for detecting outliers has been applied (Grubbs, 1969). Calculating the z-scores of the values exceeding those thresholds with respect to the time respectively light usage of that task leads to magnitudes ranging between 7.37 and 21.81, which noticeably exceed the rule-of-thumb threshold of three (Aggarwal, 2017). This further emphasizes the extremeness of those values and supports the assessment that something atypical, which

[6]For the 19 participants the energy costs do not match up to the total light usage, such that a technical issue regarding the tracking of the light usage is likely. For the non-excluded observations, the costs and light usage however match one-to-one.

does not reflect the objective of the study, must have happened. Since the occurrence of those issues is also not statistically associated with the treatments, the exclusion of the outliers is not expected to bias the results regarding the treatment effects.

Balanced sample

By design, the characteristics of the individuals should be distributed equally in the different treatments. However, to ensure that the randomization was successful and that the exclusion of certain observations did not unbalance the characteristics, equality of means across treatments was checked for the variables in Table 5 using Welch's test (Welch, 1951). In contrast to the one-way fixed effects analysis of variance, it does not require the homogeneity of group variances. As depicted in the table, none of the differences in means is significant at the 5% level. However, the differences in the age and in the time spent in the first round seem to be more pronounced, such that they will be controlled for in the following regressions. Moreover, independence between the treatment and the country of residence, the binary variables in Table 3 respectively the categorical variables in Table 4 has been tested. This analysis was conducted employing a chi-squared test for the binary variables and a Fisher's exact test with a simulated p-value for those variables with more categories, as some of these categories have limited observations. In no case, independence between the characteristics and the treatment can be rejected at the 5% significance level.

Table 5: Distribution of characteristics across treatments

	Baseline	**Environment**	**Personal**	**Both**	
Variable	Mean (Sd)	Mean (Sd)	Mean (Sd)	Mean (Sd)	**p-value**
Age	32.19 (11.68)	35.19 (13.46)	35.42 (13.16)	34.73 (14.25)	0.091
Ecological	3.71 (0.65)	3.65 (0.64)	3.72 (0.64)	3.65 (0.69)	0.759
Patience	0.13 (0.83)	-0.05 (0.83)	-0.04 (0.86)	-0.03 (0.81)	0.238
Altruism	0.06 (0.85)	-0.10 (0.80)	-0.02 (0.76)	0.05 (0.86)	0.289
Time spent round 1	751.37 (324.41)	763.12 (397.91)	727.19 (357.22)	681.09 (262.04)	0.108
Light usage round 1	140.20 (143.69)	146.79 (147.65)	173.04 (164.14)	168.02 (178.81)	0.223
No. Obs.	147/144	144/139	140/134	139/134	

For each treatment, the higher number of observations indicates the count for which data on the age and the behavior during the tasks is available, while the lower number indicates the count for which data on all the regarded variables is available. The p-values are calculated using Welch's test.

4.2 Analysis

Effects on light usage and time spent across rounds

As previously stated, the Baseline treatment serves as a control group to account for any general trends in behavior and learning/fatigue effects over the two rounds. Indeed, results in Table 6 illustrate that the time spent significantly decreases in the second round compared to the first, while light usage remains constant. This indicates the existence of a learning effect enabling participants to complete the tasks in round two faster while maintaining constant effort and thus constant light usage. It is therefore important to analyze the treatment effects in the other groups relative to this Baseline learning pattern. Subsequent analyses will examine the differences in light usage and time spent between rounds across treatments to isolate the impacts of the varied cost structures.

Table 6: Change in behavior across rounds for the Baseline group

	Round 1	**Round 2**	
Variable	Mean (Sd)	Mean (Sd)	**p-value**
Time spent	751.37 (324.41)	585.91 (223.17)	<0.001
Light usage	140.20 (143.69)	141.43 (146.44)	0.942
Number of tasks without light	4.14 (2.93)	4.25 (2.97)	0.752
Number of tasks with ample light	0.77 (1.70)	1.35 (2.16)	0.010
No. Obs.	147	147	

A task is defined to be one with ample light usage if during this task the light has been used for at least 90% of the time needed for the completion.

Furthermore, the general variation in behavior across the 16 tasks is examined. Since the difficulty of the tasks increases during each of the two rounds, a change in behavior during the progress of each round is expected for all treatments. This shift in behavior takes place and is depicted in Figures 2 and 3. Figure 2 illustrates the time spent in seconds for each task of the two rounds. The two rounds are separated by a vertical bar after the eighth task. Correspondingly, Figure 3 shows the light usage in seconds for each task. As could be expected, the individuals are using the light to a greater extent, the more strenuous the tasks get. This increase happens even though the intensified difficulty is also met with a longer time spent on most tasks. Only for the most difficult task, the time spent compared to the previous task is decreased on average in all treatments. This indicates that the individuals are not perfectly offsetting the increased strenuousness with light usage but that a general willingness to increase their effort in order to limit the costs is present. However, the difficulty of the most demanding task might be so high that investing more time instead of using more light is not an attractive approach for many. Furthermore, Figure 3 illustrates that the above-mentioned time reduction of the baseline group is also present in the other groups, as all graphs on the right side of the vertical bar are significantly shifted to lower values. Comparing the two sides of Figure 3 suggests that for no group a significant increase in light usage has happened while for some groups a decrease might have taken place.

Treatment effects on light usage and time spent

To analyze those results further and examine the impacts of the varied cost structures in depth, the differences in behavior between the first and second round are examined by treatment and compared to the Baseline differences. First, the change in total light usage in seconds between the two rounds is investigated. The results are depicted in the first two

Figure 2: Time spent for the completion of the tasks across rounds

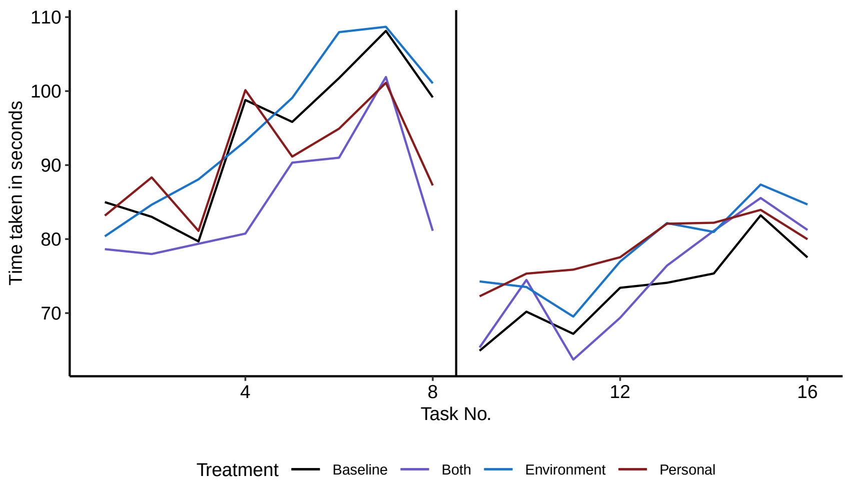
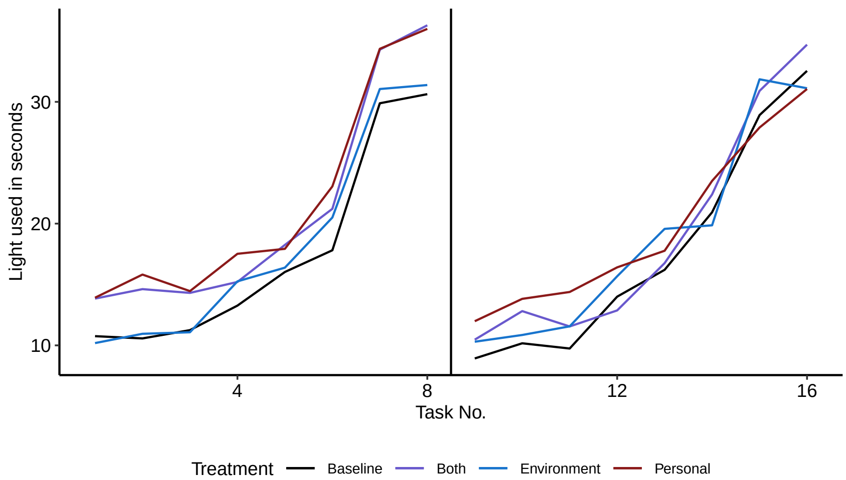


Figure 3: Light usage over time



columns of Table 7. While the rise in the environmental costs, which reflect the externality of consuming energy, does not alter the light usage significantly, raising the personal costs, alone or in conjunction with an increase in environmental costs, significantly reduces the amount of light used. Depending on the specification, this reduction lies between 8.6 % - 10.1% of the initial light usage in the first round for the Both as well as the Personal treatment group. In both specifications, an equality of the effects of the Both and Personal treatments cannot be rejected using an F-test (p-values: 0.937, 0.963). However, both coefficients are statistically significantly different from the coefficient on the Environment treatment at a 5% significance level. To analyze the occurrence of synergy or crowding-out effects, it has additionally been tested whether the sum of effects of the Environment and Personal treatments is equal to the effect of the Both treatment. Using an F-test, this cannot be rejected in any of the two specifications (p-values: 0.852, 0.891). Hence, no synergy or crowding-out effects are visible.

This decrease in light usage however comes at the cost of an increased expenditure of time as depicted in the third and fourth columns of Table 7. In the specification including the control variables, raising both cost components results in an increase in time spent of 7.4% and raising only the personal costs in an increase of 8.0%. Again, the equality of the coefficients on those two treatments cannot be rejected (p-values: 0.937, 0.963), but they are significantly different from the coefficient on the Environment treatment at a 5% level. Also, the adding up of the effects of changing only one cost component to the effect of changing both simultaneously cannot be rejected (p-values: 0.852, 0.891), even though in this situation and especially in the specification with controls, the sum is noticeably bigger than the effect of the Both treatment. The main reason for this is that also the Environment treatment leads to an increased expenditure of time that is significant at the 10% level. However, in the hereinafter elucidated specification without edge cases, this significance does not remain. Since there is also no decreased light usage which might explain a higher expenditure in time, presumably not too much meaning should be given to this effect.

Table 7: Regression results regarding the change in behavior

	Change in light		Change in time	
	(1)	(2)	(1)	(2)
(Intercept)	1.23	1.34	-165.47^{***}	200.45^{***}
	(5.61)	(8.76)	(20.49)	(32.54)
treatmentBoth	-16.80^{**}	-14.47^{*}	81.63^{***}	50.34^{**}
	(8.05)	(7.76)	(29.39)	(22.92)
treatmentEnvironment	2.77	1.88	31.82	38.86^{*}
	(7.97)	(7.70)	(29.13)	(22.69)
treatmentPersonal	-17.45^{**}	-14.83^{*}	67.52^{**}	58.00^{**}
	(8.03)	(7.76)	(29.34)	(22.87)
age		0.60^{**}		-0.53
		(0.25)		(0.62)
sum light round 1		-0.14^{***}		
		(0.02)		
sum time round 1				-0.46^{***}
				(0.02)
R^2	0.02	0.09	0.02	0.41
Adj. R^2	0.01	0.09	0.01	0.40
No. obs.	570	570	570	570

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Edge cases

The amounts of the monetary endowments have been selected in such a way that at the end of each round, no endowment should have been completely consumed. However, 12 individuals utilized more light than anticipated. To increase the robustness of our results, the behavior of all individuals completely depleting at least one of their two endowments in at least one round has been analyzed. If someone fully depletes any of the two endowments, the incentives to save energy are altered since the additional consumption of light is not as costly as before. This however complicates the identification of the costs' effect as the cost structure is not only altered by the different marginal costs but also by the circumstance that some additional consumption might be completely costless. Moreover, the occurrence of those edge cases might be correlated with the treatment status since increasing the costs leads to an earlier depletion of the endowments, and since due to the treatments altering the light usage behavior, the likelihood of using up an endowment varies over the treatments. Due to this circumstance, an exclusion of the edge cases from the outset is not appropriate. However, examining the behavior of the individuals entering the edge case paired with a comparison of the results without those individuals to the general results can strengthen the validity of the results.

Of the 12 individuals fully depleting at least one of the endowments, half do so only during the last tasks of the round such that their period of reduced costs of light usage is limited. Moreover, five of the remaining six individuals have used the light for at least 90% of the time they spend on the task page for all the eight tasks in that round, no matter whether they have already depleted any endowment or not. Because no one has used up their endowment in the initial four rounds and the early rounds are simpler without light, heavy usage of light at the start suggests that individuals would have used it extensively by the end, even if their endowment had not been depleted. Moreover, the average light usage for all those six individuals is less increased compared to the control group's mean in the rounds after having emptied the pot than in the rounds before. This suggests that no heavy increase due to the depletion of at least one endowment has happened. As an additional robustness check, the regressions in Table 7 have been run without the individuals entering the edge case. As depicted in Table 9 in the appendix, this does not greatly change any coefficients and even shows a reduced significance level for the Both and Personal treatment coefficients. As mentioned above, it removes the single significance of the Environment treatment.

No effect from rising environmental costs

One explanation for the absence of an effect of the Environment treatment on light usage might be the individuals not caring about the environment in general. However, since the mean ecological attitude, which is measured on a scale from one to five with five representing the most ecological attitude, is 3.7 with a median of 3.8, at least the elicited preferences identify many individuals with pro-environmental attitudes. A different explanation might be that the individuals question the donations' realness. However, Prolific offers high transparency regarding the payments and rights of participants (Palan and Schitter, 2018) which should ensure a strong trust in the authenticity of the donation. This should have been even further strengthened by the promised reception of a donation receipt after the donation had been made. Moreover, due to the explicitness of the donation's effect, namely the planting of a tree for every £0.82, even smaller donation amounts can create a meaningful and tangible effect. Therefore, contrasted to the more dispersed and intangible externalities of energy use in reality, the donation's tangibility might even enlarge any effects of varying intensity of environmental costs. Furthermore, Ghesla *et al.* (2020) have shown that incentivizing households by the planting of a tree can increase their energy savings such that a valuation for the planting of a tree should be present. Therefore, neither the participants' attitudes nor the donation's characteristics work as explanations for the absent effect.

What could explain the absence of an effect of increased environmental costs is the presence of those costs being the main driver in behavior and not their amount. This would be consistent with the results of Charness *et al.* (2016), Imas (2014) and Tonin and Vlassopoulos (2015) which all point in the direction that introducing donations into the remuneration scheme increases the exerted effort, but varying the amount of donations does not affect effort. Therefore, also in this dissimilar energy setting with a pro-environmental donation, introducing the donation might have an effect, but as examined, increasing the amount does not. Since in reality, there are however essentially always externalities related to energy use and at least part of them are also known by the consumers, the only way in which the externalities vary is in their markedness but never in their existence. Analogous to the above-mentioned studies, there might be a shift in the perception of energy saving when a social dimension is included, a ''warm-glow" (Andreoni, 1989) of energy saving arises. However, the impetus is the social dimension per se and not the amount of benefits incurred by society, i.e. pure altruism. Lack of synergy or crowding-out effects, along with statistically equal coefficients for the Personal and Both treatments, suggests that raising personal costs does not seem to change the hierarchy between warm-glow and pure altruism. Similarly, increasing environmental costs does not appear to affect the significance of a personal economic benefit. Hence, those components seem to be independent of each other. Since there is however no treatment without any donations as there is no energy consumption without externalities, this hypothesis can neither be proved nor falsified, but it has limited relevance for guidance in any case.

Heterogeneous treatment effects

However, the treatments may have had a differential effect depending on the importance individuals placed on the donation amount. When the donation is weighted more heavily in general, the magnitude may become more impactful. Given the pro-environmental nature of the donation, individuals with a more pro-environmental orientation could likely place greater importance on it. To analyze this hypothesis, heterogeneous treatment effects are examined based on participants' environmental orientation surveyed in the experiment. Additionally, a screener question, which requested the individuals to select a specific answer, was embedded in the 15 environmental questions. The failure to answer the screener correctly was utilized in the analysis to exclude any individuals whose environmental preference results are possibly influenced due to a lack of attention and thus not elicited correctly. This encompasses 21 individuals. For the analysis of the heterogeneous treatment effects, the surveyed environmental orientation has been standardized to have a mean of zero and a standard deviation of one, such that the interaction effects represent deviations from an average ecological orientation and the coefficients reflect an increase in the ecological orientation by one standard deviation.

As Table 8 shows, no differentiated behavior depending on the environmental attitude is visible. Even if the impact can vary based on how much importance someone places on the donation, as shown by their ecological attitude, it seems like higher environmental costs do not change behavior here. However, regressing the initial use of light and the time spent in the first round on the standardized ecological variable shows that individuals with an ecological orientation one standard deviation above the mean orientation use 18.1 seconds less light (p-value: 0.006) while spending 30.8 seconds longer (p-value: 0.029). Therefore, a correlation between general behavior and environmental attitude is present, but this correlation does not extend to a change in behavior due to subsequent cost changes. The differentiated behavior in the first round might be explained by the presence of the pro-environmental donation which motivates more environmentally individuals more strongly. Hence, the introduction of the warm-glow effect might be more potent for those individuals. But again, since no treatment without donations has been run, this aspect is out of the scope of this paper. It would however be in line with the results of Tonin and Vlassopoulos (2015). They have shown that introducing donations in the context of completing a task requiring effort has an effort-increasing effect on individuals who have volunteered for a charity in the previous twelve months but it has no significant effect on those who have not. Since their donations did not necessarily go to a pro-environmental cause, a parallel could be drawn between the pro-environmental and pro-volunteering individuals.

Other individual characteristics like patience, altruism, and household income may in principle moderate treatment effects, though this was not the case in our study setting. For instance, one could expect more patient participants to save light by taking more time on tasks or altruistic participants to be moved by donation amounts. However, results for heterogeneous effects along these dimensions were statistically weak and sometimes even produced counterintuitive patterns.[7] While more altruistic participants did save more light on average in the second round, they saved comparably less when donation costs rose. Participants with higher patience did not significantly alter their behaviors. Higher-income earners saved less light overall but reacted more strongly to the environmental treatment. Though inconclusive, these irregular results could stem from interactions be-

[7]The respective regression outputs for these characteristics are included in the Appendix in Tables 11,12 & 13.

tween motivational driers, or from environmental costs crowding out intrinsic motivations among some participants. Given the lack of statistical power, asserted explanations are speculative. Further research is needed to untangle how specific attributes shape heterogeneous responses to incentives and costs.

Table 8: Heterogeneous treatment effects regarding the ecological attitude

	Change in light		Change in time	
	(1)	(2)	(1)	(2)
(Intercept)	-0.21	-0.21	-159.85^{***}	235.71^{***}
	(5.72)	(9.03)	(21.31)	(34.00)
treatmentBoth	-15.74^{*}	-14.59^{*}	79.43^{**}	46.90^{**}
	(8.25)	(7.96)	(30.73)	(23.65)
treatmentEnvironment	7.70	6.01	27.86	30.13
	(8.14)	(7.87)	(30.33)	(23.30)
treatmentPersonal	-18.49^{**}	-14.76^{*}	58.14^{*}	52.04^{**}
	(8.21)	(7.94)	(30.56)	(23.51)
ecologicals	-6.83	-6.82	5.55	11.96
	(5.66)	(5.45)	(21.06)	(16.16)
treatmentBoth×ecologicals	5.68	2.82	-3.13	10.97
	(7.91)	(7.62)	(29.46)	(22.57)
treatmentEnvironment×ecologicals	12.19	9.69	-24.07	-22.33
	(8.19)	(7.90)	(30.49)	(23.42)
treatmentPersonal×ecologicals	8.87	3.51	-39.05	-20.92
	(8.22)	(7.97)	(30.61)	(23.56)
age		0.63^{**}		-0.87
		(0.26)		(0.64)
sum light round 1		-0.15^{***}		
		(0.02)		
sum time round 1				-0.49^{***}
				(0.03)
R^2	0.03	0.10	0.02	0.43
Adj. R^2	0.02	0.09	0.01	0.42
No. obs.	532	532	532	532

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

5 Conclusion

This paper sheds new light on the complex interplay between extrinsic and intrinsic pricing incentives in motivating energy conservation. Our online lab experiment reveals several core findings.

First, increasing personal costs of energy usage significantly reduces consumption, consistent with past evidence on responsiveness to price changes. Using this effect, policymakers could leverage instruments like smart meters to design and implement dynamic pricing schemes that incentivize energy conservation. This could involve variable rates based on time of day, encouraging consumers to shift consumption to off-peak hours, or incentive programs rewarding energy-saving behaviors. This could lead to increased public awareness and participation in energy conservation efforts.

Second, heightened externality costs did not significantly alter usage in our experiment. This contrasts with prior research showing intrinsic and prosocial incentives can independently shape effort and behaviors. One potential explanation lies in the nature of introducing a prosocial dimension versus marginally increasing its magnitude. Our results align with studies finding donations to increase effort irrespective of size. Hence, emphasizing negative environmental externalities might change how people perceive energy conservation when first implemented but gradual adjustments might not significantly affect decisions.

Lastly, we find no significant synergies or interference when both personal and externality prices rise in tandem. Their combined effect equals the sum of individual impacts, aligning with an additive pattern. Thus, no negative effect can be observed when extrinsic and intrinsic incentives are implemented simultaneously.

Our approach does face limitations in generalizability due to the online experimental setting. Further research could test these incentive patterns in field experiments and explore their persistence over time. Extending the externality manipulation to include treatments without any externality pricing could also help isolate its introduction versus size effects. Nonetheless, our findings suggest personal costs strongly incentivize energy savings, while externality pricing requires careful design to motivate conservation. Policies should consider complementary approaches, rather than view these incentives as substitutes. With further research, a better understanding of their interplay can be reached to get the optimal mix of incentives for energy conservation.

References

Abeler, J., Falk, A., Goette, L. and Huffman, D. (2011). Reference points and effort provision. *The American Economic Review*, **101** (2), 470–492.

Aggarwal, C. C. (2017). *An introduction to outlier analysis*. Springer.

Alt, M., Bruns, H., DellaValle, N. and Murauskaite-Bull, I. (2022). Synergies of interventions to tackle climate change-a meta-analysis. *Available at SSRN 4305865*.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence. *Journal of political Economy*, **97** (6), 1447–1458.

Araujo, F. A., Carbone, E., Conell-Price, L., Dunietz, M. W., Jaroszewicz, A., Landsman, R., Lamé, D., Vesterlund, L., Wang, S. W. and Wilson, A. J. (2016). The slider task: An example of restricted inference on incentive effects. *Journal of the Economic Science Association*, **2**, 1–12.

Augenblick, N., Niederle, M. and Sprenger, C. (2015). Working over Time: Dynamic Inconsistency in Real Effort Tasks *. *The Quarterly Journal of Economics*, **130** (3), 1067–1115.

Bastos, P., Castro, L., Cristia, J. and Scartascini, C. (2015). Does energy consumption respond to price shocks? evidence from a regression-discontinuity design. *The Journal of Industrial Economics*, **63** (2), 249–278.

Benndorf, V., Rau, H. A. and Sölch, C. (2019). Minimizing learning in repeated real-effort tasks. *Journal of Behavioral and Experimental Finance*, **22**, 239–248.

Buckley, P. and Llerena, D. (2022). Nudges and peak pricing: A common pool resource energy conservation experiment. *Journal of Behavioral and Experimental Economics*, **101**, 101928.

Charness, G., Cobo-Reyes, R. and Ángela Sánchez (2016). The effect of charitable giving on workers' performance: Experimental evidence. *Journal of Economic Behavior Organization*, **131**, 61–74.

—, Gneezy, U. and Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, **149**, 74–87.

Dolan, P. and Metcalfe, R. (2015). Neighbors, knowledge, and nuggets: two natural field experiments on the role of incentives on energy conservation. *Becker Friedman Institute for Research in Economics Working Paper*, (2589269).

Drews, S., Exadaktylos, F. and van den Bergh, J. C. (2020). Assessing synergy of incentives and nudges in the energy policy mix. *Energy Policy*, **144**, 111605.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G. and Jones, R. E. (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised nep scale. *Journal of social issues*, **56** (3), 425–442.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018). Global Evidence on Economic Preferences*. *The Quarterly Journal of Economics*, **133** (4), 1645–1692.

—, —, —, Huffman, D. and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences.

Fanghella, V., Ploner, M. and Tavoni, M. (2021). Energy saving in a simulated environment: An online experiment of the interplay between nudges and financial incentives. *Journal of Behavioral and Experimental Economics*, **93**, 101709.

Gabaix, X. (2019). Behavioral inattention. In *Handbook of behavioral economics: Applications and foundations 1*, vol. 2, Elsevier, pp. 261–343.

Gans, W., Alberini, A. and Longo, A. (2013). Smart meter devices and the effect of feedback on residential electricity consumption: Evidence from a natural experiment in northern ireland. *Energy Economics*, **36**, 729–743.

Ghesla, C., Grieder, M., Schmitz, J. and Stadelmann, M. (2020). Pro-environmental incentives and loss aversion: A field experiment on electricity saving behavior. *Energy Policy*, **137**, 111131.

Giamattei, M., Yahosseini, K. S., Gächter, S. and Molleman, L. (2020). Lioness lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association*, **6**, 95–111.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, **11** (1), 1–21.

Hilton, D., Charalambides, L., Demarque, C., Waroquier, L. and Raux, C. (2014). A tax can nudge: The impact of an environmentally motivated bonus/malus fiscal system on transport preferences. *Journal of Economic Psychology*, **42**, 17–27.

Horowitz, J. L. (2001). The bootstrap. In *Handbook of econometrics*, vol. 5, Elsevier, pp. 3159–3228.

Imas, A. (2014). Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics*, **114**, 14–18.

Ito, K., Ida, T. and Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, **10** (1), 240–267.

Knief, U. and Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, **53** (6), 2576–2590.

Lezzi, E., Fleming, P. and Zizzo, D. J. (2015). *Does it matter which effort task you use? A comparison of four effort tasks when agents compete for a prize*. Working Paper series, University of East Anglia, Centre for Behavioural and Experimental Social Science (CBESS) 15-05, School of Economics, University of East Anglia, Norwich, UK.

Matthews, G. and Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology Section A*, **55** (2), 659–686, pMID: 12047065.

Mizobuchi, K. and Takeuchi, K. (2013). The influences of financial and non-financial factors on energy-saving behaviour: A field experiment in japan. *Energy Policy*, **63**, 775–787.

Murakami, K., Shimada, H., Ushifusa, Y. and Ida, T. (2022). Heterogeneous treatment effects of nudge and rebate: Causal machine learning in a field experiment on electricity conservation. *International Economic Review*, **63** (4), 1779–1803.

Palan, S. and Schitter, C. (2018). Prolific. ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, **17**, 22–27.

Panzone, L. A., Ulph, A., Zizzo, D. J., Hilton, D. and Clear, A. (2021). The impact of environmental recall and carbon taxation on the carbon footprint of supermarket shopping. *Journal of Environmental Economics and Management*, **109**, 102137.

Prolific (2023). Academic panel.

Savage, S. J. and Waldman, D. M. (2008). Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Econometrics*, **23** (3), 351–371.

Sloot, D. and Scheibehenne, B. (2022). Understanding the financial incentive conundrum: A meta-analysis of the effectiveness of financial incentive interventions in promoting energy conservation behavior. *Renewable and Sustainable Energy Reviews*, **168**, 112761.

Smith, V. L. (2003). Constructivist and ecological rationality in economics. *American economic review*, **93** (3), 465–508.

Tonin, M. and Vlassopoulos, M. (2015). Corporate philanthropy and productivity: Evidence from an online real effort experiment. *Management Science*, **61** (8), 1741–2011.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, **185** (4157), 1124–1131.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, **38** (3/4), 330–336.

Werthschulte, M. (2023). Present focus and billing systems: Testing 'pay-as-you-go' vs. 'pay-later'. *Journal of Economic Behavior Organization*, **212**, 108–121.

Zhao, Q., Shen, Y. and Li, M. (2015). Control and bidding strategy for virtual power plants with renewable generation and inelastic demand in electricity markets. *IEEE Transactions on Sustainable Energy*, **7** (2), 562–575.

Appendix

Regression results excluding the edge cases

Table 9: Regression results regarding the change in behavior excluding edge cases

	Change in light		Change in time	
	(1)	(2)	(1)	(2)
(Intercept)	1.23	4.03	-165.47^{***}	199.60^{***}
	(5.43)	(8.53)	(20.54)	(33.20)
treatmentBoth	-16.81^{**}	-16.53^{**}	83.45^{***}	49.56^{**}
	(7.86)	(7.55)	(29.74)	(23.27)
treatmentEnvironment	0.31	-1.28	30.89	37.03
	(7.75)	(7.48)	(29.35)	(22.96)
treatmentPersonal	-17.11^{**}	-15.65^{**}	65.93^{**}	56.97^{**}
	(7.83)	(7.53)	(29.62)	(23.16)
age		0.58^{**}		-0.57
		(0.25)		(0.64)
sum light round 1		-0.15^{***}		
		(0.02)		
sum time round 1				-0.46^{***}
				(0.02)
Edge cases excluded	Yes	Yes	Yes	Yes
R^2	0.02	0.10	0.02	0.40
Adj. R^2	0.01	0.09	0.01	0.40
No. obs.	558	558	558	558

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 10: Heterogeneous treatment effects regarding the ecological attitude excluding edge cases

	Change in light		Change in time	
	(1)	(2)	(1)	(2)
(Intercept)	-0.21	1.82	-159.85^{***}	234.96^{***}
	(5.56)	(8.79)	(21.33)	(34.49)
treatmentBoth	-16.99^{**}	-16.85^{**}	80.64^{***}	46.90^{**}
	(8.07)	(7.72)	(30.96)	(23.87)
treatmentEnvironment	5.73	3.38	28.55	28.73
	(7.94)	(7.63)	(30.47)	(23.47)
treatmentPersonal	-18.24^{**}	-15.74^{**}	56.42^{*}	50.55^{**}
	(8.04)	(7.71)	(30.85)	(23.76)
ecologicals	-6.83	-6.73	5.55	11.97
	(5.50)	(5.26)	(21.09)	(16.22)
treatmentBoth×ecologicals	2.86	0.48	-0.10	15.78
	(7.84)	(7.50)	(30.10)	(23.12)
treatmentEnvironment×ecologicals	12.31	10.05	-29.76	-24.13
	(8.00)	(7.67)	(30.69)	(23.64)
treatmentPersonal×ecologicals	9.19	3.32	-38.29	-20.47
	(7.99)	(7.70)	(30.66)	(23.66)
age		0.64^{**}		-0.89
		(0.25)		(0.66)
sum light round 1		-0.16^{***}		
		(0.02)		
sum time round 1				-0.49^{***}
				(0.03)
Edge cases excluded	Yes	Yes	Yes	Yes
R^2	0.03	0.12	0.02	0.42
Adj. R^2	0.02	0.10	0.01	0.41
No. obs.	523	523	523	523

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

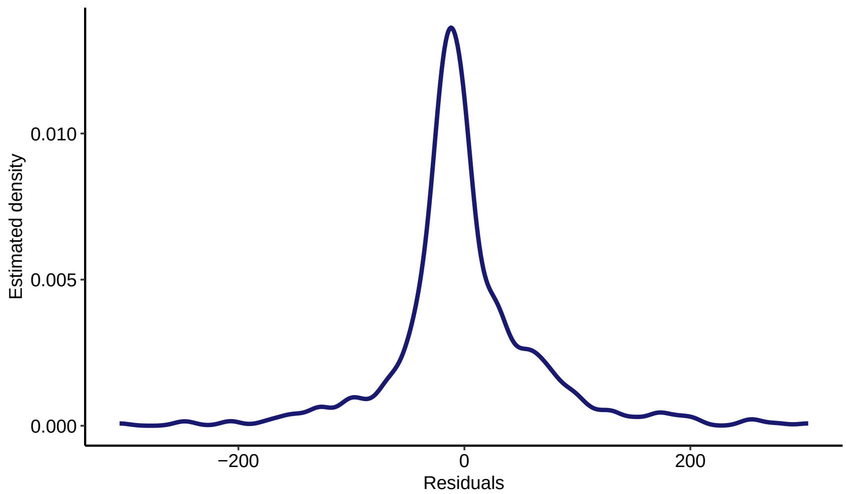
Robustness check using a bootstrap procedure

Since the distributions of the residuals of the conducted regressions have longer tails than would be the case for a normal distribution, additional robustness checks have been carried out as the normality of the residuals is a prerequisite for the conduction of ordinary least squares estimations. An example of the residuals' estimated density is depicted in Figure 4. This estimation has been done using a kernel density estimation. The shape of the density is similar for all regressions. As can be seen, the distribution is relatively symmetrical and bell-shaped with longer tails. As pointed out by Knief and Forstmeier (2021), most violations of the normality assumption do not lead to problems regarding the p-values, except when the sample size is very small or the distribution of the predictor is heavily skewed. Since none of this is the case in the regarded situation, sufficient weight can be put on the estimates' significance. However, to back up the robustness even further, calculations of the significance levels have also been done using a bootstrap approach.

The applied procedure follows the theoretical underpinnings of e.g. Horowitz (2001). For this, 100,000 bootstrap samples of the same size as the original sample have been created by sampling with replacement from the original data set. Those bootstrap sets are used to approximate the distribution of the t-statistic such that the normality assumption regarding the residuals becomes redundant. For this approximation, the respective regression has been estimated for each bootstrap set. Since in this bootstrap setting, the estimates obtained using the original sample are the analog to the population parameters in the normal linear regression setting, the deviation of the bootstrap estimates from the original estimates divided by the estimated standard errors are utilized for estimating the t-statistics' distributions. To calculate the p-value for each coefficient, the fraction of the 100,000 bootstrap samples for which the absolute value of the original t-statistic is smaller than the absolute value of the estimated one is regarded.

The by this procedure obtained results show only one slight deviation regarding the significance levels of the previously analyzed linear regressions. This concerns the first column of Table 8. Here, the bootstrap procedure points to a significance level of the coefficient on the Both treatment even at the 1% level instead of at the 5% level. Hence, the general results' robustness is affirmed.

Figure 4: Estimated density of the residuals of the regression in the second column of Table 7



Further checks for heterogeneity

Table 11: Heterogeneous treatment effects regarding the altruistic orientation

	Change in light		Change in time	
	(1)	(2)	(1)	(2)
(Intercept)	1.06	0.26	-161.10^{***}	223.46^{***}
	(5.68)	(8.97)	(21.31)	(33.98)
treatmentBoth	-17.25^{**}	-15.89^{**}	84.88^{***}	47.10^{**}
	(8.22)	(7.95)	(30.82)	(23.80)
treatmentEnvironment	6.06	4.72	25.30	26.52
	(8.09)	(7.84)	(30.35)	(23.40)
treatmentPersonal	-20.32^{**}	-16.89^{**}	57.67^{*}	50.49^{**}
	(8.13)	(7.88)	(30.49)	(23.52)
altruism	-18.88^{***}	-16.27^{**}	18.01	4.89
	(6.57)	(6.36)	(24.65)	(18.96)
treatmentBoth×altruism	14.87	13.90	-32.30	-30.35
	(9.46)	(9.13)	(35.46)	(27.27)
treatmentEnvironment×altruism	18.51^{*}	15.53^{*}	-69.58^{*}	-40.78
	(9.76)	(9.42)	(36.59)	(28.14)
treatmentPersonal×altruism	-0.23	-1.50	-5.93	-8.53
	(10.08)	(9.72)	(37.79)	(29.02)
age		0.62^{**}		-0.60
		(0.26)		(0.64)
sum light round 1		-0.14^{***}		
		(0.02)		
sum time round 1				-0.49^{***}
				(0.03)
R^2	0.05	0.12	0.02	0.43
Adj. R^2	0.04	0.11	0.01	0.42
No. obs.	530	530	530	530

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 12: Heterogeneous treatment effects regarding the patience

	Change in light		Change in time	
	(1)	(2)	(1)	(2)
(Intercept)	0.19	0.56	-158.90^{***}	232.38^{***}
	(5.80)	(9.04)	(21.55)	(33.88)
treatmentBoth	-15.99^{*}	-15.11^{*}	77.47^{**}	43.41^{*}
	(8.31)	(8.01)	(30.89)	(23.67)
treatmentEnvironment	6.96	5.15	27.84	28.60
	(8.19)	(7.92)	(30.46)	(23.31)
treatmentPersonal	-18.63^{**}	-15.45^{*}	55.94^{*}	51.20^{**}
	(8.26)	(7.98)	(30.71)	(23.52)
patience	-4.84	-6.91	-5.80	-8.33
	(6.91)	(6.66)	(25.70)	(19.62)
treatmentBoth×patience	7.89	5.30	-21.70	-4.17
	(10.11)	(9.74)	(37.61)	(28.69)
treatmentEnvironment×patience	2.48	3.06	-26.56	-16.91
	(9.86)	(9.48)	(36.67)	(27.96)
treatmentPersonal×patience	8.28	12.04	29.83	57.02^{**}
	(9.77)	(9.42)	(36.33)	(27.74)
age		0.62^{**}		-0.70
		(0.26)		(0.64)
sum light round 1		-0.15^{***}		
		(0.02)		
sum time round 1				-0.50^{***}
				(0.03)
R^2	0.03	0.10	0.02	0.43
Adj. R^2	0.01	0.09	0.01	0.42
No. obs.	532	532	532	532

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 13: Heterogeneous treatment effects regarding the income

	Change in light		Change in time	
	(1)	(2)	(1)	(2)
(Intercept)	-1.61	0.14	-169.36^{***}	195.79^{***}
	(5.78)	(8.91)	(21.30)	(33.07)
treatmentBoth	-14.17^{*}	-13.32^{*}	81.40^{***}	52.83^{**}
	(8.32)	(8.07)	(30.65)	(23.86)
treatmentEnvironment	6.59	5.81	31.18	44.85^{*}
	(8.24)	(8.01)	(30.35)	(23.64)
treatmentPersonal	-14.58^{*}	-12.53	78.01^{**}	66.31^{***}
	(8.30)	(8.07)	(30.59)	(23.83)
highIncome	46.28^{**}	38.44^{*}	59.47	49.30
	(23.29)	(22.61)	(85.79)	(66.61)
treatmentBoth×highIncome	-15.72	-9.36	38.16	-10.06
	(32.97)	(31.97)	(121.45)	(94.37)
treatmentEnvironment×highIncome	-59.86^{*}	-59.11^{*}	-3.99	-68.20
	(32.17)	(31.17)	(118.50)	(92.05)
treatmentPersonal×highIncome	-44.53	-35.38	-181.34	-132.39
	(32.97)	(31.97)	(121.44)	(94.31)
age		0.53^{**}		-0.50
		(0.25)		(0.63)
sum light round 1		-0.13^{***}		
		(0.02)		
sum time round 1				-0.47^{***}
				(0.02)
R^2	0.03	0.09	0.02	0.41
Adj. R^2	0.02	0.08	0.01	0.40
No. obs.	564	564	564	564

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

WIK Wissenschaftliches Institut für
Infrastruktur und Kommunikationsdienste GmbH
Rhöndorfer Str. 68
53604 Bad Honnef, Germany
www.wik.org

ISSN 2750-5448 (Online)