Promoting Energy-Efficient Behavior by Depicting Social Norms in a Recommender Interface

ALAIN STARKE, Human-Technology Interaction Group, Eindhoven University of Technology, The Netherlands and Marketing & Consumer Behaviour Group, Wageningen University & Research, The Netherlands

MARTIJN WILLEMSEN, Jheronimus Academy of Data Science & Human-Technology Interaction Group, Eindhoven University of Technology, The Netherlands

CHRIS SNIJDERS, Human-Technology Interaction Group, Eindhoven University of Technology, The Netherlands

How can recommender interfaces help users to adopt new behaviors? In the behavioral change literature, social norms and other nudges are studied to understand how people can be convinced to take action (e.g., towel re-use is boosted when stating that '75% of hotel guests' do so), but most of these nudges are not personalized. In contrast, recommender systems know what to recommend in a personalized way, but not much HCI research has considered how personalized advice should be presented to help users to change their current habits.

We examine the value of depicting normative messages (e.g., '75% of users do X'), based on actual user data, in a personalized energy recommender interface called 'Saving Aid'. In a study among 207 smart thermostat owners, we compared three different normative explanations ('Global', 'Similar', and 'Experienced' norm rates) to a non-social baseline ('kWh savings'). Although none of the norms increased the total number of chosen measures directly, we show that depicting high peer adoption rates alongside energy-saving measures increased the likelihood that they would be chosen from a list of recommendations. In addition, we show that depicting social norms positively affects a user's evaluation of a recommender interface.

CCS Concepts: • **Information systems → Decision support systems**; • **Human-centered computing → Human computer interaction (HCI)**; *User models.*

Additional Key Words and Phrases: Behavioral Change, Energy Conservation, Rasch Model, Recommender Systems, User Experience

ACM Reference Format:

Alain Starke, Martijn Willemsen, and Chris Snijders. 2021. Promoting Energy-Efficient Behavior by Depicting Social Norms in a Recommender Interface. *ACM Trans. Interact. Intell. Syst.* 0, 1, Article 1 (2021), 32 pages. https://doi.org/10.1145/XXXXXXX

1 INTRODUCTION

Recommender interfaces seek to present content that fits user preferences [29]. In doing so, they can explain why certain items are presented [10, 58], for example by highlighting that other users have also bought a certain product. While recommenders in leisure domains (e.g., movies) are optimized to promote *any* item, some recommenders wish to promote *specific* items that support behavioral change [20, 54], for example, in domains such as healthy eating and

Authors' addresses: Alain Starke, a.d.starke@tue.nl, Human-Technology Interaction Group, Eindhoven University of Technology, Eindhoven, The Netherlands , Marketing & Consumer Behaviour Group, Wageningen University & Research, Wageningen, The Netherlands; Martijn Willemsen, m.c.willemsen@tue.nl, Jheronimus Academy of Data Science & Human-Technology Interaction Group, Eindhoven University of Technology, The Netherlands; Chris Snijders, c.c.p.snijders@tue.nl, Human-Technology Interaction Group, Eindhoven University of Technology, Eindhoven, The Netherlands.

Manuscript submitted to ACM

energy conservation [23, 56, 59]. For obvious reasons, recommending something specific is less likely to be successful and, therefore, social explanations of recommendations are often used to 'nudge' users (cf. [49, 57]), triggering social comparison mechanisms that might help to convince users [21, 44]. For example, highlighting that 65% of other users have bought a healthy product in an online supermarket, might persuade a user to also do so.

Studies in psychology have analyzed how social norms can effectively promote specific, one-size-fits-all environmental behaviors (e.g., [4, 13, 44]). A good example is the work of Goldstein et al. [25], who persuaded tenants of hotel rooms to re-use their towel by highlighting that '75% of others guests have done so', instead of emphasizing the environmental benefits of doing so. Such descriptive social norms have yet to be tested for a larger set of energy-saving measures. In fact, digital nudges are rarely used in personalized interactive systems [30], nor in recommender systems that support behavioral change [20, 53, 56]. Trying to convince users of energy-saving measures through social comparisons in energy recommender systems is challenging though, because energy-saving measures that yield high kWh savings are quite 'unpopular' [56]. For example, solar PV has only been installed on top of 13% of Dutch households [19], and "13% of users have solar PV installed" is not very convincing when presented as a normative message. For such messages to work, one needs at least a majority percentage to convince others. Our aim is to analyze whether we can use social comparisons to create a majority norm that can promote 'unpopular but useful' energy-saving measures [25, 44].

A nudging message that uses a majority norm can be created even for unpopular energy-saving measures, by highlighting the behavior of a specific group of peer users. For example, the adoption rate of Solar PV among more experienced users is much higher than the average rate of 13% [19, 56], and possibly exceeds 50% among users with a strong energy-saving attitude [56]. This would allow for a convincing, yet truthful majority norm message: "55% of experienced users (like you) have solar PV installed". Adoption rates for different kinds of users can be obtained by using the psychometric Rasch model [33], which has been used in work on energy recommender systems [54, 56]. Rasch differentiates between users in terms of their attitudinal strength and between energy-saving measures in terms of their frequency of use, so that both "users like you" and "energy-saving measures similar to this one" have actual meaning. That is, we use the Rasch model to deliver personalized recommendations that use majority norm nudges to convince users to take more energy-saving measures. In addition, depicting high norms scores might persuade users to select specific measures, including relatively unpopular (i.e., low frequency of use), which tend to be energy-efficient (i.e., high kWh savings), as well as to select those that are perceived as effortful (cf. [46, 53]).

1.1 Objectives

This is the point of departure for this paper. We blend social norms and recommender systems to help users attain their energy-saving goals, designing social explanations to signal a majority norm in a personalized advice context. We present an energy recommender interface named 'Saving Aid', which generates a list of household energy-saving measures that is tailored towards a user's energy-saving attitude through the psychometric Rasch model. In a between-subject web study, we then use the Rasch model to craft and depict specific normative message alongside energy-saving measures that highlight either the adoption rate of all users (Global Norms: '60% of users do this'), or that of peer users with specific attitudinal strengths (Similar norms: '60% of users similar to you do this'; Experienced norms: '60% of users who perform more measures than you do this').

We posit the following research questions. We examine changes in choice behavior due to the depiction of social norms, as well as explore whether other commonly used energy-saving attribute play role (e.g., kWh savings, perceived effort). We differentiate between 'overall' changes in choice behavior (i.e., total number of chosen energy-saving measures, kWh savings, and the difficulty of chosen measures), changes in what measure is chosen from a recommendation list due to presented content (i.e., whether users choose different energy-saving measures due to presented norm scores, while controlling for other measure attributes, such as perceived effort), and changes in how users evaluate a recommender interface (e.g., changes in user satisfaction):

- **RQ1:** Do social norms increase the number of chosen energy-saving measures or kWhs saved, and does this differ across different norms and different energy-saving attitudes?
- **RQ2:** Do social norms and other measure attributes affect which energy-saving measures are chosen from within a recommendation list?
- **RQ3:** To what extent do social norms affect a user's evaluation of an energy recommender interface?

2 LITERATURE REVIEW

This review focuses on work in environmental psychology and nudging that involve descriptive, social norms. We discuss the mechanisms of descriptive norms in psychological literature, contextualize them in the HCI domain, and formulate expectations for our web study. In doing so, we explain how the psychometric Rasch model is used to personalize energy-saving advice, as well as to craft effective social norms for our user study.

2.1 Nudges in a Personalized Context

Changes in a decision environment (i.e., 'choice architecture') that lead to predictable behavior are 'nudges' [57]. Notables examples include highlighting a default choice or using normative messages (e.g., 'most users do X') [25, 31]. The use of nudges and persuasive messages is rather uncommon in personalized interactive systems. For example, while recommender systems typically provide decision support by optimizing what to recommend [29, 40], nudges focus on how such content should be presented. This way, nudges can shift user preferences, which is also illustrated by studies on explanations in recommender interfaces [12, 58]. For example, if a recommender explains that a user's peers have chosen specific items, this might steer a user's preferences towards these items, even if they have a worse fit according to the recommender system [12].

2.2 Descriptive Norms in Energy Conservation

To date, recommender systems and most HCI studies have examined conservation decisions [36, 54, 59], but only in a social vacuum [1, 44]. While a few studies have applied social eco-feedback [24, 45], in which users are compared to their peers (e.g., your neighbors consume 3000kWh annually [1, 4]), its effects on a user's behavior are often limited [6]. The majority of HCI studies have yet to adopt the theoretical and empirical evidence from environmental psychology that explaining behaviors in terms of relevant peer groups and descriptive norms can affect one's energy-saving behavior and decision-making [25, 27, 44, 51].

A convincing message that affects preferences is one that highlights a majority norm [14]. Showing that a rather large proportion of peers performs a certain behavior [13, 27, 38], can trigger or promote socially-desirable behavior [12, 63]. Two mechanisms underlie this effect: compliance (i.e., the propensity to act consistent with presented norms) and conformity (i.e., adapting one's behavior to match an apparent majority) [14]. Compliance refers to one responding to a direct request to act consistently with presented norms [14], while conformity describes how a behavior is adapted to meet that of an apparent majority. Both compliance and conformity can fulfill one's need for accuracy or appropriateness in behavior or decision-making, for it can alleviate uncertainty surrounding a certain behavior [13]. For instance, individuals may want to gain the approval of others when it comes to pro-social behaviors, such as engaging in recycling if many others do so too [47].

For the design of the current study, we highlight work from Goldstein et al. [25] on the use of social norms to promote environmental behavioral. They show that hotel guests are more inclined to re-use their towels when asked to do so using descriptive norms ('join your fellow guests in helping to save the environment'), compared to a general environmental message ('help to save the environment'). Such normative messages highlight a community aspect ('75% of guests participated'), and are more convincing if they include context-rich or 'local' aspects [13, 25]. For instance, they show that referring to '75% of hotel guests', rather than '75% of citizens' is more effective, for it highlights an uncommon characteristic with the decision-maker [22, 25, 28].

Instead of only boosting a specific behavior, descriptive norms can also be used to promote a wider range of sustainable behaviors [41]. For example, customers of web shops purchase more healthy and energy-saving products, if they are explained using social norms instead of their environmental impact [3, 15]. We expect that this also applies to personalized advice in a recommender interface, when depicting normative messages alongside energy-saving measures.

2.3 Rasch Model

There is arguably a large range of norm percentages (probably anything below 50%), which will not trigger conformity [12–14]. Although it is hard to promote 'unpopular' measures such as 'Install Solar PV' [54], they typically yield relatively high kWh savings [53]. It might therefore pay off to somehow promote such measures, by making them stand out in the larger set of personalized user recommendations.

The dimensionality of energy conservation illustrates the large variety in adoption rates across measures [11, 33]. Energy-saving measures can be mapped on a one-dimensional scale using the psychometric Rasch model, based on how often these measures are performed [60]. In the context of attitude theory 'Campbell's Paradigm' [33][1], this frequency of use or adoption rate is operationalized as behavioral costs, which is defined to represent the execution difficulty of a measure, comprising different types of costs, such as money, time, and cognition [61]. This approach postulates that measures with smaller adoption rates face higher behavioral costs. For example, a study that fitted a Rasch scale of 79 energy-saving measures, shows that 92% of respondents lower the thermostat when leaving the house for a longer period [56], which has a relatively low behavioral cost level, while only 7% of respondents uses an energy-efficient heat pump, which has high behavioral costs. Moreover, while it is easy to make verbal statements about the importance of saving energy (i.e., low behavioral costs), engaging in actual behavior is much harder and arguably more representative for mapping a user's preferences in a recommender user model [53].

The characteristics of the Rasch model can be used to craft convincing social norms. An HCI study on energy recommender systems by Starke et al. [54] shows how to form a latent factor model, by asking a group of persons whether they perform a set of energy-saving measures, or not [33, 60]. Besides ordering measures on their adoption rate this way, users are also ordered with respect to how many measures they perform, which is operationalized as a person's energy-saving attitude [33]. Hence, users with stronger attitudes are assumed to perform more measures.

The adoption rates of measures are what we label as 'Global' norms. These are statements about the general population that can be presented alongside energy-saving recommendations, analogous to the norms used by Goldstein et al. [25]. For example, "55% of other users have installed weather strips on doors" [53]. As discussed in the introduction, we

[1]The Rasch model is used in this study as a mathematical formalization of Campbell's Paradigm [33]. Other attitude theories tend to suffer from attitude-behavior gaps [16], for one's evaluative attitude might be at odds with one's actual behavior. For example, one can agree that environmental action is important (i.e., holding a favorable environmental attitude), but might not actually engage in any environmental behavior. Rasch accounts for this uncertainty by describing a stochastic relation between attitude and behavior, instead of a deterministic relation.

expect normative messages such as "75% of participants use X" to signal that the majority of a population has already adopted a certain energy-saving measure, and are therefore expected to be more persuasive than minority norms, such as "30% of users do X".

2.4 Crafting Personalized Social Norms

Using the Rasch Model, we can craft personalized norms that go beyond 'Global' percentages. Instead of highlighting the frequency of use among all users, the behavior of specific groups can be highlighted. This is achieved through the Rasch model, for the probability that a measure is performed by a specific user is person-dependent. This is shown in Equation 1: the probability p that a measure i is performed depends on a measure's behavioral costs δ, as well as the attitudinal strength θ of an individual n, where δ and θ are expressed in logistic scale units (logits) [33, 60]:

$$Probability\ of\ n\ performing\ i = \ln \frac{p_{ni}}{1 - p_{ni}} = \theta_n - \delta_i \tag{1}$$

For any energy-saving measure, Rasch predicts the same adoption probability for all users with a specific attitudinal strength [34], along with increasing probabilities for users with stronger attitudes. Among the larger population, we consider this probability to an adoption rate that can be communicated to a user, such as '60% of users with attitude X do this'. Hence, we can craft personalized normative messages based on peer users with either similar or stronger attitudes. Not only could higher norm scores across an entire recommendation list persuade users to choose more energy-saving measures, it could also help to make 'unpopular' measures, which have a relatively low 'Global' adoption rate and high behavioral costs, more appealing. This could, in turn, persuade users to choose measures that have relatively high kWh savings (e.g., Solar PV, which has a low adoption rate), or to choose measures that are subject other unattractive attributes, such as perceived effort [46, 52].

How can peer users with 'similar or stronger attitudes' be translated to a convincing normative message? Literature on advice-taking highlights relevant 'advice sources' that can be used for this purpose. Mentioning a specific peer group is shown to affect choice and advice acceptance [8], suggesting two important advice source characteristics for our work. First, similarity in relevant attitudes can increase the extent to which advice is considered or liked [8, 50]. The Rasch scale allows the design of 'Similar' norms alongside recommendations, which can show higher adoption rates than global norms, especially for users with stronger attitudes. For example, users with a strong attitude might be presented the 'Global' norm "20% of users have installed radiator reflectors" [54], while the 'Similar' norm would be "60% of users like you have installed radiator reflectors".

A second characteristic is a peer user's perceived expertise [8, 32]. Expert advice is less likely to be ignored than suggestions from novices [7, 8, 32]. In this study, we assume peers to possess such higher expertise if they perform more measures ('Experienced' norms), thus having stronger attitudes and higher adoption rates. For example, where 'Similar' norms would report "55% of users like you do X", 'Experienced' norms at an attitude θ that is +1 logit stronger than the user report an adoption rate of 78%.

Combining different advice sources and adoption rates, we craft three different normative messages:

- Global norms: "X% of users perform this measure."
- Similar norms: "Y% of users who perform similar measures as you, perform this measure."
- Experienced norms: "Z% of users who perform more measures than you, perform this measure."

Table 1. Recommendation scenario to illustrate what norm percentages are presented for each norm, and how this depends on the user's attitudinal strength. We imagine that there are two users: User 1 has relatively weak energy-saving attitude, User 2 has a relatively strong attitude. If each user is presented a measure that tailored towards their attitude ($\delta = \theta$), then these are the presented norm percentage for each norm condition.

	Presented Norm Percentages (for attitude-tailored advice: $\delta = \theta$)		
	Global Norms	Similar Norms (= User θ)	Experienced Norms (= User $\theta + 1$)
User 1: $\theta_1 = -1$	72%	50%	75%
User 2: $\theta_2 = +1$	30%	50%	75%

2.5 Global vs Person-Dependent Norm Scores

The percentages for our normative messages are determined using the Rasch model. To show how they depend on a user's energy-saving attitude, we present a recommendation scenario in Table 1. Suppose there are two users and that User 1 has an attitude $\theta_1 = -1$, which is weaker than User 2 ($\theta_2 = +1$), which are each presented a measure with behavioral costs δ equal to their attitude θ (in line with [56]). As a result, User 1 is shown a measure with lower behavioral costs than User 2.

Table 1 shows that 'Global' norm percentages depend on the user's attitude. User 1 has a relatively weak attitude and is therefore presented a 'popular' measure with a high 'Global' adoption rate (i.e., 70%). User 2 has a stronger attitude and, therefore, her attitude-tailored measure has a lower 'Global' adoption rate of 30%, which is not very convincing. In contrast, the adoption rates of the personalized 'Similar' or 'Experienced' norms do not depend on the 'Global' adoption rate, but the user's attitudinal strength and the measure's behavioral costs. Therefore, they are identical for both users (resp. 50% and 75%), and thus lead to a more convincing norm for User 2, compared to 'Global' norms.

Table 1 shows what normative messages are most likely to be the most effect what types of users. Based on the adoption rates, we expect users with stronger attitudes (i.e., User 2) to choose more measures when facing 'Similar' norms, while users with weaker attitudes (i.e., User 1) do so for 'Global' norms. It is possible that the higher degree of similarity signalled by a 'Similar' norm message could overcome differences in norm percentages with 'Global' norms [25]. Nonetheless, another study by Yaniv et al. [62] argues that inexperienced users (i.e., with a weak attitude) are more likely to rely on majority advice (i.e., a 'Global norm %') than similar advice, while experienced individuals (i.e., with a strong attitude) rely on similar peers.

Table 1 also suggests the additional benefit of higher adoption rates for 'Experienced' norms, compared to 'Similar'. Although the persuasiveness of expertise (i.e., "others who perform more measures than you") may be mitigated because of the reduced similarity, we expect that the higher adoption rates for Experienced norms (75%) across an entire recommendation list will be more persuasive than similar norms (50%). This could particularly apply to the adoption of measures that face high levels of behavioral costs or perceived effort [62].

2.6 Perception of Descriptive Norms

Besides evaluating behavior, it is also useful to understand how users perceive such a descriptive norm. Studies in environmental psychology teach us that how an individual evaluates environmental aspects can determine behavioral outcomes [2, 18]. For instance, social proof of others performing a particular behavior might lower the thresholds towards performing it [21].

Previous recommender studies have similarly highlighted the importance of perceptions in explaining the user experience [37], for they allow us to understand why a change in a particular system aspect increases the user experience. For example, Starke et al. [54] show that tailored recommendation lists with low levels of behavioral costs (δ) are more likely to be perceived as feasible and, in turn, show stronger perceived support, higher levels of user satisfaction, and more energy-efficient choices [54]. Likewise, we expect descriptive norms to lower the behavioral thresholds to choose and, eventually, adopt energy-saving measures, which is assessed through perceived feasibility, perceived support, and subsequent user satisfaction.

2.7 Research Expectations

Based on the discussed literature, we summarize the expectations for our user study per research question. First, we examine whether social norms affect the total number of chosen measures and kWhs saved, across users with different energy-saving attitudes (RQ1). In line with [25], we test three different interfaces that depict normative messages (i.e., 'Global', 'Similar', and 'Experienced' norms) alongside energy-saving measures in a recommender system, and compare their effectiveness to a non-social baseline (i.e., kWh Saving Score). We formulate the following expectations:

- Social norms increase the number of energy-saving measures chosen by users, across all attitudinal strengths.
  - Users with a weak energy-saving attitude choose more measures if a 'Global' norm is depicted instead of a 'Similar' norm, and vice versa for users with a strong energy-saving attitude.
  - Users choose more measures if they are explained with 'Experienced' norms rather than 'Similar' norms.
- Social norms increase the amount of kWh savings per chosen measure, as well as the average behavioral level, particularly for users with strong energy-saving attitudes.

Second, we investigate whether social norms and other measure attributes affect which energy-saving measures are chosen from within a recommendation list (RQ2). Based on the reviewed literature and the recommendation scenario in Table 1, we expect the following outcomes:

- The presented norm percentage increases the likelihood that a measure is chosen from a recommendation list.
- Measures with high levels of perceived effort are more likely to be chosen when accompanied by high norm percentages, such as majority norms.

Finally, we examine whether social norms affect a user's evaluation of our recommender interface (RQ3). Based on previous energy recommender research, we expect users to perceive and evaluate recommender interfaces that depict social norms more favorably, compared to those that emphasize the environmental impact.

3 METHOD

We investigated to what extent descriptive norms boosted the adoption of a heterogeneous set of tailored energy-saving measures. We first collected data in a pre-study to validate our one-dimensional construct, used to personalize both advice and norms. Thereafter, we designed our energy recommender interface called 'Saving Aid' and performed an online user study on our normative intervention.

3.1 Pre-study: Setting up a Rasch Scale for Personalized Norms

To generate recommendations based on the Rasch model, we designed a survey that was part of different study [53]. Participants were asked to disclose their current energy-saving behavior, indicating for 13 to 25 randomly sampled energy-saving measures (out of a database of 134) whether they performed them or not ('yes' or 'no').

We used dichotomous responses from 555 participants (50.6% male) with a mean age of 43.4 years (SD = 19.7) to fit a one-dimensional measurement scale of 134 energy-saving measures. An tabulation of the scale is reported in Appendix A, in Table 5. Each measure was assigned a distinct behavioral cost level, which formalized how likely a user would be to perform a particular measure [33]. In terms of adoption rates, the scale ranged from 94% to 1%.

Furthermore, Table 5 also shows how the estimated kWh savings of each measure are distributed across the scale. Although higher kWh savings seemed to be more prevalent 'higher up the scale' (i.e., for higher behavioral cost levels), it was possible to perform measures with moderately high kWh savings across the entire scale. This is also depicted in Figure 1, which shows a small increase in the average kWh savings for higher behavioral cost levels.

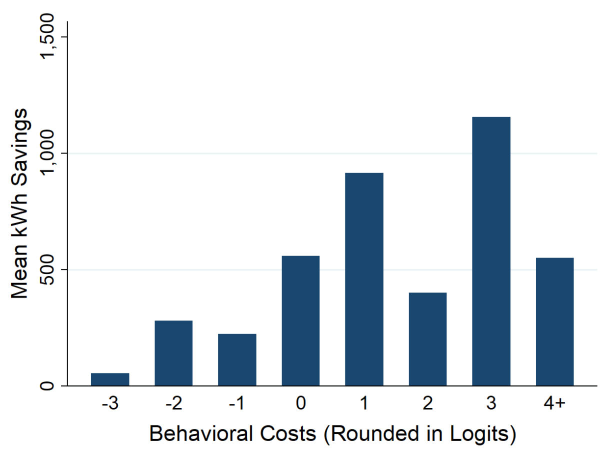

Fig. 1. kWh savings across the Rasch construct, averaged per behavioral cost level.

3.1.1 Fit Statistics. Table 5 also describes the scale's infit statistics (for mathematical details, see [9]). Overall, the scale's item parameters were determined reliably (α = .95, M = .05, SD = 1.57), as all measures fitted the construct by meeting the prescribed 'infit' criteria [9]. Due to an item separation of 4.51, we could reliably discern 4 to 5 strata of behavioral costs.

3.1.2 Perceived Effort. The same pre-study also collected data on how effortful participants perceived measures to be [53]. Although a measure's perceived effort decreased the likelihood that a measure was chosen in previous studies [46], we expected that depicting high norm scores might help to increase that likelihood. A sub-sample of the participants (N = 304) was presented a 4-point scale alongside each measure, on which they could indicate whether executing a measure would require either 'very little effort', 'little effort', 'fairly some effort', or 'a lot of effort'. The mean response per measure (304 users, rating 25 measures each) is listed in the Appendix, Table 5. We observed a moderate to strong correlation between a measure's perceived effort and a its behavioral costs: $r(134) = 0.59$, $p < 0.001$.

3.2 'Saving Aid' Recommender Study

Following our pre-study, we set up an online user study in collaboration with a Dutch energy supplier (i.e., Eneco). We compared four different recommender interfaces, of which three depicted social norms alongside energy-saving advice and one the kWh savings values.

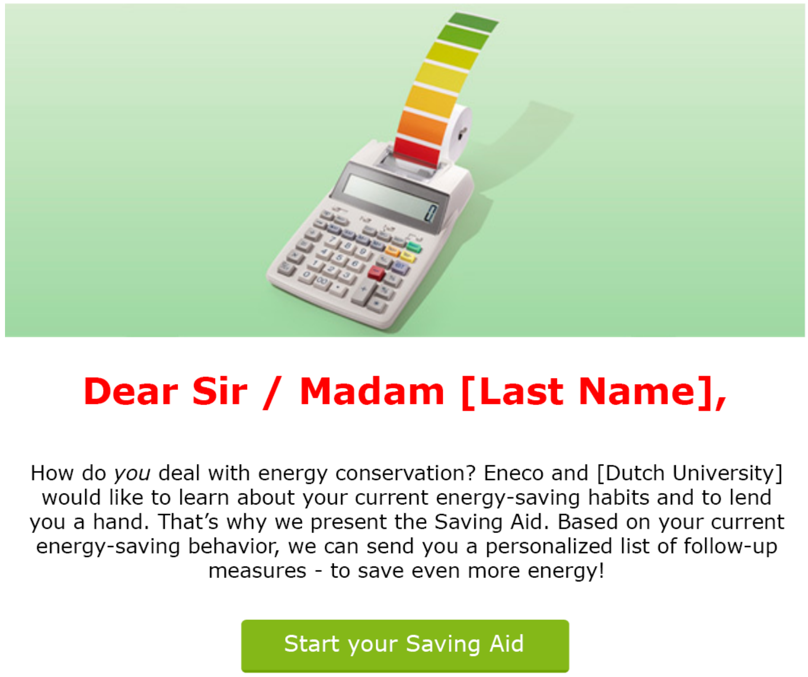

Fig. 2. Excerpt from the email template sent to customers of Dutch energy supplier Eneco who owned a smart thermostat.

3.2.1 Participants. Members of a consumer panel at Eneco were invited to use our 'Saving Aid' recommender system to find and select appropriate energy-saving measures to take in the households. Panel members, which were all smart thermostat owners, were sent a formal email invitation, of which an English translation is depicted in Figure 2. This panel was considered a good target group for our study, as they were able the improve energy efficiency in one's household beyond simple behavioral curtailment, as they were predominantly homeowners.

In total, 217 participants used our 'Saving Aid' and filled out the evaluation questionnaire. However, we excluded ten participants for either completing the study in less than three minutes, indicating to not trust the website, or showing no variation in the evaluation questionnaire. Eventually, we considered a sample of 207 participants (M = 53.5 years, SD = 14.0) that comprised predominantly males (87%). Among our participants, only 26.6% owned the home they lived in, while the majority lived in a town house (58.5%).

3.2.2 Procedure. To estimate each user's attitude, we randomly sampled 13 energy-saving measures from across the behavioral costs scale. These were presented sequentially to users, who indicated whether they performed them or not ('yes' or 'no'). Subsequently, we inquired on the user's housing situation to filter irrelevant measures from the recommendation list.

Afterwards, we presented each user a list of nine energy-saving recommendations, whose behavioral costs were tailored towards the user's estimated energy-saving attitude ($\theta \approx \delta$). In addition, the measures were ordered in terms of their estimated kWh savings. Figure 3 depicts the top-2 measures of such a list, presenting a measure's name, a short description, a score or percentage, and a (norm) explanation. Recommendations were sampled between the adoption probabilities of 18% to 75%, which ranged from -1.5 to $+1$ logit in terms of the attitude-behavioral costs difference.

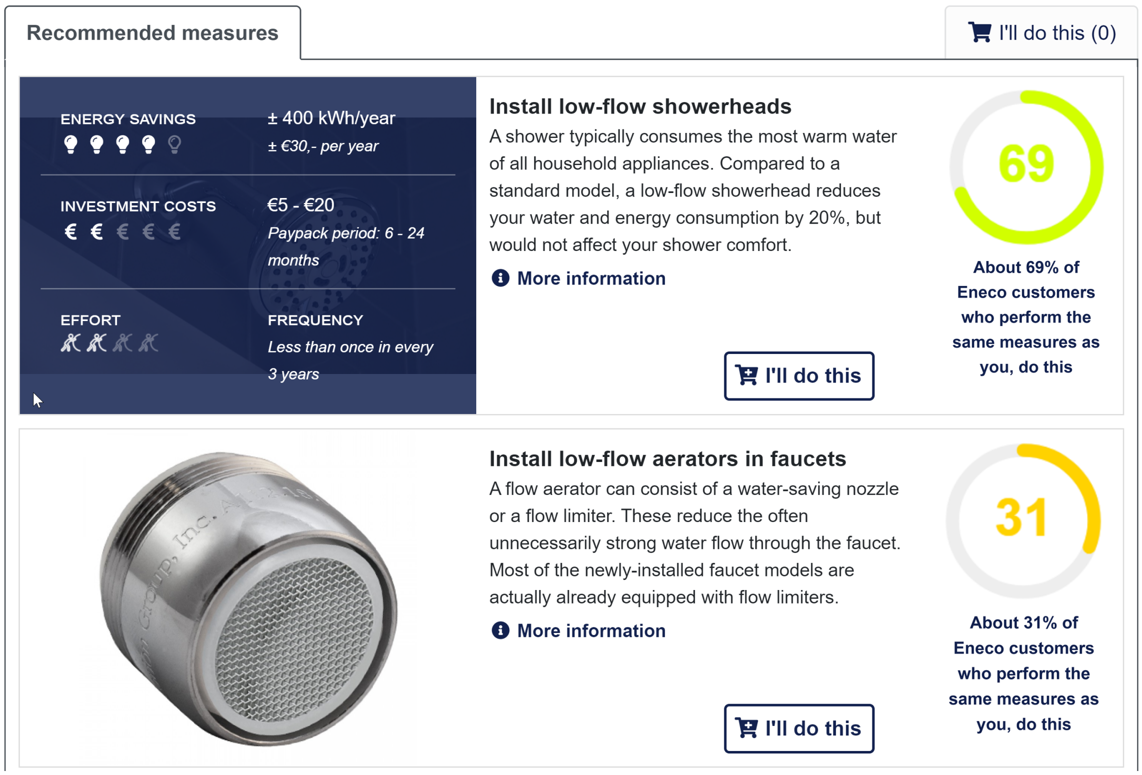

Fig. 3. Our 'Saving Aid' energy recommender interface (NL: 'Besparingshulp.nl'), translated to English. Depicted are the name and a short description of the top-2 recommendations in our interface (e.g., at the top: 'Install low-flow showerheads'), out of a total of nine recommendations. Users could select any number of measures they would like to perform, by clicking 'I will do this'. Measures are sorted from high to low kWh savings. On the left, users could hover a measure's image to inspect additional attributes: kWh savings (scaled from 1 to 5 light bulbs), the annual savings (in €), investment costs (in €), payback period (from 'less than a month' to 'never'), effort, and behavioral frequency. Depending on the condition, the numbers on the right either show a score or a norm percentage. Depicted here is a 'Similar' norm.

We asked users to select any number of recommended measures that they wished to perform. Users could hover for 'more info' to see other commonly used attributes of an energy-saving measure, such as its frequency and kWh savings. Figure 3 portrays this on the left-hand side of the top measure.

After interacting with our 'Saving Aid' interface, we inquired on the user's subjective evaluation of the system. To this end, users were presented statements on 7-point Likert scales. Finally, users could share demographic details and disclose their email address to receive information on chosen measures.

3.2.3 Research Design. The presented score and explanation depicted alongside each measure in the list is subject to four between-subject conditions. In line with Goldstein et al. [25], we compared three norm explanations to an environmental baseline:

(1) 'Savings' score (baseline): We presented a 'Saving Score' of 0 to 100, where 100 represented the highest kWh savings in the list.

(2) 'Global' norm: The adoption rate of measures on the scale, which ranged from 2% to 98%, explained as: "XX% of other customers do this."
(3) 'Similar' norm: The user's adoption probability (between 18% and 75%), explained as: "XX% of other customers who perform the same measures as you, do this."
(4) 'Experienced' norm: The adoption probability for an attitude 1 logit above the current user, which fell between 37.8% and 88%. It was explained as: "XX% of customers who perform more measures than you, perform this measure."

3.3 Measures

3.3.1 Choice Variables. We examined a user's choice behavior through two analyses. First, we considered the total number of chosen measures by a user per condition [RQ1], as well as the amount of chosen kWh savings per measure and the average behavioral cost level of chosen measures (i.e., attitude-cost difference). Second, we predicted the likelihood that a specific measure was chosen in each condition using on the presented norm scores [RQ2]. In the same model, we considered a measure's perceived effort, as well as explored possible interaction effects.

3.3.2 Attributes & Characteristics. To address our research questions, we dichotomized each user's energy-saving attitude to discern between users with weak and strong attitudes. Figure 4 depicts the distribution of attitudinal strengths in our sample, which were estimated at discrete levels. In previous studies that used a Rasch scale, a cut-off would be placed at $\theta = 0$ [9], but this would lead to very uneven groups in the current study ($N_{weak} = 48$, vs $N_{strong} = 159$). However, a median or mean split (*Median* = 0.42, *M* = 0.52, *SD* = 0.79) would neither lead to balanced groups[2], nor would it properly differentiate between weak and strong attitudes. To balance representativeness both factors, we instead placed the cut-off at $\theta \leq 0.25$ (see Figure 4), creating a group of 82 users with a weak attitude and a group of 125 users with a strong attitude.

Other attributes are presented in our 'Saving Aid' interface (cf. Figure 3). We used the presented score as a continuous measure ('Savings Score' or norm %'s) to assess its impact on the probability that a measure was chosen (RQ2). Other attributes could be inspected in the interface by hovering an energy-saving measure. From these, we included a measure's perceived effort in our analyses, as we examined whether social norms could boost the adoption of effortful measures.

3.3.3 User Evaluation Aspects. To examine whether users evaluated recommender interfaces that depicted social norms more positively than non-social one (RQ3), we inquired on different user evaluation aspects. After interacting with the 'Saving Aid', users were presented questionnaire items on a 7-point Likert scale about the *Perceived Feasibility* of the presented recommendations, their *Perceived Support* from the system, and the user's satisfaction with the chosen measures (i.e., *Choice Satisfaction*). All items were based on earlier research of Knijnenburg and Willemsen [37], and were eventually submitted to a confirmatory factor analysis, as part of a Structural Equation Model. The results are described in Table 4, and discussed in the results section.

We also included two user characteristics in our user evaluation analysis. Besides discerning between users with weak and strong energy-saving attitudes, we also inquired on a user's environmental concern, scoring all 15 items of the revised NEP scale [18] on a 7-point Likert scale. We found that the scale had an acceptable internal consistency ($\alpha = .78$). However the path model was built using only 6 items to optimize the fit of the Structural Equation Model.

[2] The group of users that was estimated to have an attitude θ of 0.42 was rather large: 25.6%. For obvious reasons, labeling this group as having a 'weak' or 'strong' attitude would lead to unbalanced groups either way.

4 RESULTS

We investigated to what extent social norms affected user choices and evaluation of an attitude-tailored list of energy-saving measures, compared our kWh savings baseline. After presenting manipulation checks, we first examined the total number of choices, chosen kWh savings, and average chosen behavioral cost level across our normative conditions to the baseline (RQ1). Second, we predicted whether the likelihood that a measure was chosen from a recommendation list was affected by the presented norm score, compared to the effects in the baseline (RQ2). Third, we investigated whether the different normative messages affected a user's perception of the system (RQ3).

4.1 Manipulation Checks

4.1.1 Presented Norm Scores. We examined whether the presented scores and percentages were in line with our intended manipulations. As outlined in Table 1, we intended 'Global' norms to yield higher norm %'s for users with weak attitudes than in the 'Similar' condition, and vice versa for users with strong attitudes. Moreover, we intended 'Similar' norms to have a median score around 50%, while the median score for 'Experienced' norms was designed to fall around 75%, across all attitudinal strengths.

Figure 5 depicts the distribution of presented scores per condition, across weak (in blue) and strong (in red) attitudes. It shows that the Savings Score roughly captured all possible scores with a median score of 60, while the normative conditions had narrower distributions, which was expected. However, it shows only a minor difference in presented median scores between 'Global' norms (53%) and 'Similar' norms (49%) for users with weak attitudes, which was much

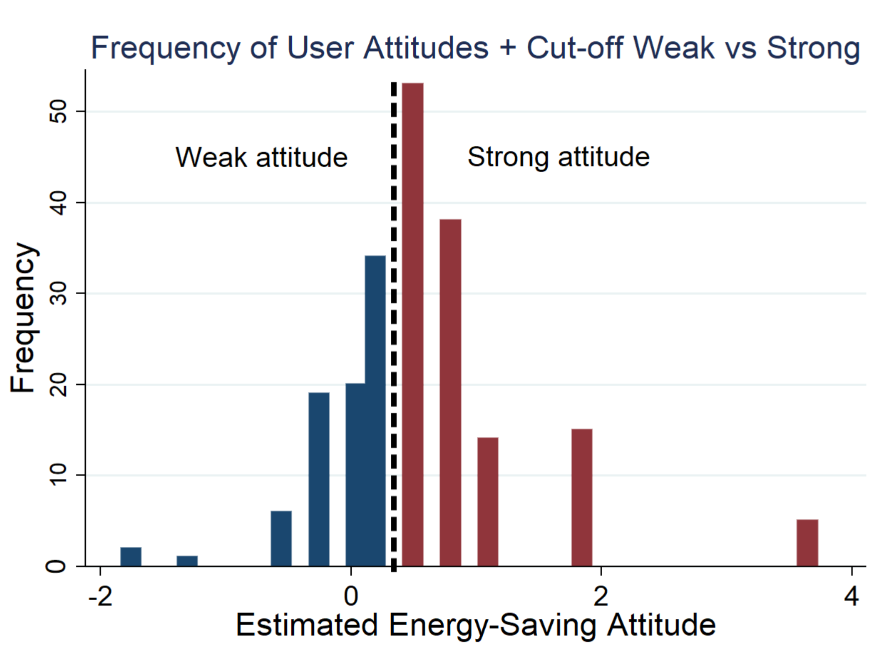

Fig. 4. Histogram depicting the distribution of energy-saving attitudes in our sample. It also depicts the cut-off between weak and strong energy-saving attitudes, placed at $\theta = 0.25$.

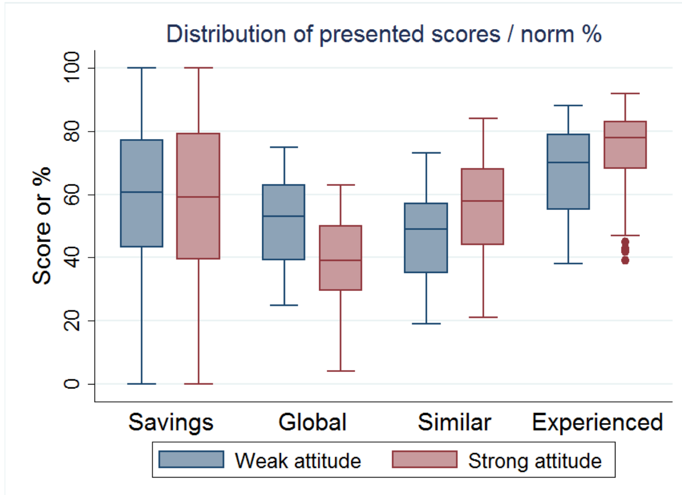

Fig. 5. Box plot of the presented scores or norm %'s in our recommender interface, across conditions and attitudinal strength, presented as a manipulation check. Scores presented to users with weak attitudes are depicted in blue, scores presented to users with strong attitudes in red. As intended, the Savings condition included all scores, while the distributions were narrower and more selective in other conditions.

smaller than our intended manipulation (72% vs. 50%). This made it less likely for any effect to surface between 'Global' and 'Similar' norms for users with a weak attitude. In contrast, the difference in median scores for strong attitudes between 'Global' (39%) and 'Similar' (58%) was consistent with our intended manipulation (30% vs. 50%).

Furthermore, Figure 5 shows there were small differences in the presented median scores between weak and strong attitudes in the 'Similar' ($Weak$ = 49% vs $Strong$ = 58%) and 'Experienced' conditions ($Weak$ = 70% vs $Strong$ = 78%). Nonetheless, the difference between the median scores of 'Similar' (53%) and 'Experienced' (75%) was in line with our intended manipulation.

4.1.2 Validation of the used Rasch construct. To validate the representativeness of the Rasch scale used for this study (cf. Table 5), we inferred a new Rasch construct based on data collected in this study. Without reporting detailed fit statistics, we found the Rasch scale to have an adequate model fit (item reliability $\alpha = 0.85$; person reliability $\alpha = 0.69$). We compared the behavioral cost levels of the new construct to the one used in the current study, performing a pairwise correlation analysis. In spite of both research populations being rather different, it revealed a strong correlation: $r(134) = 0.82$, $p < 0.001$, which showed that the two constructs were comparable and had to a large extent a similar order. We only found that a few measures would significantly shift in terms of their behavioral cost level. For example, because all users in the current sample were smart thermostat owners, the measure 'install a centralized temperature system with zone controls & thermostats' changed from +3.33 to −2.95 logit.

4.2 Overall Choice Behavior (RQ1)

4.2.1 Total Number of Chosen Measures. We investigated whether the depiction of descriptive norms increased the number of energy-efficient choices, compared to our non-social baseline (RQ1). We first addressed this question through a multilevel logistic regression analysis to estimate whether the likelihood that a measure was chosen. Table 3, Model 1 (reported in Section 4.3) examined whether a measure was more likely to be chosen in each normative condition,

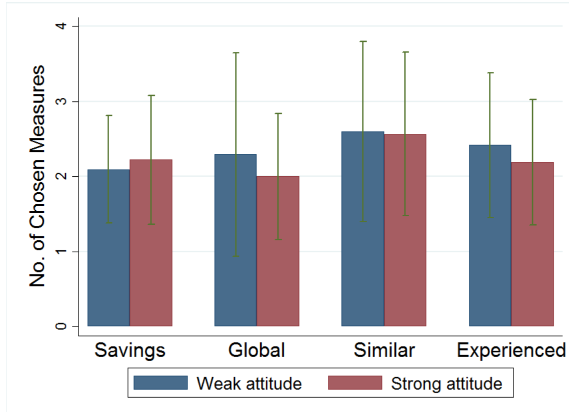

Fig. 6. Total number of chosen measures, per condition and attitude strength. Error bars are 1 $S.E.$

compared to the the baseline. We found no differences between the normative conditions and the Saving Score baseline (all p-values > 0.05): not for 'Global' norms ($OR = .83$, $S.E. = .33$), neither for 'Similar' norms ($OR = 1.17$, $S.E. = .46$), and nor for 'Experienced' norms ($OR = 1.02$, $S.E. = .39$).

This result is illustrated in Figure 6. It depicts small, but non-significant differences in the number of chosen measures between the Savings baseline and each norm condition, indicating that social norms did not lead to changes between conditions. Based on the norm scores, we expected that users with a weak attitude would choose more measures when facing 'Global' norms compared to 'Similar' norms, and vice versa for users with a strong attitude. Although Figure 6 depicts a small difference between the two types, Kruskal-Wallis tests of ranks revealed that these were not significant: not between 'Global' and 'Similar' norms for users with a weak attitude: $H(1, 36) = 0.267$, $p = 0.61$, nor for users with a strong attitude: $H(1, 65) = 0.45$, $p = 0.50$. In addition, based on the norm scores, we expected that depicting 'Experienced' norms would lead to more choices than presenting 'Similar' norms, but we found no significant difference between the two: $H(1, 105) = 0.028$, $p = 0.87$.

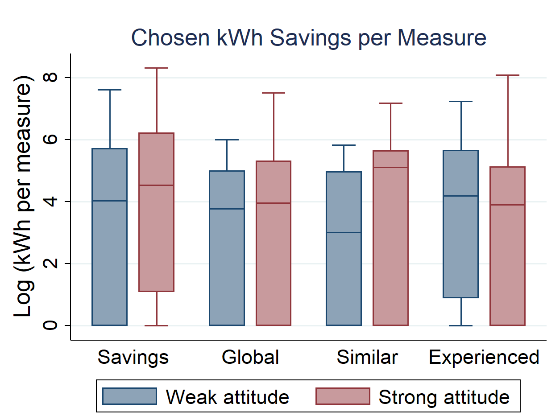

Fig. 7. Box plot of the log transformed chosen kWh savings per measure, divided across conditions, as well as between weak (in blue) and strong (in red) attitudinal strength.

4.2.2 Chosen kWh Savings per Measure. We further expected that social norms could boost the overall kWh savings chosen. We present our results in Figure 7 and compared differences in chosen kWh savings per measure, across conditions and attitudinal strengths. Kruskal-Wallis tests provided no evidence for differences in chosen kWh savings between the baseline (M = 401, SD = 843) and the normative conditions (M = 193, SD = 393); not for 'Global' norms: $H(1, 100) = 2.21$, $p = 0.14$; neither for 'Similar' norms: $H(1, 97) = 0.65$, $p = 0.42$; nor for 'Experienced' norms: $H(1, 103) = 0.79$, $p = 0.37$. As we neither found differences across different attitudinal strengths, this suggested that our normative condition did not increase overall kWhs saved, as indicated by a user's choice behavior.

4.2.3 Behavioral Costs. Finally, we examined whether the behavioral cost levels of chosen measures differed across conditions and attitudes (i.e., weak vs strong). We used the difference between a user's attitude and the behavioral costs of a chosen measure: the 'attitude-cost difference'. A positive difference would indicate that users had chosen relatively challenging measures for their attitude level, while a negative difference difference would suggest that users had selected relatively easy ones. We expected that social norms might be more effective to persuade users to select challenging measures, particularly for users with a strong energy-saving attitude, leading to a positive attitude-cost difference.

Table 2 presents two multilevel linear regression models, clustered at the user level. Model 1 examined whether this led to different choices in the norm conditions, compared to the Saving Score baseline. We found that both the 'Global' and 'Similar' norm conditions positively affected the chosen 'attitude - behavioral cost' difference of chosen measures, compared to choices in the baseline: $\beta = 0.29$, $p < 0.01$, for the Global norm condition; $\beta = 0.22$, $p < 0.05$, for Similar norms. This suggested that users in those conditions were more likely to choose measures with higher behavioral costs, which were located 'further up' the Rasch scale. In contrast, such an effect was not observed for the 'Experienced' condition.

Table 2. Multilevel linear regression models predicting the relative behavioral costs level (difference between attitude - behavioral costs) of chosen measures, clustered at the user level. Energy-saving attitude discerns between weak and strong, norm condition dummies are compared to the Saving Score baseline. β represents the regression coefficient. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

	Model 1	**Model 2**
	β (*S.E.*)	β (*S.E.*)
Global Condition	.29 (.10)**	.22 (.11)*
Similar Condition	.22 (.10)*	.071 (.10)
Experienced Condition	.17 (.099)	.14 (.092)
Energy-saving attitude		.25 (.12)*
Attitude X Global		.0032 (.15)
Attitude X Similar		.071 (.14)
Attitude X Experienced		-.0068 (.14)
Constant	-.082 (0.72)	-.14 (.069)*
R^2 (overall)	.029*	.12***

Table 2, Model 2 examines whether the effects of Model 1 were affected by a user's energy-saving attitude. We found that users with strong energy-saving attitudes were more likely to choose measures with relatively higher behavioral costs, across all conditions: $\beta = 0.25$, $p < 0.05$. The observed effects of the presented norms in Model 1 were reduced in Model 2, as only Global norms still positively affected the chosen behavioral cost level ($p < 0.05$). Moreover, we neither found interaction effects between attitude and the presented norm interface.

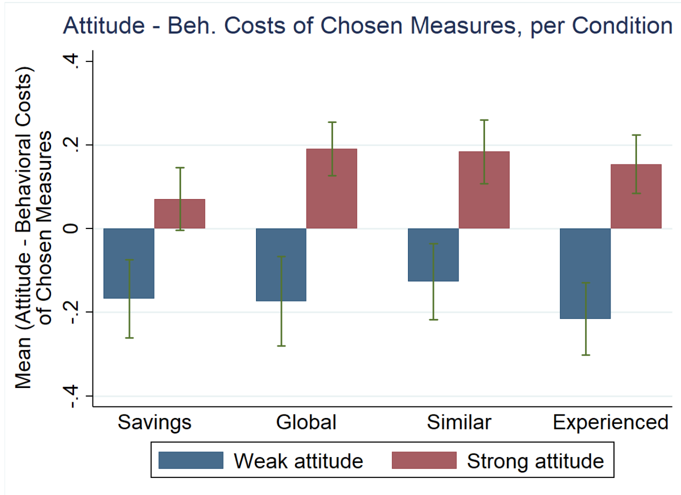

Fig. 8. Mean relative behavioral cost level of chosen measures (i.e., Attitude − Behavioral Costs), divided per condition and attitudinal strength.

To better understand the results reported in Table 2, please refer to Figure 8. It depicts how users with stronger energy-saving attitudes had also chosen measures with with behavioral costs above their own attitude, yielding a positive 'attitude-cost difference'. In contrast, users with weak attitude had chosen 'below' their own attitude in all conditions. Regarding specific conditions, the average behavioral cost level of chosen measures in the 'Global' ($M = 0.24$) and 'Similar' ($M = 0.13$) conditions was higher than in the baseline ($M = -.075$). In addition to the models reported in Table 2, we also examined whether demographical factors (i.e., age, income, etc.) and housing characteristics affected the current results, but we found no significant effects.

4.2.4 Conclusion. Overall, we found that the use of descriptive norms did not persuade users to choose more energy-saving measures in total, compared to our kWh savings baseline. Nor did we observe a change in the overall kWh savings selected by users. It seemed that adding normative explanations to a personalized list of energy-saving recommendations did not boost the overall choice behavior in terms of sustainability. Instead of main effects, it seemed that more 'within list' effects had occurred (cf. Section 4.3).

We did find overall chosen in terms of the relative difficulty of chosen measures for some normative conditions, in particular 'Global' Norms. This suggested that normative explanations, or other social explanations for that matter, are more successful in persuading users to select 'challenging' measures, compared to a factual explanation (i.e. kWh savings). Moreover, we found a main effect of energy-saving attitude on the behavioral cost level of chosen measures, suggesting that more experienced users could be presented measures with relatively high behavioral costs compared to their attitudinal strength, while those with a weaker attitude should be presented relatively easy measures.

4.3 Choice Behavior for Individual Measures (RQ2)

We further investigated whether social norms affected choices for individual measures, compared to a non-social baseline (RQ2). The results of this analysis are reported in Table 3. Model 2 addresses whether measures with higher norm scores were more likely to be selected in our interface, which we expected. It examined whether the presented

Table 3. Three multilevel logistic regression analyses predicting the choice probability per measure, clustered at the user level. The 'Main Effects' examine whether *any* measure was more likely to be chosen in a norm condition, compared to the baseline, effectively comparing the total number of measures chosen (Model 1). The 'Within List' effects examined for each measure whether the likelihood it was chosen was affected by the presented norm score (Models 2+3) and its perceived effort (Model 3), also comparing the effect in each norm condition to the baseline. Reported are odds ratios ($OR < 1$ implies a negative effect, $OR > 1$ a positive effect) and standard errors ($S.E.$). $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

	Model 1	**Model 2**	**Model 3**
	OR ($S.E.$)	OR ($S.E.$)	OR ($S.E.$)
Main Effects (RQ1)			
Global Condition	0.83 (.33)	0.82 (.35)	0.71 (.29)
Similar Condition	1.17 (.46)	1.20 (.53)	1.02 (.42)
Experienced Condition	1.02 (.39)	1.06 (.63)	0.88 (.35)
Within list (RQ2)			
Interface score		0.78 (.32)	1.19 (.56)
Score X Global		3.80 (2.22)*	2.43 (1.58)
Score X Similar		2.26 (1.27)	1.19 (.75)
Score X Experienced		2.99 (1.66)*	1.46 (.90)
Perceived effort			0.68 (.21)
Effort X Global			1.42 (.56)
Effort X Similar			1.03 (.41)
Effort X Experienced			0.70 (.28)
Score X Effort			0.15 (.14)*
Score X Effort X Global			6.29 (7.59)
Score X Effort X Similar			6.46 (7.68)
Score X Effort X Experienced			12.52 (14.93)*
Constant	-1.58 (.28)***	0.20 (.059)***	0.23 (.068)***

norm score affected the likelihood that a measure was chosen, compared per norm condition (e.g., Score X Global) to the effect in the baseline. Second, in addition to the norm score, Table 3, Model 3 considers a measure's perceived effort ('Effort') and the interaction between the two ('Score X Effort') on the likelihood that a measure is chosen.

4.3.1 Norm Percentages. Table 3, Model 2 shows, for the baseline, that depicting higher Savings Scores did not affect the probability that a measure was chosen: $OR = 0.78$, $p = 0.533$. This pointed out that higher kWh savings did not persuade users to choose a measure. Figure 9 illustrates this as well, as the proportion of chosen measures (depicted in green) did not differ across kWh Savings scores.

In line with our expectations, Table 3, Model 2 provides evidence that showing higher norm scores increased the likelihood that a measure was chosen, compared to the effect of presenting high kWh savings. The effect of Score in the 'Global' ($OR = 3.80$, $p = 0.022$) and 'Experienced' conditions ($OR = 2.99$, $p = 0.049$) increased the likelihood that a measure was chosen, while no such effect was found for Score in the 'Similar' condition ($OR = 2.26$, $p = 0.15$), all compared to effect in the baseline. This suggested that high 'Global' and 'Experienced' norm percentages were more likely to persuade a user to choose measure, compared to presenting a measure's kWh savings. While we could not make such assertions for 'Similar' norms, the effect points in a similar direction.

We explored what norm percentage cohorts were more likely to persuade a user to choose a measure. Figure 10 shows two levels at which the proportion of chosen measures seemed to increase: around 20% (an increase from 0.07 to

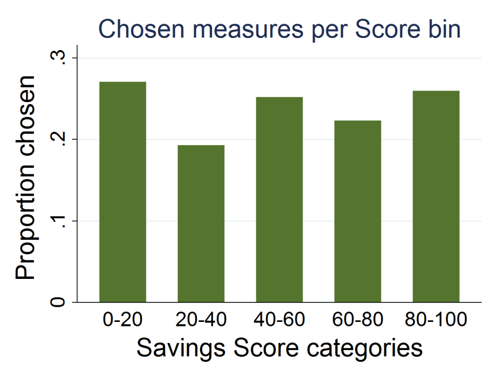

Fig. 9. Depicted in green are the proportions of chosen measures in the baseline condition (among those presented), per Savings Score category.

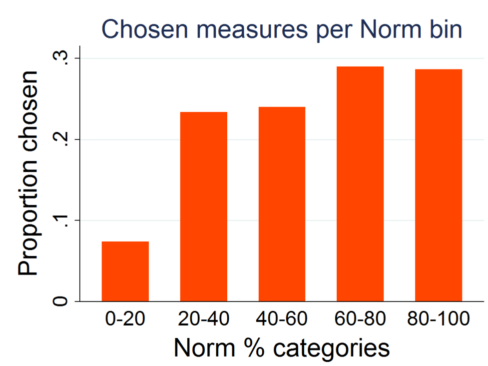

Fig. 10. Proportion of chosen measures in the normative conditions, per norm percentage category. Data is averaged across all categories (Global, Similar, and Experienced).

0.23) and at 60% (an increase from 0.24 to 0.29). This suggested that normative messages below 20% discouraged users from choosing them, while measures that depicted norm percentages over 60% seemed the most likely to be chosen, which was most common among 'Experienced' norms.

4.3.2 Perceived Effort. To expand the results found in Table 3 Model 2, we also analyzed a model that included the presented norm score, a measure's perceived effort, and the interaction between the two. We expected that presenting high norm scores in our interface would overcome a measure's perceived effort level, increasing the likelihood that a measure was chosen.

Table 3, Model 3 reports the results. We observed no significant within-list effects for score, nor were significant effects observed for perceived effort between the norm conditions and the baseline ($p > 0.05$ for all effects). Model 3 did reveal interesting interaction effects. We found that an interaction between score and effort in the baseline negatively affected the likelihood that a measure was chosen: $OR = 0.15$, $p = 0.041$. This suggested that measures with high kWh savings were more likely to be chosen if they had low levels of effort, while this likelihood decreased if a measure had high levels of perceived effort. The latter, high effort and high savings, was however far more common among the set of energy-saving measures used in this study.

We examined further interaction effects between score and perceived effort for the normative conditions. Table 3, Model 3 reveals a non-significant difference in choice likelihood between the baseline and both the 'Global' and 'Similar' conditions ($OR \approx 6.3$, $p > 0.05$). Even though this showed that high norm scores did not significantly persuade users to choose more effortful measures, the OR was comparable to that of the baseline, suggesting that the negative baseline effect was reduced (i.e. users only choosing measures with high kWh savings if effort was low). Furthermore, we did find a significant increase in the choice likelihood for the 'Experienced' condition: $OR = 12.52$, $p = 0.034$, suggesting that high norm percentages, explained in terms of experienced peers, increased the probability that an effortful measure was chosen, rather than those with low effort. The odds ratio of this positive effect was two times larger than the negative effect in the baseline, showing that higher 'Experienced' norm scores could persuade users to choose effortful measures.

4.3.3 Conclusion. We examined whether depicting social norms in an energy recommender interface also what measures were chosen from a list of recommendations (RQ2). We found that recommendations were more likely to be chosen if they were presented alongside high norm scores or percentages, suggesting that they stand out from a list of tailored measures. In particular, it seemed that presenting 'Experienced' norms alongside effortful measures could increase the likelihood that they were chosen, while explaining measures in terms of their kWh savings led users to choose relatively low-effort measures. Although the previous section did not report any changes in the overall choice behavior, the current section showed that social norms in a personalized context were capable of promoting specific measures in a recommendation list.

4.4 User Evaluation of the Saving Aid (RQ3)

Finally, we examined whether perceptions of the recommender system differed between conditions, and whether this affected, in turn, choice behavior and satisfaction (RQ3). We organized the objective constructs, subjective constructs, and relevant interactions into a path model using Structural Equation Modeling in MPlus [37, 43]. To do so, we first performed a confirmatory factor analysis, after which we tested a fully saturated model and performed stepwise removal of non-significant relations.

4.4.1 Confirmatory Factor Analysis. We submitted all items in our evaluation questionnaire, described in Table 4, to a confirmatory factor analysis. We had to drop perceived support for our subsequent structural equation model (SEM) analysis, as it could not be reliably discerned from the choice satisfaction aspect, violating divergent validity [37]. Both the feasibility ($\alpha = .72$) and choice satisfaction aspects ($\alpha = .87$) had an acceptable internal consistency and met the standards for convergence validity ($AVE > 0.5$), as prescribed by [37].

4.4.2 Structural Equation Model. Figure 11 depicts the final path model of Structural Equation Model. It had excellent fit statistics, indicating there was little unexplained variance: $\chi^2(63) = 71.548$, $p = 0.76$, $CFI = 1.000$, $TLI = 1.006$, $RMSEA = 0.000$, 90%-CI: [0.000,0.028]. The final model met the prescribed guidelines for discriminant validity [37].

Table 4. Results of the confirmatory factor analysis on user experience. Items without loading were removed from the final model. Perceived Support was omitted due to high cross-loadings with Choice Satisfaction [37]. AVE denotes the average variance explained by an aspect, α represents Cronbach's Alpha.

Aspect	Item	Loading
Perceived Feasibility	The recommended measures are hard to perform.	−.808
	I do not have the possibility to perform the recommended measures.	−.566
$AVE = .54$	The recommended measures are applicable in my home environment.	
$\alpha = .72$	It takes little effort to perform the recommended measures.	.738
Perceived Support	I want to use the Saving Aid more often.	
	The Saving Aid was useless.	
	I make better choices using the Saving Aid.	
	The Saving Aid has made me more aware of my energy behavior.	
	I would recommend the Saving Aid to others.	
	The Saving Aid was useful to find appropriate measures.	
	I could easily choose measures with the Saving Aid.	
Choice Satisfaction	I like the measure I have chosen.	.690
	I think I chose the best measures from the list.	.659
$AVE = .67$	I would enjoy performing the chosen measures.	.746
$\alpha = .87$	The chose measures exactly fit my preferences.	.820

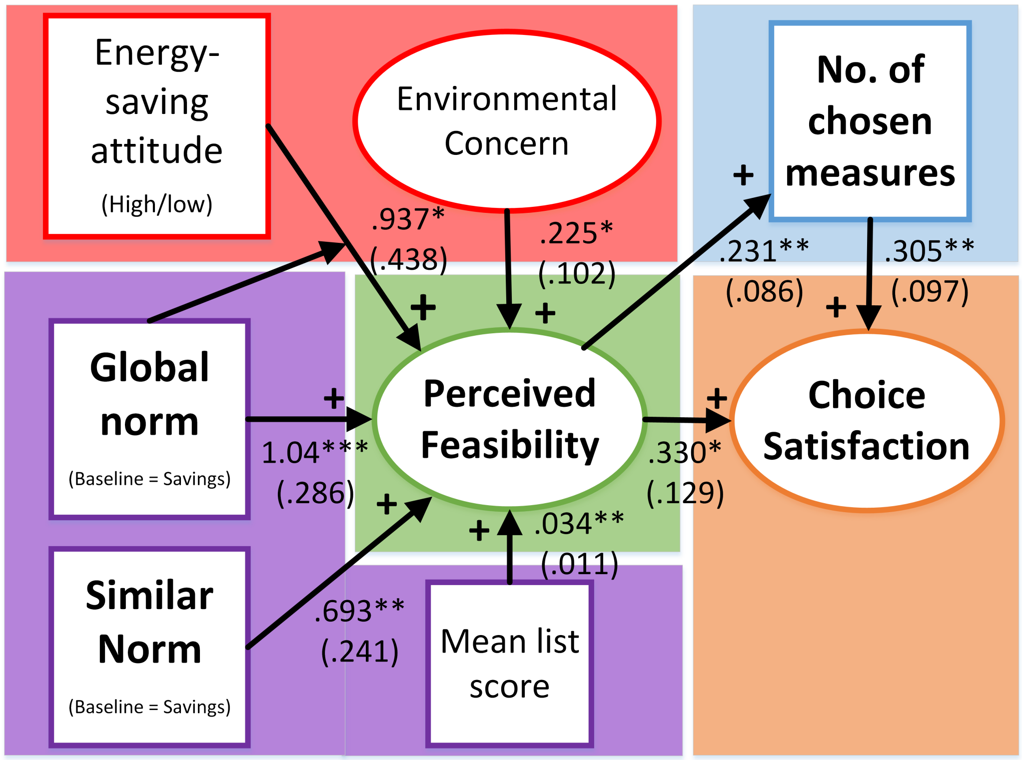

Fig. 11. Structural Equation Model (SEM). Numbers on the arrows represent the β-coefficients, standard errors are between brackets. Effects between subjective constructs are standardized and resemble correlations. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

4.4.3 Perceived Feasibility and Choice Behavior. We expected that the different social norm interfaces would be perceived as more feasible than the kWh savings baseline. Figure 4 partially confirms this, as both the 'Global' norm ($\beta = 1.04$) and 'Similar' norm conditions ($\beta = 0.693$) positively affected a recommendation list's perceived feasibility compared to the baseline. Moreover, the interaction between the user's energy-saving attitude and a 'Global' norm on a user's perceived feasibility ($\beta = 0.937$) matched our expectations that the user's attitudinal strength would determine how the 'Global' norms would be evaluated.

In contrast, no such effect on feasibility was observed for 'Experienced' norms ($\beta = -.286$, $p = .310$; not depicted in Figure 11). Although the results in Table 3 suggested that 'Experienced' norms convinced users to choose effortful measures, they did not increase the perceived feasibility of the presented measures altogether.

In turn, Figure 11 shows that perceived feasibility positively affected the number of measures chosen by a user. A bootstrapped test of indirect effects from the 'Global' condition towards the number of chosen measures was significant: $\beta = 0.240$, 95%-*CI*: $[0.020, 0.460]$, $p = 0.033$, while the effect from the 'Similar' condition to the number of chosen measures was not significant: $\beta = 0.160$, 95%-*CI*: $[-0.009, 0.329]$, $p = 0.064$.

4.4.4 Choice Satisfaction. The SEM model in Figure 11 shows two positive effects on choice satisfaction: perceived feasibility ($\beta = 0.330$) and the number of chosen measures ($\beta = 0.305$). This suggested that choosing feasible measures, as well as more measures positively affected how they were evaluated. In terms of indirect effects, the path from 'Global' norms to choice satisfaction ($\beta = 0.417$, 95%-CI: [0.041, 0.793], $p = 0.030$), as well as the path from the 'Similar' norm condition ($\beta = 0.278$, 95%-CI: [0.034,0.521], $p = 0.026$) was significant, showing that the positive effects of these norms on satisfaction were mediated by feasibility.

Besides the interface effects, we found that the mean presented recommendation list score positively affected the list's feasibility perception ($\beta = 0.034$). This confirmed that higher scores, regardless of the source (i.e., kWh savings or norm), were related to higher levels of feasibility. Figure 11 also depicts that a user's environmental concern positively affected perceived feasibility ($\beta = 0.225$), showing that users who attributed greater concern towards their role in protecting the environment, indicated that the recommended measures were more feasible to perform.

4.4.5 Conclusion. We examined whether user perceptions of recommendations were affected by the depiction of social norms (RQ3). The path model showed that 'Global' and 'Similar' norms increased the perceived feasibility of recommendations, relative to the Savings baseline, while no such effect was found for 'Experienced'. This suggested that the effectiveness of descriptive norms did not simply boil down to high percentages, but that the advice source played a role, possibly through similarity rather than expertise.

Furthermore, our path model shows that higher levels of feasibility increased both the number of chosen measures, as well as choice satisfaction. Our tests of indirect effects confirmed that most of the paths from the 'Global' and 'Similar' norms to both the 'number of chosen measures' and choice satisfaction were mediated by feasibility. This indicated that explaining energy-saving recommendations in terms of normative messages affected how users perceived them (i.e. making them seem more feasible to perform) and, in turn, increased the number of measures that users had selected (i.e., a proxy for behavioral intention), as well as their choice satisfaction levels.

5 DISCUSSION

We have investigated to what extent different social norm nudges affect choice behavior in an energy recommender system. We have translated the findings of a well-known social psychology study by Goldstein et al. [25], which used descriptive norms to promote towel re-use, to an HCI context. Specifically, we have investigated whether the merits of descriptive norms still apply in a choice environment where a more diverse, yet personalized set of energy-saving measures is presented.

Our results show that normative explanations affect user decision-making *within* lists of tailored recommendations, in the context of energy conservation. Although normative messages have not lead to more energy-efficient choices in total, as investigated in [RQ1], we find that presenting normative explanations, as well as comparatively high norm scores can boost the adoption of specific energy-saving measures (RQ2). Moreover, the depiction of social norms also positively affects a user's evaluation of a recommender system (RQ3), in terms of the perceived feasibility and choice satisfaction. These findings underline that social proof [21] and implied norms can act as persuasive nudges, both through majority preferences (in this study: the presented norm percentages), as well as specific peers (in this study: the different norm sources) – even in the context of personalized advice. Such norms are found to be more effective than presenting additional information about key attributes of the recommended items (in this study: kWh savings). Moreover, social norms also affect what types of measures are chosen, in terms of relevant energy-saving attributes: a measure's behavioral costs (i.e., execution difficulty) and its perceived effort.

More generally, we show that person-dependent nudges (in our case: social norms) are beneficial to contexts where the advice itself is also personalized. For nudging researchers (i.e., behavioral economists [57]), this implies that it could be fruitful to move beyond one-size-fits-all persuasion, for its effectiveness might be lost when the content of an intervention is aligned with a user's preferences [35, 54]. For recommender system scholars, the merit of our study lies in the effectiveness of personalized nudges in recommender interfaces to shift user preferences. This is particularly important to recommender domains where self-actualization and behavioral goals play a role [20], as users in those domains often consider what their peers are doing [13, 39]. Earlier approaches in behavioral recommender domains (i.e., energy and health) show that is hard for users to seek behavioral change, as most algorithms reinforce their current preferences [48, 53]. The use of nudges, such as social norms, could alleviate those issues.

5.1 Influence of Study Design

We further discuss the study results in more detail, by first examining the influence of our study design. Contrasting with the findings in Goldstein et al. [25], the use of descriptive norms did not lead to an overall increase in the total number of chosen energy-saving measures, compared to a baseline that emphasized kWh savings (RQ1). We discuss a number of possible causes for this different outcome through our study design.

For one, the decision contexts are different, in terms of the number of presented norms and measures. The observed behavior in Goldstein et al. [25] is rather straightforward, as it only promotes towel re-use by means of a door hanger in a hotel room. In contrast, our study comprises a set of attitude-tailored energy-saving measures in a web shop study, presenting multiple descriptive norms simultaneously. While the change from a hotel room to a web-based interface might not have impacted the results, the fact that our recommender has simultaneously presented multiple measures with different norm percentages could have led users to make more comparative judgments. For example, users could have been influenced by the presented interface score rather than the norm source. This is supported by our findings that, in the 'Global' and 'Similar' norm conditions, measures were more likely to be chosen from recommendation lists if they had a comparatively high score, while revealing no differences in the main effect between norm conditions.

The results in our study clearly show that, contrary to what has been suggested in a meta-review [1], simply generalizing the effectiveness of descriptive norms for a simple, one-time behavior (e.g., towel re-use) to all types of measures is not a representative statement. Hence, potential adopters of a norm-based approach should seriously consider the nature of the behavior that is being promoted, whether it is energy conservation or other behaviors in the recommender domain that involve effort, such as healthy eating [48, 56]. While the effects of descriptive norms have been rather consistent in both movie and social recommender systems [12, 26, 49], these are domains distinct from energy and health, as their behaviors (e.g., clicks) face few behavioral thresholds (e.g., no financial costs).

Finally, we wish to emphasize that our study has used a strict baseline. By framing treatment energy-saving measures in terms of their energy-saving measures, we have attempted to reflect the study design of Goldstein et al. [25], rather than adhering to the more traditional social psychology study design that employs a no-treatment control group [1]. Hence, we have evaluated the effectiveness of our normative manipulations critically, which is consistent with recommender system research, where novel algorithms and interfaces are benchmarked against commercial applications or state-of-the-art technology [37].

5.2 Differences in Norms

Besides differences in choice behavior as a result of the presented norm percentages, we also discuss to what extent the type of descriptive norms played a role. Goldstein et al. [25] propose that 'provincial' norms (i.e. 'local' norms as 'Similar' and 'Experienced') are more effective in changing user behavior than the more 'Global' norms, for they share very specific and context-rich characteristics with the recipient of an environmental appeal. While each normative condition in this study leverages some similarity with the user, the 'Global' and 'Similar' norms are arguably the most ‘provincial’, as they specifically emphasize the similarities with the user. In contrast, the ‘Experienced’ norm also emphasizes a difference by pointing out that other customers ‘performed more measures than [the user]’.

Nonetheless, our path model shows effects consistent with a context-rich, provincial norm explanation. The ‘Global’ and ‘Similar’ norm conditions produce higher levels of perceived feasibility compared to the Savings baseline, while no such effect is found for the ‘Experienced’ condition. That analysis suggests that the increase in feasibility can be attributed to the norm source rather than the score, as the ‘Experienced’ norm presented the highest percentages. This would suggest that users are influenced by the principle of ‘similar others are doing it, therefore I can do this too’, a more general heuristic for choice.

The relevance of this finding on feasibility lies in the indirect effects of ‘Global’ and ‘Similar’ norms on choice satisfaction, which are mediated by feasibility. The use of such normative explanations has not only increased the perceived feasibility of the recommended measures, but has also led to higher levels of choice satisfaction, compared to users in the kWh Savings baseline. This increase in choice satisfaction might be important to ultimately spur behavioral change, as it can, in turn, persuade users to re-use a recommender system at a later stage [37]. Moreover, previous studies have shown that higher levels of choice satisfaction lead to a higher likelihood that users actually implement chosen measures [54].

Although our ‘Experienced’ norms have not increased feasibility compared to the kWh savings baseline, our analysis on choice in recommendation lists (cf. Table 3) reveals that it can boost preferences for high-effort measures. Unlike in the baseline condition, where only low-effort measures are more likely to be chosen for high (kWh) Saving Scores, higher ‘Experienced’ norm percentages seem to boost the selection of high-effort measures. This finding nicely shows that *personalized* nudges rather than one-size-fits-all persuasion can improve the effectiveness of a recommender system, as the persuasiveness of the ‘Experienced’ norm is specific to effortful measures.

5.3 Limitations

There might be some concerns about the use of self-reported behavior and choice as our behavioral indicators. Although we are aware that self-report can be an inaccurate measurement method, it should have not had a large impact on the study’s results, for we have only examined differences in choice behavior between randomly assigned conditions.

Furthermore, it is possible that some users have also chosen measures that they already performed. Particularly for the so-called ‘curtailment’ measures (i.e., highly frequent behaviors [56] with no or little investment costs), users may have chosen a certain measure to indicate that they want to ‘keep doing’ something, such as turning off the lights after leaving a room. However, since we have mostly made comparison across condition, we expect the impact of this issue to be small. If any, the amount of chosen kWh savings per measure could even be larger, for curtailment measures tend to yield relatively low energy savings (cf. Table 5 and [17, 56]).

Finally, the used sample might not be representative for the broader population. The sample comprises energy supplier customers with a smart thermostat, a group that happens to be composed of mostly males with relatively strong energy-saving attitudes. Although this might limit the extent to which our results could be generalized to the broader population, our randomly-assigned between-subject research design reduces the impact of using such a specific sample population. Hence, we have examined the effectiveness of different personalized normative interfaces. Nonetheless, it would be useful to replicate this research among a more representative population, to check whether our findings on specific normative explanations still apply. Since the study has been conducted in the Netherlands, it is possible that some of the energy-saving determinants (e.g., a measure’s perceived effort) show small variations across different countries [56, 60].

5.4 Future Directions

5.4.1 Mitigating Climate Change. Our findings show how current approaches to household energy conversation promotion could be improved. Although some personalized campaigns are in place [56], relatively many are still one-size-fits-all. For example, whereas many government information pages describe best practices, they could also implement a tool similar to the 'Saving Aid' (cf. Figure 3), for which personalization brings forth relatively little costs in terms of time. Furthermore, other interventions only focus on the use of social norms for a single metric or behavior [25, 51, 55], for example by comparing the overall energy use of one household with its neighbors [4].

We expect that a combination of personalization and social norm interventions is likely improve the overall effectiveness of energy-saving promotion, in terms of actual behavioral change. However, beyond energy-efficient choices, it requires further research to understand to what extent it actually lowers the threshold of 'getting started'. For example, some people defer from making energy-saving choices, for they believe that they are the only engaging in it – the so-called 'sucker effect' [27]. Follow-up research should make clear to what extent observing the behavior of others alongside personalized advice, also in a recommender interface such as ours, could help to lower the thresholds to actual energy-saving behavior. Previous research has provided evidence that users who evaluate a recommender interface positively, are also much more likely to implement chosen measures [54].

5.4.2 Applications of Social Norms in Other Recommender Domains. The current paper has employed social norms to support behavioral change for a user's 'better self' [20]. While such persuasion techniques may be somewhat paternalistic, it could be argued for the energy domain that the user ultimately benefits, due to a lower energy bill and a positive environmental impact. However, the implications of this study also reach beyond the energy domain, for it has been among the first to apply nudges in a personalized advice interface. For example, in food, most recommender systems only focus on a user's current eating habits [23, 42], while a user might have certain eating goals that can be attained more easily by the use of social nudges [3, 52].

There are also domains in which norms can easily backfire or lead to arguably unethical situations. For instance, in the context of news recommender systems [5], normative explanations of news articles could reinforce partisan readership, if a user observes fellow democrats or conservatives consuming certain articles. Designers of recommender interfaces should always consider whether it could be harmful to a user if she 'follows the herd', and to what extent reinforcing such behavior through persuasive messaging exacerbates this.

Nonetheless, we wish to repeat that our study has focused on nudging within a personalized list of recommendations. Since the presented items already fit a user, this might mitigate possible 'herd behavior'. Moreover, the recommendation algorithm used in this paper (i.e., based on the Rasch model) is less biased towards popular items [52], for it focuses on the relation between the user and an item, based on its execution difficulty (i.e., behavioral costs) and novelty (i.e., execution probability) [56]. We think that this study can serve as a starting for various recommendation interfaces, in which social explanations are presented alongside personalized items.

ACKNOWLEDGMENTS

We thank Eneco for their collaboration. This work was made possible by the Research Talent Program with project number 406-14-088, which is financed by the Netherlands Organization for Scientific Research (NWO). The paper was finalized with support of the Niels Stensen Fellowship.

REFERENCES

[1] Wokje Abrahamse and Linda Steg. 2013. Social influence approaches to encourage resource conservation: A meta-analysis. *Global Environmental Change* 23, 6 (Dec 2013), 1773–1785. https://doi.org/10.1016/j.gloenvcha.2013.07.029

[2] Wokje Abrahamse, Linda Steg, Charles Vlek, and Talib Rothengatter. 2005. A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology* 25, 3 (Sep 2005), 273–291. https://doi.org/10.1016/j.jenvp.2005.08.002

[3] Silvio Aldrovandi, Gordon D. A. Brown, and Alex M. Wood. 2015. Social norms and rank-based nudging: Changing willingness to pay for healthy food. *Journal of Experimental Psychology: Applied* 21, 3 (2015), 242–254. https://doi.org/10.1037/xap0000048

[4] Hunt Allcott. 2011. Social norms and energy conservation. *Journal of public Economics* 95, 9-10 (2011), 1082–1095.

[5] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (Jun 2015), 1130–1132. https://doi.org/10.1126/science.aaa1160

[6] Mary L. Barreto, Agnieszka Szóstek, Evangelos Karapanos, Nuno J. Nunes, Lucas Pereira, and Filipe Quintal. 2014. Understanding families' motivations for sustainable behaviors. *Computers in Human Behavior* 40 (Nov 2014), 6–15. https://doi.org/10.1016/j.chb.2014.07.042

[7] Michael H. Birnbaum and Steven E. Stegner. 1979. Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology* 37, 1 (1979), 48–74. https://doi.org/10.1037/0022-3514.37.1.48

[8] Silvia Bonaccio and Reeshad S. Dalal. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101, 2 (Nov 2006), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

[9] Trevor G. Bond and Christine M. Fox. 2006. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences, Second Edition*. Psychology Press.

[10] P. Bonhard and M. A. Sasse. 2006. 'Knowing me, knowing you' — Using profiles and social networking to improve recommender systems. *BT Technology Journal* 24, 3 (Jul 2006), 84–98. https://doi.org/10.1007/s10550-006-0080-3

[11] Hilary S. Boudet, June A. Flora, and K. Carrie Armel. 2016. Clustering household energy-saving behaviours by behavioural attribute. *Energy Policy* 92 (May 2016), 444–454. https://doi.org/10.1016/j.enpol.2016.02.033

[12] Chien Chin Chen, Shun-Yuan Shih, and Meng Lee. 2016. Who should you follow? Combining learning to rank with social influence for informative friend recommendation. *Decision Support Systems* 90 (Oct 2016), 33–45. https://doi.org/10.1016/j.dss.2016.06.017

[13] Robert B. Cialdini and Noah J. Goldstein. 2004. Social Influence: Compliance and Conformity. *Annual Review of Psychology* 55, 1 (2004), 591–621. https://doi.org/10.1146/annurev.psych.55.090902.142015

[14] Robert B. Cialdini and Melanie R. Trost. 1998. *Social influence: Social norms, conformity and compliance.* McGraw-Hill, 151–192.

[15] Christophe Demarque, Laetitia Charalambides, Denis J. Hilton, and Laurent Waroquier. 2015. Nudging sustainable consumption: The use of descriptive norms to promote a minority behavior in a realistic online shopping environment. *Journal of Environmental Psychology* 43 (Sep 2015), 166–174. https://doi.org/10.1016/j.jenvp.2015.06.008

[16] Christian Dienes. 2015. Actions and intentions to pay for climate change mitigation: Environmental concern and the role of economic factors. *Ecological Economics* 109 (2015), 122–129.

[17] Thomas Dietz, Gerald T. Gardner, Jonathan Gilligan, Paul C. Stern, and Michael P. Vandenbergh. 2009. Household actions can provide a behavioral wedge to rapidly reduce US carbon emissions. *Proceedings of the National Academy of Sciences* 106, 44 (March 2009), 18452–18456. https://doi.org/10.1073/pnas.0908738106

[18] Riley E. Dunlap, Kent D. Van Liere, Angela G. Mertig, and Robert Emmet Jones. 2000. New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale. *Journal of social issues* 56, 3 (2000), 425–442.

[19] Heleen Ekker. 2020. Aantal huizen met zonnepanelen tikt 1 miljoen aan. https://nos.nl/artikel/2345132-aantal-huizen-met-zonnepanelen-tikt-1-miljoen-aan.html Last accessed 7 January 2021.

[20] Michael D. Ekstrand and Martijn C. Willemsen. 2016. Behaviorism is Not Enough: Better Recommendations Through Listening to Users. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, 221–224. https://doi.org/10.1145/2959100.2959179

[21] Leon Festinger. 1954. A theory of social comparison processes. *Human relations* 7, 2 (1954), 117–140.

[22] Ferdinando Fornara, Giuseppe Carrus, Paola Passafaro, and Mirilia Bonnes. 2011. Distinguishing the sources of normative influence on proenvironmental behaviors: The role of local norms in household waste recycling. *Group Processes & Intergroup Relations* 14, 5 (Sep 2011), 623–635. https://doi.org/10.1177/1368430211408149

[23] Jill Freyne and Shlomo Berkovsky. 2010. Intelligent food planning: personalized recipe recommendation. In *Proceedings of the 15th international conference on Intelligent user interfaces*. 321–324.

[24] Jon Froehlich, Leah Findlater, and James Landay. 2010. The Design of Eco-feedback Technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, 1999–2008. https://doi.org/10.1145/1753326.1753629

[25] Noah J. Goldstein, Robert B. Cialdini, and Vladas Griskevicius. 2008. A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research* 35, 3 (Oct 2008), 472–482. https://doi.org/10.1086/588568

[26] Ido Guy. 2015. *Social Recommender Systems*. Springer US, 511–543. https://doi.org/10.1007/978-1-4899-7637-6_15

[27] Michel J. J. Handgraaf, Margriet A. Van Lidth de Jeude, and Kirstin C. Appelt. 2013. Public praise vs. private pay: Effects of rewards on energy conservation in the workplace. *Ecological Economics* 86 (Feb 2013), 86–92. https://doi.org/10.1016/j.ecolecon.2012.11.008

[28] Fritz Heider. 1958. *The Psychology of Interpersonal Relations*. Psychology Press.

[29] Anthony Jameson. 2013. Choice Architecture for Human-Computer Interaction. *Foundations and Trends® in Human–Computer Interaction* 7, 1–2 (2013), 1–235. https://doi.org/10.1561/1100000028

[30] Mathias Jesse and Dietmar Jannach. 2021. Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports* 3 (2021), 100052.

[31] Eric J. Johnson and Daniel Goldstein. 2003. Do Defaults Save Lives? *Science* 302, 5649 (Nov 2003), 1338–1339. https://doi.org/10.1126/science.1091721

[32] Helmut Jungermann and Katrin Fischer. 2005. Using expertise and experience for giving and taking advice. *The routines of decision making* 1 (2005), 157–173.

[33] Florian G. Kaiser, Katarzyna Byrka, and Terry Hartig. 2010. Reviving Campbell's Paradigm for Attitude Research. *Personality and Social Psychology Review* 14, 4 (Jan 2010), 351–367. https://doi.org/10.1177/1088868310366452

[34] Florian G. Kaiser and Carmen Keller. 2001. Disclosing situational constraints to ecological behavior: A confirmatory application of the mixed Rasch model. *European Journal of Psychological Assessment* 17, 3 (2001), 212–221. https://doi.org/10.1027//1015-5759.17.3.212

[35] Maurits Kaptein, Panos Markopoulos, Boris de Ruyter, and Emile Aarts. 2015. Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles. *International Journal of Human-Computer Studies* 77 (May 2015), 38–51. https://doi.org/10.1016/j.ijhcs.2015.01.004

[36] BP Knijnenburg, MC Willemsen, and R Broeders. 2014. Smart sustainability through system satisfaction: tailored preference elicitation for energy-saving recommenders. In *20th Americas Conference on Information Systems (AMCIS 2014), August 7-9, 2014, Savannah, Georgia, United States*. AIS/ICIS, 1–15.

[37] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. *Evaluating recommender systems with user experiments.* Springer, 309–352. http://link.springer.com/chapter/10.1007/978-1-4899-7637-6_9

[38] Robert E. Kraut, Paul Resnick, Sara Kiesler, Moira Burke, Yan Chen, Niki Kittur, Joseph Konstan, Yuqing Ren, and John Riedl. 2012. *Building successful online communities: Evidence-based social design.* Mit Press. http://books.google.nl/books?hl=nl&lr=&id=lIvBMYVxWJYC&oi=fnd&pg=PR7&dq=Designing+from+theory:+Using+the+social+sciences+as+the+basis+for+building+online+communities.&ots=zXI_dmk7Oz&sig=OfmPw3lvG5uhQv3XUNC5ZOk2Hlw

[39] Jennifer Mankoff, Deanna Matthews, Susan R. Fussell, and Michael Johnson. 2007. Leveraging Social Networks To Motivate Individuals to Reduce Their Ecological Footprints. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS '07)*. IEEE Computer Society, 87–. https://doi.org/10.1109/HICSS.2007.325

[40] Sean M. McNee, John Riedl, and Joseph A. Konstan. 2006. Being Accurate is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, 1097–1101. https://doi.org/10.1145/1125451.1125659

[41] Tobias Mirsch, Christiane Lehrer, and Reinhard Jung. 2017. Digital Nudging: Altering User Behavior in Digital Environments. In *Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017)*. Association for Computing Machinery, New York, NY, USA, 634–648. https://wi2017.ch/de/proceedings

[42] Cataldo Musto, Christoph Trattner, Alain Starke, and Giovanni Semeraro. 2020. Towards a knowledge-aware food recommender system exploiting holistic user models. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 333–337.

[43] L. K. Muthén and B. Muthén. 2015. Mplus. *The comprehensive modelling program for applied researchers: user's guide* 5 (2015).

[44] Jessica M. Nolan, P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2008. Normative social influence is underdetected. *Personality and social psychology bulletin* 34, 7 (2008), 913–923.

[45] Petromil Petkov, Suparna Goswami, Felix Köbler, and Helmut Krcmar. 2012. Personalised Eco-feedback As a Design Technique for Motivating Energy Saving Behaviour at Home. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI '12)*. ACM, 587–596. https://doi.org/10.1145/2399016.2399106

[46] Wouter Poortinga, Linda Steg, Charles Vlek, and Gerwin Wiersma. 2003. Household preferences for energy-saving measures: A conjoint analysis. *Journal of Economic Psychology* 24, 1 (2003), 49–64.

[47] P Wesley Schultz. 1999. Changing behavior with normative feedback interventions: A field experiment on curbside recycling. *Basic and applied social psychology* 21, 1 (1999), 25–36.

[48] Hanna Schäfer and Martijn C. Willemsen. 2019. Rasch-based tailored goals for nutrition assistance systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces - IUI '19*. ACM Press, 18–29. https://doi.org/10.1145/3301275.3302298

[49] Amit Sharma and Dan Cosley. 2013. Do social explanations work?: studying and modeling the effects of social explanations in recommender systems. In *Proceedings of the 22nd international conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1133–1144. http://dl.acm.org/citation.cfm?id=2488487

[50] Janet A. Sniezek and Lyn M. Van Swol. 2001. Trust, Confidence, and Expertise in a Judge-Advisor System. *Organizational Behavior and Human Decision Processes* 84, 2 (Mar 2001), 288–307. https://doi.org/10.1006/obhd.2000.2926

[51] Henk Staats, Paul Harland, and Henk AM Wilke. 2004. Effecting durable change a team approach to improve environmental behavior in the household. *Environment and Behavior* 36, 3 (2004), 341–367.

[52] Alain Starke. 2019. RecSys Challenges in achieving sustainable eating habits. In *HealthRecSys'19: Proceedings of the 4th International Workshop on Health Recommender Systems, co-located with 13th ACM Conference on Recommender Systems (HealthRecSys'19)*. Association for Computing Machinery, New York, NY, USA, 29–30.

[53] A.D. Starke. 2019. *Supporting energy-efficient choices using Rasch-based recommender interfaces*. Ph.D. Dissertation. Eindhoven University of Technology. https://research.tue.nl/nl/publications/supporting-energy-efficient-choices-using-rasch-based-recommender
[54] Alain Starke, Martijn Willemsen, and Chris Snijders. 2017. Effective User Interface Designs to Increase Energy-efficient Behavior in a Rasch-based Energy Recommender System. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17)*. ACM, 65–73. https://doi.org/10.1145/3109859.3109902
[55] Alain Starke, Martijn C. Willemsen, and Chris Snijders. 2015. Saving Energy in 1-D: Tailoring Energy-saving Advice Using a Rasch-based Energy Recommender System.. In *DMRS*. CEUR, 5–8. https://pdfs.semanticscholar.org/df2c/10101581ccd33d91890ec39c048786f89a0a.pdf
[56] Alain D. Starke, Martijn C. Willemsen, and Chris C. P. Snijders. 2020. Beyond "one-size-fits-all" platforms: Applying Campbell's paradigm to test personalized energy advice in the Netherlands. *Energy Research & Social Science* 59 (Jan 2020), 101311. https://doi.org/10.1016/j.erss.2019.101311
[57] Richard H. Thaler and Cass R. Sunstein. 2008. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
[58] Nava Tintarev and Judith Masthoff. 2012. Evaluating the Effectiveness of Explanations for Recommender Systems. *User Modeling and User-Adapted Interaction* 22, 4–5 (Oct 2012), 399–439. https://doi.org/10.1007/s11257-011-9117-5
[59] Sabina Tomkins, Steven Isley, and Lise Getoor. 2018. Sustainability at Scale: Towards Bridging the Intention-Behavior Gap with Sustainable Recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems*. Association for Computing Machinery, New York, NY, USA, 214–218.
[60] Jan Urban and Milan Ščasný. 2016. Structure of Domestic Energy Saving: How Many Dimensions? *Environment and Behavior* 48, 3 (Apr 2016), 454–481. https://doi.org/10.1177/0013916514547081
[61] Theo M. M. Verhallen and Rik G. M. Pieters. 1984. Attitude theory and behavioral costs. *Journal of Economic Psychology* 5, 3 (Sep 1984), 223–249. https://doi.org/10.1016/0167-4870(84)90024-2
[62] Ilan Yaniv, Shoham Choshen-Hillel, and Maxim Milyavsky. 2011. Receiving advice on matters of taste: Similarity, majority influence, and taste discrimination. *Organizational Behavior and Human Decision Processes* 115, 1 (May 2011), 111–120. https://doi.org/10.1016/j.obhdp.2010.11.006
[63] Haiyi Zhu and Bernardo A. Huberman. 2014. To Switch or Not To Switch: Understanding Social Influence in Online Choices. *American Behavioral Scientist* 58, 10 (Sep 2014), 1329–1344. https://doi.org/10.1177/0002764214527089

A APPENDIX: RASCH SCALE OF ENERGY-SAVING MEASURES

Table 5. Tabulation of the Rasch scale of energy-saving measures i. Described are names, behavioral cost levels (θ_i), the infit statistics (MNSQ denotes Mean Square, ZSTD the standardized mean), kWh savings per year (i.e., 'kWh'), and perceived effort (i.e., 'EF').

			Infit statistics		*Attributes*	
i	**Measure**	θ_i	**MNSQ**	**ZSTD**	**kWh**	**EF**
1	Wash only full loads of laundry	-3.09	1.07	0.31	30	1.73
2	Cook with pots & pans the same size as the heating element	-2.78	0.84	-0.44	5	1.33
3	Turn off or down heating/cooling system when going away for several days	-2.64	1.15	0.60	115	1.36
4	Repair leaky faucets	-2.53	1.01	0.12	70	2.73
5	Cool hot food before putting it in the refrigerator	-2.46	1.04	0.24	2	1.45
6	Cover pots & pans when cooking	-2.4	1.01	0.13	5	1.25
7	Refrain from installing an air conditioner unless necessary for safety	-2.38	0.85	-0.48	400	1.56
8	Do not leave your exhaust hood on when not in use	-2.22	1.08	0.41	18	1.13
9	Run a dishwasher only when full, but not overloaded	-2.17	0.92	-0.25	40	1.55
10	Take a shower instead of a bath	-2.13	1.07	0.38	400	1.54
11	Install double-pane windows	-2.00	0.90	-0.42	2500	3.44
12	Replace incandescent light bulbs with CFLs	-1.96	0.96	-0.13	140	2.13
13	Hang/air dry laundry	-1.94	1.21	1.11	290	2.22
14	Rake leaves with a garden rake instead of a leaf blower	-1.94	1.15	0.74	36	2.71
15	Switch off the coffee machine completely	-1.82	1.01	0.14	80	1.10
16	Install a laptop (instead of a desktop computer)	-1.80	1.19	1.11	40	2.21
17	Use natural light in the daytime	-1.75	0.87	-0.70	150	1.44
18	Decide what you want from the refrigerator before opening the door	-1.72	0.87	-0.86	40	1.50
19	Refrain from using a screensaver	-1.71	0.99	0.00	20	1.13
20	Wash laundry at a low(er) temperature	-1.70	0.97	-0.10	105	1.30
21	Boil only as much water as you need	-1.64	0.93	-0.35	30	1.50
22	Refrain from using an electric blanket	-1.62	1.21	1.20	60	1.63
23	Use blankets (instead of a heater)	-1.61	1.07	0.49	1000	1.58
24	Replace all incandescents with CFLs or LEDs	-1.42	1.00	0.06	30	1.85
25	Turn off monitors when not in use	-1.38	0.93	-0.52	45	1.57
26	Open blinds/curtains/drapes/shades at night when cooling your home	-1.30	1.13	1.00	1000	1.92
27	Thaw food in a refrigerator or sink	-1.28	1.00	0.05	10	1.71
28	Turn off bathroom exhaust fans 20 minutes after bathing	-1.28	0.95	-0.34	12	1.78
29	Dry only full loads of laundry	-1.23	1.20	1.45	50	1.89

	Continuation of Table 5	θ_i	**MNSQ**	**ZSTD**	**kWh**	**EF**
30	Switch off the dishwasher immediately after use	-1.08	0.97	-0.19	70	1.44
31	Install a light switch at both ends of hallways	-1.03	1.07	0.59	17	2.67
32	Air rooms for 20 minutes a day	-1.03	0.93	-0.52	250	1.64
33	Turn off air conditioners in unoccupied rooms	-1.00	0.90	-0.73	100	1.67
34	Towel/air dry hair instead of using electric hair dryer	-0.94	1.20	1.71	30	1.22
35	Iron multiple garments in a row	-0.93	0.92	-0.61	2	1.38
36	Set freezers to the warmest food-safe temperature (-18°C)	-0.91	1.18	1.58	50	1.27
37	Cook with a frypan (instead of an oven)	-0.85	1.09	0.76	250	2.07
38	Set the thermostat to 14°C at night during heating season	-0.80	0.95	-0.43	1250	1.89
39	Turn off computers when not in use	-0.75	1.01	0.09	100	1.89
40	Set refrigerators to the warmest food-save temperature (4°C)	-0.73	0.95	-0.37	50	1.00
41	Place refrigerator contents to allow for good air circulation	-0.69	0.93	-0.62	2	1.92
42	Install nylon brush seals or spring flaps on exterior door keyholes	-0.67	0.95	-0.41	25	2.22
43	Set the thermostat 1°C lower when heating your home	-0.66	1.02	0.26	1100	1.50
44	Turn off air conditioners when leaving the house	-0.66	1.17	1.62	50	1.80
45	Refrain from using portable electric heaters to heat large spaces	-0.65	1.21	1.67	800	1.33
46	Use the dishwasher's eco-program	-0.62	0.95	-0.41	85	1.60
47	Scrape food scraps off dishes prior to loading them into the dishwasher	-0.61	0.87	-1.29	2	1.64
48	Install an energy efficient washing machine	-0.45	0.99	-0.06	200	2.62
49	Stir-fry food	-0.44	1.02	0.23	20	1.60
50	Install weather strips on windows	-0.43	1.00	-0.01	460	2.63
51	Do maintenance on boiler or geyser	-0.41	0.88	-1.22	120	2.00
52	Install thermal mixer taps	-0.40	0.97	-0.37	70	2.06
53	Caulk & seal exterior walls	-0.37	0.90	-0.93	250	2.25
54	Defrost fridge/freezer	-0.36	0.87	-1.69	50	2.61
55	Make coffee without a hotplate	-0.35	0.90	-0.91	25	1.54
56	Use rechargeable batteries	-0.33	1.06	0.75	0	2.09
57	Buy an energy-efficient fridge-freezer	-0.30	0.99	-0.04	160	2.42
58	Unplug devices, appliances, and chargers when not in use.	-0.25	0.95	-0.54	25	1.76
59	Install an energy efficient dishwasher	-0.20	0.92	-0.89	70	2.33

	Continuation of Table 5	θ_i	MNSQ	ZSTD	kWh	EF
60	Trim bushes with garden shears instead of an electric trimmer	-0.15	0.96	-0.39	5	2.44
61	Get rid of a second refrigerator	-0.13	1.07	0.73	240	2.22
62	Use an energy efficient TV	-0.09	0.93	-0.72	120	2.67
63	Insulate roofs	-0.08	0.94	-0.67	8000	3.71
64	Set the thermostat to 60°C to 65°C on hot water storage systems, and 50°C on instantaneous hot water systems	-0.06	0.98	-0.16	900	1.54
65	Empty/replace vacuum cleaner filter bags regularly	-0.05	0.80	-2.39	5	2.07
66	Position refrigerators to allow for air circulation around their coils	0.02	0.94	-0.64	25	2.88
67	Insulate cavity walls	0.03	0.90	-1.08	8000	3.40
68	Enable the power management features of computers	0.04	1.00	0.05	40	1.73
69	Descale coffee machines and electric kettles	0.04	1.04	0.48	10	2.05
70	Air clothes instead of washing them	0.06	0.93	-0.89	30	1.89
71	Take advantage of the night-time tariff	0.07	0.84	-1.97	0	2.22
72	Install weather strips on doors	0.08	0.94	-0.68	120	2.00
73	Switch off the computer with a power strip	0.11	1.00	0.01	175	2.17
74	Use renewable energy	0.13	1.13	1.32	0	2.25
75	Check the pressure in your boiler	0.17	0.81	-2.07	0	1.86
76	Insulate heat ducts	0.20	0.90	-1.10	100	2.92
77	Seal any holes in insulation with low-expansion spray foam	0.21	1.04	0.45	1000	3.00
78	Reduce the duration of your showers	0.21	0.88	-1.38	185	2.33
79	Keep exhaust hood filters clean	0.24	0.78	-2.71	2	2.25
80	Replace dimmer switches	0.25	1.21	1.92	60	1.55
81	Install floor insulation	0.26	1.08	0.78	2450	3.60
82	Install an energy efficient freezer	0.26	0.95	-0.55	190	2.44
83	Install a drain waste water heat recovery system	0.26	1.14	1.60	750	2.38
84	Install energy efficient light fixtures	0.27	0.92	-1.03	50	2.31
85	Maintain clean, air-tight refrigerator door seals	0.31	1.01	0.13	25	2.29
86	Put rugs on the floor to contain heat	0.32	1.07	0.73	60	2.86
87	Install low-flow showerheads	0.36	1.03	0.38	400	2.20
88	Install smaller cisterns on toilets	0.38	0.80	-2.08	0	2.71
89	Clean floors with a broom (instead of a vacuum cleaner)	0.39	1.23	2.44	25	2.25
90	Install an energy efficient computer monitor	0.43	0.92	-0.74	20	1.79
91	Do the dishes manually	0.48	1.25	3.00	160	2.75
92	Plant deciduous trees around your home	0.50	1.06	0.59	1000	3.08
93	Insulate the attic, including the trap/access door	0.55	0.71	-3.11	6400	3.64

	Continuation of Table 5	θ_i	MNSQ	ZSTD	kWh	EF
94	Install underfloor heating (in a well-insulated house only)	0.68	1.03	0.32	640	3.87
95	Use the TV one hour less every day	0.71	1.13	1.34	75	2.00
96	Switch off the washing machine completely	0.72	1.10	0.95	95	1.78
97	Insulate ceilings	0.74	0.92	-0.78	7400	4.20
98	Install an energy-efficient dryer	0.76	0.90	-0.86	210	2.43
99	Install low-flow aerators in faucets	0.81	1.02	0.23	500	2.45
100	Vent radiators regularly	0.84	0.89	-1.03	300	2.78
101	Turn off the water while soaping up during showers	0.88	0.90	-0.84	400	2.58
102	Set your TV to energy efficient settings	0.89	0.96	-0.29	50	1.78
103	Replace your water heater if it is more than 7 years old	0.89	0.98	-0.09	800	3.15
104	Lower the boiler temperature	0.90	1.00	0.05	40	1.78
105	Insulate hot water tank if it's warm to the touch	0.91	0.95	-0.38	130	2.70
106	Tumble dry t-shirts briefly instead of ironing them	0.91	1.14	1.04	25	2.20
107	Turn off the hot water heater system when going away for a few days	0.94	0.86	-1.09	20	2.50
108	Use a tablet instead of a laptop/desktop	1.01	0.88	-0.92	80	2.00
109	Descale the washing machine	1.01	0.98	-0.19	10	2.08
110	Install a motion sensor for indoor/outdoor lights	1.04	0.90	-0.81	25	2.64
111	Switch off the computer screen when downloading	1.09	1.07	0.70	6	1.53
112	Install exterior wall insulation (house wrap)	1.29	1.16	1.05	1700	4.19
113	Switch off the refrigerator when on holiday	1.35	1.12	0.85	20	3.00
114	Install solar powered garden lights	1.36	1.00	0.07	50	1.91
115	Install solar panels	1.49	1.00	0.04	2000	3.33
116	Install an induction stove instead of a natural gas stove/range	1.51	1.10	0.74	185	3.73
117	Turn off the oven 10 minutes early	1.55	0.79	-1.65	10	1.79
118	Install door closers	1.72	1.14	0.86	220	2.67
119	Install a remote controlled thermostat	1.76	0.98	-0.07	1000	2.88
120	Use electric blankets (instead of a heater)	1.94	1.22	0.99	740	2.31
121	Install heat-resistant radiator reflectors between exterior walls and radiators	2.03	1.05	0.35	900	2.47
122	Mow your lawn with a push reel mower rather than an electric mower	2.12	1.11	0.61	50	3.00
123	Install a timer on your boiler	2.24	0.98	-0.02	500	2.50
124	Clean lights & light fittings regularly	2.40	0.83	-0.71	2	2.50
125	Set timers on space heaters	2.74	1.08	0.34	50	1.93
126	Use a pressure cooker	2.92	1.11	0.47	30	2.29

	Continuation of Table 5	θ_i	**MNSQ**	**ZSTD**	**kWh**	**EF**
127	Install a home energy usage feedback system to identify excess base-load	2.97	0.99	0.08	750	2.75
128	Install a solar boiler	3.20	1.16	0.59	1850	3.17
129	Install a centralized temperature system with zone controls & thermostats	3.33	1.05	0.26	3100	3.92
130	Install a heat pump system (when heating your home using electricity)	3.56	1.07	0.30	1100	3.00
131	Clean refrigerator coils regularly	3.66	0.75	-0.58	15	3.67
132	Use a hot-fill washing machine	3.90	1.04	0.23	40	3.13
133	Install a shower water usage feedback system	3.93	0.90	-0.10	200	2.75
134	Erect a small wind turbine	6.48	1.00	0.00	1400	4.23

End of Table