Public review for

Beyond clustering: rethinking the segmentation of energy consumers when nudging them towards energy-saving behavior

Merkourios Karaliopoulos, Leonidas Tsolas, Maria Halkidi, Iordanis Koutsopoulos, Stephanie Van Hove, Peter Conradie

This paper explores how interventions, or "nudges", can be used to promote energy-saving behaviors. The proposed approach segments participants into one of the six energy consumer groups based on energy-consuming profiles instead of using traditional data-driven clustering. The authors collect data using web-based surveys from around 5000 people, followed by user segmentation into six energy consumer profiles based on 15 constructs. Recommendations, or "nudges", are created through analysis of each of the six energy profiles based on socio-demographic characteristics and disseminated to participants.

The reviewers find the motivations for this work clear and convincing. The mathematical formulation helps in the understanding of the proposed methods. A large population was studied through a widely disseminated survey, which helps to further substantiate the claims. Reviewers also agree that the revision has improved the clarity of the paper and adequately explained the rationales of various design choices, such as methods used and semantics of thresholds. Reviewers also appreciate the additional detail on measuring sociopsychological constructs presented in the appendices.

Areas for improvements include 1. example questions and responses to the survey, as well as how scores are reported or rated; 2. case study or simulation results showing quantitative energy savings if using the proposed consumer segmentation method with nudging interventions; and 3. more detailed comparisons of proposed clustering methods to state-of-the-art.

Overall, the reviewers and the editor find this paper to be a significant contribution to this area of research and recommend acceptance.

Public review written by
Xiaofan (Fred) Jiang
Columbia University, USA

Beyond clustering: rethinking the segmentation of energy consumers when nudging them towards energy-saving behavior

MERKOURIS KARALIOPOULOS, LEONIDAS TSOLAS, IORDANIS KOUTSOPOULOS, Athens University of Economics and Business, Greece
MARIA HALKIDI, University of Piraeus, Greece
STEPHANIE VAN HOVE, PETER CONRADIE, Ghent University, Belgium

Besides technological innovations in energy production and management technologies, the fight against climate change requires fundamental changes in our energy consumption behavior. Behavioral interventions are key to this process, especially when tailored to different energy consumer segments accounting for their socio-demographic profiles, socio- psychological characteristics and energy consumption practices.

In this work, we propose a novel approach to energy consumer segmentation that facilitates the choice of (nudging) interventions for each segment. We call it intervention-driven energy consumer profiling since it explicitly considers upfront the set of interventions that can be delivered to energy consumers and defines profiles that can be readily matched with them. The profiles are specified as combinations of socio-psychological factors with implications for energy-saving behavior and are parameterized by thresholds that measure how strongly these factors are represented in each profile. One profile represents ideal energy-savers, whereas each of the remaining five profiles shares one or two distinct features that serve as barriers towards energy-saving behavior and/or prescribe specific type of nudging interventions for strengthening such behavior. We use the responses of users to a European-wide online survey to formulate and solve an optimization problem for these thresholds and then assign the survey respondents to the profiles. Finally, we analyze them also in terms of socio-demographic variables and recommend appropriate nudging interventions for them.

Additional Key Words and Phrases: clustering, nudging, segmentation, energy saving, behavioral interventions, optimization

1 INTRODUCTION

In recent years, ambitious targets have been set worldwide for reducing the effects of climate change, with a view towards eliminating CO_2 emissions by 2050. These targets motivate investments and research in several technologies such as energy renewable sources, electric cars, green transportation and energy-efficient buildings. However, it is widely acknowledged that any energy innovation needs to be coupled with fundamental changes in our energy consumption behavior, from (e-)waste reduction to energy conservation on a daily basis, such as heating and cooling our homes and work places; and this remains a challenging task [30].

One promising approach towards sustainable behavioral change are *behavioral interventions*. The term denotes initiatives and techniques that go beyond typical policy tools (*e.g.*, subsidies or tax deductions) and leverage findings from behavioral sciences as to how human behavior emerges out of values, norms, habits, but also social processes and cognitive biases. Nudging interventions, in particular, or "nudges" aim to "alter people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" [28]. Nudges typically modify what is called the choice architecture, *i.e.*, the way the decision/choice alternatives are presented to the end users. For instance, consider a user at the very moment that he/she increases the thermostats' target temperature setting. An energy-saving nudge, called *just-in-time prompt with loss framing*, may be realized through a popup message at his/her mobile or the thermostat display warning against the consequences of this action for the environment and/or the household budget.

Typically, nudges leverage cognitive biases, which are systematic deviations from rational judgment [10]. Within the energy sustainability domain, the most frequently leveraged biases are the availability heuristic and the herd-instinct bias [6] by means of real-time [11] or non-real-time [3][4] feedback and social comparison features [3], respectively. Common to most of these experimental studies is that nudges are applied "horizontally" to all households participating in the experiment, without exploring how appropriate an intervention is for a particular household owner or tenant.

On the contrary, in studies like [7][13][27] [29], distinct segments of energy consumers are identified and interventions (*e.g.*, marketing campaigns) are tailored to each of those segments. The de facto approach to the segmentation task is *clustering*, a common unsupervised learning technique for data analysis. Clustering comes with a rich toolbox of methods and algorithms and the energy consumer segments readily emerge as user groups sharing "similar" *features*, such as socio-demographic profiles, socio-psychological characteristics and energy consumption practices. On the negative side, these groups are not necessarily informational as to which (nudging) interventions are most appropriate for them, in particular when the original feature space is engineered as part of the clustering process, *e.g.*, with techniques such as Principal Component Analysis.

We experienced this shortcomings of clustering ourselves when we applied it to the dataset of this study: a rich set of features about energy consumers and their households, collected through a large Europe-wide online survey. Therefore, and as main contribution of this work, we propose an alternative approach to energy consumer segmentation that we call *intervention-driven* energy consumer profiling. Rather than letting data determine the energy consumer groups with questionable value for identifying interventions, we specify upfront the energy consumer profiles. Specifically, we first identify the set of nudging interventions that are relevant to energy consumers. Next, we define energy consumer profiles that can be readily matched with one or more of those interventions drawing

Authors' addresses: Merkouris Karaliopoulos, Leonidas Tsolas, Iordanis Koutsopoulos, Athens University of Economics and Business, 76, Patision Str., Athens, Greece, 10434; Maria Halkidi, University of Piraeus, 80, Karaoli Dimitriou Str., Piraeus, Greece, 18534; Stephanie Van Hove, Peter Conradie, Ghent University, De Krook, Miriam Makebaplein 1, Ghent, Belgium, 9000.

on socio-psychological *constructs*[1] that are measured in the survey. Identification of respondents with a profile requires that their scores in constructs that are relevant to the profile fall within certain ranges called *compatibility intervals*. These compatibility intervals are distinct for each (construct, profile) pair and their endpoints are parameterized by threshold values. Finally, we allocate the survey respondents into one or more of these energy consumer profiles by formulating and solving a non-linear optimization problem over these threshold parameters.

We end up with six distinct energy consumer profiles: Environmentally conscious and well-informed energy consumers, Concerned but comfort-oriented energy consumers, Concerned but lacking awareness energy consumers, Materialistic energy consumers escaping personal responsibility, Prone to social influence energy consumers, and Indifferent energy consumers. The first group represents ideal energy-savers, whereas energy consumers in each of the other five groups share one or two distinct features that serve as barriers towards their energy-saving intentions and/or prescribe specific type of interventions for strengthening these intentions.

The rest of the paper is organized as follows. Section 2 presents the survey, its underlying human behavior models and the set of constructs it measures. Our segmentation approach and the assignment of the survey respondents to the resulting six energy consumer profiles is detailed in Section 3. We elaborate on those profiles in Section 4, analyzing them also in terms of socio-demographic characteristics and recommending types of nudges for each one of them. We contrast our study against related work in literature in section 5 and conclude in section 6 outlining future work on the validation of our approach against real data about energy consumption.

2 SURVEY METHODOLOGY AND DATA COLLECTION

A large-scale online survey has been carried out from February till July 2021. The main part of the survey participants were recruited through a network of European consumer organisations. An additional sample of 1000+ Flemish respondents were mobilized by a private data collection organization. The survey was available online in 15 languages[2].

2.1 Underlying model and measured constructs

The primary goal of the survey was to gain insights to the main determinants of energy consumers' behavior. To this end, it measured a number of sociopsychological constructs (variables) originating from three different theoretical frameworks of human behavior. The main one is the Theory of Planned Behaviour (TPB), as introduced in [2]. The general TPB model argues that there exist three factors that determine whether an individual adopts a specific behavior or not: her *attitude* towards this behaviour, *subjective norms* concerning that behaviour and her perception of how difficult it is to enact this behavior (*perceived behavioral control*). In our case, we let the attitude construct depend on four antecedent variables: *financial* and *environmental concerns*, the awareness about energy-saving practices (*energy awareness*) and possible concerns about the discomfort that the adoption of energy-saving behavior brings about (*loss of comfort*).

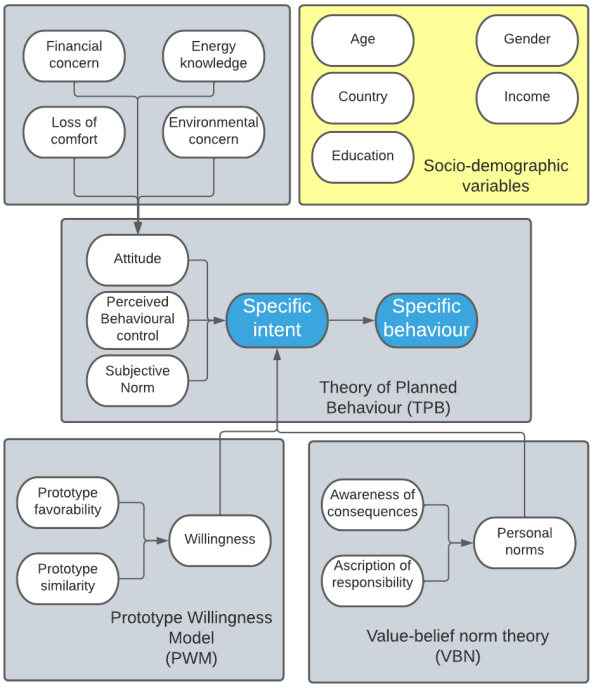


Fig. 1. Complete research model with constructs from three theoretical frameworks, the TPB model, the VBN-theory and the Prototype Willingness model and the sociodemographic variables under consideration.

This extended TPB model is enriched with constructs from the Value Belief Norm (VBN) theory [24][26]. Contrary to the individualistic emphasis of TPB, VBN allows for collective/societal returns of pro-environmental behavior. The primary construct is *personal norms*, defined as feelings of moral obligation to engage in a behavior. According to [22], the activation of personal norms in an individual depends on two antecedents: the *awareness of consequences* of her behavior and a feeling of responsibility for environmental problems (*ascription of responsibility*).

Finally, a third model, the Prototype Willingness Model (PWM) [8] prescribes that the individual's *willingness* to adopt a certain behavior depends on images (or prototypes) that she associates with that behavior. More specifically, this willingness depends on two antecedent variables: *prototype favorability*, denoting how favorably an individual perceives someone who engages in the behavior at hand; and *prototype similarity*, denoting how similar she anticipates herself to be to someone behaving that way. The resulting composite research model underlying the survey is shown in Fig. 1.

2.2 Survey structure and measured constructs

Each construct in the four dark-grey blocks of Fig. 1 is measured by one or more *items*, typically three to five, which are statements prompting the survey participants to respond to what extent they

[1] A construct is a hypothesized cause for certain behavior and in survey research it is what one wishes to measure using survey questions.
[2] English, Dutch, French, Italian, Portuguese, Croatian, Greek, German, Lithuanian, Latvian, Romanian, Slovenian, Slovak, Spanish, and Bulgarian.

identify with them. The responses are measured on a 5-point semantic differential scale ranging from 1 = strongly disagree to 5 = strongly agree, except for the Attitude construct, which is measured on a 7-point scale. The 15 constructs of interest are listed and briefly explained in Table 1. Details about the survey items and their internal consistency analysis are given in appendix A.

Additional survey questions gather sociodemographic information about the survey participants (gender, age, education, income, household type), their residence characteristics together with possible energy production facilities, their practices regarding the use of electric appliances and heating and their interest in platforms that monitor and control energy consumption in real time[3].

Table 1. Sociopsychological constructs measured in the survey

Abbreviation	Definition
ASCR_RESP	*Ascription of responsibility*: Acceptance of personal responsibility for energy saving
ATT	*Attitude*: Personal stance against energy-saving behavior
CONSEQ_AWARE	*Awareness of consequences*: Self-awareness about the consequences of energy waste
ENERGY_AWARE	*Energy awareness*: Self-awareness about ways to save energy (*e.g.*, operational modes of electric appliances)
ENV_CONCERN	*Environmental concern*: Concern about environmental matters and climate change
FIN_CONCERN	*Financial concern*: Concern about the financial implications of energy saving
INT_GEN	*General intent*: Overall determination to save energy
INT_SPEC	*Specific intent*: Determination to save on heating-related energy consumption in winter
LOSS_COMFORT	*Loss of comfort*: Concern about the discomfort resulting from saving energy
PERS_NORM	*Moral norm*: Feeling of moral obligation to reduce energy
PBC	*Perceived behavioral control*: Anticipated control over energy-saving behavior
PROT_FAV	*Prototype favorability*: Favorability of energy-saver persona
PROT_SIM	*Prototype similarity*: Self-identification with the energy-saver persona
SN	*Subjective norm*: Beliefs about whether peers and people of importance to the person think (s)he should engage in energy saving
WILL	*Willingness*: Behavioral willingness to become more energy-efficient

2.3 User participation and sample size for data analysis

Overall, 7,089 people opened the web page of the survey. Out of those, 954 (13.46%) dropped out after reading the introduction and 536 (8.74%) after reading the privacy statement, while 689 answered incorrectly one or both quality control questions that were added to the survey to test whether the respondents thoroughly read the survey questions (see appendix B for details). After further cleaning for incomplete responses and responses from countries outside Europe, 3,129 filled-out questionnaires were retained for data analysis.

[3]The full survey along with all the items that respondents were prompted to respond to can be found at https://www.nudgeproject.eu/report-profiling-of-energy-consumers-psychological-and-contextual-factors-of-energy-behavior/

3 SEGMENTATION OF ENERGY CONSUMERS

The purpose of our study is to segment the 3,129 survey respondents (interchangeably: energy consumers, users) in an intervention-ready way that facilitates the choice of (nudging) interventions for each segment. Two methods have been explored to this end, clustering analysis and a novel intervention-aware profiling scheme.

3.1 Clustering analysis

The first method is based on clustering. With clustering or cluster analysis the survey responses are organized into groups called clusters in such a way that responses in each cluster are similar to each other and dissimilar to responses in other clusters with respect to a given set of features. The main advantage of clustering is that it provides a solid analytical framework to automatically generate clustering structures. On the other hand, these structures need to be further analyzed to extract hints for matching interventions to each cluster.

The original feature set for the clustering experiments consists of the 15 energy-related constructs of Table 1. This choice is in line with reported practices in literature of segmentation studies [20][27], where the sample set is segmented on the basis of socio-psychological constructs and socio-demographic variables are only used in a second step, to describe the derived clusters. Although the resulting clustering structures are balanced and reasonably fit, the relative ranking of the 15 features across all clusters is identical. This is a pattern that persists across experiments with different clustering algorithms (k-means/medians, hierarchical), different parameterizations of those algorithms (number of clusters, similarity measure) and under different manipulations of the feature set (feature selection and transformation techniques), as shown in appendix C. Overall, clustering generates energy consumer segments that are not at all informative as to which type of intervention is more appropriate for them.

3.2 Intervention-aware profiling of of energy consumers

With clustering, we start from the energy consumers, we group them into clusters and then ask which interventions could be applicable to those clusters. The alternative is to start from the set of nudging interventions that we could realize and deliver, and then formulate energy consumer profiles for which those interventions could be effective. The question is then how many energy consumers actually match those profiles. The process comprises four steps:

3.2.1 Step 1: Enumeration of nudging interventions. The nudging interventions in our case are strongly technology-mediated leveraging smartphone apps and web dashboards. Hence, a natural starting point for structuring our arsenal of interventions is the typology proposed in [6]. On the other hand, this typology has its origin

within the Human Computer Interaction (HCI) domain. Reflecting on its relevance for motivating energy conservation within households, we ended up other times expanding it and other times filtering out elements, ending up with the following four nudge categories: *facilitating*, *reinforcement*, *social influence* and *confrontation* nudges.

Facilitating nudges: Nudges of this kind aim to minimize individuals' physical or mental effort to adopt energy efficient behavior. They try to render the desired behavior the preferable option and trigger it as automatic response [12], *e.g.*, by promoting it to a *default* [14].

Reinforcement nudges: Reinforcement nudges attempt to place desired behaviours at the fore of individual's thinking [6]. Nudges of this type primarily address the availability heuristic, a systematic cognitive bias [32], through feedback. Feedback may concern the household's real-time energy consumption in kWh, the resulting financial cost, and/or impact on environment (carbon emission). It is often provided through energy-management systems, in-home displays, mobile apps and web portals [1][21][11]. On the other hand, timely reminders upon acting (just-in-time prompts) and ways to either make conservation fun (hedonic goal activation) or provoke emotional reaction to energy consumption (empathy instigation) have been largely overlooked in the context of energy saving, although they have been leveraged in other aspects of environmental resource preservation [25][31].

Social influence nudges: These nudges are among the most applied nudges in energy conservation studies. They target people's desire to comply with what is socially anticipated as the norm. They usually consist in providing feedback on the energy consumption *of others* (friends, relatives, neighbors), hence enabling *social comparison* [5][16]. Alternatively, they motivate people to set energy conservation goals (*goal setting*) and publicly commit to energy-saving behavior, *e.g.*, [23].

Confrontation nudges: Confrontation nudges, as the name suggests, try to prevent an intended user's action by prompting her to iterate on the negative impact of this action. These nudges are inspired by findings suggesting that emphasizing the potential loss out of adopting environmentally non-friendly behavior (loss framing) is often more effective than focusing on what is gained by an environmentally friendly behavior (gain framing) [19].

Examples of operationalizing these nudges through a smartphone app are given in appendix D.

3.2.2 Step 2: Specification of energy consumer profiles. Each energy consumer profile can be originally specified by (a) a set of relevant constructs (*profile constructs*), which form a subset of the full construct set in Table 1; and, (b) a qualitative (verbal) description of what we consider profile-compatible expression for each profile construct. For a pair of (profile, profile construct)=(p,x), we have identified the following four options as most useful for describing how strongly x is expressed in p: strongly, non-strongly (weakly or moderately), weakly and non-weakly (moderately or strongly).

Then, we quantify these qualitative measures by leveraging the normalized scores (at the [0,1] scale) of the survey respondents in the survey items that measure each construct in the 1-to-5 Likert scale. With the help of four parameters θ_1, θ_2, θ_3, θ_4, taking values in [0,1] and hereafter called *threshold types*, we interpret the four qualitative measures of a construct's expression to four types of *compatibility intervals*, *i.e.*, score ranges:

i) *Compatibility intervals of type* [θ_1,1], with *type 1 thresholds* $\theta_1 \in$ [0.7,0.85], marking high scores in the construct and capturing strong expression of a construct.

ii) *Compatibility intervals of type* [0,θ_2], with *type 2 thresholds* $\theta_2 \in$ [0.6,0.75], marking low or average scores in the construct and capturing non-strong (moderate or weak) expression of a construct.

iii) *Compatibility intervals of type* [θ_3,1], with *type 3 thresholds* $\theta_3 \in$ [0.45,0.65], marking average or high scores in the construct and capturing non-weak (moderate or strong) expression of a construct.

iv) *Compatibility intervals of type* [0,θ_4], with *type 4 thresholds* $\theta_4$taking values in $[0.3, 0.5]$, marking low scores in the construct and capturing weak expression of a construct.

Eventually, energy consumer profiles are specified by their profile constructs and corresponding compatibility intervals so that an energy consumer satisfies a profile p as long as his/her scores in *all* constructs of p lie within the corresponding compatibility intervals. We could then, in principle, search for arbitrary energy consumer profiles within the survey data. In our case, the choice is dictated by the nudging interventions that we can deliver (see step 1) and has resulted in the specification of six energy consumer profiles of energy consumers. We describe those profiles qualitatively below and summarize their quantitative specification, in terms of constructs and compatibility intervals, in Table 2. Note that the actual values of the thresholds involved in this specification will emerge as the solutions of an optimization problem, as detailed in section 3.2.3.

1) Environmentally conscious and well-informed energy consumers. These are the idealistic energy savers in [27]. Energy saving sets a favorable paradigm for this profile, whereby solid knowledge about the climate change and its consequences is combined with high interest in taking actions to reduce energy consumption. Regular reminders of the environmental issues and the importance of energy saving would suffice to keep the interest of this energy consumer profile alive and turn their intentions into action. These reminders could be educational material, brief update letters about the energy situation or general-purpose marketing campaigns on social media.

In formal terms, this profile involves compatibility intervals of type 1 for six constructs: $CONSEQ\_AWARE$, $ENV\_CONCERN$, $ASCR\_RESP$, $PROT\_FAV$, INT_SPEC, INT_GEN.

2) Concerned but comfort-oriented energy consumers. Consumers in this profile still intend to save energy, but they are not willing to sacrifice comfort-wise, *e.g.*, to wear more clothes and live in a cooler house as a result of setting the thermostat at a lower temperature during winter. On the other hand, they are concerned about the monetary cost of energy consumption, letting space for interventions that point to the financial implications of energy saving.

In terms of profile constructs and compatibility intervals, this profile is specified by compatibility intervals of type 3 in $LOSS\_COMF$, type 2 in INT_SPEC and type 2 in $FIN\_CONCERN$ and INT_GEN.

3) Concerned but lacking awareness energy consumers. Lack of energy awareness serves as a barrier for energy-saving behavior for consumers in this profile. Whereas they are concerned about the consequences of high energy consumption and they understand the risks for the environment, they are not familiar with actual energy-saving practices.

Table 2. Energy consumer profiles, profile constructs and compatibility intervals

Energy consumer profile	**Profile constructs and corresponding** compatibility intervals
Environmentally conscious and well-informed	$\theta_1 \leq CONSEQ\_AWARE \leq 1$, $\theta_1 \leq ENV\_CONCERN \leq 1$, $\theta_1 \leq ASCR\_RESP \leq 1$ $\theta_1 \leq PROT\_FAV \leq 1$, $\theta_1 \leq INT\_SPEC \leq 1$, $\theta_1 \leq INT\_GEN \leq 1$
Concerned but comfort-oriented	$\theta_3 \leq LOSS\_COMF \leq 1$, $\theta_1 \leq FIN\_CONCERN \leq 1$ $0 \leq INT\_SPEC \leq \theta_2$, $\theta_1 \leq INT\_GEN \leq 1$
Concerned but lacking awareness	$\theta_3 \leq CONSEQ\_AWARE \leq 1$, $0 \leq ENERGY\_AWARE \leq \theta_4$ $\theta_3 \leq ENV\_CONCERN \leq 1$ $0 \leq INT\_GEN \leq \theta_2$
Materialistic escaping personal responsibility	$0 \leq ASCR\_RESP \leq \theta_4$, $\theta_1 \leq FIN\_CONCERN \leq 1$, $0 \leq INT\_GEN \leq \theta_2$
Prone to social influence	$\theta_3 \leq SN \leq 1$, $0 \leq INT\_SPEC \leq \theta_2$, $0 \leq INT\_GEN \leq \theta_2$
Indifferent	$0 \leq PBC \leq \theta_4$, $0 \leq PROT\_SIM \leq \theta_4$, $0 \leq INT\_SPEC \leq \theta_2$, $0 \leq INT\_GEN \leq \theta_2$

Two types of interventions are appropriate for this profile. The first one consists in the provision of tips, either online or offline, so that people gradually learn and adopt a more energy-efficient behavior. The second one, suitable for users who are less willing to learn through energy-saving tips, includes configuring devices/appliances to operate at energy-friendly defaults.

This profile specification involves compatibility intervals of type 3 for the *CONSEQ_AWARE* and *ENV_CONCERN* constructs, type 4 for *ENERGY_AWARE*, and type 1 for *INT_GEN*.

4) Materialistic energy consumers escaping their personal responsibility. This energy consumer profile is marked by lower-than-average intentions to save energy, combined with a weak sense of personal responsibility and high concern about the financial ramifications of their behavior. The latter concerns are the principal focus point of the interventions, as an antidote to their lack of self-responsibility. Establishing a sense of responsibility would probably require large-scale interventions at the education domain.

This profile is described by compatibility intervals of type 4 in the *ASCR_RESP* construct, type 1 in *FIN_CONCERN* and type 2 in the *INT_GEN* construct.

5) Prone to social influence energy consumers. Despite the relatively low intention to adopt heating-related energy-saving behavior, energy consumers with that profile appear more vulnerable to normative prescriptions originating from their social environment. Clearly, these energy consumers should be treated with interventions that try to leverage the social pressure and social comparison effects.

This energy consumer profile involves compatibility intervals of type 2 in the *INT_SPEC* and *INT_GEN* constructs and a compatibility interval of type 3 in the *SN* construct.

6) Indifferent energy consumers. The low perception of behavioral control and self-efficacy is the Achilles' heel of energy consumers in this profile. These users do not trust themselves with respect to their capacity to engage in energy-saving activities. Interventions for energy consumers with this profile could include practical tips or use of default settings for the operation of energy devices.

The formal specification of this energy consumer profile combines compatibility intervals of type 4 for the *PBC* and *PROT_SIM* constructs and type 2 for the *INT_SPEC* and *INT_GEN* constructs.

As shown in Table 2, the six energy profiles are parameterized by the four thresholds, $thr_1 - thr_4$, which define the four types of compatibility intervals for the construct scores. In the next step, we describe how we determine those values and subsequently assign energy consumers to one or more of the six profiles.

3.2.3 Step 3: Assignment of survey respondents to energy consumer profiles. Let $\mathcal{P}$ be the set of the six energy consumer profiles. For each profile $p \in \mathcal{P}$, C_p is the set of profile constructs and $I(c, p)$ the compatibility interval of construct c for profile p. Such compatibility intervals can be defined for every (construct, profile) pair. When a construct is a profile construct, the compatibility interval is one of the four types described in 3.2.2; otherwise, it is the [0,1] interval.

Let also $\mathcal{U}$ be the set of survey respondents, with $U = |\mathcal{U}|$, and $\mathcal{C}$ the set of constructs in Table 1. If $s_{uc}, u \in \mathcal{U}, c \in \mathcal{C}$ denotes the normalized score of user u in construct c, we can define two sets of binary variables, $\{x_{up}\}$ and $\{z_u\}, u \in \mathcal{U}, p \in \mathcal{P}$ as follows:

$$x_{up} = \begin{cases} 1 & \text{if } s_{uc} \in I(c, p)\ \forall c \in C_p \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

determines whether survey respondent u satisfies the specification of profile p, and

$$z_u = \begin{cases} 1 & \text{if } \exists\, p \in \mathcal{P} \text{ s.t. } x_{up} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

marks whether the survey respondent u can be assigned to *at least* one of the six energy consumer profiles.

If vector $\boldsymbol{\theta} = [\theta_1, \theta_2, \theta_3, \theta_4] \in [0, 1]^4$ is the vector representation of the four decision variables, then the problem we face,hereafter called (OPT), is how to

$$\max_{\boldsymbol{\theta}} \quad \sum_{u \in \mathcal{U}} z_u \tag{3}$$

$$s.t. \quad \theta_i \geq \theta_j,\ \forall i, j = 1, 2, 3, 4, with\ i < j \tag{4}$$

$$\sum_{u \in \mathcal{U}} x_{up} \geq \alpha U,\ p \in \mathcal{P} \tag{5}$$

$$\underline{\theta_i} \leq \theta_i \leq \overline{\theta_i},\ i \in \{1, 2, 3, 4\} \tag{6}$$

In (OPT), the goal is to maximize the number of survey respondents assigned to one or more of the six energy consumer profiles. These users can then be subject to the respective interventions that motivated those profiles. Constraints (4) and (6) are inherited from the definition of the four thresholds and the four types of compatibility intervals in section 3.2.2, $\underline{\theta_i}$ and $\overline{\theta_i}$ denoting the range of possible values for θ_i, as prescribed there. Finally, in (5) parameter $\alpha \in [0, 1]$ sets a lower bound on the number of survey respondents

Table 3. Compatibility of energy consumers with one or more profiles

1 profile	2 profiles	3 profiles	4 profiles	5 profiles
1,180	643	262	43	4

that should be assigned to a profile to give it non-negligible mass. The default value for α is 0.03, corresponding to 95 users.

(OPT) is a non-linear optimization problem. Although its characterization and the identification/design of efficient algorithms for it have independent theoretical interest, the problem size (3,129 users, 6 profiles, 15 constructs) is such that it does not prohibit the exhaustive enumeration. We have let the θ_1-θ_4 parameters vary in steps of 0.01 across their ranges [0.7,0.85], [0.65,0.75], [0.45,0.65] and [0.3, 0.5], respectively, and computed the (OPT) objective value for each feasible combination of their values. The optimal solution was found to be ($\theta_1 = \theta_2 = 0.75, \theta_3 = \theta_4 = 0.5$).

With these values of the four parameters 2, 132 out of 3, 129 consumers are assigned to at least one profile. The number of compatible profiles for each of those 2, 132 users is given in Table 3. Their majority (1,180 or 55.27%) can only be assigned to one profile, whereas the rest (952 or 44.73%) are *multihomed*, *i.e.*, they match the specifications of more than one profile. The largest portion of those (643 or 30.2% of all survey respondents) are compatible with two profiles, while a small part could be identified as members of four (43 or 2%) and even five (4 or 0.2%) energy consumer profiles.

Eligible for the six profiles are 529, 477, 507, 425, 1,041 and 465 users, respectively. If it is required to assign each user to a single profile, *e.g.*, when only one intervention is feasible per user due to budget limitations, we can order the profiles in any arbitrary way and parse them sequentially, eliminating in each step users who have previously been assigned to another profile. For instance, ranking profiles in the order of their presentation in section 3.2.2, we come up with 529 (24.81%), 400 (18.76%), 440 (20.64%), 259 (12.15%), 392 (18.39%), and 112 (5.25%) users, respectively. Nevertheless, one should bear in mind that there are 6!=720 such orders each resulting in a different partition of survey respondents to the six profiles.

3.2.4 Step 4: Assignment of remaining survey respondents to energy consumer profiles. By the end of the third step, 2, 132 survey respondents are assigned to at least one profile. To do the same for the remaining 997 users, we apply a variant of the nearest centroid classifier [15]. First, for each profile we compute the average scores in the profile constructs (rather than all 15 constructs) over all users assigned to the profile. This essentially yields the scores of the profiles' centroids. Then, we assign each of the remaining 997 users to the profile with the nearest centroid. The distance of remaining users to each centroid is computed as the 'cityblock' (Manhattan) distance over the respective profile constructs and is normalized by the number of the profile constructs. At the end of this step, the six profiles include 917 (29.31%), 733 (23.43%), 497 (15.88%), 311 (9.94%), 499 (15.95%) and 172 (5.49%) users, respectively.

4 CHARACTERIZATION OF ENERGY CONSUMER PROFILES AND MATCHING WITH NUDGES

Figure 2 plots how the six energy consumer profiles score in each of the 15 constructs in Table 1. When we compare Fig. 2 to its counterpart at the end of step 3 (see Fig. 8 in appendix E), namely before

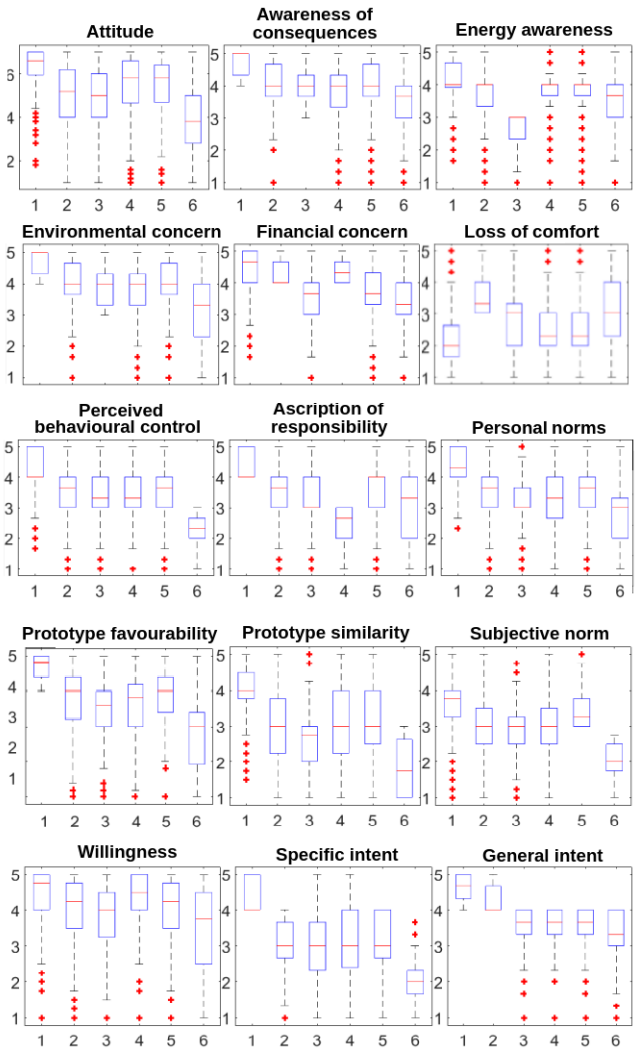


Fig. 2. Box plots of the per profile scores (y-axis) in the 15 constructs of Table 1. The scores are averages over the respondents assigned to each profile on the original 1-5 Likert scale (except for the ATT construct, which is measured on the 1-7 Likert scale): 3,129 survey respondents, after step 4.

we apply the nearest centroid classifier to the remaining 997 users, we note a few more outliers, as expected, but, otherwise, the per profile median scores and their relative rank remain practically intact. In what follows, we extract distinct energy consumer profiles (personas) out of these classes, also considering how they score in sociodemographic variables, and outline nudging interventions that are applicable to them. Further details on the sociodemographic characterization of the six energy consumer profiles are given in appendix F. Figure 9 plots the distributions of gender, age and education level across the six profiles and reports the outcomes of two-sample t-tests and Kolmogorov-Smirnov tests on the similarity of their means and distributions, respectively.

Environmentally conscious and well-informed energy consumers set a benchmark in terms of energy-saving behaviour. They

score high not only in the profile constructs, but also in other constructs identified as important motivators for energy-saving behaviour. Besides their concern about environmental matters and awareness about the consequences of irresponsible energy-related behaviour, they are well informed about ways to save energy. They bear a strong sense of personal responsibility for acting in an energy-friendly manner and possible sacrifices in terms of comfort do not stand as barriers to this end. These consumers are males and females, in percentages that closely resemble the overall gender distribution of the survey respondents (Male: 51.14%, Female: 48,86%, see Fig. 9(b) in section F.2 in the appendices). Energy consumers with this profile are a couple of years younger than the overall age average and enjoy a noticeably higher-than-average educational status, 3 out of 4 of them having acquired at least a Bachelor's level degree.

The average consumer in this segment hardly needs any nudging-type intervention treatment. The main requirement is to keep him/her sensitized about energy saving and the positive consequences of his/her behavior. This could be achieved through simple *reinforcement nudges* such as the (real-time) provision of informational tips about energy-saving practices.

Concerned but comfort-oriented energy consumers are a very distinct segment of energy consumers. Their overall intention to adopt an energy-saving behaviour is high (in fact: the 2nd highest after the 1st benchmark profile) and this is supported by high concern about the environment and good understanding of the risks involved in energy-wasting. Nevertheless, these intentions are clearly weaker when the question is about energy-saving with respect to heating and cooling. Namely, the possible comfort sacrifice due to a slightly lower temperature as thermostat's setting in winter (or a higher one during summer) appears to be much less tolerable for this than any other energy consumer profile.

There is no gender or age bias in this energy consumer segment when compared to the overall survey population, except for an over-representation of the 31-43 age group (every third consumer in this segment belongs to this age group, see appendix F). Education-wise, these consumers exhibit the second best scores on average, clearly above the global average.

Confrontation nudges could be applied to this segment. For instance, we could real-time prompt the consumer to consider the consequences of an action that implies higher energy consumption, *e.g.*, an increase in the target temperature of the thermostat or the air conditioner. These prompts should insist on the extra cost the action incurs, projecting it in terms of higher energy bills at monthly/annual level.

Concerned but lacking awareness energy consumers form one of the four energy-consumer segments, whose stated intentions to save energy can be strengthened, both specifically with respect to heating and, more generally, with respect to other energy-consuming activities, such as the use of electric appliances or lighting (see the box plots of Specific and General intent in Fig. 2). The lack of knowledge about practical ways to save energy serves as barrier for an unconditionally positive attitude towards energy-saving, which exists, even less strongly than in the first benchmark segment.

A consumer in this segment is more probably female rather than male and his/her education status is no better or worse than what is evidenced in the overall survey population. Almost every second consumer in this segment is 18-43 years old, implying that educating the younger generations about energy-saving should remain high in the list of possible interventions.

This group of consumers could be nudged through just-in-time prompts: energy-saving tips and recommendations exactly upon the time they mingle with devices' setting as to how to reduce energy consumption and protect the environment. Alternatively, *facilitating* nudges of the default type could save the user from the "burden" of learning what is appropriate and what is not. These consist in turning energy-friendly operational settings of devices (thermostat, air conditioning equipment) into operational defaults.

Materialistic energy consumers escaping personal responsibility form the second energy consumer segment that lags in overall energy-saving intentions. Neither concern about the environment, nor knowledge about ways to save energy are missing in their case. However, whereas they claim awareness of the consequences that increasing energy demand bears for the environment and the society, they do not accept their own share of responsibility to act on this. On the other hand, and this gives some hope for their treatment, they demonstrate high concern for the height of their energy bills and the monetary fingerprint of energy consumption.

Males are marginally over-represented in this segment, which also tends to be older than the average, with more than half its population exceeding 57 years. This age bias is also reflected in the lower educational status of these consumers; more than half of them have not obtained a degree from a higher education institution.

Prone to social influence energy consumers attribute high value to the fact that people they deem important in their lives approve and support energy-saving, which sets a strongly favourable behavioural prototype. Hence, this form of indirect social pressure serves as facilitator of energy- saving in their case.

Consumers in this segment share the gender bias (more males) with the previous segment but tend to be younger and better educated than them. Their education status is the most representative of the overall population of survey respondents.

Nudging that enables social comparison can be considered for this group of consumers. We could leverage different means (such as written text, diagrams printed on a paper to online social platforms, dynamic query-response systems) to facilitate the comparison with other users (friends, neighbors, consumers of similar demographic characteristics). Besides, goal-setting programs can be used as ways to elicit consumers' commitment to save upon what they consume.

Indifferent energy consumers demonstrate strikingly low intentions for energy saving. They doubt their own capacity to adopt energy-saving behaviour as well as any impact this can have on energy-saving overall. They are nowhere close to the energy-saver prototype (which they do not find favourable anyway) and they do not perceive social pressure to adopt energy-saving behaviour.

An indifferent energy consumer is male or female as often as any energy consumer but tends to be older than him/her. This explains, in part, the fact that only half the energy consumers in this segment have acquired a degree from a higher level institution.

Since this group of consumers is characterized by the lowest levels of environmental concern and energy awareness as well as the lowest pressure from norms of any kind, both facilitating and reinforcement interventions apply. We could consider using tips,

notifications, marketing campaigns, to sensitize this group of users and overcome their reservations about the efficacy of their behavior. Moreover, we could turn energy-friendly operational settings of devices (thermostat, air conditioning equipment) into defaults, assisting, thus, the user to adopt energy-saving practices.

5 RELATED WORK

The segmentation of people considering sociopsychological determinants of behavior is acknowledged as a prerequisite for the design of targeted interventions that can have an impact on human behavior [7]. Several studies tried to apply this principle in the domain of energy efficiency, relying almost exclusively on survey data.

The study in [27] considers several constructs rooted in the VBN theory and the TPB model together with its antecedents, as we do, whereas it does not include any constructs from the PWM model. The analyzed data are collected from 1, 292 Swiss households through emailed questionnaires and address the energy efficiency of participants' behavior in various domains such as food, transportation and households. The authors apply hierarchical clustering with Euclidean distance as similarity measure on 17 features, including 11 sociopsychological constructs and variables capturing actual behavior across the different domains. They come up with six clusters of energy consumers, a few of which (partly) resemble the profiles we came up with in section 3. Interestingly, the clusters tend to rank uniformly across 8 sociopsychological constructs, which is what we came across with clustering as well (see section 3.1). The study also highlights interventions for the six segments, these interventions being mainly marketing and policy strategies (monetary incentives, subsidies) rather than of the behavioral/nudging type we consider.

In [33] the focus is on the energy efficiency behavior of employees in a major UK national rail operator. This is an exploratory factor analysis study: the constructs that serve as behavior-determinants are not selected beforehand but emerge out of the responses to the survey items, applying PCA with varimax rotation and Kaiser normalization. The six composite constructs that become input to the clustering analysis are termed Technology Adoption Norms, Benefit Evaluation, Energy Intentions, Goal Flexibility, Energy Awareness and Energy Self-Efficacy; some of them (*e.g.*, Energy Awareness) do have a clear correspondence with our 15 constructs in Table 1 and others do not. The 628 survey respondents are eventually grouped into five clusters, which are analysed further to extract recommendations for how energy efficiency directives should be channeled through the firm's organizational structure.

Both the scope and the segmentation approach are slightly different in [13]. Therein, the feature set for the clustering of 1, 119 Chinese households consists of the Big Five personality traits: extraversion, agreeableness, conscientiousness, neuroticism and openness. The number of clusters is set to four after applying Hierarchical clustering and then a second clustering round (k-means) partitions the survey sample into (i) positives, (ii) temperates, (iii) conservatives, and (iv) introverts. Through statistical ANOVA tests and Structural Equation Modeling (SEM), the four clusters are found adequately different with regard to an extended variant of the TPB model (see section 2.1), enriched with the Personal Norms construct. This study does not address possible interventions at all.

We have also worked with survey data and emphasized sociopsychological constructs in the energy consumer segmentation process; in fact, our construct set is the richest, encompassing constructs from the three behavioral models in Fig. 1. And, similar to [27] and [33], the energy consumer segmentation aims at tailoring interventions for the derived segments, which, in our case, are of the nudging type and delivered through mobile apps and web dashboards. Methodologically, however, we depart from all three studies in not relying on clustering analysis for the segmentation task. Our ad hoc segmentation method favors experience and intuition about energy consumer characteristics over the automation of unsupervised learning and helps us derive energy consumer segments that are more instructive as to which nudges are appropriate for them.

6 CONCLUDING REMARKS

The goal of our work is to come up with a segmentation of energy consumers that will be most informative about applicable nudging interventions. Although clustering is the *de facto* approach to such studies, its application to our survey data yields groups of energy consumers without distinct differentiation characteristics that could guide the choice of nudges. Hence, we follow an alternative approach: starting from the set of candidate nudges, we specify upfront energy consumer profiles that clearly match specific kinds of nudges. These profiles are interpreted in concrete parameterized score ranges in the constructs measured in the survey so that the maximal assignment of survey respondents to them is cast as a nonlinear optimization problem. We end up with six energy consumer profiles, we characterize them with respect to their energy-saving intentions and recommend possible types of nudging interventions for strengthening these intentions.

We should note that this study measures behavior drawing on survey data, *i.e.*, self-reported data. As such, these data capture how the respondents perceive their own behavior and not necessarily how they actually behave. On top of that, it is known that survey participants tend to avoid/suppress behaviors or attitudes that are considered socially undesirable in their responses in favor of more socially desirable ones, in an expression of the *social desirability response bias* [18]. The ultimate validation of our profiling study has to be carried out against real data recording actual behaviors of energy consumers. We are currently working towards this direction in the context of the EU H2020 NUDGE project (https://www.nudgeproject.eu/the-project/). Field trials are ongoing in five European countries to test a broad set of interventions that nudge users towards energy saving in various activities, such as heating, lighting and electric vehicle charging.

ACKNOWLEDGMENT

This paper is based on research that was conducted within the project NUDGE ("Nudging consumer towards energy efficiency through behavioral science", Grant Agreement No. 957012), funded by the European Commission in the context of the Horizon 2020 Framework Programme for Research and Innovation (EU H2020).

REFERENCES

[1] Wokje Abrahamse, Linda Steg, Charles Vlek, and Talib Rothengatter. 2007. The effect of tailored information, goal setting, and tailored feedback on household

energy use, energy-related behaviors, and behavioral antecedents. *Journal of Environmental Psychology* 27, 4 (2007), 265–276.
[2] Icek Ajzen. 1991. The theory of planned behavior. *Organizational Behavior and Human Decision Processes* 50, 2 (1991), 179–211.
[3] Hunt Allcott and Todd Rogers. 2014. The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review* 104, 10 (October 2014), 3003–37.
[4] Mark A. Andor, Andreas Gerster, Jörg Peters, and Christoph M. Schmidt. 2020. Social Norms and Energy Conservation Beyond the US. *Journal of Environmental Economics and Management* 103 (2020), 102351.
[5] Alec Brandon, John A. List, Robert D. Metcalfe, Michael K. Price, and Florian Rundhammer. 2019. Testing for crowd out in social nudges: Evidence from a natural field experiment in the market for electricity. *Proceedings of the National Academy of Sciences* 116, 12 (2019), 5293–5298.
[6] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. 1–15.
[7] Elisabeth Engl, Peter Smittenaar, and Sema K. Sgaier. 2019. Identifying population segments for effective intervention design and targeting using unsupervised machine learning: an end-to-end guide. *Gates Open Res* 3, 21 (Oct. 2019).
[8] Meg Gerrard, Frederick Gibbons, Amy Houlihan, Michelle Stock, and Elizabeth Pomery. 2008. A dual-process approach to health risk decision making: The Prototype Willingness Model. *Developmental Review* 28 (03 2008), 29–61.
[9] Brian Hopkins and J. G. Skellam. 1954. A New Method for determining the Type of Distribution of Plant Individuals. *Annals of Botany* 18, 70 (1954), 213–227.
[10] Daniel Kahneman. 2011. *Thinking, fast and slow.* Farrar, Straus and Giroux, New York.
[11] Adnane Kendel, Nathalie Lazaric, and Kevin Mar©chal. 2017. What do people learn by looking at direct feedback on their energy consumption? Results of a field study in Southern France. *Energy Policy* 108 (2017), 593–605.
[12] Siegwart Lindenberg and Esther K. Papies. 2019. Two Kinds of Nudging and the Power of Cues: Shifting Salience of Alternatives and Shifting Salience of Goals. *International Review of Environmental and Resource Economics* 13, 3-4 (2019), 229–263.
[13] Xuan Liu, Qian-Cheng Wang, Izzy Yi Jian, Hung-Lin Chi, Dujuan Yang, and Edwin Hon-Wan Chan. 2021. Are you an energy saver at home? The personality insights of household energy conservation behaviors based on theory of planned behavior. *Resources, Conservation and Recycling* 174, 105823 (November 2021).
[14] Claire-Michelle Loock, Thorsten Staake, and Frédéric Thiesse. 2013. Motivating Energy-Efficient Behavior With Green Is: An Investigation of Goal Setting and the Role of Defaults. *MIS Quarterly* 37, 4 (2013), 1313–1332.
[15] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval.* Cambridge University Press, Cambridge, UK.
[16] Erica Myers and Mateus Souza. 2020. Social comparison nudges without monetary incentives: Evidence from home energy reports. *Journal of Environmental Economics and Management* 101 (2020), 102315.
[17] EU NUDGE project consortium. 2021. Profiling of energy consumers: psychological and contextual factors of energy behavior. Deliverable D1.1, [Online]: https://www.nudgeproject.eu/report-profiling-of-energy-consumers-psychological-and-contextual-factors-of-energy-behavior/.
[18] Donna M. Randall and Maria F. Fernandes. 1991. The Social Desirability Response Bias in Ethics Research. *Journal of Business Ethics* 10, 11 (1991), 805–817.
[19] Aja Ropret Homar and Ljubica Knezevic Cvelbar. 2021. The effects of framing on environmental decisions: A systematic literature review. *Ecological Economics* 183 (2021), 106950.
[20] John R Rossiter and Larry Percy. 1987. *Advertising and promotion management.* McGraw-Hill Book Company.
[21] Paul Wesley Schultz, Mica Estrada, Joseph Schmitt, Rebecca Sokoloski, and Nilmini Silva-Send. 2015. Using in-home displays to provide smart meter feedback about household electricity consumption: A randomized control trial comparing kilowatts, cost, and social norms. *Energy* 90 (2015), 351–358.
[22] Shalom H. Schwartz. 1977. Normative Influences on Altruism. Advances in Experimental Social Psychology, Vol. 10. Academic Press, 221–279.
[23] Henk Staats, Paul Harland, and Henk A. M. Wilke. 2004. Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household. *Environment and Behavior* 36, 3 (2004), 341–367.
[24] Linda Steg and Judith de Groot. 2010. Explaining prosocial intentions: Testing causal relationships in the norm activation model. *British Journal of Social Psychology* 49, 4 (2010), 725–743.
[25] Linda Steg, Goda Perlaviciute, Ellen van der Werff, and Judith Lurvink. 2014. The Significance of Hedonic Values for Environmentally Relevant Attitudes, Preferences, and Actions. *Environment and Behavior* 46, 2 (2014), 163–192.
[26] Paul C. Stern, Thomas Dietz, Troy D. Abel, Gregory A. Guagnano, and Linda Kalof. 1999. A Value-Belief-Norm Theory of Support for Social Movements: The Case of Environmentalism. *Human Ecology Review* 6 (1999), 81–97.
[27] Bernadette Sütterlin, Thomas A. Brunner, and Michael Siegrist. 2011. Who puts the most energy into energy conservation? A segmentation of energy consumers based on energy-related behavioral characteristics. *Energy Policy* 39, 12 (December 2011), 8137–8152.
[28] Richard Thaler and C. Sunstein. 2009. *NUDGE: Improving Decisions About Health, Wealth, and Happiness.* Penguin.
[29] Daniel Robert Thomas, Shalu Agrawal, S.P. Harish, Aseem Mahajan, and Johannes Urpelainen. 2020. Understanding segmentation in rural electricity markets: Evidence from India. *Energy Economics* 87 (March 2020). https://doi.org/10.1016/j.eneco.2020.104697
[30] John Thgersen. 2021. Consumer behavior and climate change: consumers need considerable assistance. *Current Opinion in Behavioral Sciences* 42 (2021), 9–14. Human Response to Climate Change: From Neurons to Collective Action.
[31] Verena Tiefenbeck, Lorenz Goette, Kathrin Degen, Vojkan Tasic, Elgar Fleisch, Rafael Lalive, and Thorsten Staake. 2018. Overcoming Salience Bias: How Real-Time Feedback Fosters Resource Conservation. *Manage. Sci.* 64, 3 (March 2018), 1458–1476.
[32] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131.
[33] Rupert Zierler, Walter Wehrmeyer, and Richard Murphy. 2017. The energy efficiency behaviour of individuals in large organisations: A case study of a major UK infrastructure operator. *Energy Policy* 104 (2017), 38–49.

APPENDICES

A MEASURING THE SOCIOPSYCHOLOGICAL CONSTRUCTS IN THE SURVEY

The fifteen constructs in Table 1 reflect attitudinal, motivational and behavioural characteristics and originate from the three theoretical frameworks of human behavior that are briefly presented in section 2.1: the Theory of Planned Behaviour (TPB) [2], the Value-Belief-Norm Theory (VBN) [26] and the Prototype Willingness Model (PWM) [8].

A.1 Items and scales of measurement

Specific and General intent: In the survey, participants were asked to imagine a concrete energy-saving action, *i.e.*, "saving energy by lowering the temperature setting in winter". This energy-saving activity has been determined based on its prevalence across Europe and its substantive impact on energy conservation. Moreover, the more tangible the situation, the better respondents can assess their behavior in that particular situation. On the other hand, to be able to estimate how specific and general intent relate, we also introduced in the survey a construct that measures general intent to save energy at home. Both intent constructs consisted of three items with "I intend to save energy at home/by lowering the temperature setting in winter" as an exemplary item.

TPB constructs: All TPB constructs except for Attitude were measured on a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. More specifically:

- *Attitude* was measured with five 7-point semantic differential scales, which have been informed by general and topic-related research: useless – useful, foolish – wise, disadvantageous – advantageous, ineffective – effective, dull – interesting (Webb et al., 2013).
- *Subjective norm* was measured by four items, *e.g.*, 'Most people who are important in my life would approve that I save energy by lowering the temperature setting in winter'
- *Perceived behavioral control* was measured by three items. An indicative item is 'I have the capabilities to save energy by lowering the temperature setting in winter'.

VBN constructs: All VBN-constructs have been measured on a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree:

- *Moral norm* was covered by three items [1] with "I feel morally obliged to reduce my energy use, regardless of what other people do" as one of the items.
- *Ascription of responsibility* was addressed by three items [1]. One of the items was: 'I take joint responsibility for the depletion of energy resources'.
- *Awareness of consequences* was measured by three items: "Energy conservation contributes to a reduction of global warming" [1]; "The increasing energy demand is a serious problem for our society"; "The increasing shortage of energy sources is a serious problem for our society".

PWM constructs: The PWM constructs were measured as follows:

- For *Prototype favourability* the respondents were asked to rate the favorability of the energy-saver persona on a 5-point scale (1 = not at all to 5 = totally) using five adjectives: conscious, progressive, smart, green, responsible (Van Gool et al., 2015).
- *Prototype similarity* was assessed with four items (Elliott et al., 2017) on a five-point scale. An example item was "Do you resemble the typical person who saves energy by lowering the temperature setting in winter?" (1 = no to 5 = yes).
- *Willingness* was measured by asking responders to specify how frequently they perform four specific actions (Frater et al., 2017) on a 5-point Likert scale from 1 = Extremely unlikely to 5 = Extremely likely. The actions were: "You lower the temperature setting in all unused rooms when you are at home all day"; "You lower the temperature setting when you leave home"; "You keep the doors closed to prevent heat loss"; "You go to sleep and you lower the temperature setting".

Extended TPB model constructs: Finally, each of the four antecedent variables behind the attitude construct consisted of three items and was measured on a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree.

- Example item for *Financial concern*: "I pay attention to energy-saving tips to reduce my electricity bills" (Chen et al., 2017).
- Example item for *Loss of comfort*: "Energy conservation means I have to live less comfortably" [1][27]
- Example item for *Energy knowledge*: "I know energy-saving methods well" (Dianshu et al., 2010; Wang et al., 2014)
- Example item for *Environmental concern*: "I am very concerned about the environment"(Kilbourne Pickett, 2008)

The full list of items used in the NUDGE survey can be found in [17].

A.2 Reliability of constructs

We performed Cronbach's α analysis for all constructs in our model. The consensus in literature is that the Cronbach's α value for any particular construct should be higher than 0.7 to render its measurement reliable. Cronbach's α values for all 15 constructs in our model were satisfactory, the lowest value being .77 for Willingness, well above the customary 0.7 threshold. Table 4 reports the results of the internal consistency tests for each of the 15 constructs we measure in the survey.

Table 4. Cronbach's α values of all constructs in Table 1

Model construct	Cronbach value
Specific intent	0,90
Subjective Norm	0.83
Attitude	0,91
Perceived behavioral control	0,82
Prototype favorability	0,92
Prototype similarity	0,95
Willingness	0,77
Financial concern	0,80
Loss of comfort	0,90
Energy knowledge	0,94
Environmental concern	0,82
Awareness of consequences	0,78
Ascribing responsibility	0,93
Moral norm	0,80
General intent	0,84

Additional references for this section:

Chen, C. fei, Xu, X., Day, J. K. (2017). Thermal comfort or money saving? Exploring intentions to conserve energy among low-income households in the United States. Energy Research and Social Science, 26, 61–71

Dianshu, F., Sovacool, B. K., Vu, K. (2010). The barriers to energy efficiency in China: Assessing household electricity savings and consumer behavior in Liaoning Province. Energy Policy, 38(2), 1202–1209

Elliott, M. A., McCartan, R., Brewster, S. E., Coyle, D., Emerson, L., Gibson, K. (2017). An application of the prototype willingness model to drivers' speeding behaviour. European Journal of Social Psychology, 47(6), 735–747

Frater, J., Kuijer, R., Kingham, S. (2017). Why adolescents don't bicycle to school: Does the prototype/willingness model augment the theory of planned behaviour to explain intentions? Transportation Research Part F: Traffic Psychology and Behaviour, 46, 250–259

Kilbourne, W., Pickett, G. (2008). How materialism affects environmental beliefs, concern, and environmentally responsible behavior. Journal of Business Research, 61(9), 885–893

Van Gool, E., Van Ouytsel, J., Ponnet, K., Walrave, M. (2015). To share or not to share? Adolescents' self-disclosure about peer relationships on Facebook: An application of the Prototype Willingness Model. Computers in Human Behavior, 44, 230–239

Webb, D., Soutar, G. N., Mazzarol, T., Saldaris, P. (2013). Self-determination theory and consumer behavioural change: Evidence fromahousehold energy-saving behaviour study. Journal of Environmental Psychology, 35, 59–66

B SURVEY QUALITY CONTROL QUESTIONS

The inclusion of these quality control questions has been motivated by the work of Meade and Craig (2012). Especially in long surveys

composed of a lot of constructs (our survey has a median completion time of 21:36 minutes and contains 59 items), attentiveness to careless responses is recommended. Bogus items (an item with a clear true answer) are among the most sensitive methods to identify careless responses (Meade Craig, 2012). We decided to include two bogus items in a series of psychological statements being part of the third and last construct of the survey. Respondents were notified of the control items and were literally asked to click on 'disagree' (or 'strongly disagree'). The items were correctly answered by 86.8% and 77.6% of the respondents, respectively.

Meade, A. W., Craig, S. B. (2012). Identifying careless responses in survey data. Psychological Methods, 17(3), 437–455.

C CLUSTERING EXPERIMENTS AND RESULTS

C.1 Methodology:

The cluster analysis starts with the preprocessing of data, which involves the imputation of missing values in the dataset and the normalization of feature (*i.e.*, construct) scores. Then, a clustering algorithm is selected and parameterized with respect to the feature set and, when required, the number of clusters to be used in the cluster analysis.

Data preprocessing: The imputation of missing values is relevant to 31 records (0.99% of the sample), where at least one question is left unanswered in the 15 constructs of interest. The k-nearest neighbors imputation algorithm is used for this purpose. For each missing value in a record, the 10 nearest neighbors with a non-missing value in the unanswered feature are identified based on the Euclidean distance and the missing value is set to the weighted average of those 10 values. The weights assigned to each neighbor are inversely proportional to its distance from the record at hand.

Next, the feature scores of the dataset are normalized with the min-max scaling technique so that that they all end up in the range [0,1]. If x_{uc} is the score of record u on feature c, the normalized score is given by

$$f(x_{uc}) = \frac{x_{uc} - \min_u x_{uc}}{\max_u x_{uc}} \tag{7}$$

Algorithm selection and parameterization: We experimented with hierarchical clustering algorithms, both divisive and agglomerative, and the k-means/k-centers algorithm. These algorithms are further parameterized/configured by the following parameters:

(i) *Number of clusters.* By default, the clustering algorithms take the number of clusters as an input. To identify the best clustering structure, we evaluated the results based on standard cluster validity indices such as the Silhouette index.

(ii) *Distance measure.* Different distance measures were tested, including the Euclidean, Manhattan/taxicab, and cosine distance.

(iii) *Feature engineering.* We experimented with both feature selection and feature transformation. The feature selection process was carried out in two different ways. The first one relied on the use of the Hopkins statistic [9]. The second one involved the manual selection of features with intuitive matching with interventions.

With feature transformation, the original feature set was subject to dimensionality reduction using the Principal Component Analysis (PCA), which yielded a set of new features called Principal Components (PCs).

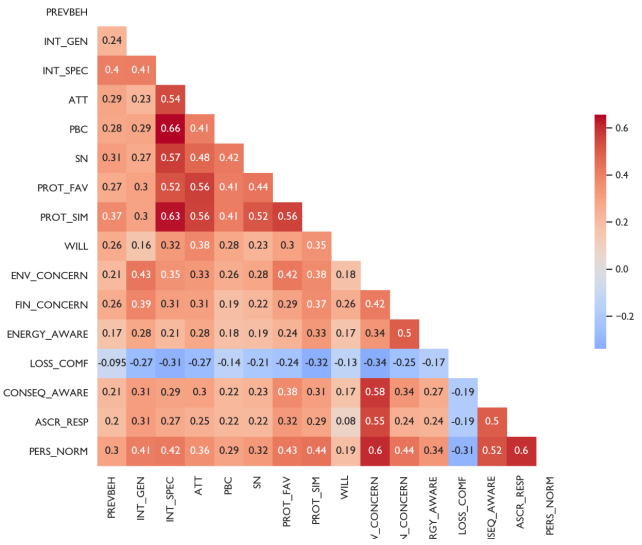


Fig. 3. Pairwise correlation matrix of the 15 constructs in Table 1

C.2 Clustering results

Our experiments spanned a large area of the parameterization space and, in almost all cases, they yielded clustering structures that score adequately in terms of clustering fitness and balance. Nevertheless, they exhibit a very particular persistent pattern in how clusters scores rank in the different features: the ranking of clusters across almost all features is identical. This pattern of strongly correlated score rankings is in agreement with the quite high positive pairwise feature correlations observed during feature analysis, as shown in Fig. 3, but it does not provide much information regarding which intervention would be most appropriate for each energy consumer cluster.

We demonstrate the outcome of two indicative experiments carried out with the k-means algorithm and Euclidean distance as the (dis)similarity measure. The first one employed manual feature selection with input feature set: {ASCR_RESP, CONSEQ_AWARE, ENV_CONCERN, PERS_NORM}. The input number of clusters was five since that was the number that maximized the Silhouette index in Fig. 4. The distribution of scores and the ranking across clusters pattern is presented in Fig. 5.

The second experiment employed feature transformation with PCA. More specifically, the original feature space of 15 features is transformed into a feature space of dimension 4. The number of clusters ranged from two to five and the maximum average Silhouette score was obtained for three clusters. The results are provided in Fig. 6 and confirm the same pattern of per cluster scores across all features. The pattern is repeated in experiments with the hierarchical agglomerative clustering algorithm, different distance metrics and choices of feature sets.

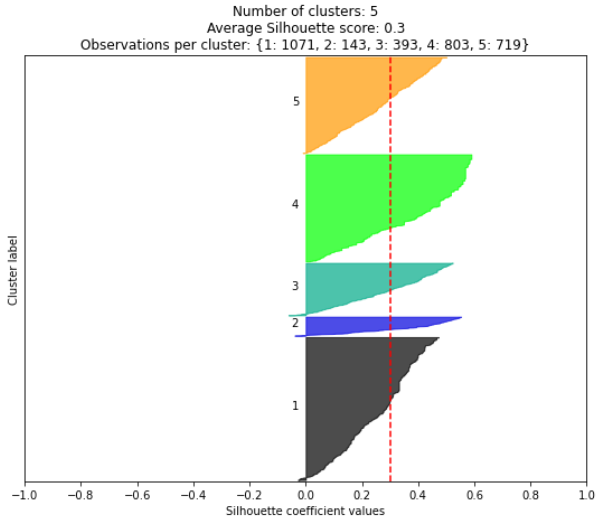


Fig. 4. Silhouette scores of 5 clusters in the 15 sociopsychological constructs of the feature set and 3 additional sociodemographic variables: k-means, Euclidean distance.

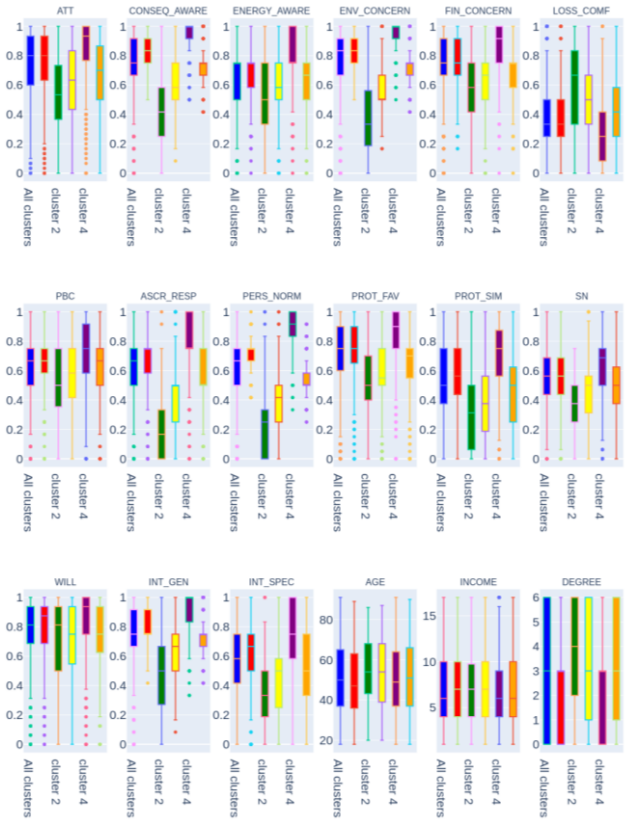


Fig. 5. Scores of 5 clusters in the 15 sociopsychological constructs of the feature set and 3 additional sociodemographic variables: k-means, Euclidean distance, manual feature selection.

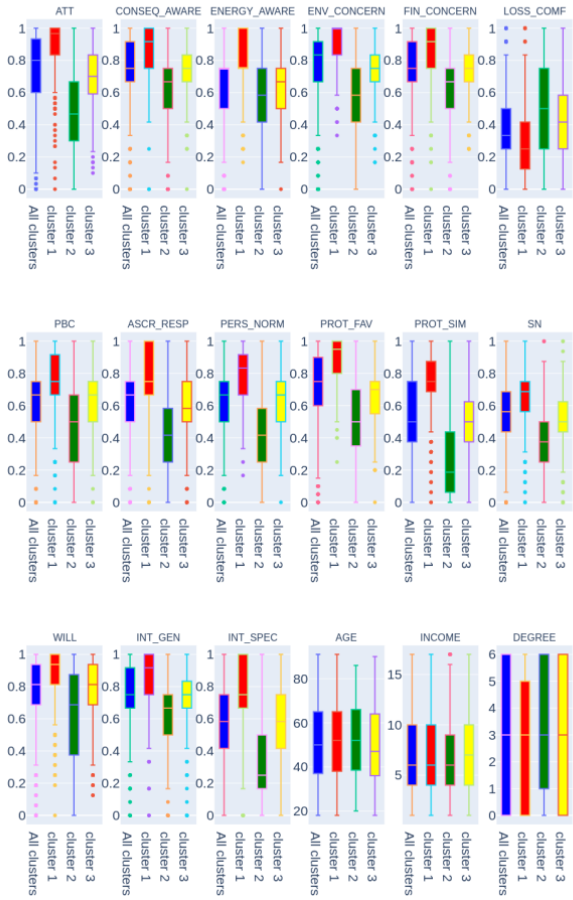


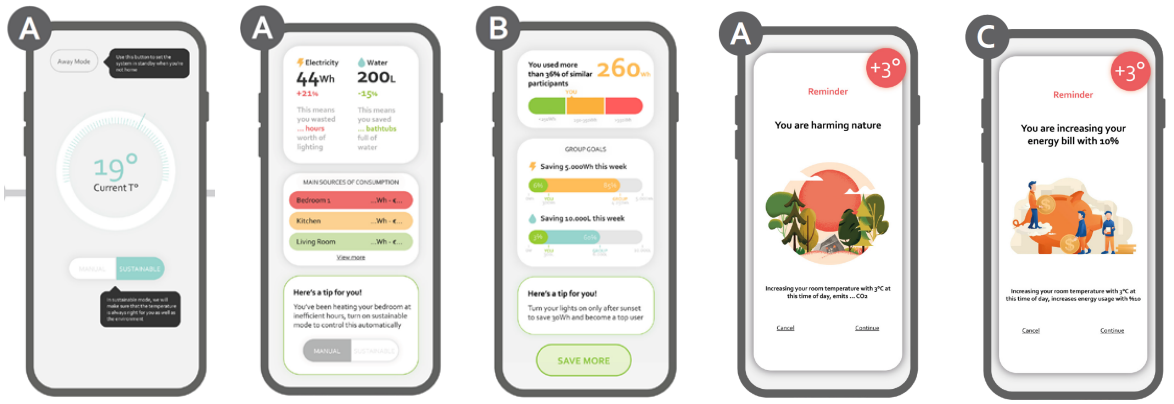
Fig. 6. Scores of 3 clusters in the 15 sociopsychological constructs and 3 sociodemographic variables: k-means, Euclidean distance, feature transformation with PCA.

D OPERATIONALIZING NUDGING INTERVENTIONS

We elaborate on how the nudges could be operationalized through smartphone apps. Figure 7 lists smartphone mockups that exemplify how different types of nudges could be delivered to different profiles of users through a smartphone app.

In Fig. 7(a), the app implements the default mechanism, a facilitating nudge. The temperature of 19° is set as the default setting, which can be chosen by the user with a single click. Hence, it is made easier for the user to adopt pro-environmental behavior; of course, the user is still presented with the option to switch to 'manual' mode and manually adjust the temperature setting.

As mentioned in section 3.2.1, reinforcement nudges attempt to bring desired behaviors to the fore in individuals' thinking [6]. In Fig. 7(b) we show an energy dashboard that informs users about their energy consumption and saving results. Wherever possible, this information can be personalized and contextualized to make it more eligible and maximize its impact.



(a) Facilitation (b) Reinforcement (c) Social Influence (d) Confrontation-ENV (e) Confrontation-FIN

Fig. 7. Examples of operationalizing different types of nudges through a smartphone app.

The dashboard in Figure 7(c) applies a social influence nudge. The first widget compares the households' electricity consumption with that of other households, which could be selected with either geographic (neighborhood, municipality) or other criteria (house area, family size). It then shows how the household rank would improve if its residents saved a certain amount of energy and provides tips to them how to do this.

The mockups in Fig. 7(d) and (e) demonstrate confrontation nudges in action. When a user interacts with the phone to increase the thermostat setting, the app invokes a screen that reminds the user of the negative consequences of his/her intended action. The exact context of the screen (graphic, text) depend on the users. If they have strong environmental concerns and sensitivities, the screen could be something like Fig. 7(d); if they are more concerned about the financial aspects of excess consumption, the nudge could be realized through a screen like in Fig. 7(e).

E CONSTRUCT SCORES IN THE SIX ENERGY CONSUMER CLASSES AT THE END OF STEP 3

Figure 8 plots how each profile of energy consumers scores in each of the 15 constructs in Table 1. These scores are averages over the responses of those 2132 users who have been assigned to one of the six energy-saving behavior profiles during step 3 of the classification process described in section 3.2.2 of the main paper; namely, before the nearest centroid classifier was applied to the remaining 997 survey respondents.

F SOCIODEMOGRAPHIC CHARACTERISTICS OF THE SIX ENERGY CONSUMER CLASSES

F.1 Gender

Overall, the six profiles do not exhibit significant differentiation with respect to the gender distribution. For three profiles, the Environmentally conscious and well-informed energy consumers, the Concerned but comfort-oriented energy consumers, and the Indifferent energy consumers, the portion of females and males is approximately identical with the one in the overall dataset (48.86% and 51.14%, respectively).

For the other three profiles, males are marginally over-represented, as shown in Fig. 9(a). The two-sample t-test for the proportion of males in each of the three profiles and the overall dataset are marginally rejected (57.42% males, p = 0.024 for the Concerned but lacking awareness energy consumers, 56.14% males with p=0.035 for the Materialistic energy consumers escaping their personal responsibility and 53.85% males with p=0.035 for the Prone to social influence energy consumers).

F.2 Age

The within-profile age distributions deviate from the one in the total dataset. At 5% significance level the two-sample Kolmogorov-Smirnov test rejects the hypothesis that the within-profile age distribution and the overall (average age $M = 50.41$) are identical (samples of the same underlying distribution) for all energy consumer segments except for the Concerned but comfort-oriented energy consumers ($M = 49.02, p = 0.09$): strongly for the Environmentally conscious and well-informed energy consumers ($M = 48.74, p = 0.004$), the Materialistic energy consumers escaping personal responsibility ($M = 55.52, p = 0.000$), and the Prone to social influence energy consumers ($M = 53.49, p = 0.004$) and marginally for the Concerned but lacking awareness energy consumers ($M = 48.36, p = 0.027$) and the Indifferent energy consumers ($M = 53.49, p = 0.04$).

Figure 9(b) shows that Materialistic energy consumers escaping personal responsibility tend to be older, with more than half of them exceeding the age of 57 years, whereas the Prone to social influence energy consumers exhibit similar mass concentration in the interval 44-69 years old, prevailing in the ages 44-56. The lack of self-confidence is also prevalent in middle and high ages, 7 out of 10 users in this class being older than 44 years. On the other hand, almost half the Concerned but lacking awareness energy consumers are found in the two youngest groups [18-43], implying that educating the younger generations about energy-saving should remain high in the list of possible interventions.

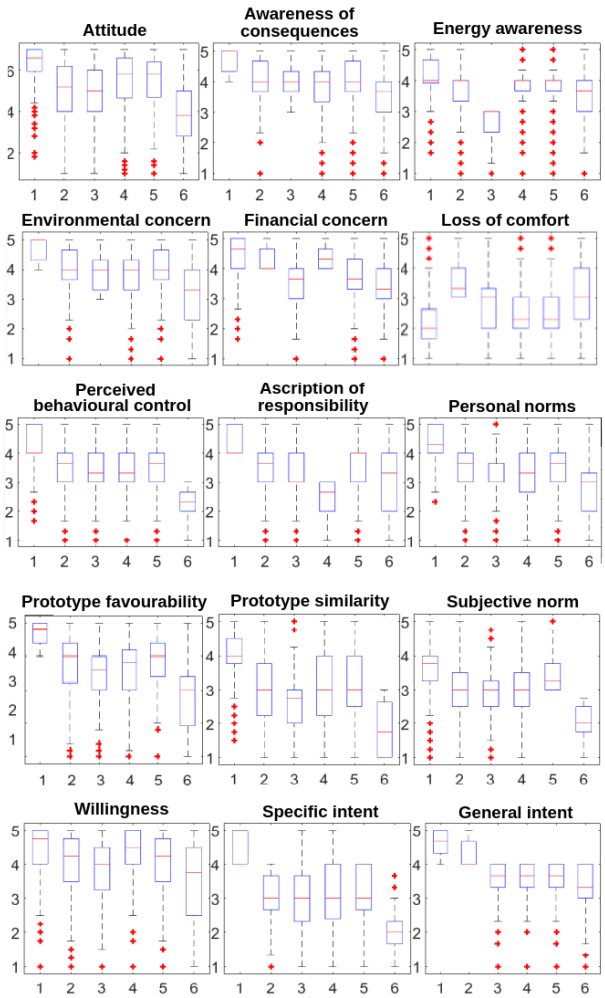


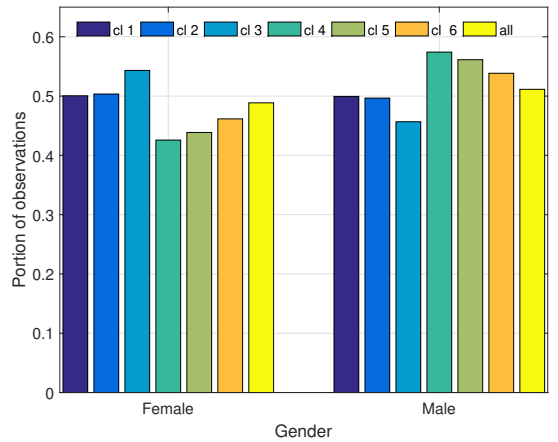
Fig. 8. Box plots of the per class scores (y-axis) in the 15 constructs of Table 1. The scores are averages over the respondents assigned to each class on the original 1-5 Likert scale (except for the ATT construct, which is measured on the 1-7 Likert scale): 2132 survey respondents, as classified in 3.2.3.

F.3 Education degree

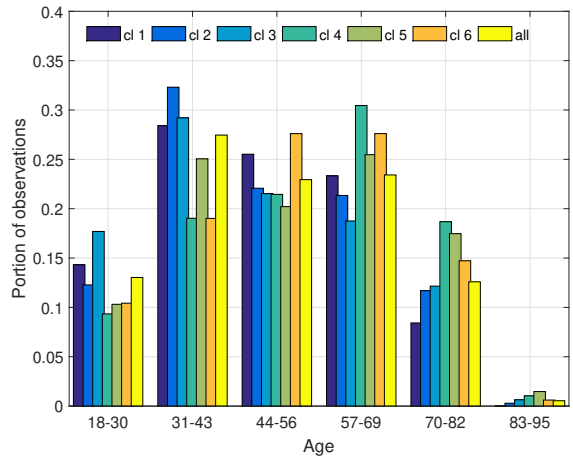
The education level of respondents is measured as an ordinal number on a scale of 0 to 6: None, Primary education, Lower secondary education, Upper secondary education, Bachelor's or equivalent level, Master's or equivalent level and Doctoral or equivalent level.

Applying the two-sample Kolmogorov-Smirnov test at 5% significance level to assess the hypothesis that the profile-level distributions of education's degree is identical to the one of the overall sample (average M = 3.972):

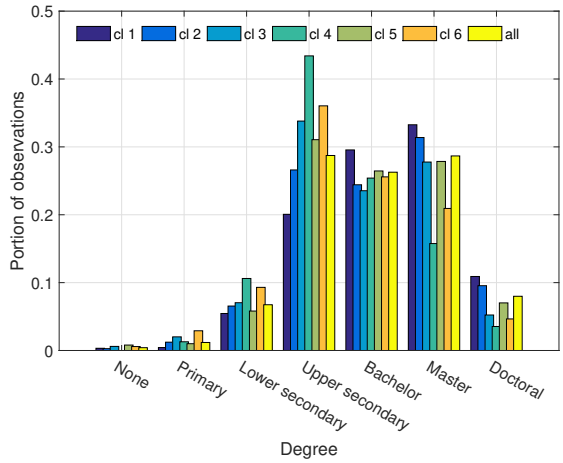
- The hypothesis is rejected for the Environmentally conscious and well-informed energy consumers (M = 4.21, $p < 0.0001$), the Materialistic energy consumers escaping personal responsibility (M = 3.54, $p<0.0001$) and the Indifferent energy consumers ($p=0.02$). Note that the average educational status of the 1st well-behaving cluster is noticeably



(a) Gender



(b) Age



(c) Education level

Fig. 9. Distributions of sociodemographic indicators across the six energy consumer profiles and the overall population of 3192 survey respondents; in the legend, cl1-cl6 denote energy consumer profiles 1 to 6.

higher than the average status in the overall sample, whereas it is the other way round (noticeably lower educational status) with the Materialistic energy consumers escaping personal responsibility and the Indifferent energy consumers (p=0.02).

- The hypothesis cannot be rejected for the other three groups, *i.e.*, the Concerned but comfort-oriented energy consumers (M =4.03, p = 0.22), the Concerned but lacking awareness energy consumers (M=3.82, p=0.058), the Prone to social influence energy consumers (M=3.93, p=0.99).

Looking at Fig. 9(c), the following remarks are worth making:

- 3 out of 4 Environmentally conscious and well-informed energy consumers have at least a Bachelor's level degree. The respective proportions are 2 out of 3 for the Concerned but comfort-oriented energy consumers and 3 out of 5 for the Prone to social influence energy consumers.
- On the other extreme, more than half of the Materialistic energy consumers escaping personal responsibility and 1 out of 2 Indifferent energy consumers have not obtained a degree from a higher education institution. This obviously relates to the fact that these two energy consumer groups involve the oldest (on-average) consumers.