Energy Research & Social Science 91 (2022) 102734

Contents lists available at ScienceDirect

Energy Research & Social Science

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

Boosting or nudging energy consumption? The importance of cognitive aspects when adopting non-monetary interventions

Nicolas Caballero [a,b,*], Matteo Ploner [c]

[a] Eurac Research, Institute for Renewable Energy, Italy
[b] Department of Economics, Ruhr-Universität Bochum, Germany
[c] University of Trento, Department of Economics and Management, Cognitive and Experimental Economics Lab (CEEL), Italy

ARTICLE INFO

JEL classification:
C91
D03
D04
D12
D19
D91
Q40

Keywords:
Energy management
Online experiment
Nudge
Boost
Energy justice

ABSTRACT

Identifying effective behaviour-change interventions to promote energy conservation in the residential sphere has been the topic of extensive empirical research. While existing literature has advised several successful interventions, their context-dependency is still an open question. Furthermore, existing evidence has primarily focused on trialling nudges, that is, interventions that influence behaviour directly by changing aspects of the decision environment and circumventing cognitive bias. Boosts, which instead aim to influence behaviour by fostering the competences of decision-makers and correcting bias, are still under-researched in this domain.

We present the results of an online experiment where we compare the effects of a nudge-like, and a boost-like intervention on decisions in an incentive-compatible energy management task. These interventions are trialled in relatively high income and low income populations. Finally, we repeat the experiment with the same participants after removing the interventions.

Our results show that income is a significant determinant of performance in the task, with the higher income cohort performing better than the lower income counterpart. However, this difference is largely explained by underlying idiosyncratic factors, namely the level of cognitive competences of participants. Furthermore, both boosting and nudging approaches brought energy savings close to the ceiling of achievable goals, but the boosting approach proved more challenging for participants with lower cognitive competences. Finally, we report evidence of intertemporal spillovers.

We conclude by highlighting directions of future research to further assess the interplay between intervention choice and cognitive aspects in the field, to design effective behaviour-change policies in an ethical and targeted manner.

1. Introduction

Energy demand from buildings accounted for 28 % of global energy-related CO_2 emissions in 2019, when including indirect emissions from upstream power generation [1]. As a consequence of the COVID-19 pandemic, new modes of smart working are emerging, which require increased energy consumption at home and a shift in behavioural patterns [2]. This is expected to add further importance to the role of household energy decisions in shaping our climate future. Furthermore, as we decarbonise the energy grid and move towards a more decentralised, weather-dependent one, changes in the way we manage energy resources at home are going to be crucial to support a successful energy transition towards net zero carbon targets [3].

Policymakers face the difficult task of designing successful interventions that foster virtuous energy consumption from individuals in the residential sphere. Against this backdrop, behavioural economics has emerged as a key discipline to inform policy-making, by providing evidence supporting many cost-effective interventions to promote more pro-environmental consumption paths, including energy conservation [4], shifting peak demand [5,6], and investing in energy efficiency [7]. Despite the popularity of these approaches, questions on their effectiveness remain. There exist competing frameworks on the design of non-monetary interventions to influence behaviour, the two key ones being *nudging* (which encourages targeting decision environments to impact behaviour directly) and *boosting* (which encourages targeting the competences of decision-makers). The key difference between these two

* Corresponding author at: Via Alessandro Volta 13/A, NOI Techpark, Bolzano 39100, Italy.
E-mail address: nicolas.caballero@eurac.edu (N. Caballero).

https://doi.org/10.1016/j.erss.2022.102734
Received 5 January 2022; Received in revised form 18 June 2022; Accepted 30 June 2022
Available online 21 July 2022

approaches is in the way they deal with cognitive bias. Boosts aim to correct biases, while nudges aim to circumvent them via careful choice architecture. Furthermore, considering increasing awareness on justice aspects of energy provision and consumption [8], it is important that policymakers evaluate the effectiveness of these behavioural tools relative to socio-economic factors and their cognitive implications. This aligns with calls for increased scrutiny in the literature regarding the context-dependency of different tools [9].

In this study, we take an experimental approach to shed light on these issues. In an online experiment, we frame a multiple-round allocation task as an energy management problem. Participants are endowed with virtual energy resources and must experiment with different allocations in order to find an optimal energy mix which maximises their experimental payoff. In this environment, we introduce two interventions to help participants better manage their resources and optimise allocations quickly. These interventions are designed, based on the existing literature and their potential for field replicability, to reflect a more nudge-like and boost-like strategy. In order to assess context-dependency of these tools in the present set-up, we take a stratified randomisation approach and split our sample based on self-reported household income. Finally, we repeat the experiment in absence of interventions a week after each session, shedding light on potential intertemporal spillovers.

Our results show that income is an important factor to consider when determining which intervention approach is preferable to target specific groups. In particular, while all interventions were equally effective in the high income cohort, they have differing implications in the low income cohort, with boosting leading to detrimental effects in the short-run when compared to a control group. However, our results further highlight that it is not income per se that impacts performance across the different interventions, but rather underlying psychological factors correlated with income, specifically cognitive competences. This suggests that individuals in different income groups might exhibit different decision processes when making energy management decisions, and that these processes should be considered when deciding which behaviourally-informed approach to adopt. We also report interesting intertemporal implications associated to the two types of intervention, indicating that a complementary approach may be best-suited to achieve both short-term behavioural shifts and long-term habit formation. Finally, we highlight the need for further field applications to confirm our results in real-life energy scenarios and draw out policy recommendations.

The present research contributes to the literature in three ways. Firstly, it introduces a new type of boost, namely Fast and Frugal Trees, as a potential tool to optimise energy decisions in the residential sphere. Secondly, it empirically compares the effectiveness of the two approaches in identical settings, contributing to the literature contrasting boosts to nudges [10–12]. Thirdly, our focus on income levels and associated cognitive function contributes to the discussion of how to design behaviour change interventions that are effective to target vulnerable groups [13].

The paper is structured as follows. Section 2 presents the current state of the literature. Section 3 introduces our specific research questions and the methods to address them. In Section 4 we report our results, and in Section 5 we discuss them in light of the existing literature and study limitations, as well as recommend directions for future research to inform policy action. Section 6 concludes the paper.

2. Literature review

2.1. Energy consumption decisions through the behavioural economics lens

A growing body of literature adopts insights from behavioural economics to study energy consumption choices [14]. Under this perspective, consumer choices on energy use are affected by heuristics and biases, similar to what happens in many other economic domains [15]. The underlying idea of the heuristics-and-biases approach is that individuals are imperfect decision-makers who systematically deviate from optimal decision patterns. Rather than making complex cost-benefit analyses, as is assumed by the traditional economic perspective of rational choice theory [16], individuals use "rules of thumb" or "heuristics" that, although often outperforming rational choice models under uncertainty [17], can at times lead to predictable errors [18].

Frederiks et al. [14] identify a number of biases that affect household energy use and lead to a clear "value-action" gap between people's environmental beliefs (e.g., that climate change has associated negative consequences) and their actions (e.g., the reduction of domestic energy use [19]). For example, people often exhibit a status quo bias and are reluctant to change their energy habits irrespective of the information they are provided with [20,21]. Even when individuals explicitly state their pro-environmental preferences, such as favouring energy from renewable sources, they often fail to take necessary action in this direction [22]. This gap has been discussed as a potential area where non-monetary interventions can be particularly effective to align pro-environmental attitudes with actual behaviour [23].

Crucially, this literature has also been used to describe behavioural patterns that trap low income individuals in cycles of energy vulnerability. Behaviour has been recognised as a relevant driver of energy poverty [24], and a number of behavioural biases linked to living in scarcity conditions have been proposed to describe inefficient behavioural patterns that impede the vulnerable from taking active control of their energy resources [25]. For example, conditions of scarcity have been linked to increasing issues of temporal discounting [26], leading the vulnerable to exhibit excessive impatience that results in poor household energy decisions such as inefficient consumption [27]. More generally, literature on scarcity theory suggests that resource scarcity can lead to sub-optimal decision-making due to its impairment on cognitive function [28], a consequence of the cognitive load that results from facing difficult economic decisions routinely. The recent review by de Brujin & Antonides [29] highlights overall empirical support for the theory, though recognising some inconsistencies.

2.2. Tools to optimise energy consumption: nudges and boosts

Two distinct approaches to behavioural non-monetary interventions can be identified in the literature, namely nudges and boosts. Both of these aim to promote virtuous behaviour, but they differ in relation to (amongst others) (i) the assumptions they make on the role of heuristics in decision-making processes and (ii) the approach they undertake to change behaviour.

Nudges, formalised by Thaler and Sunstein [30] in their seminal book, are defined as "any aspect of the choice architecture that alters people's behaviour in a predictable way without forbidding any options or significantly changing their economic incentives" (p.6). Choice architecture is the process of designing different ways in which choices are presented. Therefore, nudges are a type of non-fiscal, non-coercive behaviourally-informed intervention that consists of changing aspects of how choices are presented to promote a specific decision. Different types of nudges have been tested to promote energy conservation both in households [31] and workplaces [32]. Nudges have also been tested to reduce peak demand consumption [33], and proposed as tools to increase uptake of solar energy by households [34].

Maybe the most notable example of a nudge is the setting of a default choice. Defaults leverage the status quo bias, and set the desirable choice as a pre-defined choice. In the domain or residential energy use, defaults have been tested to promote greener energy mixes [20,35], and encourage conservation [36].

Setting default thermostat temperatures could be interpreted as a real world application of default nudges, with the aim of conserving energy and improving thermal comfort [37]. Research on programmable thermostats for example has found that they have great energy savings potential [38], but their effectiveness is limited by the willingness of users to engage with them and set accurate settings [39]. In this regard, programmable thermostats could be considered a technological innovation that provide users with the capacity to “self-nudge” [40], but their willingness to do so may remain an issue for real world applicability of these measures. Approaches hoping to replicate the impact of these interventions in a lab environment therefore must internalise this semi-optimising nature, as we do in our study.

An alternative approach to nudges is that of boosts. These are behaviourally-informed interventions that target competences rather than behaviour directly [41,42]. Differently from nudges, which start from the implicit assumption that biases are difficult to overcome, boosts attempt to contain cognitive pitfalls to foster desirable behaviours. A key goal for boosts is to enable individuals to exercise their own agency [42,43], regardless of what the optimal course of action may be.

Applications of boosts are less common in the literature, but a list of examples can be found in [43]. Amongst these, fast-and-frugal trees (FFTs) stand-out as a promising avenue to boost competences in uncertain environments, as is the domain of residential energy use. FFTs, a type of intervention used extensively in medical decision-making [44], are highly-accessible, logical decision trees that can be used to follow a series of rules of thumb that culminate in prescribing a decision via answering several yes/no questions. These binary questions are related to predictor variables, and prescribe a specific decision (or an “exit” to the tree) at each node, with a final exit at the final node [45]. An example in the domain of medical decision-making can be found in Fig. 1. Furthermore, FFTs find a close real-world connection to widely distributed energy management “tips” for households.[1]

To our knowledge, there exists only two examples in the literature that explicitly consider boost-like interventions in the domain of residential energy-use. With respect to consumption, Lazaric & Toumi [47], in a field application, provide consumers with information about problems related with energy consumption and offer practical advice to households on how to reduce it. This type of information is in principle similar to that which could be used to design a FFT for use in a real-world setting. Another notable contribution is that of Blasch et al. [48] who, in the context of an online experiment, study whether individuals exposed to decision-support tools are more likely to identify efficient household appliances, finding positive results.

2.3. Justice considerations of nudges and boosts

In recent years, increasing attention has been paid to the evaluation of interventions in terms of energy justice [49]. Energy justice is the idea of applying justice principles to energy-related issues, including policy, consumption, and more [8].

Consequently, behaviour-change interventions in the field of residential energy use have also been evaluated under this lens, most notably by DellaValle & Sareen [10]. In their paper, they evaluate ethical implications of behaviourally-informed interventions concerning the extent to which (i) people's goals are known (information problem), (ii) targeted people are initially endowed with cognitive skills and motivation (multi-dimensional problem), and (iii) policy designers are error-prone and benevolent (political economy problem). They argue that understanding these factors is “crucial for assessing the ethical implications of interventions in contexts that drive unfair distributional patterns, and affect individual capabilities, recognition and participation” (pg. 4). At the core of these issues lay considerations regarding the agency of decision-makers when faced with nudges and boosts.

Taking as an example the multi-dimensional problem, if we assume, as suggested by scarcity theory, that vulnerable individuals are prone to be cognitively loaded as a consequence of living in strenuous conditions [28], then it may be particularly effective to use a nudge-like intervention, such as a default, to enable the efficient use of energy resources in the least cognitively demanding way possible. However, while this intervention may be particularly effective in promoting desirable behaviour, it is not focused on enabling individuals to apply their own agency, which can be particularly important for vulnerable populations [50]. In this regard, a boost may be more desirable.

Furthermore, as a boost requires an individual's active participation to be effective, they are generally considered more transparent than nudges [51,52]. This means they can effectively act as assurances against bad-faith policy actions that may exploit behavioural biases to nudge individuals in directions that are sub-optimal for them.

Overall, the effectiveness and ethical considerations of nudges and boosts must be considered relative to their distributional consequences, so that policymakers are able to take a targeted approach that responds to the needs of different energy consumers. In light of these ethical considerations, we agree with the claim of Santos Silva [51] who states “The choice of a regulatory strategy should rather be made on a case-by-case basis and boosting should ideally precede nudging” (pg. 11).

3. Materials and methods

3.1. Research questions

Given the present state of the literature on behaviourally-informed non-monetary energy management interventions, we are interested in addressing three critical gaps: (i) what is the context-dependency of intervention success (with relation to income and associated cognitive function)? (ii) Can a new type of intervention (namely boosts) be a successful direction for future empirical research? (iii) Can we identify any difference in persistence effects once the interventions are removed? Accordingly, we define three research questions that help us address these gaps in our setup adopting an incentive-compatible virtual energy management task:

Q1. Do participants recruited from different income groups perform differently in our task?

Q2. Is a boost approach effective at improving performance in our task compared to a nudge approach?

Q3. Is a boost approach associated with higher intertemporal spillovers in performance in our task compared to a nudge approach?

With relation to Q1, we note that our interests are not to directly assess the impact of resource availability on behaviour. Instead, we are interested in studying how the implications of belonging to different income groups reflect on optimising behaviour in this context, with a particular focus on cognitive function. For this reason, we opted against experimentally controlling for resource endowment and instead rely on self-reported measures of income to split our sample, as detailed below.

3.2. Design

We present here the design of our experiment, starting with the experimental task, then the set-up, and finally, the treatments.

3.2.1. Task

Participants move through 3 simulated “working days” (12 rounds each, henceforth each set of 12 rounds is referred to as a Day) consisting of 2 phases each (for an illustration, see Fig. 2). Phase 1 (consisting of 4 rounds) is an effort task where participants are asked to generate the resources they will need to take part in Phase 2. This is done to foster the legitimacy of the experimental endowment [53]. Phase 2 (consisting of 8 rounds) is a framed allocation task meant to simulate energy management decisions in a virtual setting. Choices gathered in Phase 2 are our main observational unit.

In Phase 1, participants are instructed to count the number of 0s in a set of 7 × 7 tables comprised of 0s and 1s. For each correctly counted

[1] A recent example can be found in the following NEA report: https://www.nea.org.uk/wp-content/uploads/2020/10/English-Coronavirus-Energy-Saving-Tips-002.pdf.

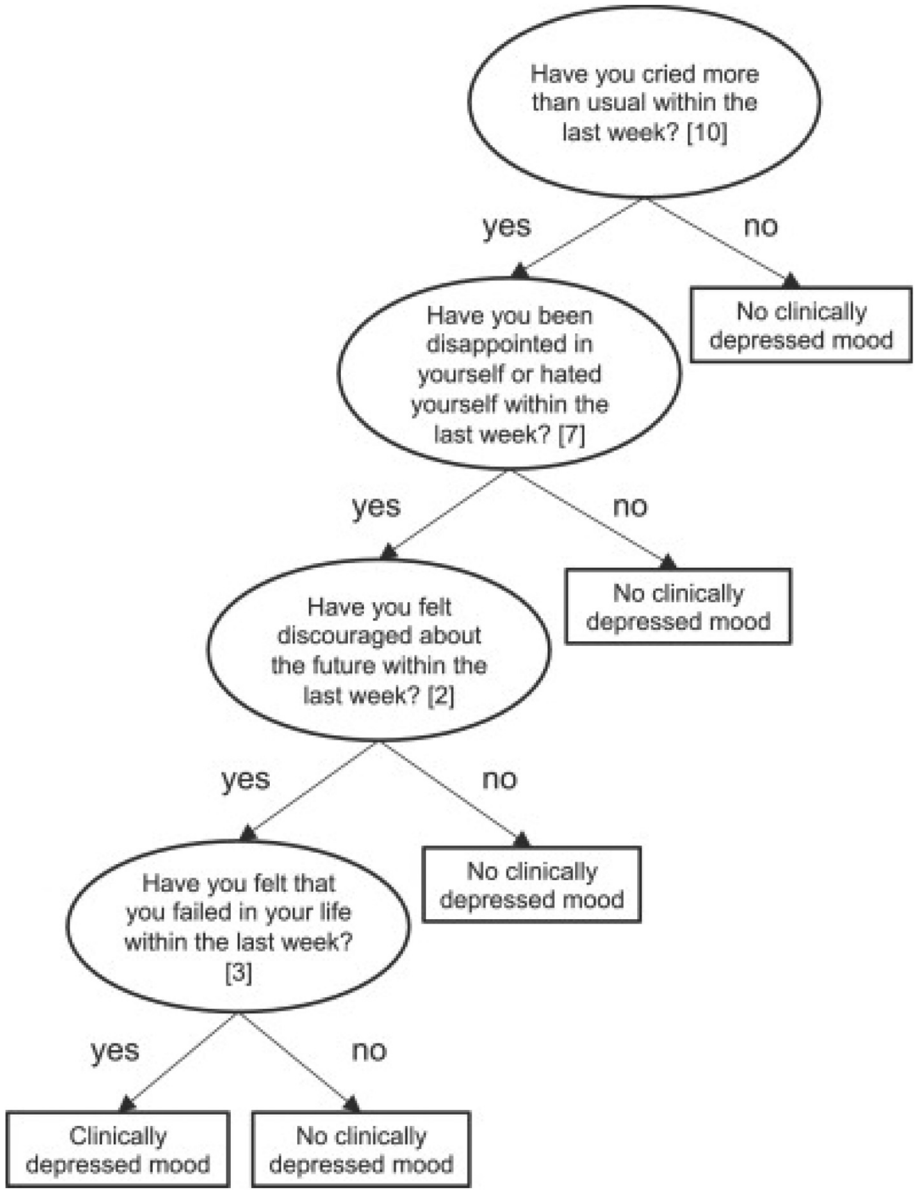


Fig. 1. A fast-and-frugal tree for diagnosing clinically depressed mood. Figure from Jenny et al. [46].

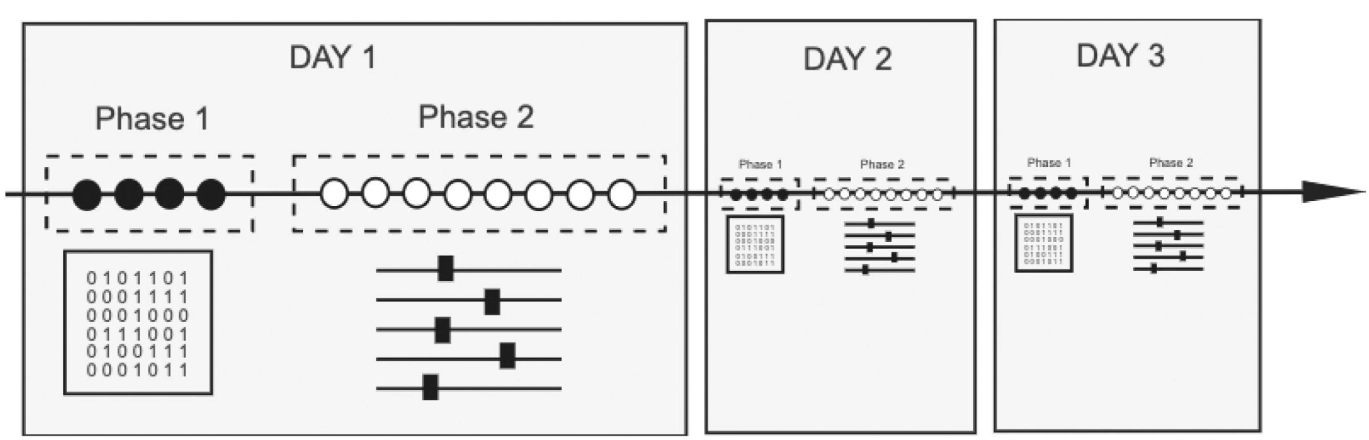


Fig. 2. Structure of the experiment.

table, participants gather 100 “energy consumption points” (ECP) to be used in Phase 2. Participants cannot progress to Phase 2 unless they answer all tables correctly, meaning that by the start of Phase 2 in each Day, every participant begins with the full endowment of 400 ECP.

In Phase 2 participants are required to complete an incentive-compatible energy management task, adapted from Casal et al. [54]. In this task, participants must allocate the ECP to “heat-up” 5 rooms in a virtual apartment across 8 time intervals, or “hours” as they were framed. In practice, this is a repeated allocation task where participants allocate ECP to 5 rooms through a slider (Fig. 3). The task instructions[2] are also framed to increase the likelihood that participants link performance in the task with real-life energy management decisions [55]. This was done to support the study's objective of generating empirical evidence that can support field testing of similar interventions.

Participants are informed that each point allocated is used to purchase a certain amount of energy in terms of kWh, with an exchange rate of 1 ECP = 0.05 kWh. They are further informed that they could assign a total of 20 ECP per slider, purchasing a maximum amount of 1.0 kWh of energy in each room, for each time-slot. Crucially, they are informed that the energy purchased would translate into an amount of earnings of up to 450 experimental currency units (ECU) per time-slot, and that they can claim a cash reward with these earnings.[3] Participants receive instant feedback on their allocation choices, which updates as they interact with the on-screen sliders. They also receive instant feedback in relation to how these choices translated into ECU.

Participants are informed that the maximum amount of earnings could be achieved by identifying an efficient mix of ECP across the rooms. The “efficient allocation” is the optimal allocation of points across all 5 sliders that maximises overall earnings (450 ECU) in a particular time-slot. This efficient allocation changes every Day, meaning it lasts for 8 time-slots of Phase 2, and switches once participants start Phase 2 again in the following Day. This efficient allocation is exogenously set by the researchers and unknown to participants. The payoff function for each slider is adapted from Casal et al. [54], and not shared with participants.

The participant's goal during Phase 2 is, therefore, to identify this efficient mix of allocations across the 8 time-slots of each Day. Faced with uncertainty on what is the optimal allocation, participants should rely on experimenting with their allocations following a trial-and-error approach.

Before moving on, there are two points related to learning in Phase 2 that are worth noting. First, in the baseline condition, participants are informed that the initial allocation of sliders does not follow any specific criterion and would change at the beginning of each time-slot. This sampling of initial allocations is exogeneously set by the experimenters so that each participant in the baseline condition faces the same sequence of initial conditions. This was done to control for any differences in ECU between subjects that could occur from an individual random sampling of allocation distributions. However, these starting allocations would change in each time-slot and participants would not know how they would be set in the following time-slot. Second, in order to allow for learning with this “random” sampling of distributions, participants would receive information between time-slots on their current round allocations, last round allocations, and realised earnings in both cases.

Ideally, participants should try small variations of different allocations, identifying how their allocation decisions in each room impact overall earnings. To stagger learning, we also introduce a time limit of 50 s per time-slot. This time was chosen to be generous enough to give participants the time to make their decisions but not test all possible allocation choices in a single time-slot. The 50-s mark was selected following pilot testing.

Once Phase 2 finished, participants started Phase 1 in the following Day, until the end of Day 3. After this, participants completed a short survey meant to collect a number of demographic characteristics and cognitive competences (more information in Appendix B). At the end of the main task, a Phase 2 round was chosen randomly by the experiment software for payoff.

3.2.2. Experimental set-up

We ran two waves of the task, inviting a number of participants from the first wave (Wave 1) to participate in the repetition (Wave 2).

During Wave 1 we ran a between-subjects 2-by-3 design, totalling 6 experimental conditions, as can be seen in Table 1. We performed a controlled manipulation along two dimensions: intervention and income level. During this stage, a number of sessions were ran with individuals belonging to relatively “higher income” households and individuals belonging to relatively “lower income” households. This income-based split allows us to address Q1.

To address Q2, we then randomised assignment of participants to each treatment group. This includes two groups exposed to interventions, and one control. Finally, to address Q3, in Wave 2 we invited participants from the intervention groups to repeat the task (with the optimising parameters changed). However, during this wave, they did not have access to the interventions they were previously exposed to. As can be seen in Table 2, Wave 2 sessions had a 2-by-2 design, with variation occurring from Wave 1 treatment and income group.

3.2.3. Treatments

Within each income group, participants were randomly allocated into three treatments: control, nudge/default, and boost/FFT.

Participants in the control treatment completed the baseline experimental task as described above, with no intervention to help them optimise their decisions. These participants were also excluded from Wave 2.

Participants in the nudge/default treatment were presented with a default ECP allocation mix at the beginning of each Day. This default allocation mix was set to achieve an earning as close as possible to 300 ECU. This provided participants with a semi-optimising default rule to follow, saving them the effort of testing different allocations to reach 300 ECU. At the beginning of each new Day, the allocation was changed to reflect the new efficient allocation mix. Participants still had control of the slider during the time-slots and could choose to change allocations to maximise their earnings further.

Finally, participants in the boost/FFT treatment completed the task under the same conditions as the control treatment, but were presented with an experimentally-designed FFT (Fig. 4). This FFT was designed to provide participants with easily implementable rules of thumb to apply in the uncertain decision-environment, allowing them to uncover the optimal allocation of resources sooner by fostering a sensible trial-and-error approach with the sliders. Specifically, the tree is designed to follow a one-bounce heuristic [56,57]. It prescribes the decision-maker to continue to increase or decrease initial allocations until earnings stop improving, at which point the decision-maker is instructed to move on to the next slider. The FFT was presented initially during instructions to Phase 2, and could be accessed through a tab at all times (when choosing allocations, and between time-slots). Data collected in the main experiment show that only in 1.7 % of the cases was the FFT information accessed after the instructions and that this happened mainly in early time-slots. Indeed, in the first time-slot, the rate is 12.2 %.

Our data in Wave 1 show that, overall, 16.5 % of the choices in nudge/default condition match the initial value given to the participants. When taking the same measure for the conditions baseline and boost/FFT, where initial configurations are random, we obtain that 9.8 % of the choices match the initial value. Thus, in nudge/default the persistence of the initial configuration is much higher than in the other

[2] All instructions can be found in Appendix D.

[3] The rate chosen to convert ECU to real currency was 1 ECU = 0.013 GBP. If they reached 450 ECU in a time-slot therefore their earnings for that time-slot would be 5.85 GBP. Participants were informed about this conversion rate.

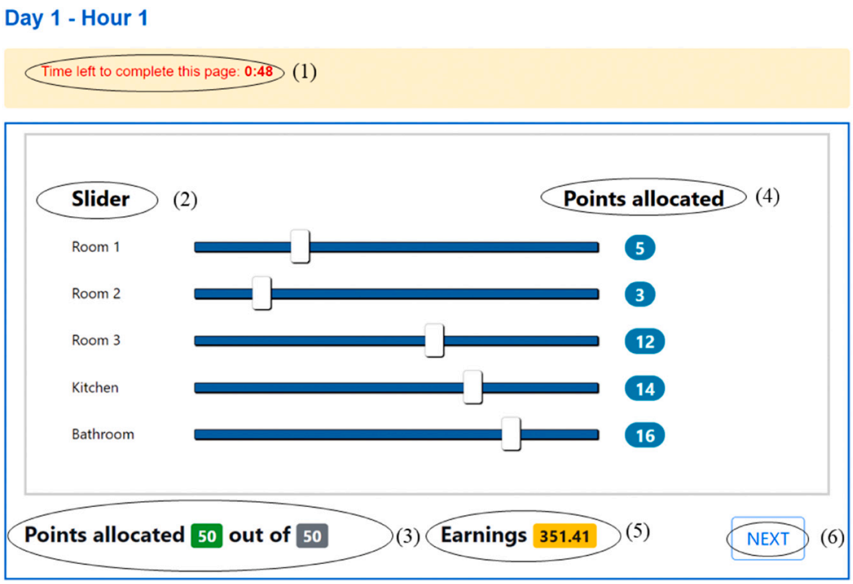


Fig. 3. The task interface in Phase 2. (1) Time limit: once no time is left, the current allocation would auto-select and participants move to the results page. (2) Sliders: participants would use the sliders to select their allocation of ECP in each room. (3) Points allocated (total): Feedback on points allocated overall. (4) Points allocated: Feedback on points allocated per room. (5) Earnings: Immediate feedback on how overall allocation translates into ECU earnings. (6) Next: Before time was up, participants pressed this button to progress.

Table 1
Manipulation set-up for main task in Wave 1.

Wave 1	Control	Nudge/default	Boost/FFT
High income	HI-C	HI-N	HI-B
Low income	LI-C	LI-N	LI-B

Table 2
Manipulation set-up for main task in Wave 2.

Wave 2	Pre-boost	Pre-nudge
High income	HI-PB	HI-PN
Low income	LI-PB	LI-PN

conditions, suggesting that participants understood the informative content of the values on the screen. Concerning the time spent in the task, 16.3 %, 14.5 %, and 11.6 % of the allocations were submitted when time was over (timeout) in treatments boost/FFT, nudge/default, and control, respectively. Amongst those who did not experience a timeout, the longest average time on the page was registered for boost/FFT (28 s), followed by control (25 s), and nudge/default (21 s). According to non-parametric tests (Wilcoxon rank-sum test), the individual-level average time spent on a page is significantly different between boost/FFT and the other two treatments (both p-values < 0.011). At the same time, no statistical difference is observed when comparing control and nudge/default (p-value $= 0.237$). Thus, individuals consistently spend more time in the more cognitively demanding setting, while defaults do not deliver a significant advantage in terms of time. In terms of average time on page by income class, no significant differences are observed in boost/FFT and control (both p-values > 0.26). In nudge/default those in the high income group seem to spend more time on average on the task than their low income counterparts (25 and 23 s, respectively), but the difference is statistically only marginally significant (p-value $= 0.066$).

3.2.4. Elicitation of cognitive competences

We should expect to see differences in important characteristics between income groups, as income will undoubtedly be confounded with a variety of different aspects, including demographics, preferences, and, importantly, cognitive function. We embrace here the conceptualisation of cognitive function adopted by Schilbach et al. [58] (under the umbrella term of "mental bandwidth") who describe it as the cognitive ability to perform higher-level decisions and behaviours.[4] Crucially, under this lens, cognitive function is seen as a limited resource, one which can be depleted as a result of the cognitive load that results from living in strenuous conditions.

The cognitive function of individuals across income groups is indeed a key aspect in our set-up: we assume low income individuals will be exposed to environments more cognitively demanding, leading to cognitive load negatively impacting their competences [e.g., 28]. We expect reduced competences to then affect performance in the task. Therefore, it will be important to elicit them.

In the post-task survey, we use a modified cognitive reflection test [59], and numeracy skills test [60] to elicit competences. The first is an easily-implementable measure of executive control often used in experimental settings, while the second is a general measure of numeracy, shown to be negatively correlated with cognitive load [61]. These two measures were chosen in large part due to their practicality and quick implementation. While we acknowledge that more precise measures of cognitive function could have been implemented (such as Raven's Matrices test [62]), we believe our elicitation strategy is suitable to assess the general level of cognitive competences of participants at the end of an already lengthy task. We then average the performance over 6 survey items to generate an average cognitive score for participants from 0 (all questions answered incorrectly) to 6 (all questions answered correctly).

3.3. Behavioural predictions

Based on our review of the literature, we can make preliminary predictions as to the direction of results.

With relation to the role of income, following predictions of scarcity theory, we predict that individuals in the low income cohort will be more cognitively loaded, leading to worse performance:

P1 Individuals in the low income group will perform worse than their high income counterparts.

To test our prediction we explicitly refer to the elicited cognitive

[4] In particular, the concept of cognitive function is seen as encapsulating cognitive capacity (closely related to fluid intelligence) and executive function.

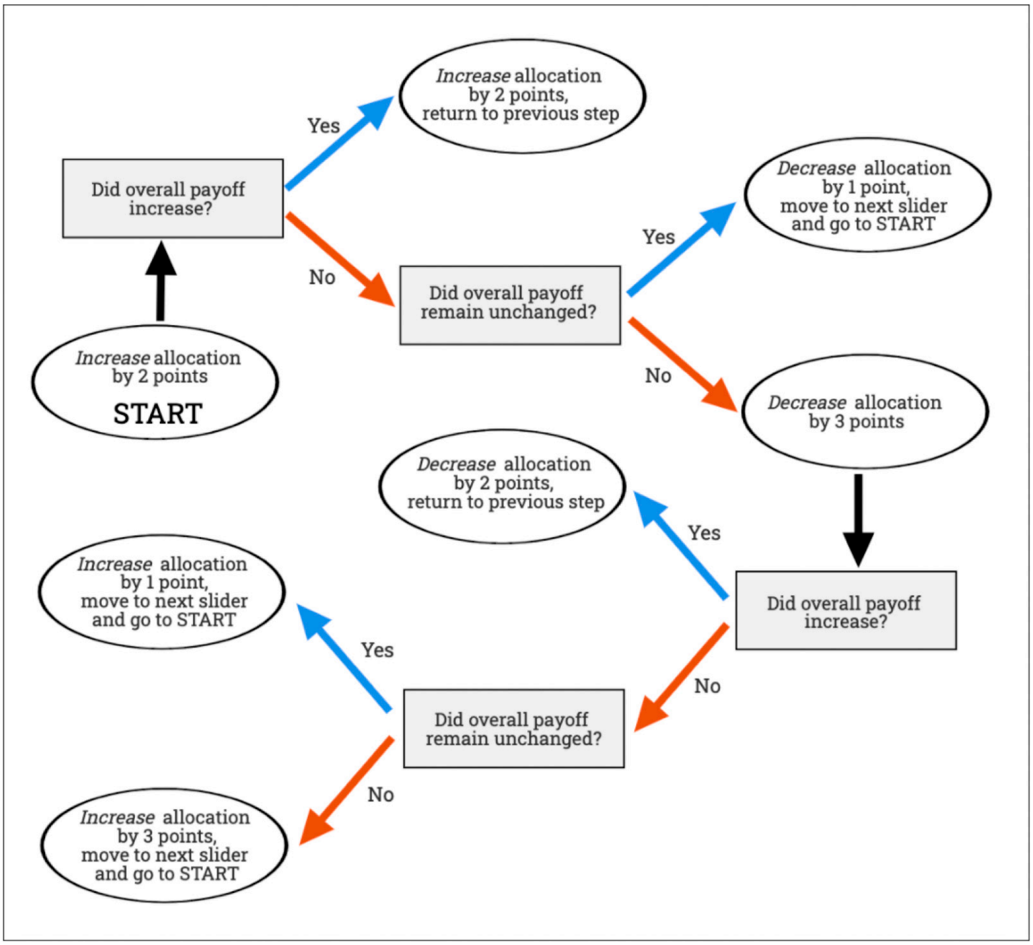


Fig. 4. The FFT designed for the task and presented to participants in the boost group. We note that the fact that the second node prescribes two actions instead of one is unusual for an FFT. Given our set-up, this implementation was necessary in order to allow us to present a single tree, instead of two, which may have added complexity. However, as one of the actions then leads to the following node, it cannot be considered an "exit" to the tree. Therefore, we believe the tree can still be categorised as an FFT.

score as a potential mediator of the effect of income. We acknowledge that different incentives to perform well in the task between income groups might also be relevant. Individuals in the low income group might be proportionally more motivated to do well than their high income counterpart. However, past experimental work suggests that it is unlikely to be significant [63].

With regards to the effect of interventions, we make the following prediction for Wave 1:

P2 Individuals exposed to the default will perform better in Wave 1 than individuals in the control and FFT groups.

This prediction is based on the large existing literature on the effectiveness of default interventions in the residential energy domain [20,23]. With regards to boosts, while existing research does highlight their success in similar online environments [48], their theoretical basis implies most of the benefits of these interventions are accrued once competences are mobilised. This process is likely less immediate than that of defaults.

Boosting may however be more conducive to the internalisation of optimisation strategies than nudging, which leads to the following prediction.

P3 Individuals previously exposed to the FFT will perform better in Wave 2 than individuals previously exposed to the default.

This prediction is again based on the theoretical basis of boosts which, acting on competences rather than environment, are expected to lead to lasting behaviour change even once removed [42]. Furthermore, the prediction is in line with findings from van Roekel et al. [12] who find preliminary evidence of a boost intervention being more persistent than a nudge intervention on hand hygiene protocol compliance. Nudges in the residential energy sphere have shown some evidence of persistence [64], but this is largely a result of habit formation. Our set-up abstracts from this aspect of real-life, so we do not expect the impact of the default to be persistent in our case.

3.4. Procedure

The experiment was pre-registered[5] and ran throughout the month of March 2021, with multiple sessions taking place for both income groups, and Wave 2 repetitions taking place exactly 1 week after Wave 1 for each session. This was done to maximise participation, assuming most participants would still be active on the platform after a week, and relatively familiar with the task (although instructions were repeated).

The experiment was programmed and conducted with the support of oTree [65]. The platform Prolific was used to recruit participants, and its pre-screening filters were used to categorise participants into low and high income groups based on self-reported household budgets. In order to make this categorisation, we limited our pool of potential participants to only those with a UK residence to allow us to split the sample on the basis of a national median split of household budget. The UK was chosen as it is the European country with most active participation on Prolific.

According to the Office for National Statistics, the median household disposable income in the UK for 2020 was £30,800. Using the income brackets available in Prolific, we classified households living in the UK as low income when they reported a net income less than £20,000, and as high income when they reported net incomes between £40,000 and £60,000. The large gap in reported household budgets between the two

[5] The pre-registration can be found at: https://osf.io/h6pnw.

groups was intentionally set to ensure the largest possible difference in income between groups.

Furthermore, in order to avoid observing extreme behaviours in the task owning to different household situations and energy needs that may affect preferences, we further restricted the sample to participants belonging to 3–4 person households. Finally, to ensure maximum engagement with the task and optimisation of the software we also imposed the restriction that the task had to be accessed from a computer (rather than a tablet or smartphone). We note that due to the online nature of our experiment, we do not have full control over the environment in which individuals make decisions. We cannot exclude that participants were distracted while going through the task.

Our sample size was defined according to an ex-ante power analysis for two-group independent sample *t*-test. Given the absence of directly comparable studies from which to infer the effect size, we assumed an effect size between 0.40 and 0.50 (a "medium" effect according to Cohen [66]). With a conventional power of 0.80 and an error of 0.05, we computed our target sample size to be around 70 observations per treatment. This meant that for Wave 1 we aimed to recruit 420 participants. In the end, 429 participants were recruited in Wave 1, 216 of them taking part in the high income session and 213 in the low income session. Of these, 206 participated again in the Wave 2 repetition. This is short of our goal of 280 recruited for Wave 2 (70 per 4 treatments in our 2-by-2 design in Wave 2). We were unfortunately limited in our recruitment efforts by the willingness of participants to return to participate, which may be particularly difficult in an online platform. The anticipation of possible attrition also motivated our design choice to not have a pre-intervention observational wave where none of the participants were exposed to the intervention. Although we concede that this would have been a superior design, we expected that a three-wave repeated online experiment would lead to an unacceptable drop-off rate.

4. Results

In what follows, we analyse the data from Wave 1 and Wave 2 sessions separately, presenting in turn our results. We first report descriptive summary statistics of earnings, followed by non-parametric statistical tests of our hypotheses, finally detailing the results of a regression approach.

4.1. Wave 1

4.1.1. Descriptive statistics

In Table 3, we report a few descriptive statistics about per-round earnings by income level and treatment. This outcome follows directly from choices of the participants in the allocation task and will be our main unit of observation.

Median values are close to the upper limit of 450 in all conditions for the high income group. The share of choices in correspondence to the maximum payoff is 41.6 %, 42.8 %, and 45.0 % in FFT, default, and control, respectively. In the low income group, median choices are higher than in the high income group only in default, where they equate to the maximum achievable payoff. The share of choices in correspondence to the maximum payoff is 32.5 %, 38.2 %, and 51.8 % in FFT, control, and default, respectively.

Table 3
Descriptive statistics of earnings in Wave 1.

Income	Treatment	N	Mean	Median	SD
High	FFT	1704	387.946	440.270	93.573
High	Control	1752	373.967	440.270	106.299
High	Default	1728	387.838	440.270	90.533
Low	FFT	1632	346.320	384.265	111.209
Low	Control	1752	356.366	422.340	113.013
Low	Default	1728	384.472	450.000	104.429

Note: the reported statistics refer to per-round earnings in ECU across groups.

Looking at differences in means across rows, we observe distinctions in treatment effects between high and low income groups. At both income levels participants in the default treatment perform relatively better than those in control, on average (3.64 % difference in high income-default relative to control, 7.59 % difference in low income-default relative to control). However, when it comes to the FFT treatment, participants in the high income group benefit marginally from the intervention, on average (3.67 % difference in high income-FFT relative to control), while it actually appears to backfire for those in low income group (−2.86 % difference in low income-FFT relative to control).

The difference in the effectiveness of the boosting approach between income groups is further confirmed by Fig. 5. The figure reports on individual-level average earnings across different experimental conditions and Days. Specifically the distribution of earnings is captured by boxplots grouped by Income Level, Day, and Treatment.

The graph highlights how in all treatments there are learning effects, with performances improving over the duration of the task. The performance of the high income group is relatively stable across treatments. In contrast, the low income group displays more considerable differences, with participants exposed to the FFT generally performing worse than those exposed to the default and those in control. The backfiring effect of the boost in the low income group is mainly concentrated on Day 1, with the gap relative to high income participants and other treatments shrinking over time.

4.1.2. Statistical tests

Our first statistical analysis approach involves running non-parametric tests on average per-round earnings (in ECU, we refer to this measure as performance) at the individual level to identify differences between groups.[6] Details about the statistical tests are reported in Appendix A.1.

The tests show that there are significant differences in performance across income levels for those participants under the FFT treatment ($p < 0.05$). Interestingly, we can reject the null hypothesis of no difference in distributions between treatments (namely when considering differences in FFT-default exclusively) for subjects in the low income group, but not for those in the high income group. This further highlights (following our descriptive analysis) that there is a large treatment gap for subjects in low income group (the FFT seems to backfire, whereas the default is slightly effective in optimising behaviour) and a smaller one for subjects in high income group.

Finally, the differences between the first and last day of activity are statistically significant for every treatment condition and for both income levels (all p-values < 0.001). Thus, learning dynamics in behaviour are relevant in our task, irrespective of the experimental condition and individual features.

4.1.3. Regression analysis

Using a regression approach we are able to control for factors varying at the individual level, complementing the above analysis. To control for repeated observations, we implement linear mixed models with random effects at the individual level. In our model specifications we take per-round earnings as the dependent variable.

We take a nested model approach. Starting from our most parsimonious model (Model 1) we include only treatment effects, income effects, and learning effects proxied by the variable Day. The model is therefore specified as follows:

$$Y_{it} = \alpha + \beta_1 Income_i + \beta_2 Treatment_i + \beta_3 Day_t + \varepsilon_{it} \quad (1)$$

[6] For differences across income and treatments, we adopted a two-sample Wilcoxon rank sum test on individual-level averages. For differences across days, we adopted a Wilcoxon signed rank test on individual-level averages. Details about the set of tests performed is reported in Table 7 in A.1.

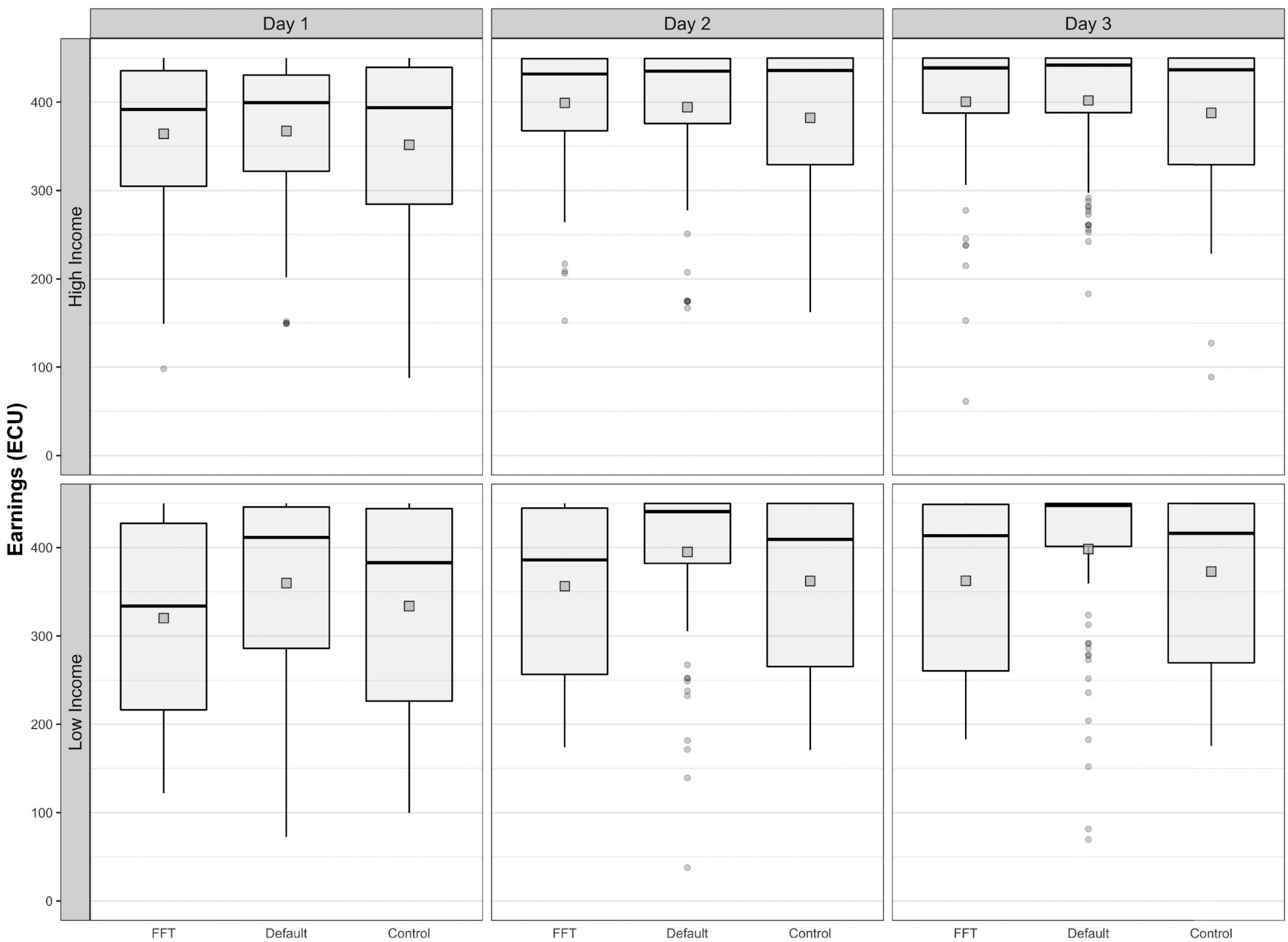


Fig. 5. Earnings by day, income level, and treatment.
Note: the reported statistics refer to mean per-round earnings in ECU.

where Y_{it} stands for earnings of participant i in round t, $Income_i$ is a factor variable that captures the income group a subject belongs to, $Treatment_i$ is a factor variable that captures what intervention group a participant belongs to, and Day_t is another factor variable that represents which day each round is a part of. The term ε is the stochastic error term. The effects of interest to address our research questions with this model are β_1 and β_2, which capture the effects of income-level and treatment assignment, respectively. The results from estimation of this model can be found in column (1) of Table 4.

From Model (1), income-level appears to be a significant determinant of performance in the task, with those in the high income bracket performing better than those in the low income one. Being exposed to the FFT does not have a noticeable effect on performance, but receiving the default does seem to lead to slightly improved performance. Crucially, there are significant learning effects at play in both days.

In Model (1), however, we do not control for potentially relevant individual differences in control variables. In particular, we are interested in assessing whether our measure of cognitive competences impacts performance in the task. Furthermore, we also control for age and gender.[7] When controlling for additional control variables (column (2) in Table 4), a clearer picture emerges on treatment and income effects.

Income is no longer significant, suggesting that the effect of income was moderated by another variable, most likely cognitive score, which has a highly significant impact on performance ($p < 0.01$) and significantly differs for the two income groups (see Appendix C).

The impact of the default becomes only marginally significant, indicating that when controlling for individual differences the default rule for sliders has a weaker impact on the performance. Learning effects remain substantially significant even when controlling for the demographic characteristics. Unlike what was reported in Section 4.1.2, the regression analysis does not detect any significant difference amongst individuals differing in income levels who are exposed to FFTs. This is most likely because the regression controls for other relevant factors, specifically cognitive competences.

We performed a mediation analysis of the relationship between income, cognitive score, and performance. In our path analysis model, we explicitly allowed the effect of income on performance to be mediated by cognitive score. The analysis performed on average performance values at the individual level shows that cognitive score fully mediates the effect of income, as no significant direct effect of income on performance is identified ($p = 0.202$).

We also consider interaction effects between our key variables of interest (income, treatment) and other significant variables from estimates in column (2). In column (3) we report the estimates of a model encompassing interactions between key variables, as well as with cognitive score and learning effects. Overall, interaction effects appear not to be significant, with the only exception given by a weakly positive

[7] In an exploratory analysis we also controlled for several individual-specific characteristics, attitudes and preferences collected in the questionnaire (Appendix B). Adding these control variables did not substantially change the regression results.

Table 4
Wave 1 regression models. linear mixed models with random effects at the individual level. The dependent variable is per-round performance in the task.

	(1)	(2)	(3)
(Intercept)	330.972 (8.513)***	303.033 (15.457)***	305.850 (19.795)***
Income high	20.915 (8.517)*	8.999 (7.389)	9.921 (12.579)
FFT	2.647 (10.457)	4.298 (8.909)	−13.521 (23.708)
Default	21.446 (10.364)*	15.810 (8.829)°	24.809 (24.367)
Day 2	31.984 (1.310)***	31.984 (1.310)***	29.377 (2.250)***
Day 3	37.854 (1.310)***	37.854 (1.310)***	37.791 (2.250)***
Age		−1.711 (0.309)***	−1.697 (0.310)***
Male		29.867 (7.729)***	29.810 (7.777)***
Cogn. score		20.713 (2.266)***	19.980 (3.836)***
Income High:FFT			−1.362 (18.522)
Income High: Default			−5.306 (17.681)
FFT:Day 2			6.188 (3.216)°
FFT:Day 3			1.503 (3.216)
Default:Day 2			1.777 (3.187)
Default:Day 3			−1.262 (3.187)
FFT:Cogn. score			4.003 (5.612)
Default:Cogn. score			−1.592 (5.318)
AIC	112,930.030	112,783.932	112,753.494
BIC	112,987.927	112,863.541	112,891.000
Num. obs.	10,272	10,272	10,272
Num. groups: prolificID	428	428	428

** $p < 0.01$.
*** $p < 0.001$.
* $p < 0.05$.
° $p < 0.1$.

impact on Day 2 for FFTs. Relative to the estimate in column (2), the main effect of the default becomes not significant, while Day and cognitive score remain significant determinants of task performance and are not mediated by treatment.

4.2. Wave 2

We now move on to discuss the results from the analysis of Wave 2 data: 206 out of the 429 participants who completed Wave 1 also completed the second wave (48 %). Given that only some of the Wave 1 participants accepted our invitation to repeat the task, our results in this section must be interpreted in light of our reduced sample size.

We first analyse the potential carry-over effect between waves by testing intertemporal spillovers, i.e., by comparing individual earnings in contiguous rounds of Wave 1 and Wave 2. Then we focus on earnings in the three days of Wave 2, following the same structure of analysis of Wave 1.

4.2.1. Intertemporal spillovers

For our analysis, we take earnings in the last round of Wave 1 and in the first round of Wave 2 and compute the percentage change at the individual level. A positive change would capture a lasting effect of the intervention, while a negative change would capture a decay in the effect. In FFT, 45.9 % of participants obtain the same earnings across waves, 35.2 % register a positive change, and 18.9 % worsen their performance. In Default, 45.2 % of participants receive the same earnings across waves, 23.8 % report a positive change, and 30.9 % worsen their performance. Thus, in both conditions, most participants get at least the same earnings they got in the previous wave. Summary statistics of % changes in earnings of contiguous rounds across waves are reported in Table 5. As a robustness check, we also compared average

Table 5
Individual-level % change in earnings across waves.

Income	Treatment	N	Mean	Median	SD
High	FFT	64	14.1	0.0	59.6
High	Default	37	2.6	0.0	18.1
Low	FFT	58	22.7	0.0	76.1
Low	Default	47	4.0	0.0	26.9

Note: the reported statistics refer to % changes between waves in individual-level per-round earnings in terms of ECU.

earnings on Day 3 of Wave 1 and on Day 1 of Wave 2, and the analysis delivered similar results to those reported here.

The largest positive performance changes are observed for those previously in the FFT group. Wilcoxon signed-rank tests show that the % changes are significantly different from zero for those exposed to the FFT (high income, p-value = 0.047; low income, p-value = 0.008) but not for those exposed to defaults (high income, p-value = 0.513; low income, p-value = 0.915). Thus, the analysis cannot reject the hypothesis that interventions have a positive carry-over effect, with a more substantial delayed learning associated to those previously in the boosting condition. Most likely, this is because those exposed to the default took full advantage of the learning possibilities already in Wave 1, with less room for improvement than those exposed to the FFT.

4.2.2. Earnings

Fig. 6 provides a description of individual-level average earnings distribution in the three days of Wave 2 by income levels.

As the figure shows, median earnings are close to the upper bound limit of 450 each day for both treatment groups and income levels. However, in the low income group, outcomes are more dispersed, and this affects average returns. In the high income group, the share of choices in correspondence to the maximum payoff is 58.6 % and 69.7 % in default and FFT, respectively. In contrast, in the low income group, the share of choices in correspondence to the maximum payoff is 68.9 % and 50.7 % in default and FFT, respectively. Unlike in Wave 1, most allocations achieve the maximum outcome in Wave 2, further testifying to learning spillovers.

Non-parametric tests (see Table 8 in A.2) show that low income participants tend to perform significantly worse ($p < 0.05$) than high income participants when they previously participated in the boost condition, but no significant difference is observed for the nudge condition. The positive impact of being exposed to a default for the low income is also corroborated by the statistically significant difference between the performance of this group and that of those exposed to the FFT. Finally, low income participants significantly improved their performance from Day 1 to Day 3, irrespective of the treatment they were exposed to. Differently, high income participants show a significant improvement only when exposed to the boosting condition in Wave 1.

We use the same regression approach as for Wave 1, with the difference that some of the variables are re-specified to allow for the 2-by-2 set-up instead of 2-by-3. For example, our treatment variable becomes a dummy that takes the value of 1 if the subject was previously in the nudge condition, and 0 if previously in the boost condition. Accordingly, we need to interpret our treatment effect as the difference in performance predicted from the model from subjects previously in the default treatment relative to those previously in the FFT treatment.

As illustrated by Table 6, learning effects are important in Wave 2, similar to Wave 1, with participants improving their performance over time, even though the effect seems to weaken progressively. In Models 1 and 2, we do not identify significant treatment effects, suggesting subjects performed relatively homogeneously in the task regardless of what treatment they were exposed to previously. Once again, differences in income initially present in Model 1 disappear once we control for other variables, particularly cognitive score. When considering Model 3, however (including interaction terms), we find a marginally significant

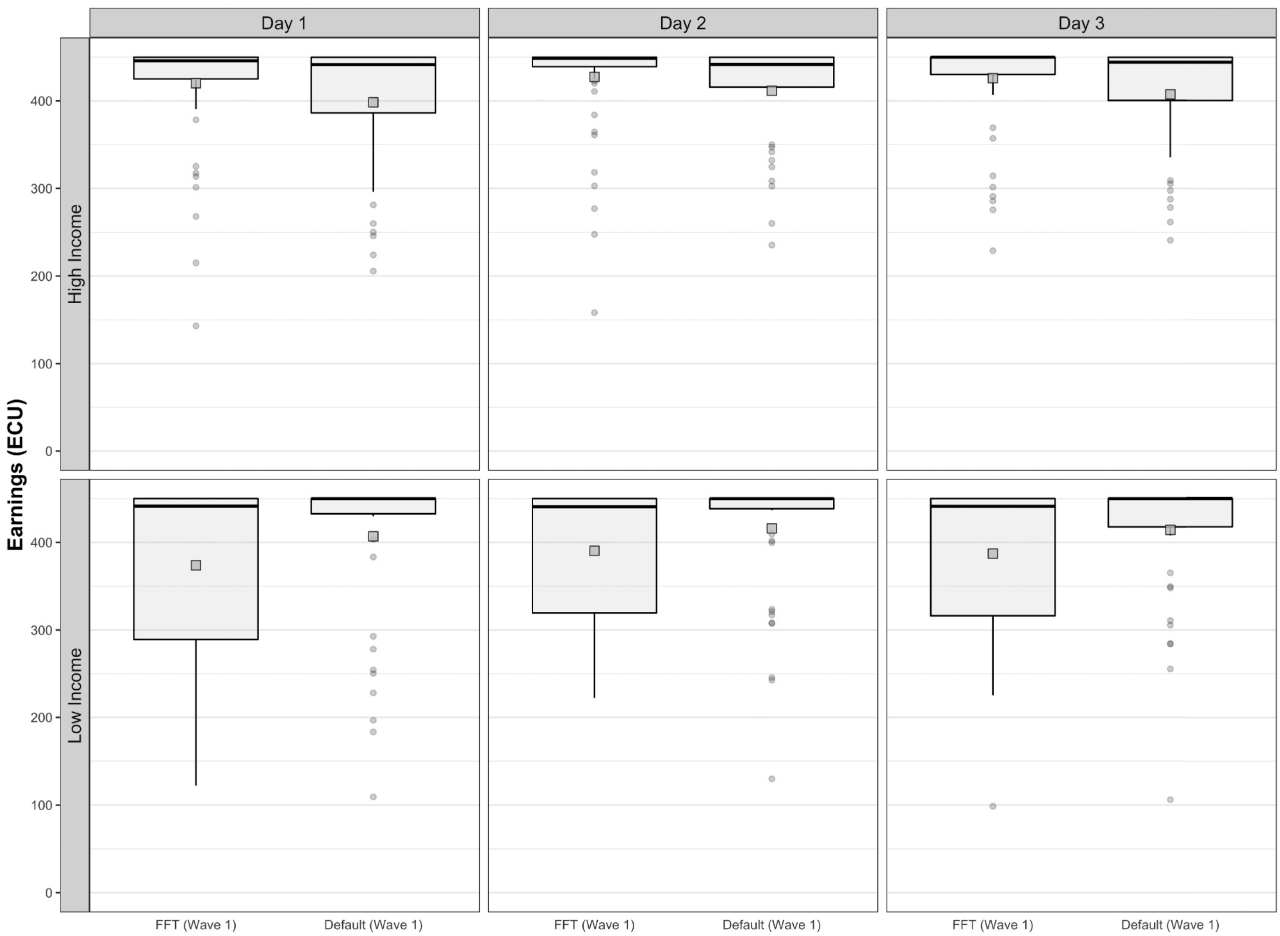


Fig. 6. Earnings by day, income level and treatment in Wave 1.
Note: the reported statistics refer to mean Wave 2 per-round earnings in terms of ECU.

Table 6
Wave 2 regression models. Linear mixed models with random effects at the individual level. The DV is per-round performance in the task.

	(1)	(2)	(3)
(Intercept)	387.099 (8.241)***	346.301 (19.147)***	331.807 (20.672)***
Income high	21.627 (9.845)*	10.501 (8.894)	12.749 (12.026)
Default	6.120 (10.015)	4.549 (8.906)	45.693 (23.399)
Day 2	11.415 (1.363)***	11.415 (1.363)***	11.657 (1.771)***
Day 3	8.933 (1.363)***	8.933 (1.363)***	9.360 (1.771)***
Age		−0.768 (0.368)*	−0.775 (0.367)*
Male		13.244 (9.276)	13.272 (9.247)
Cogn. score		17.444 (2.707)***	20.800 (3.672)***
Income High: Default			−14.068 (18.158)
Default:Day 2			−0.594 (2.774)
Default:Day 3			−1.047 (2.774)
Default:Cogn. score			−8.806 (5.324)°
AIC	51,168.816	51,108.951	51,092.962
BIC	51,214.357	51,174.010	51,184.045
Num. obs.	4944	4944	4944
Num. groups: prolificID	206	206	206

** $p < 0.01$.
*** $p < 0.001$.
* $p < 0.05$.
° $p < 0.1$.

effect of the interaction term between default and cognitive score. This, taken together with the main effect of default, means that the relative impact of the default intervention in Wave 1 is more pronounced for those endowed with lower cognitive capacities.

5. Discussion

Our results suggest that, in the context of our incentive-compatible virtual energy management task: (i) individuals in the high income bracket tend to perform better than individuals in the low income bracket, but this is mainly due to differences in cognitive competences; (ii) defaults are overall more effective in promoting efficient choices than FFTs, especially amongst low income subjects, but considerable learning effects are present for both interventions; (iii) there are intertemporal spillovers between waves, with performances at the restart of Wave 2 equal, in condition nudge, or even better, in condition boost, than performances at the end of Wave 1.

In what follows, we assess each result in the context of the present study's strengths and limitations. Following this, we discuss directions for further research based on our findings, and give indications as to what policy prescriptions could follow from a field validation of results.

5.1. Discussion of results and limitations

Firstly, the result that income-level variability in the performance of participants is moderated by cognitive competences is compatible with the literature on scarcity and its impacts on economic decision-making [67]. Under the lens of this theory, our results can be interpreted as an indication that lower levels of income result in reduced cognitive function (reflected in lower cognitive scores), causing vulnerable individuals to make sub-optimal decisions. Of course, income may well be confounded with a variety of other potentially relevant aspects for task performance. Perhaps individuals in low income groups are more likely to live in smaller houses with less rooms to heat, which could systematically impact their behaviour. Our results cannot conclusively state that cognitive function is the only reason behind differences in performance amongst income groups in our set-up, but they do highlight that it is an important aspect worth further exploration.

Secondly, the result that defaults are more effective than FFTs in optimising performance is also subject to some caveats. The negative effects of boosting are mostly clustered amongst participants in the low income group and primarily in earlier rounds. Receiving the FFT seems to negatively impact performance in the low income group relative to the control, possibly a result of increased perceived complexity of the task and misunderstandings related to the use of the FFT. When considering the high income group in isolation, however, performance is much more stable across all three treatments. Once more, this can be interpreted as a manifestation of the predictions of scarcity theory. Lower income individuals living in cognitively strenuous situations have more to gain from a nudge-like intervention that saves cognitive resources than individuals endowed with higher income [10]. Engaging with the boosting intervention can lead to more cognitive strain, an approach that may backfire in vulnerable populations while minimally impacting performance for higher income groups. This heterogeneity of effects in relation to income could prescribe taking a targeted approach when designing behaviourally-informed strategies, mindful of decision processes across different populations.

It must also be noticed that significant learning effects are present for both interventions and that, eventually, performances of participants in the low income group tend to converge to the maximum achievable output even when exposed to the FFT. In our set-up, the maximum earnings ceiling was relatively low and the time limit reasonably constraining, which is obviously detrimental for the effectiveness of boosts which require higher cognitive engagement than nudges. The evidence that by the last rounds even the boost led to improved performance for the low income cohort, considering all the constraining factors of our set-up, is rather remarkable, and suggests that the competences fostered by these tools may be internalised quickly.

Finally, the intertemporal spillovers between waves suggest that skills acquired through experience can be internalised. Counter to our initial prediction, the boosting intervention seems less prompt than the default intervention in optimising behaviour, but it still generates relevant spillover. We did not expect the default to be persistent due to the fact that our set-up abstracts from habit formation. The fact that we do observe persistence is likely a result of the fact that the default in our set-up effectively lowered the cost of experimenting with the sliders in the limited time, (as participants started from a sure baseline of 300ECU). However, the limited time frame of the task, together with the short time between the two waves, calls for caution when attempting to generalise these findings to longer time horizons. Furthermore, as boosts are best suited to foster competences that are useful in a variety of contexts, the fact that Wave 2 was specified in quite similar terms to Wave 1 likely limited the effectiveness of this approach. If Wave 2 added new aspects to the task, beyond simply changing the optimising parameters, it is possible that the boost would have been more effective and our prediction on intertemporal spillovers would have been realised.

All our results must be considered in relation to the context in which they are generated. Similar to a conventional lab experiment, we have a large control over the decision environment, leading to findings with high internal validity. For example, because we set the optimal allocation of energy resources exogenously, we do not need to make assumptions as to what the optimal level of energy consumption is for an individual. However, a disadvantage of this approach relative to field implementation is that external validity remains an issue. Moreover, our results are solely indicative of how nudges and boosts, specified as defaults and FFTs, impact behaviour in this context. Nudges and boosts are large conceptual categorisations that include different sub-types of interventions, all of which could have different behavioural implications. Furthermore, there are likely to be features of real-life energy consumption decisions that are relevant for the effectiveness of these tools that are not captured in our virtual context, i.e.: environmental concerns.

In light of these considerations, to inform policy in real-world settings, our interventions would require further experimentation in a field context and consider many different types of nudges and boosts. Nonetheless, considering the high cost of implementing these interventions in the field, we believe our results can provide guidance that motivates a pilot implementation in the field.

5.2. Policy implications and further research

While the laboratory nature of our results calls for caution in suggesting policy action at this stage, we believe it is important to outline directions for further research to test their replicability in a real-world setting. We also briefly indicate what policy implications a field validation of our results could have.

As mentioned, our interventions were designed in part based on their feasibility in a field context. Defaults acted similarly to programmable or smart thermostats, while the FFTs were intended to simulate the effects of "energy saving tips" included in smart meters or delivered as part of energy training programmes. To our knowledge there exist no experimental literature assessing the relative efficacy of the two interventions. Furthermore, these interventions are not generally analysed in relation to the frameworks of nudging and boosting, which limits the potential for this vast literature to inform real-world applications. A first direction of further research would therefore be to test these interventions in the field, designing and analysing them in relation to the fundamental concepts of nudging and boosting. This would go a long way in extending our knowledge of boosts and nudges and their impacts on energy management decisions.

Given the increasingly important role of energy justice in discussions about behaviour change, and given that our results highlight the importance of considering how decision processes are affected by cognitive aspects when choosing interventions, we suggest that both lab and field research could provide further insights in this direction. Lab experiments that closely control for cognitive load could establish the relationship between cognitive functions and sub-optimal decisions in a similar task, while field experiments trialling similar boost and nudge interventions could replicate our stratified randomisation design to explore the inter-linkages between intervention effect and level of cognitive function in real-world scenarios, using a wider array of elicitation methods.

We further acknowledge that the specific FFT we adopt here might be cognitively demanding. While we provided participants multiple opportunities to engage with the FFT (they could access it without time limit at different stages), the cognitive demand of the FFT makes our intervention less comparable to "energy saving tips", which are typically designed as accessible tools. Furthermore, there is an important relationship between an individual's comprehension of a boosts, and its effectiveness. Recent evidence from Woike et al. [68], testing boosts and other interventions in a pandemic-like simulation, shows that self-assessed levels of boost comprehension were predictive of their success. Future research testing boosts in a residential energy domain should employ less cognitively demanding tools, perhaps shorter FFTs with more contextualised advice. This could further increase the realism of the tool for real-world implementation and lead to more effective boosting approaches.

Turning now to potential policy implications, a field validation of our results could point at the importance of taking a targeted approach when designing behaviour change interventions. Furthermore, the implications of our results could be applied to design interventions that go beyond the residential energy domain, such as to promote water and landfill reductions.

At first glance, our results seem indicative that interventions with the aim of optimising behaviour for a general population should be designed more as nudges than boosts, because the former are less likely to require cognitive strain to engage with and could therefore be more inclusive. This would be compatible, albeit in a different context, with findings from Bradt et al. [11] who find that nudges are more effective than boosts in in increasing WTP for flood insurance. However, the evidence that even in the low income group the impact of boosts kicks in over rounds (leading to indistinguishable performance from control) could have important implications for the role of different behaviourally-informed interventions when targeting vulnerable groups. As discussed previously, taking a boosting approach has agency-related benefits that nudges do not have. Policymakers must balance the agency benefits of boosts against their impact on the target behaviour in the relevant population. Our results seem to suggest that boosts, if applied to individuals in vulnerable conditions, could initially be detrimental to improved energy management, but in time may become increasingly effective while also being more agency-preserving than a nudge. This relies on the assumption that decisions related to energy management have a relatively low opportunity cost and that learning is possible in this environment. Nudges may instead be useful as short-term interventions, perhaps when an immediate change in behaviour is necessary to accommodate a new technological innovation (for example, retrofitting the social housing stock). Even better, a field validation of our results might suggest that policymakers willing to adopt a behaviourally-informed approach must also adopt measures to reduce scarcity in the population altogether. Not only would such a measure have obvious benefits, but it would allow them to adopt boosting as an effective strategy that is both agency-preserving and transparent.

6. Conclusion

Integrating behaviourally-informed interventions that improve residential energy management into demand-side strategies will become increasingly important. In this context, consumer behaviour is expected to assume a central role, and promoting the effective management of energy resources throughout the day will be a key objective. Mindful of this, policy-makers are tasked with uncovering the most effective tools to promote virtuous energy use, and the frameworks provided by behavioural economics can provide valuable insights for policy-makers to decide what tools are best.

Our study introduces a novel approach to test different types of behaviour change interventions in an online experimental setting. In this set-up we explore whether income-level impacts identification of optimal energy management strategies. We also test different types of interventions (boosts and nudges) and study their distributional impacts on optimising behaviour as well as their persistence effects. Our results show that income is a relevant predictor of performance in our task, although the effect is driven by differences in cognitive competences between income groups. Furthermore, we find that nudges are the preferred strategy to obtain an immediate improvement in performance in our task, but that boosts are equally effective in all populations by the last rounds of the experiment. However, boosts do seem to be associated with increased task complexity, which has a backfiring impact on performance in low income populations in the initial rounds. Finally, both interventions seem persistent over time.

Data declaration

The anonymous data and R code are available at https://osf.io/n6q9d/.

Ethics declaration

The research involves adult human participants who previously registered on Prolific platform to voluntary participate in online studies. Participant could freely withdraw from the study at any time. The study involves the collection of personal data that are treated anonymously and will not be traced back to the originator. The study does not involve the collection of sensitive personal data and does not involve the tracking of participants.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The anonymous data and R code are available at: https://osf.io/n6q9d/.

Acknowledgments

We are grateful to Marco Tecilla from CEEL for extensive support in designing the software interface and back-end. We would also like to thank Nives DellaValle, Valeria Fanghella and all participants in the CEEL seminar for useful feedback on an early-stage design of the experiment. Finally, we would like to thank participants from the ICEP 2021 panel session "Challenges and opportunities in Energy Management" for their insightful comments on the work.

Funding

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions, Innovative Training Networks, Grant Agreement No 812730.

Appendix A. Statistical tests

A.1. Wave 1

Table 7
Statistical tests Wave 1. The tests applied in this case all refer to Wilcoxon rank sum tests with continuity correction. All test results in bold are statistically significant at the 5 % level at least.

Test	Statistic	*p*.value
Difference across Income (given treatment)		
FFT	1894.0	**0.029**
Default	2828.5	0.346

(*continued on next page*)

Table 7 (*continued*)

Test	Statistic	*p*.value
Control	2473.5	0.456
Difference across treatments (given low income)		
FFT|default	1828.0	**0.010**
Control|default	2997.5	0.144
Control|FFT	2221.0	0.282
Difference across treatments (given high income)		
FFT|default	2609.5	0.831
Control|default	2659.0	0.904
Control|FFT	2547.5	0.862
Differences between Day 1 and Day 3 (given low income and treatment)		
FFT	144	<**0.001**
Default	194	<**0.001**
Control	265	<**0.001**
Differences between Day 1 and Day 3 (given high income and treatment)		
FFT	279	<**0.001**
Default	230	<**0.001**
Control	320	<**0.001**

A.2. Wave 2

Table 8
Statistical tests Wave 2. The tests applied refer to Wilcoxon rank sum tests with continuity correction, except for the tests on difference in days which refers to Wilcoxon signed rank test with continuity correction (paired data).

Test	Statistic	*p*.value
Difference across Income (given treatment)		
Pre-FFT	1446	**0.034**
Pre-default	1040	0.119
Difference across treatments (given low income)		
Pre-FFT|pre-default	1041.5	**0.035**
Difference across treatments (given high income)		
Pre-FFT|pre-default	1008	0.212
Differences between Day 1 and Day 3 (given low income and treatment)		
Pre-FFT	142	**0.002**
Pre-default	143	**0.042**
Differences between Day 1 and Day 3 (given high income and treatment)		
Pre-FFT	261	**0.011**
Pre-default	114	0.026

Appendix B. Background information

The following pieces of background information were collected in a questionnaire administered at the end of the main task.

- *Cogn_score*: given by the sum of correct answers in
  - a cognitive reflection task:
    * a) A bat and a ball cost $1.10 in total. The bat costs $ 1.00 more than the ball. How much does the ball cost? Provide the answer in cents.
    * b) If it takes five machines 5 min to make five widgets, how long would it take 100 machines to make 100 widgets?
    * c) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half the lake?
  - a numeracy task:
    * a) Imagine that we rolled a fair, six-sided die 1000 times. Out of 1000 rolls, how many times do you think the die would come up even (2, 4, or 6)?
    * b) In the BIG BUCKS LOTTERY, the chances of winning a $10,000 prize are 1 %. What is your best guess about how many people would win a $10,000 prize if 1000 people each buy a single ticket to BIG BUCKS?
    * c) In the ACME PUBLISHING SWEEPSTAKE, the chance of winning a car is 1 in 1000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car?

- *Age*: respondents' age in years
- *Male*: self-reported gender (Male, Female, or Other). In the regression analysis, we adopt a binary specification: if male = 1, otherwise = 0
- *Hours_home*: hours, on average, spent at home in a weekday
- *Retired*: at least one retired person living full-time in the household
- *Children*: at least one child under the age or 16 living in the household
- *Time_pref*: willingness to give up something that is beneficial today in order to benefit more from that in the future? (0: extremely unwilling - 10: extremely willing)
- *Numb_occupants*: people who live full-time in the household
- *Gender_comp*: women over the age of 20 living in the household
- *Altruism*: willingness to give to good causes without expecting anything in return? (0: extremely unwilling - 10: extremely willing)
- *Trust*: assumption that people have only the best intentions (0: completely disagree - 10: completely agree)
- *Env_score*: sum of scores from the following questions
  - "I think of myself as an environmentally-friendly consumer" (0: completely disagree - 10: completely agree)
  - "I think of myself as someone who is very concerned with environmental issues" (0: completely disagree - 10: completely agree)

Table 9 shows a few descriptive statistics about the pieces of information gathered from the final questionnaire for each experimental condition. As shown by the table, observable characteristics are overall balanced across treatments, with a significant difference across treatments only for variable *Numb_occupants* (Kruskal-Wallis tests).

Table 9
Descriptive statistics of background variables. The *p*-values are obtained from Kruskal-Wallis rank sum tests comparing variables across treatments.

Variable	Control			FFT			Default			*p*-Value
	N	Mean	SD	N	Mean	SD	N	Mean	SD	
Age	146	33.171	12.254	139	36.065	12.269	144	33.771	11.299	0.117
Altruism	146	7.377	2.241	139	7.180	2.148	144	6.931	2.258	0.262
Children	146	0.445	0.499	139	0.417	0.495	144	0.507	0.502	0.372
Cogn_score	146	3.897	1.656	139	3.993	1.653	144	4.028	1.681	0.726
Env_score	146	13.377	3.950	139	12.446	3.929	144	13.222	3.555	0.086
Gender_comp	146	1.164	0.665	139	1.194	0.563	144	1.181	0.665	0.581
Hours_home	146	28.041	32.284	139	44.086	163.681	144	24.944	23.928	0.541
Male	146	0.349	0.478	139	0.381	0.487	144	0.465	0.501	0.130
Numb_occupants	146	3.247	0.944	139	3.065	1.092	144	3.285	0.966	0.024
Retired	146	0.130	0.338	139	0.187	0.391	144	0.139	0.347	0.310
Time_pref	146	7.788	1.759	139	7.475	1.771	144	7.410	1.606	0.072
Trust	146	4.884	2.218	139	5.094	2.196	144	5.132	2.467	0.384

It is important to note that some of the answers were non-plausible. Specifically, in *Age* one participant reported being 2 years old, for *Gender_comp* two participants reported a negative number, and for *Hours_home* 56 participants reported a value larger than 24.

Appendix C. Cognitive capacity

We measure cognitive capacity as the sum of the score in a numeracy test and in a cognitive reflection test (see description in Appendix B). Table 10 reports on descriptive statistics about scores in the tests for the two income levels.

Table 10
Results in cognitive tasks for high and low income participants. The numeracy and CRT results are the number of correct answers out of 3, while the cognitive score is an aggregation of the two tests, so number of correct answers out of 6.

	Numeracy task		Cognitive reflection task		Cognitive score	
Income	Mean	SD	Mean	SD	Mean	SD
High	2.532	0.700	1.662	1.159	4.194	1.546
Low	2.319	0.846	1.427	1.218	3.746	1.737

Participants in the high-income bracket are generally characterized by higher scores, and a Wilcoxon Rank Sum test shows that the difference is statistically significant. (p-value = 0.009).

Appendix D. Experiment instructions

Phase 1 Instructions

The following instructions only refer to Phase 1. Instructions for Phase 2 will be provided once you complete this Phase.

In this Phase you are tasked with generating the energy points (EP) necessary to take part in Phase 2.

You are asked to count the number of 0 in different tables such as the one below. For each solved table, you will earn 100 EP. Your goal is to collect 400 ECP by solving 4 tables. You must answer all tables correctly in order to progress to Phase 2. Phase 1 will repeat at the start of each new Day.

The EP you earn in this Phase will be needed in Phase 2 and converted in a bonus payment that will depend upon choices made in Phase 2.

0	0	0	0	1	1	1	1	1	0
1	0	1	0	0	0	1	0	1	0
1	0	1	0	0	1	0	1	0	0
0	0	1	1	0	0	0	0	1	1
0	0	0	1	1	1	0	0	0	0
0	1	0	0	1	1	1	1	1	0
0	1	0	0	1	1	1	1	1	1

Continue

Fig. 7. Instructions for effort task displayed in Phase 1 of Day 1.

Phase 2 Instructions

By completing Phase 1, you earned the right to participate in Phase 2. In Phase 2 you have to use the EP earned in Phase 1 to heat up 5 virtual rooms 8 times. More precisely, in Phase 2 there are 8 timeslots and in each timeslot you have to decide how much energy to purchase for heating in each of the 5 rooms. You can think of the timeslots as hours of a day.

You purchase energy by allocating a certain number of EP to each room through a slider. The allocation will determine how much energy you purchase in each room by a rate of 1 point = 0.05 kWh. You can allocate to each slider up to 20 EP, which is a maximum of 1.0 kWh of energy purchased per room. Across all rooms you need to allocate exactly 50 EP in each timeslot, or purchase 2.5 kWh every hour.

According to the purchasing choices you make, you can earn an additional bonus payment. The additional bonus will be expressed in virtual currency (ECU): each ECU will be converted in 0.013 GBP at the end of the experiment. A random timeslot will be chosen at the end of Day 3 and the ECU you have earned in that timeslot will determine your payment.

The total earning of each timeslot is computed as the sum of the ECUs obtained from your choices in each room: the maximum total earning in each hour (which is reachable by distributing 50 EP) is equal to 450 ECU or 5.85 GBP. You will get immediate feedback on the number of ECU from your choices across all rooms displayed under the sliders.

There is only 1 efficient level of energy per room (the level of energy that will maximize your earnings from that room), and only 1 efficient mix of choices which maximizes your earnings across rooms. The efficient level of energy in each room stays the same within a day but changes between days, ie: if the efficient level of energy is 0.75kWh (15 EP) in room 3 in Day 1, the same might not be true in Day 2. Your goal is to find the efficient mix of energy across rooms each day, maximizing therefore your earnings.

In every new timeslot, your starting choices will be selected randomly. You can adjust them how you choose using the sliders.

You have 50 seconds to make your choices in each timeslot, after which you will be shown the results for that timeslot and moved on to the next.

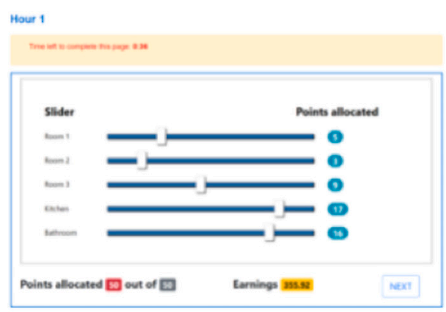


Fig. 8. Baseline instructions for main task displayed in Phase 2 of Day 1 in treatments control and boost.

Phase 2 Instructions

By completing Phase 1, you earned the right to participate in Phase 2. In Phase 2 you have to use the EP earned in Phase 1 to heat up 5 virtual rooms 8 times. More precisely, in Phase 2 there are 8 timeslots and in each timeslot you have to decide how much energy to purchase for heating in each of the 5 rooms. You can think of the timeslots as hours of a day.

You purchase energy by allocating a certain number of EP to each room through a slider. The allocation will determine how much energy you purchase in each room by a rate of 1 point = 0.05 kWh. You can allocate to each slider up to 20 EP, which is a maximum of 1.0 kWh of energy purchased per room. Across all rooms you need to allocate exactly 50 EP in each timeslot, or purchase 2.5 kWh every hour..

According to the purchasing choices you make, you can earn an additional bonus payment. The additional bonus will be expressed in virtual currency (ECU): each ECU will be converted in 0.013 GBP at the end of the experiment. A random timeslot will be chosen at the end of Day 3 and the ECU you have earned in that timeslot will determine your payment.

The total earning of each timeslot is computed as the sum of the ECUs obtained from your choices in each room: the maximum total earning in each hour (which is reachable by distributing 50 EP) is equal to 450 ECU or 5.85 GBP. You will get immediate feedback on the number of ECU from your choices across all rooms displayed under the sliders.

There is only 1 efficient level of energy per room (the level of energy that will maximize your earnings from that room), and only 1 efficient mix of choices which maximizes your earnings across rooms. The efficient level of energy in each room stays the same within a day but changes between days, ie: if the efficient level of energy is 0.75kWh (15 EP) in room 3 in Day 1, the same might not be true in Day 2. Your goal is to find the efficient mix of energy across rooms each day, maximizing therefore your earnings.

In each new day, your initial choices will be selected to get you close to 300 ECU during the timeslots of that day. You may still adjust your choices if you so choose, using the sliders.

You have 50 seconds to make your choices in each timeslot, after which you will be shown the results for that timeslot and moved on to the next.

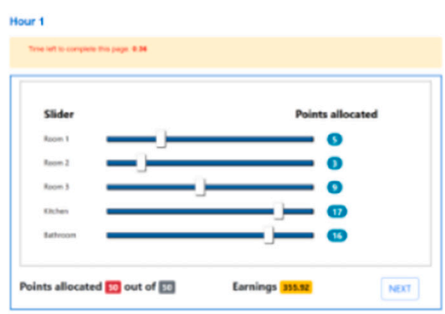


Fig. 9. Default instructions for main task displayed in Phase 2 of Day 1 in treatment nudge.

Some Tips

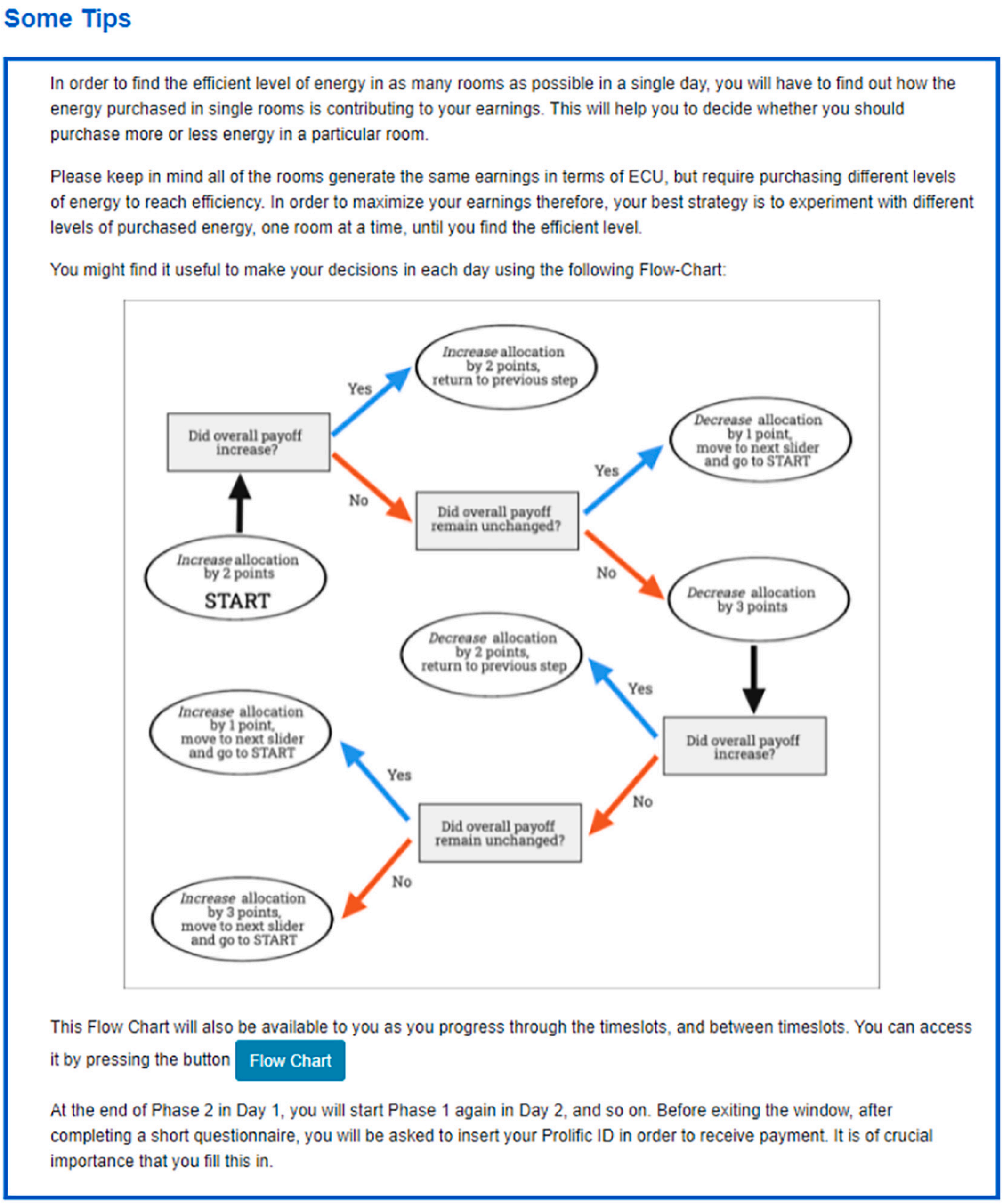
In order to find the efficient level of energy in as many rooms as possible in a single day, you will have to find out how the energy purchased in single rooms is contributing to your earnings. This will help you to decide whether you should purchase more or less energy in a particular room.

Please keep in mind all of the rooms generate the same earnings in terms of ECU, but require purchasing different levels of energy to reach efficiency. In order to maximize your earnings therefore, your best strategy is to experiment with different levels of purchased energy, one room at a time, until you find the efficient level.

At the end of Phase 2 in Day 1, you will start Phase 1 again in Day 2, and so on. Before exiting the window, after completing a short questionnaire, you will be asked to insert your Prolific ID in order to receive payment. It is of crucial importance that you fill this in.

Continue

Fig. 10. Further instructions for control and default treatments displayed in Phase 2 of Day 1.



Some Tips

In order to find the efficient level of energy in as many rooms as possible in a single day, you will have to find out how the energy purchased in single rooms is contributing to your earnings. This will help you to decide whether you should purchase more or less energy in a particular room.

Please keep in mind all of the rooms generate the same earnings in terms of ECU, but require purchasing different levels of energy to reach efficiency. In order to maximize your earnings therefore, your best strategy is to experiment with different levels of purchased energy, one room at a time, until you find the efficient level.

You might find it useful to make your decisions in each day using the following Flow-Chart:

This Flow Chart will also be available to you as you progress through the timeslots, and between timeslots. You can access it by pressing the button Flow Chart

At the end of Phase 2 in Day 1, you will start Phase 1 again in Day 2, and so on. Before exiting the window, after completing a short questionnaire, you will be asked to insert your Prolific ID in order to receive payment. It is of crucial importance that you fill this in.

Fig. 11. Further instructions for boost treatment displayed in Phase 2 of Day 1.

References

[1] T.M. Khanna, G. Baiocchi, M. Callaghan, F. Creutzig, H. Guias, N.R. Haddaway, L. Hirth, A. Javaid, N. Koch, S. Laukemper, A. Löschel, J.C. Minx, M.D.M. Zamora Dominguez, A multi-country meta-analysis on the role of behavioural change in reducing energy consumption and CO2 emissions in residential buildings, Nature Energy 6 (9) (2021) 925–932, https://doi.org/10.1038/s41560-021-00866-x. https://www.nature.com/articles/s41560-021-00866-x#citeas.

[2] S. Bielecki, T. Skoczkowski, L. Sobczak, J. Buchoski, U. Maciag, P. Dukat, Impact of the lockdown during the COVID-19 Pandemic on electricity use by residential users, number: 4 Publisher: Multidisciplinary Digital Publishing Institute, Energies 14 (4) (2021) 980, https://doi.org/10.3390/en14040980, https://www.mdpi.com/1996-1073/14/4/980.

[3] M. of the Energy Research Partnership, How behaviour change will unlock Net Zero, 2021. https://erpuk.org/project/how-behaviour-change-will-unlock-net-zero/.

[4] M.A. Andor, K.M. Fels, Behavioral economics and energy conservation – a systematic review of non-price interventions and their causal effects, Ecological Economics 148 (2018) 178–210, https://doi.org/10.1016/j.ecolecon.2018.01.018. https://www.sciencedirect.com/science/article/pii/S092180091731039X.

[5] E.V. Hobman, E.R. Frederiks, K. Stenner, S. Meikle, Uptake and usage of cost-reflective electricity pricing: insights from psychology and behavioural economics, Renew. Sust. Energ. Rev. 57 (2016) 455–467, https://doi.org/10.1016/j.rser.2015.12.144. https://www.sciencedirect.com/science/article/pii/S1364032115015270.

[6] N. Good, Using behavioural economic theory in modelling of demand response, Appl. Energy 239 (2019) 107–116, https://doi.org/10.1016/j.apenergy.2019.01.158. https://www.sciencedirect.com/science/article/pii/S0306261919301965.

[7] S. Holladay, J. LaRiviere, D. Novgorodsky, M. Price, Prices versus nudges: what matters for search versus purchase of energy investments? J. Public Econ. 172 (2019) 151–173, https://doi.org/10.1016/j.jpubeco.2018.12.004. https://www.sciencedirect.com/science/article/pii/S0047272718302317.

[8] K. Jenkins, D. McCauley, R. Heffron, H. Stephan, R. Rehner, Energy justice: A conceptual review, Energy Research & SocialScience 11 (2016) 174–182, https://doi.org/10.1016/j.erss.2015.10.004. https://www.sciencedirect.com/science/article/pii/S2214629615300669.

[9] E. Heiskanen, K. Matschoss, S. Laakso, E.-L. Apajalahti, Chapter 4.1 - A critical review of energy behaviour change: The influence of context, in: M. Lopes, C. H. Antunes, K.B. Janda (Eds.), Energy and Behaviour, Academic Press, 2020, pp. 391–417, https://doi.org/10.1016/B978-0-12-818567-4.00015-6.
[10] N. DellaValle, S. Sareen, Nudging and boosting for equity?Towards a behavioural economics of energy justice, Energy Research & Social Science 68 (2020), 101589, https://doi.org/10.1016/j.erss.2020.101589. https://www.sciencedirect.com/science/article/pii/S221462962030164X.
[11] J. Bradt, Comparing the effects of behaviorally informed interventions on flood insurance demand: an experimental analysis of 'boosts' and 'nudges', Publisher: Cambridge University Press, Behavioural Public Policy (2019) 1–31, https://doi.org/10.1017/bpp.2019.31, https://www.cambridge.org/core/journals/behavioural-public-policy/article/comparing-the-effects-of-behaviorally-informed-interventions-on-flood-insurance-demand-an-experimental-analysis-of-boosts-and-nudges/47B35632F9BE3F60D20CA7BF31B4A13A.
[12] H.V. Roekel, J. Reinhard, S. Grimmelikhuijsen, Improving hand hygiene in hospitals: comparing the effect of a nudge and a boost on protocol compliance, publisher: Cambridge University Press, Behavioural Public Policy 6 (1) (2022) 52–74, https://doi.org/10.1017/bpp.2021.15, https://www.cambridge.org/core/journals/behavioural-public-policy/article/improving-hand-hygiene-in-hospitals-comparing-the-effect-of-a-nudge-and-a-boost-on-protocol-compliance/44EFD0FF5B87C54332342835337B7C17.
[13] N. Caballero, N.Della Valle, Tackling energy poverty through behavioral change: a pilot study on social comparison interventions in social housing districts, 0, publisher: Frontiers, Frontiers in Sustainable Cities (2021), https://doi.org/10.3389/frsc.2020.601095. URL https://www.frontiersin.org/articles/10.3389/frsc.2020.601095/full.
[14] E.R. Frederiks, K. Stenner, E.V. Hobman, Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour, Renew. Sust. Energ. Rev. 41 (2015) 1385–1394, https://doi.org/10.1016/j.rser.2014.09.026. https://www.sciencedirect.com/science/article/pii/S1364032114007990.
[15] D. Kahneman, Maps of bounded rationality: Psychology for behavioral economics, publisher: American Economic Association, The American Economic Review 93 (5) (2003) 1449–1475, https://www.jstor.org/stable/3132137.
[16] G. Hodgson, On the Limits of Rational Choice Theory Accepted: 2007-09-26T12:31:18Z, University of Hertfordshire, 2003. http://uhra.herts.ac.uk/handle/2299/691.
[17] G. Gigerenzer, Why heuristics work, publisher: SAGE Publications Inc, Perspectives on Psychological Science 3 (1) (2008) 20–29, https://doi.org/10.1111/j.1745-6916.2008.00058.x. URL doi:10.1111/j.1745-6916.2008.00058.x.
[18] A. Tversky, D. Kahneman, Judgment under uncertainty: Heuristics and biases, publisher: Cambridge University Press, Science 185 (4157) (1974) 1124–1131, https://doi.org/10.1126/science.185.4157.1124, https://science.sciencemag.org/content/185/4157/1124.
[19] R. Flynn, P. Bellaby, M. Ricci, The 'value-action gap' in public attitudes towards sustainable energy: The case of hydrogen energy, publisher: SAGE Publications Ltd, The Sociological Review 57 (2_suppl) (2009) 159–180, https://doi.org/10.1111/j.1467-954X.2010.01891.x. URL doi:10.1111/j.1467-954X.2010.01891.x.
[20] D. Pichert, K.V. Katsikopoulos, Green defaults: Information presentation and pro-environmental behaviour, J. Environ. Psychol. 28 (1) (2008) 63–73, https://doi.org/10.1016/j.jenvp.2007.09.004. https://www.sciencedirect.com/science/article/pii/S0272494407000758.
[21] T.J. Brennan, Consumer preference not to choose: Methodological and policy implications, Energy Policy 35 (3) (2007) 1616–1627, https://doi.org/10.1016/j.enpol.2006.04.023. https://www.sciencedirect.com/science/article/pii/S0301421506001960.
[22] H. Allcott, S. Mullainathan, Behavior and energy policy, publisher: American Association for the Advancement of Science, Science 327 (5970) (2010) 1204–1205, https://doi.org/10.1126/science.1180775. URL doi:10.1126/science.1180775.
[23] K. Momsen, T. Stoerk, From intention to action: Can nudges help consumers to choose renewable energy? Energy Policy 74 (2014) 376–382, https://doi.org/10.1016/j.enpol.2014.07.008. https://www.sciencedirect.com/science/article/pii/S0301421514004121.
[24] A. Kearns, E. Whitley, A. Curl, Occupant behaviour as a fourth driver of fuel poverty (aka warmth & energy deprivation), Energy Policy 129 (2019) 1143–1155, https://doi.org/10.1016/j.enpol.2019.03.023. https://www.sciencedirect.com/science/article/pii/S0301421519301879.
[25] N. DellaValle, People's decisions matter: understanding and addressing energy poverty with behavioral economics, Energy Build. 204 (2019), 109515, https://doi.org/10.1016/j.enbuild.2019.109515. https://www.sciencedirect.com/science/article/pii/S0378778819305675.
[26] M.L. Sturge-Apple, J.H. Suor, P.T. Davies, D. Cicchetti, M.A. Skibo, F.A. Rogosch, Vagal tone and children's delay of gratification: Differential sensitivity in resource-poor and resource-rich environments, publisher: SAGE Publications Inc, Psychological Science 27 (6) (2016) 885–893, https://doi.org/10.1177/0956797616640269. URL doi:10.1177/0956797616640269.
[27] R. Mould, K.J. Baker, Documenting fuel poverty from the householders' perspective, Energy Research & SocialScience 31 (2017) 21–31, https://doi.org/10.1016/j.erss.2017.06.004. https://www.sciencedirect.com/science/article/pii/S2214629617301743.
[28] A. Mani, S. Mullainathan, E. Shafir, J. Zhao, Poverty impedes cognitive function, publisher: American Association for the Advancement of Science Section: Research Article, Science 341 (6149) (2013) 976–980, https://doi.org/10.1126/science.1238041, https://science.sciencemag.org/content/341/6149/976.
[29] E.-J. de Bruijn, G. Antonides, Poverty and economic decision making: a review of scarcity theory, Theory Decis. (Mar. 2021), https://doi.org/10.1007/s11238-021-09802-7.
[30] R.H. Thaler, C.R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness, in: Nudge: Improving Decisions About Health, Wealth, and Happiness, Yale University Press, New Haven, CT, US, 2008 pages: x, 293.
[31] H. Allcott, Social norms and energy conservation, J. Public Econ. 95 (9) (2011) 1082–1095, https://doi.org/10.1016/j.jpubeco.2011.03.003. https://www.sciencedirect.com/science/article/pii/S0047272711000478.
[32] C. Charlier, G. Guerassimoff, A. Kirakozian, S. Selosse, Under pressure! Nudging electricity consumption within firms. Feedback from a field experiment, publisher: International Association for Energy Economics, Energy Journal 42 (1) (Jan. 2021), https://hal.archives-ouvertes.fr/hal-02526322.
[33] B.W. Pratt, J.D. Erickson, Defeat the Peak: Behavioral insights for electricity demand response program design, Energy Res. Soc. Sci. 61 (2020), 101352, https://doi.org/10.1016/j.erss.2019.101352. https://www.sciencedirect.com/science/article/pii/S2214629619304384.
[34] A. Colasante, I. D'Adamo, P. Morone, Nudging for the increased adoption of solar energy?Evidence from a survey in Italy, Energy Research & Social Science 74 (2021), 101978, https://doi.org/10.1016/j.erss.2021.101978. https://www.sciencedirect.com/science/article/pii/S2214629621000712.
[35] M. Kaiser, M. Bernauer, C.R. Sunstein, L.A. Reisch, The power of green defaults: the impact of regional variation of opt-out tariffs on green energy demand in Germany, Ecol. Econ. 174 (2020), 106685, https://doi.org/10.1016/j.ecolecon.2020.106685. https://www.sciencedirect.com/science/article/pii/S0921800919317975.
[36] C.-M. Loock, T. Staake, F. Thiesse, Motivating energy-efficient behavior with green is: An investigation of goal setting and the role of defaults, MIS Quarterly 37 (4) (2013) 1313–1332, pages: x, 293, https://www.jstor.org/stable/43825794.
[37] N.D. Sintov, P.W. Schultz, Adjustable green defaults can help make smart homes more sustainable, number: 4 Publisher: Multidisciplinary Digital Publishing Institute, Sustainability 9 (4) (2017) 622, https://doi.org/10.3390/su9040622, https://www.mdpi.com/2071-1050/9/4/622.
[38] T.E. Malinick, N. Wilairat, J. Holmes, L. Perry, undefined, https://www.semanticscholar.org/paper/Destined-to-Disappoint, 2012.
[39] M. Pritoni, A.K. Meier, C. Aragon, D. Perry, T. Peffer, Energy efficiency and the misuse of programmable thermostats: the effectiveness of crowdsourcing for understanding household behavior, Energy Research & SocialScience 8 (2015) 190–197, https://doi.org/10.1016/j.erss.2015.06.002. https://www.sciencedirect.com/science/article/pii/S2214629615000730.
[40] G. Torma, J. Aschemann-Witzel, J. Thøgersen, I nudge myself: Exploring 'self-nudging' strategies to drive sustainable consumption behaviour, International Journal of Consumer Studies 42 (1) (2018) 141–154, https://doi.org/10.1111/ijcs.12404. URL doi:10.1023/A:1007850605129.
[41] C. Schubert, Green nudges: do they work? Are they ethical? Ecol. Econ. 132 (2017) 329–342, https://doi.org/10.1016/j.ecolecon.2016.11.009. https://www.sciencedirect.com/science/article/pii/S0921800916301860.
[42] T. Grüne-Yanoff, R. Hertwig, Nudge versus boost: how coherent are policy and theory? Minds Mach. 26 (1) (2016) 149–183, https://doi.org/10.1007/s11023-015-9367-9.
[43] R. Hertwig, T. Grüne-Yanoff, Nudging and boosting: steering or empowering good decisions, Perspect. Psychol. Sci. 12 (6) (2017) 973–986, https://doi.org/10.1177/1745691617702496.
[44] L.J. Gibbons, K. Stoddart, 'Fast and frugal heuristics': clinical decision making in the emergency department, Int. Emerg. Nurs. 41 (2018) 7–12, https://doi.org/10.1016/j.ienj.2018.04.002. https://www.sciencedirect.com/science/article/pii/S1755599X18300600.
[45] S. Hafenbrädl, D. Waeger, J.N. Marewski, G. Gigerenzer, Applied decision making with fast-and-frugal heuristics, publisher: The Society for Applied Research in Memory and Cognition, Journal of Applied Research in Memory and Cognition 5 (2) (2016) 215–231, https://doi.org/10.1016/j.jarmac.2016.04.011. URL doi:10.1016/j.jarmac.2016.04.011.
[46] M.A. Jenny, T. Pachur, S.Lloyd Williams, E. Becker, J. Margraf, Simple rules for detecting depression, Journal of Applied Research in Memory and Cognition 2 (3) (2013) 149–157, https://doi.org/10.1016/j.jarmac.2013.06.001. https://www.sciencedirect.com/science/article/pii/S2211368113000533.
[47] N. Lazaric, M. Toumi, Reducing consumption of electricity: a field experiment in Monaco with boosts and goal setting, Ecol. Econ. 191 (2022), 107231, https://doi.org/10.1016/j.ecolecon.2021.107231. https://www.sciencedirect.com/science/article/pii/S0921800921002901.
[48] J.E. Blasch, M. Filippini, N. Kumar, A.L. Martinez-Cruz, Boosting the choice of energy-efficient home appliances: the effectiveness of two types of decision support, Applied Economics 0 (0) (2022) 1–23, https://doi.org/10.1080/00036846.2021.2014395, publisher: Routledge _eprint: doi:10.1080/00036846.2021.2014395. doi:10.1080/00036846.2021.2014395.
[49] R.J. Heffron, D. McCauley, B.K. Sovacool, Resolving society's energy trilemma through the Energy Justice Metric, Energy Policy 87 (2015) 168–176, https://doi.org/10.1016/j.enpol.2015.08.033. https://www.sciencedirect.com/science/article/pii/S030142151530077X.
[50] A. Sen, Development as Freedom, 0, publisher: Frontiers, Knopf Doubleday Publishing Group, 2011. URL https://www.frontiersin.org/articles/10.3389/frsc.2020.601095/full.
[51] M.Santos Silva, Nudging and other behaviourally based policies as enablers for environmental sustainability, number: 1 Publisher: Multidisciplinary Digital Publishing Institute, Laws 11 (1) (2022) 9, https://doi.org/10.3390/laws11010009, https://www.mdpi.com/2075-471X/11/1/9.

[52] S. Reijula, R. Hertwig, Self-nudging and the citizen choice architect, 0, publisher: Frontiers, Behavioural Public Policy 6 (1) (2022) 119–149, https://doi.org/10.1017/bpp.2020.5. URL https://www.frontiersin.org/articles/10.3389/frsc.2020.601095/full, https://www.cambridge.org/core/journals/behavioural-public-policy/article/abs/selfnudging-and-the-citizen-choice-architect/F526628F7F3C7B436FA2BCBFC1FC3C76.

[53] J. Abeler, A. Falk, L. Goette, D. Huffman, Reference points and effort provision, American Economic Review 101 (2) (2011) 470–492, https://doi.org/10.1257/aer.101.2.470. URL https://www.aeaweb.org/articles?id=10.1257/aer.101.2.470.

[54] S. Casal, N. DellaValle, L. Mittone, I. Soraperra, Feedback and efficient behavior, number: 4 Publisher: Multidisciplinary Digital Publishing Institute, PLOS ONE 12 (4) (2017), e0175738, https://doi.org/10.1371/journal.pone.0175738.

[55] A. Alekseev, G. Charness, U. Gneezy, Experimental methods: when and why contextual instructions are important, J. Econ. Behav. Organ. 134 (2017) 48–59, https://doi.org/10.1016/j.jebo.2016.12.005. https://www.sciencedirect.com/science/article/pii/S0167268116302876.

[56] S. Mousavi, G. Gigerenzer, Risk, uncertainty, and heuristics, J. Bus. Res. 67 (8) (2014) 1671–1678, https://doi.org/10.1016/j.jbusres.2014.02.013. https://www.sciencedirect.com/science/article/pii/S0148296314000885.

[57] J.D. Hey, Search for rules for search, J. Econ. Behav. Organ. 3 (1) (1982) 65–81, https://doi.org/10.1016/0167-2681(82)90004-X. https://www.sciencedirect.com/science/article/pii/016726818290004X.

[58] F. Schilbach, H. Schofield, S. Mullainathan, The psychological lives of the poor, Am. Econ. Rev. 106 (5) (2016) 435–440, https://doi.org/10.1257/aer.p20161101. https://www.aeaweb.org/articles?id=10.1257/aer.p20161101.

[59] S. Frederick, Cognitive reflection and decision making, J. Econ. Perspect. 19 (4) (2005) 25–42, https://doi.org/10.1257/089533005775196732. https://www.aeaweb.org/articles?id=10.1257/089533005775196732.

[60] I.M. Lipkus, G. Samsa, B.K. Rimer, General performance on a numeracy scale among highly educated samples, Medical Decision Making: An International Journal of the Society for Medical Decision Making 21 (1) (2001) 37–44, https://doi.org/10.1177/0272989X0102100105.

[61] C. Deck, S. Jahedi, The effect of cognitive load on economic decision making: a survey and new experiments, Eur. Econ. Rev. 78 (2015) 97–119, https://doi.org/10.1016/j.euroecorev.2015.05.004. https://www.sciencedirect.com/science/article/pii/S0014292115000690.

[62] W.B. Bilker, J.A. Hansen, C.M. Brensinger, J. Richard, R.E. Gur, R.C. Gur, Development of abbreviated nine-item forms of the Raven's Standard Progressive Matrices Test, Assessment 19 (3) (2012) 354–369, https://doi.org/10.1177/1073191112446655, publisher: SAGE Publications Inc.

[63] C.F. Camerer, R.M. Hogarth, The effects of financial incentives in experiments: A review and capital-labor-production framework, Journal of Risk and Uncertainty 19 (1) (1999) 7–42, https://doi.org/10.1023/A:1007850605129. URL doi: 10.1023/A:1007850605129.

[64] H. Allcott, T. Rogers, The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation, Am. Econ. Rev. 104 (10) (2014) 3003–3037, https://doi.org/10.1257/aer.104.10.3003. https://www.aeaweb.org/articles?id=10.1257/aer.104.10.3003.

[65] D.L. Chen, M. Schonger, C. Wickens, oTree—An open-source platform for laboratory, online, and field experiments, Journal of Behavioral and Experimental Finance 9 (2016) 88–97, https://doi.org/10.1016/j.jbef.2015.12.001. https://www.sciencedirect.com/science/article/pii/S2214635016000101.

[66] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd Edition, Routledge, New York, 1988, https://doi.org/10.4324/9780203771587.

[67] A.K. Shah, S. Mullainathan, E. Shafir, Some consequences of having too little, publisher: American Association for the Advancement of Science, Science 338 (6107) (2012) 682–685, https://doi.org/10.1126/science.1222426. URL doi: 10.1126/science.1222426.

[68] J.K. Woike, S. Hafenbrädl, P. Kanngiesser, R. Hertwig, The transmission game: Testing behavioral interventions in a pandemic-like simulation, publisher: American Association for the Advancement of Science, Science Advances 8 (8) (2022), eabk0428, https://doi.org/10.1126/sciadv.abk0428. URL doi:10.1126/sciadv.abk0428.