

Smart Sustainability through System Satisfaction: Tailored Preference Elicitation for Energy-saving Recommenders

Completed Research Paper

Bart P. Knijnenburg

University of California, Irvine
Donald Bren School of Information and
Computer Sciences
bart.k@uci.edu

Martijn C. Willemsen

Eindhoven University of Technology
School of Innovation Sciences
Human-Technology Interaction group
m.c.willemsen@tue.nl

Ron Broeders

Eindhoven University of Technology
School of Innovation Sciences
Human-Technology Interaction group
r.broeders@tue.nl

Abstract

People can adopt many different energy-saving measures, but how can they be encouraged to take action? Recommender systems could offer a solution, but how recommender systems are used and perceived will depend on the level of knowledge people have regarding energy-saving measures. We test an energy-saving recommender system that uses Multi-Attribute Utility Theory (MAUT) to recommend energy-saving measures to its users. Across four experiments we test nine different preference elicitation methods for this system, and demonstrate that users' system satisfaction with each of these interfaces depends on whether they are an expert on energy-saving or a novice. Moreover, we show that system satisfaction is a driver of behavioral outcomes. In effect, a suitable preference elicitation method not only makes users more satisfied with the system; it also entices them to choose more measures with higher average savings, and makes them more satisfied with their choices as well.

Keywords

Recommender systems, preference elicitation, domain knowledge, user experience, consumer behavior, energy-saving, sustainability.

Introduction

People can adopt many different energy-saving measures, but how can they be encouraged to take action? *Recommender systems* could offer a solution, but how recommender systems are used and perceived will depend on the level of knowledge people have regarding energy-saving measures. Experts will be able to effectively express their preferences regarding energy-saving measures, and prefer advanced means of controlling the system. Novices, on the other hand, need a more simplified way of expressing their preferences.

In previous work we have shown that the user interface of a recommender system interacts with the domain knowledge of the user to determine their system satisfaction and choice satisfaction. But does this also reflect on their actual choice behavior? This is an important extension to investigate if we want a recommender system to have a lasting impact on energy-saving behavior. Therefore, this is the topic of the current paper.

Specifically, we provide an overview of our efforts to support expert and novice energy-savers with a recommender system. Across four experiments (three discussed in Knijnenburg and Willemsen (2009, 2010; Knijnenburg et al. 2011), one new) we test nine different preference elicitation methods, and demonstrate that their success—in terms of system satisfaction—depends on whether the user is an expert on energy-saving or a novice. Moreover, our previous work focused primarily on the effect of different preference elicitation methods on system satisfaction and did not analyze the type and number of measures selected. The current paper makes this important extension, and shows that system satisfaction is, in turn, a driver of *behavioral outcomes*. In effect, a suitable preference elicitation method not only makes users more satisfied with the system; it also entices them to choose more measures with higher average savings, and makes them more satisfied with their choices as well. These behavioral consequences demonstrate that selecting the correct preference elicitation method should not only be a concern for system designers, but also for managers and policy makers.

Theory and hypothesis development

Recommender systems for energy-saving measures

Although people are increasingly motivated to save energy, many have only limited knowledge regarding possible measures they can take to reduce their energy consumption, and which of these measures are most effective (e.g., Benders et al. 2006; Fischer 2008; Gardner and Stern 2008). In addition, when choosing among several energy-saving measures, people have to make tradeoffs between many underlying *attributes*. This is a difficult process because the information about different measures available online is often unstructured, incomplete and hard to compare: measures may not be described on the same attributes, or information about relevant attributes may be lacking.

Existing research on the psychology of energy conservation has shown that influencing individual energy-saving behaviors is challenging, and requires a solution that is *tailored* to the individual if one is to make a lasting impact (see e.g., Abrahamse et al. 2005; Aune 2007; Benders et al. 2006; Darby 2010; Fischer 2008; Lehman and Geller 2005; Petkov et al. 2011; Steg 2008; Wilson and Dowlatabadi 2007). For this reason, recommender systems could be the key to provide good assistance in curating and choosing energy-saving measures, as they can suggest measures that are tailored to the individuals' personal preferences. Based on preference information (e.g. the user's ratings of known items, or past purchases), recommender systems present the user with personalized recommendations that are all likely to be very relevant to the user's needs (Konstan and Riedl 2012; Xiao and Benbasat 2007, 2014) and will therefore have more impact once chosen.

The specific characteristics of recommender systems give them an advantage above and beyond current behavioral interventions such as mass media campaigns and periodic feedback on energy consumption. The effects of mass-media campaigns are often limited because of their general, non-tailored nature (McKenzie-Mohr 2000). Feedback on energy consumption is important, but typically does not advise households what measures to implement in order to improve their savings bill (Fischer 2008). Recommender systems, in contrast, can give persistent, timely, and personalized advice on what energy-saving measures to implement. Many existing websites provide energy-saving advice, and some of these are able to tailor the measures to the users' living conditions. However, to our best knowledge, there exists no system that provides personalized energy-saving recommendations based on people's stated *preferences*.

In Knijnenburg and Willemsen (2009, 2010; Knijnenburg et al. 2011) we therefore developed and tested a system that uses Multi-Attribute Utility Theory (MAUT) to recommend energy-saving measures to its users. The system contains a sample of 80 efficiency and curtailment measures¹ that span a wide range on several dimensions. For each user, the utility of a certain choice option is calculated by multiplying the values of each of its attributes with the user's weight of that attribute (Guttman and Maes 1999; Häubl and Trifts 2000). The measures with the highest utility are then recommended to the user.

¹ Curtailment measures focus on influencing repetitive behaviors, e.g. switching off devices or lowering the thermostat, whereas efficiency measures are one-shot behaviors entailing buying energy-efficient equipment, e.g. better roof insulation or solar panels (Abrahamse et al. 2005; Gardner & Stern 2008).

Preference elicitation

To calculate recommendations, our system needs to somehow discover the attribute weights that optimize the recommendations for each user. The way in which the system allows users to express their preferences, the *preference elicitation method* (PE-method), has been the topic of numerous studies in the field of recommender systems (e.g., Chen and Pu 2009; Dooms et al. 2011; Gena et al. 2011; Lee and Benbasat 2011; Sparling and Sen 2011). Particularly, the PE-method seems to have an impact on users' satisfaction with the system (Chen et al. 2009; Knijnenburg et al. 2012; Sparling and Sen 2011). Below we discuss several PE-methods that can be employed in a MAUT-based recommender system.

Attribute-based PE

Attribute-based PE is the most-used PE method for MAUT-based recommenders. In this method, users directly indicate the importance of each of the attributes with which choice options are described. Häubl and Trifts (2000) were the first to explore attribute-based PE. They found that it “allow[ed] consumers to make much better decisions while expending substantially less effort” (pp. 17–18). Olson and Widing (2002) did not find a similar increase in choice accuracy, but they did find an increase in satisfaction and a decrease in decision time.

Case-based PE

Case-based PE takes an indirect approach to discover users' attribute weights, namely by analyzing the users' evaluation of exemplary choice options (McGinty and Smyth 2006; Smyth 2007). In case-based PE, users' positive (or negative) evaluation of an example is indicative of their preferences regarding its most prominent attribute, and this evaluation is therefore translated into a higher (or lower) weight for that attribute. Studies show that case-based PE results in better decisions and higher satisfaction (Chen and Pu 2009, 2011). Note however that a system with case-based PE does not show users their actual preference settings. This makes the process of indicating one's preference less transparent, and this lack of transparency may negatively impact the acceptance of the recommendations (Kramer 2007).

Needs-based PE

Needs-based PE takes the indirect approach to PE a step further: In this method, users express their preferences not in terms of product attributes, but in terms of consumer needs (Randall et al. 2007). These needs are typically better represented by latent rather than concrete product attributes (cf. *Matrix Factorization*, (Koren et al. 2009)).

Needs-based PE resembles Edwards and Fasolo's (2001) conceptualization of the consumer's decision process: the first step in the decision process is to identify the needs a product should fulfill, and the second step is to determine which attributes it should contain to fulfill each need (Butler et al. 2006, 2008). The needs-based PE method lets users control the first step while automating the second step.

Implicit PE

Implicit PE does not require users to actively express their preferences. Instead, it infers the attribute weights as a by-product of the user's browsing behavior. When a user inspects, selects, or discards a recommended item, the system uses the attribute values of this item to update the user's attribute weights accordingly. For example, if the user inspects an item with a very low value on the “continuous effort” attribute, then the system infers that this user has a high weight for the “low continuous effort” attribute. Conversely, if the user discards an item with a high level of comfort, then the system infers that this user does not have a high weight for the “comfort” attribute.

Knijnenburg et al. (2012) found that using Implicit PE can result in more accurate recommendations, but that traditional explicit methods may result in more varied recommendations. This may be because Implicit PE runs the risk of turning recommendations into a self-fulfilling prophecy via a positive reinforcement loop (Pazzani and Billsus 2002; Smyth and McClave 2001).

Hybrid PE

To overcome the problem of narrowing, but at the same time allow the system to discover users' preferences without explicitly eliciting their preferences, Knijnenburg et al. (2012) propose a hybrid PE-method that combines the two methods. In a MAUT-based system, implicit PE can be combined with attribute-based PE, giving users the convenience of automatic preference elicitation while still allowing them to monitor and control the attribute weights. Note though that the simultaneous influence of manual and automatic updates on the attribute weights adds some complexity to this method, which might decrease its understandability.

Baseline methods: Top-N and Sort

Several studies have found that consumers find systems that provide personalized recommendations more satisfying to use than non-personalized systems (Felfernig 2007; Häubl and Trifts 2000; Knijnenburg et al. 2012), and that personalized systems significantly improve users' decision quality (Diehl et al. 2003; Häubl and Trifts 2000; Häubl et al. 2004; Hostler et al. 2005; Pedersen 2000; Vijayarathy and Jones 2001). However, Chin (2001) claims that the advantage of a personalized system is not always apparent, and he suggests that the personalized system should be tested against one or more non-personalized baseline systems. In this paper we consider two such non-personalized baseline systems. The top-N baseline simply ranks the energy-saving methods in decreasing order of popularity. The Sort baseline allows users to sort the recommendations on one of the attributes.

The moderating effect of domain knowledge

Researchers have shown that users prefer a recommender system that uses a decision strategy similar to the one they would use themselves (Aksoy 2006; Al-Natour et al. 2006). Given that users' decision strategy depends on their personal characteristics (Bettman et al. 1998), users with different characteristics (who use different strategies) may prefer different PE-methods.

Domain knowledge is a personal characteristic that may significantly influence one's decision strategy. For example, compared to novices, energy-saving experts have more knowledge about the underlying attributes of energy-saving measures, and are therefore better capable of translating their needs into attribute weights (Hutton and Klein 1999), and making complex tradeoffs between them (Shanteau 1988; Xiao and Benbasat 2007). Novices on the other hand lack the knowledge required to understand the impact of the attributes (Hutton and Klein 1999), and may thus not readily know how to express their preferences in terms of product attributes (Xiao and Benbasat 2007).

Because experts and novices differ in the way they make decisions, they arguably also prefer different PE-methods. Spiekerman and Paraschiv (2002) indeed note that many existing recommender systems fail to motivate user interaction because they limit their interaction to attribute-based PE and fail to adjust to the user's level of domain knowledge. They propose a strategy to integrate different knowledge levels in the system by offering a different interface for experts and novices. Similarly, Guttman et al. (1998) suggest that "matching the system's user interface with the consumer's manner of shopping will likely result in greater customer satisfaction." (p. 153). Following this, Randall et al. (2007) demonstrate that experts are more satisfied with a system that employs a parameter-based PE-method (a variant of attribute-based PE), while novices are more satisfied with a system that uses a needs-based PE-method. Based on this we can thus hypothesize that:

H1. Users' system satisfaction depends on the interaction between their level of domain knowledge and the PE-method of the system.

Our previous work (Knijnenburg and Willemsen 2009, 2010; Knijnenburg et al. 2011) tested H1 in an exploratory fashion for eight different PE-methods. We expand upon these tests in the current paper (see next section) and add a new study that tests a ninth PE-method (a needs-based PE-method). Since we already tested H1 in previous work, it would be a futile exercise to make *ex ante* hypotheses about which PE-method will be better for experts and which will be better for novices without resorting to post-hoc argumentation. In line with our previous work, we will discuss these effects *ex post*.

Decision behavior and choice satisfaction

Xia and Benbasat (2007) argue that the fit between the user and the recommender not only increases their system satisfaction, but also increases their intention to follow up on recommendations, and their satisfaction with their decisions. Therefore, a good fit between the user's level of domain knowledge and the PE-method of our recommender should result in users selecting a larger number of energy-saving measures, with a higher average level of savings per measure, and also a higher level of choice satisfaction. Beyond our findings in previous work, we therefore introduce the following hypotheses:

- H2. The number of measures that users select depends on the interaction between their level of domain knowledge and the PE-method of the system.
- H3. The average savings (in kWh) per selected measure depends on the interaction between users' level of domain knowledge and the PE method of the system.
- H4. Users' choice satisfaction depends on the interaction between their level of domain knowledge and the PE-method of the system.

Similar to H1, we only hypothesize generic interaction effects; we will discuss which PE-method is better for experts and which is better for novices *ex ante*.

Mediation by system satisfaction

We argued that the effects of H2-H4 occur because a match between PE-method and users' level of domain knowledge results in a more pleasant interaction. System satisfaction may thus partially or fully mediate the effects of PE-method and domain knowledge on decision outcomes and choice satisfaction. Indeed, existing frameworks for the user-centric evaluation of recommender systems contend that system satisfaction mediates the effects of PE-methods and other system aspects on users' behavioral intentions (Knijnenburg et al. 2012; Pu et al. 2011). Similarly, the perceived quality of a recommender has shown to positively influence users' decision satisfaction (Bharati and Chaudhury 2004; Knijnenburg et al. 2012).

Note however, that the *preferred* PE method (i.e. the PE method that makes the consumer more satisfied with the *system* (McNee et al. 2003)) may not be most *accurate* PE method (i.e. the PE method that allows the consumer to make better *decisions* (McNee et al. 2002)). This may mean that there may still be a residual direct interaction effect of PE-method and domain knowledge on the behavioral outcomes. Similarly, some studies have shown system satisfaction and choice satisfaction to be influenced by the recommender system independently from each other (Knijnenburg et al. 2010). Tentatively, though, we conjecture that at least a part of the effect of the match between PE-method and domain knowledge on decision behavior and choice satisfaction is mediated by system satisfaction, as expressed in the following hypotheses:

- H5. Users who are more satisfied with the system select a larger number of measures.
- H6. Users who are more satisfied with the system select measures with higher average savings.
- H7. Users who are more satisfied with the system are also more satisfied with their choices.

Choice satisfaction as a result of behavior

The purpose of our recommender system is to help users save more energy, so users are likely to have this goal in mind when they use the system. It thus stands to reason that the hypothesized effects on choice satisfaction (H4 and H7) may be mediated by users actual decision behavior: users will be more satisfied with their choices if they end up selecting more and more impactful measures.

On the other hand, not everyone may use the system with the ambitious goal to implement many impactful measures², so choice satisfaction could just as well be independent of the actual number and impact of the chosen. Yet tentatively we formulate the following hypotheses:

² E.g. some users may prefer to select many small measures, while others may prefer to select few large measures... such individual differences are why we developed a recommender system in the first place!

H8. Users who select a larger number of measures are more satisfied with their choices.

H9. Users who select measures with higher average savings are more satisfied with their choices.

Overview of hypotheses

Figure 1 summarizes the hypothesized effects. The core hypotheses in this model are H1 and H5-H7, and are signified with thicker arrows. As we argued above, H2-H4 may not reach significance if the effects of the match between domain knowledge and PE-method are indeed fully mediated by system satisfaction. Similarly, the H8-H9 may not reach significance if decision behavior does not mediate the effects on choice satisfaction.

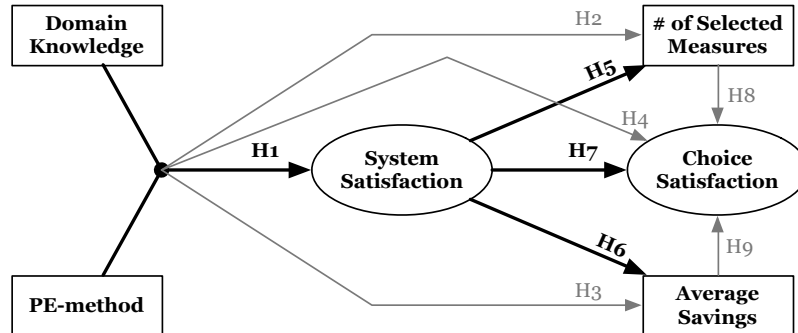


Figure 1. The Hypothesized Effects Across Our Four Studies.

Study descriptions

We tested the developed hypotheses using the data from four online user experiments with our recommender system. Three studies (1, 3 and 4) were ran as online lab studies where participants were sampled from a participant database at a Dutch university. Participants were invited and compensated to participate. The data from study 2 stems from a public field trial with the system after it was published in a local Dutch newspaper. The system contains 80 energy-saving measures defined on 8 attributes: initial effort, continuous effort, initial costs, savings in Euro/year, savings in kWh/year, return on investment, overall environmental effects, and comfort.

The interface of the system (Figure 2) consists of three parts: the top part is for preference elicitation, and differs per experimental condition. The middle part shows the recommendations (5 in studies 1 and 2; 10 in studies 3 and 4). When participants click on a recommendation the system shows them additional information about the measure, and allows the participant to choose what to do with the measure: “I don’t know yet” (default), “I want to do this”, “I’m already doing this”, and “I don’t want to do this” (only in studies 3 and 4). Choosing one of the latter three options moves the measure from the recommendations into one of the lists at the bottom of the interface. A new recommendation is then added to the bottom of the list in the middle part³. At the end of the experiment, participants can print the list of measures they want to implement.

More information about the participants in each study is displayed in Table 1. Note that study 4 was originally a 2-by-2 between-subjects experiment with 175 participants. In the second manipulation we primed participants with either a concrete or abstract mindset. The abstract mindset condition asked participants for reasons *why* they should be saving energy as a pre-experimental task. This condition was designed to make participants focus more on the sustainability goals and long term benefits of energy saving. The concrete mindset condition asked participants to already think of reasons *how* they can save energy. Our hypotheses were confirmed for the concrete mindset condition, but not for the abstract mindset condition. The reason for this lies beyond the scope of the current paper, and will be investigated extensively in a future publication. For now, we include the concrete mindset data because it is arguably

³ This, in combination with the “I don’t want to do this” option, allows users to browse through all recommendations, even without any mechanism to change the recommendations (i.e. the top-N baseline).

also the default mindset in the other studies (because the system and the task given to the users is concrete, i.e. focuses on selecting specific measures).

Indicate your preference
 Here is a list of possible needs. Indicate how important they are for you. By clicking multiple times you can change your preference.

Make a choice
 Here are several recommendations; choose those energy-saving measures from this list which you want to do, or which you are already doing.

Your savings
 Here are your selected savings!
 Show totals in ☐ euro ☐ kWh
 You have been using the system for 1 minutes. When you press stop, you will be asked a few final questions, after which you can print your savings.
 stop

| Name | Initial effort | Continuous effort | Initial costs | Savings euro/year | Savings kWh/year | Return on investment | Env. effects | Comfort |
|----------------------------------|----------------|-------------------|---------------|-------------------|------------------|----------------------|--------------|---------|
| Roof insulation | | | € 3100.00 | € 299.00 | 1424 kWh | 10 year | | |
| Laptop instead of a PC | | | € 95.00 | € 31.50 | 150 kWh | 3 year | | |
| Turn off PC when absent | | | none | € 96.28 | 458 kWh | direct | | |
| Heap laundry | | | none | € 57.75 | 275 kWh | direct | | |
| Close curtains/shutters at night | | | none | € 57.47 | 639 kWh | direct | | |
| Shower 3 minutes less | | | none | € 50.00 | 450 kWh | direct | | |
| Boiler-heated dryer | | | € 650.00 | € 100.00 | 370 kWh | 7 year | | |
| Air-dry clothes | | | none | € 60.90 | 290 kWh | direct | | |
| A++ Fridge/freezer combo | | | € 180.00 | € 73.50 | 350 kWh | 29 months | | |
| A+ Fridge/freezer combo | | | € 110.00 | € 56.70 | 270 kWh | 22 months | | |

| This is what I want to do: | |
|---|-----------------|
| Enable PC energy management | € 67.20 |
| Install LED light bulbs | € 121.06 |
| Place heat exchangers on the air vents | € 261.06 |
| Shut down boiler in the summer | € 59.85 |
| This is what I will save per year: | € 509.31 |

| This is what I already do: | |
|---------------------------------------|-----------------|
| Double glazing | € 302.00 |
| Green energy | none |
| Cook on gas instead of electric | € 75.00 |
| I was already saving per year: | € 377.00 |

| This is what I don't want to do: | |
|----------------------------------|----------|
| HR-E boiler | € 320.00 |
| Install CFL light bulbs | € 85.99 |

Figure 2. The Online Recommender System For Energy-Saving Measures (Translated From Dutch). The Explicit PE-method Is Shown

| Study | Type | # of pps | Min. interaction time | Gender | Avg. age |
|-------|------------------|----------|-----------------------|----------|----------|
| 1 | Online lab study | 90 | 10 min | 66M, 24F | 36.0 |
| 2 | Field trial | 107 | 3 min | 77M, 30F | 42.8 |
| 3 | Online lab study | 147 | 2.5 min | 79M, 68F | 40.0 |
| 4 | Online lab study | 88 | 2.5 min | 44M 44F | 28.0 |

Table 1. Participants In The Four Studies.

Manipulations

Each of the four studies explored a different set of PE-methods as its main manipulation.

Study 1

Study 1 (Knijnenburg and Willemsen 2009) tested attribute-based PE against case-based PE. The attribute-based PE-method let users explicitly assign attribute weights (Figure 3a), while the case-based PE-method let the users evaluate entire choice options (Figure 3b). Each example in the case-based PE-method represents a higher level of one of the attributes, and a positive (negative) evaluation can thus directly be translated into a higher (lower) weight on that attribute.

Study 2

For study 2 (Knijnenburg and Willemsen 2010) we developed new versions of the attribute-based and case-based PE-method to overcome some usability problems of the PE-methods in study 1. The old attribute-based PE-method let the user increase or decrease the *importance* of each attribute, which could have caused confusion for negatively phrased attributes. E.g. *increasing* the importance of “continuous effort” actually showed energy saving measures with *lower* effort levels. The new method explicitly showed the direction of the effect of clicking each button, e.g. “continuous effort” was replaced by “lower continuous effort” (Figure 3c). Similarly, the old case-based PE-method was very cluttered because it showed all the attribute values of the examples. The new method only showed the names of the examples (Figure 3d). These new PE-methods also included a “double-increase” and “double-decrease” button to reduce the amount of clicking users would have to do to change their preferences.

Study 3

Study 3 (Knijnenburg et al. 2011) included the same attribute-based PE-method as study 2 (Figure 3c), and also an implicit PE-method (which has no PE interface but instead updates the weights automatically based on users' clicking behavior) and a hybrid PE-method (which combines attribute-based PE with automatic updates of implicit PE). This study also included a top-N baseline (sorted by the popularity of the measures as observed in studies 1 and 2) and a sort baseline (initially sorted by popularity, but allowing users to sort on any attribute).

Study 4

Study 4 (new data) tested the attribute-based PE-method of studies 2 and 3 (Figure 3c) against a needs-based PE-method (Figure 3e). For the needs-based PE-method multi-dimensional scaling was applied to the selection behaviors of participants in the previous studies. This procedure resulted in two dimensions: popularity (commonly popular versus unusual energy-savings measures) and efficiency (small and elaborate versus impactful yet simple measures). These dimensions were presented as two sliders, one ranging from "popular measures" to "unique measures" and the other ranging from "every kWh counts" to "save lots of energy per Euro". More details on the development of the needs-based PE-method can be found in (Reijmer 2011).

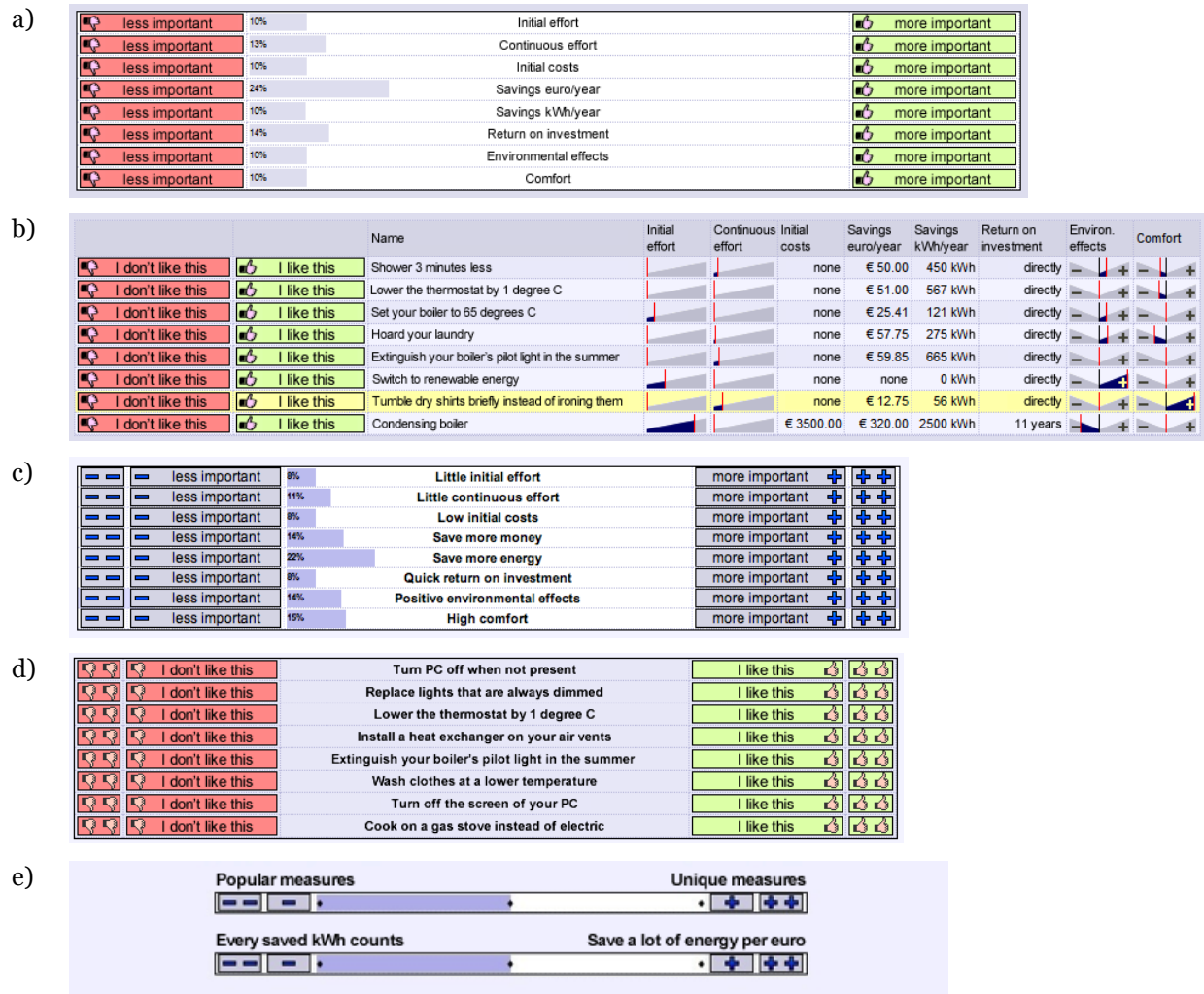


Figure 3. Screenshots Of The Different PE-methods.

Dependent variables

We measured domain knowledge by asking participants to fill out a short self-report survey in advance about their familiarity with energy-saving measures, and their perceived ability to evaluate and compare measures. We measured domain knowledge with the same 7 items in all studies except study 2, where a subset of 4 items is used. This is due to the public nature of the field trial, where we wanted to keep the pre-experimental procedures to a minimum.

At the end of the study, we measured participants' system satisfaction and choice satisfaction. System satisfaction is a positive self-relevant evaluation of a system (Hassenzahl 2005). Satisfaction is not only determined by tangible aspects, such as service quality but also by intangible ones, such as feelings of joy, fear, and frustration associated with the service experience (Johnson and Grayson 2005). The concept of system satisfaction relates to the perceived usefulness construct in TAM (Davis 1989) and the performance expectancy construct in UTAUT (Venkatesh et al. 2003). We measured system satisfaction with the same 6 items across the four studies.

Choice satisfaction is a positive self-relevant evaluation of the *outcome* of using a system (Bechwati and Xia 2003; Bollen et al. 2010; Pedersen 2000). It is related to users' decision confidence (Hostler et al. 2005; Krishnan et al. 2008; Vijayarathy and Jones 2001). Whereas system satisfaction provides users with an outcome expectation, choice satisfaction is an actual evaluation of those outcomes. We measured choice satisfaction with the same 4 items across the four studies.

| | | Study 1 | Study 2 | Study 3 | Study 4 |
|----------------------------|---|---------|------------|---------|---------|
| Domain knowledge | Average variance extracted: | 0.662 | 0.587 | 0.575 | 0.568 |
| | Correlation with System satisfaction: | -0.028 | -0.095 | -0.034 | 0.146 |
| | Correlation with Choice satisfaction: | 0.146 | -0.139 | 0.386 | 0.203 |
| | I know the energy consumption of all devices in my household | 0.826 | 0.791 | 0.641 | 0.627 |
| | I understand difference between different types of energy-saving measures | 0.863 | 0.766 | 0.764 | 0.762 |
| | I know energy-saving measures that most others haven't even heard of | 0.748 | 0.789 | 0.791 | 0.699 |
| | I know which energy-saving measures are useful to implement | 0.851 | 0.717 | 0.753 | 0.783 |
| | I am able to choose the right energy-saving measures | 0.776 | not tested | 0.858 | 0.874 |
| | I sometimes doubt whether I chose good energy-saving measures | | not tested | | |
| | I don't understand most energy-saving measures | | not tested | -0.725 | |
| System satisfaction | Average variance extracted: | 0.664 | 0.673 | 0.615 | 0.725 |
| | Correlation with Domain knowledge: | -0.028 | -0.095 | -0.034 | 0.146 |
| | Correlation with Choice satisfaction: | 0.476 | 0.547 | 0.383 | 0.648 |
| | The system made me more energy-conscious | | | 0.658 | 0.811 |
| | The system restricted my decision freedom | | | | |
| | I would use the system more often if possible | 0.888 | 0.856 | 0.842 | 0.864 |
| | I make better choices with the system | 0.738 | 0.703 | 0.806 | 0.772 |
| | The system was useless | -0.771 | -0.803 | -0.763 | -0.881 |
| | I would recommend the system to others | 0.853 | 0.905 | 0.836 | 0.921 |
| Choice satisfaction | Average variance extracted: | 0.506 | 0.543 | 0.538 | 0.607 |
| | Correlation with Domain Knowledge: | 0.146 | -0.139 | 0.386 | 0.203 |
| | Correlation with System satisfaction: | 0.476 | 0.547 | 0.383 | 0.648 |
| | I like the measures I've chosen | 0.750 | 0.697 | 0.707 | 0.823 |
| | I think I chose the best measures from the list | 0.788 | 0.687 | 0.789 | |
| | The chosen measures exactly fit my preference | 0.577 | 0.820 | 0.769 | 0.697 |
| | How many measures will you implement? | | | 0.664 | 0.810 |

Table 2. CFA Outcomes Of The Questionnaires.

Measurement validity of the constructs with a confirmatory factor analysis (CFA, Table 2). Since the items were measured on a 5-point scale, we used a WLS estimator that treats the items as ordered-categorical. We iteratively removed items with a communality < 0.350, with the exception of the third choice satisfaction item in study 1, which was retained in order to have at least 3 items for each factor.

We also measured users' decision behaviors in each study; we tracked the number of energy-saving measures they selected ("I want to do this"), and the average savings (in kWh) of these selected measures. The distribution of these two behavioral variables is highly skewed, so we applied a log transformation to make the variables more normally distributed.

Results

We tested the research model and each hypothesized path using structural equation modeling (SEM) with a weighted least squares estimator. To allow interaction with PE-method, domain knowledge is not modeled as a reflective latent construct but calculated as a weighted sum based on the CFA results. Since study 3 tests more than two PE-methods, the interaction effect is there tested with an omnibus Wald test.

The first step in the model estimation was to remove paths that are non-significant ($p > .05$) from each model. In general, models can be trimmed or built based on theoretical and/or empirical standards (Kline 2004). Since we already argued that H2-H5 and H8-H9 might not be consistently significant, trimming these paths if non-significant is justified. Figure 4 shows the resulting models and their fit indices. The interaction effects between domain knowledge and PE-methods are displayed in graphs in the models. All models had a good fit. Below we discuss the individual hypothesized effects.

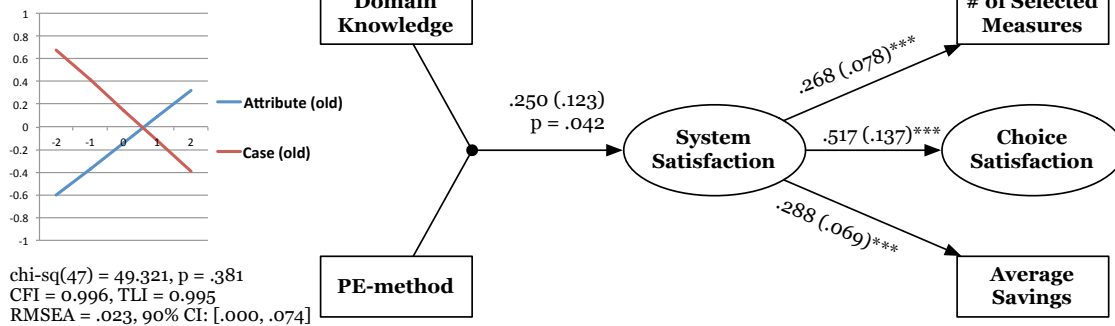
The results show that the interaction between domain knowledge and PE-method indeed has a consistently significant effect on system satisfaction (H1 is supported). Indeed, the best and worst interfaces are different for experts and for novices: with the exception of study 2 (more on this below), the attribute-based, implicit, and hybrid PE-methods seem to be better for energy-saving experts, while the case-based and needs-based PE-methods seem to be better for novices. Note though that novices are also highly satisfied with the baseline systems used in study 3.

The models also show robust support for H5-H7, but not for H2-H4. System satisfaction consistently mediates (H5-H7) the effect of the match between domain knowledge and PE-method (represented by our interaction effect) on users' decision behavior and choice satisfaction without any significant residual direct effects (H2-H4). The only exception to this is that system satisfaction had no effect on average savings in study 4 (H6 not supported). Generally, though, the effects of a good PE-method on users' energy-saving behavior and choice satisfaction are entirely mediated by their system satisfaction: a "well-fitting" PE-method makes users more satisfied with the system, which in turn makes them choose more and larger energy-saving measures, and increases their choice satisfaction as well. Beyond previous work we thus demonstrate that a well-fitting PE method does not only have *attitudinal* but also *behavioral* consequences.

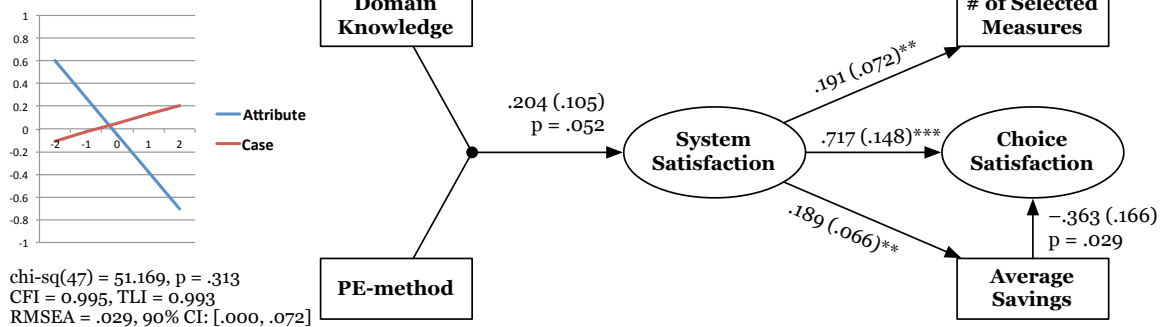
The effect of decision behavior on choice satisfaction is inconsistent (H8 not supported, H9 only in studies 3 and 4) and in some cases even contradicts our hypothesis (selecting larger savings reduces choice satisfaction in study 2; selecting more measures reduces choice satisfaction in study 3). As we have explained, this may be due to users' variety of preferences for either many small or a few large measures.

Another inconsistency is the effect of the PE-methods in study 2, which is opposite to the same effects in the other studies. The absence of detailed attribute values in the case-based PE-method in study 2 may have made it harder for novice users to make case-based trade-offs. The attribute-based PE-method was slightly changed between study 1 and 2 but was kept exactly the same as in studies 3 and 4, but with the opposite effect on satisfaction for experts and novices. The results of the attribute-based PE in study 2 therefore deviate from our theoretical predictions and from the other three studies. We have no plausible explanation for this, other than that study 2 was based on a convenience sample of field trial participants. The consistency of the effects between the other studies and the tighter experimental control we were able to exert in those studies leads us to believe that the results of study 2 are an anomaly.

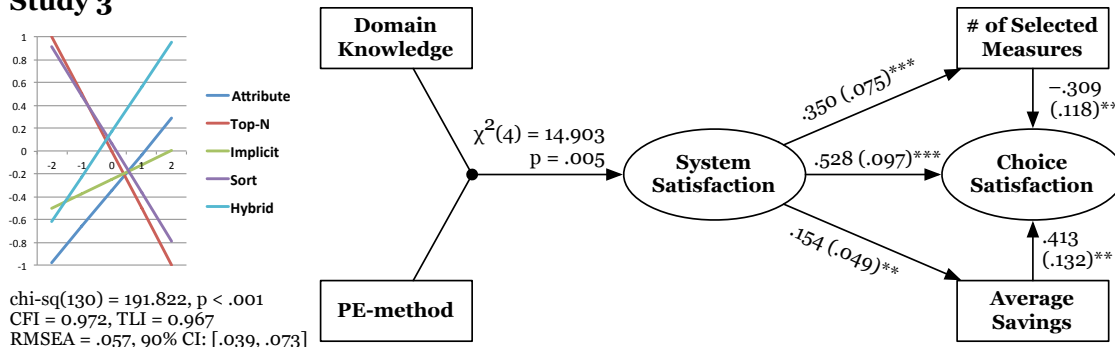
Study 1



Study 2



Study 3



Study 4

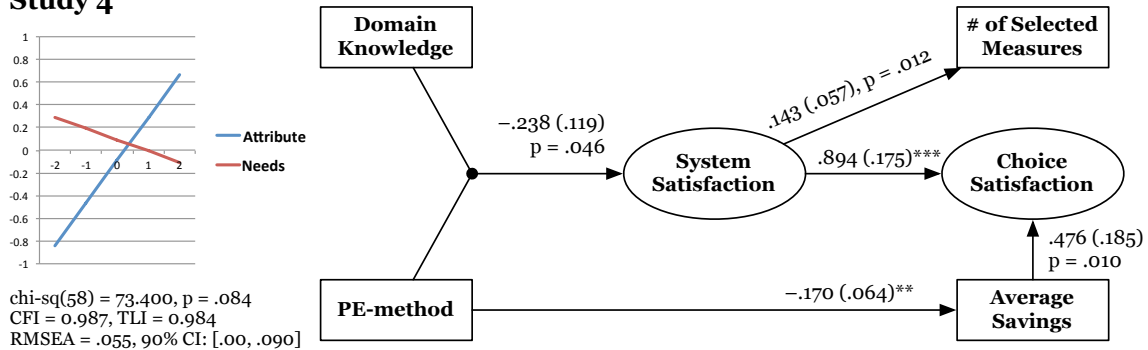


Figure 4. SEM Models With Standardized Factor Scores. ** $p < .01$, *** $p < .001$. The Graphs Show The Effect Of Domain Knowledge (x-axis) And PE-method (Different Lines) On System Satisfaction (y-axis).

Managerial implications and future work

Hundreds of energy-saving measures exist, but they are scattered across hundreds of websites and brochures. We are arguably the first to identify a choice problem in this domain: one simply cannot implement all these measures at once. We demonstrate that a recommender system can help people in choosing energy-saving measures that fit their personal needs and preferences.

More importantly, tailoring the interface of this recommender to the level of knowledge users have about energy-saving does not only increase their satisfaction with the system, as we already showed in previous work (studies 1-3) and replicated in study 4; it is also the key to make users save more energy, because they end up choosing more and more impactful energy-saving measures. This elevates the idea of “tailored preference elicitation” from a design consideration to an important managerial concern, and qualifies it as an effective tool for policy makers to influence energy consumption patterns.

Specifically, it seems that energy-saving experts prefer complex systems that allow direct control over the attributes weights (attribute-based and hybrid PE), while novices prefer systems that are tailored to their needs (needs-based PE; a new insight from study 4), provide limited control (sort) or rather no control at all (top-N). Those seeking to help consumers save a maximum amount of energy should heed these findings when developing their decision support systems.

Our study did not go beyond the selection of energy-saving measures (and the offer of a print-out); future work should provide helpful references for the implementation of the selected measures (Web shops, DIY instruction sites) and possibly even follow up on users’ intentions to implement the selected measures. The findings presented in this paper should also be tested on participants outside of The Netherlands.

Finally, we argue that the demonstrated behavioral consequences of a satisfying experience suggest a prominent role of user experience (UX) design in the development of systems for behavioral change, especially regarding the important topic of energy consumption. Future work on smart systems to support sustainable decisions should focus on creating pleasant experiences for novices and experts alike.

REFERENCES

- Abrahamse, W., Steg, L., Vlek, C., and Rothengatter, T. 2005. “A review of intervention studies aimed at household energy conservation,” *Journal of Environmental Psychology* (25:3), pp. 273–291.
- Aksoy, L. 2006. “Should Recommendation Agents Think Like People?,” *Journal of Service Research* (8:4), pp. 297–315.
- Aune, M. 2007. “Energy comes home,” *Energy Policy* (35:11), pp. 5457–5465.
- Bechwati, N. N., and Xia, L. 2003. “Do Computers Sweat? The Impact of Perceived Effort of Online Decision Aids on Consumers’ Satisfaction With the Decision Process,” *Journal of Consumer Psychology* (13:1–2), pp. 139–148.
- Benders, R. M. J., Kok, R., Moll, H. C., Wiersma, G., and Noorman, K. J. 2006. “New approaches for household energy conservation—In search of personal household energy budgets and energy reduction options,” *Energy Policy* (34:18), pp. 3612–3622.
- Bettman, J. R., Luce, M. F., and Payne, J. W. 1998. “Constructive consumer choice processes,” *Journal of consumer research* (25:3), pp. 187–217.
- Bharati, P., and Chaudhury, A. 2004. “An empirical investigation of decision-making satisfaction in web-based decision support systems,” *Decision Support Systems* (37:2), pp. 187–197.
- Bollen, D., Knijnenburg, B. P., Willemsen, M. C., and Graus, M. 2010. “Understanding choice overload in recommender systems,” in *Proceedings of the fourth ACM conference on Recommender systems*, Barcelona, Spain, pp. 63–70.
- Butler, J. C., Dyer, J. S., and Jia, J. 2006. “Using Attributes to Predict Objectives in Preference Models,” *Decision Analysis* (3:2), pp. 100–116.
- Butler, J. C., Dyer, J. S., Jia, J., and Tomak, K. 2008. “Enabling e-transactions with multi-attribute preference models,” *European Journal of Operational Research* (186:2), pp. 748–765.
- Chen, J., Geyer, W., Dugan, C., Muller, M., and Guy, I. 2009. “Make new friends, but keep the old: recommending people on social networking sites,” in *Proceedings of the 27th international conference on Human factors in computing systems*, Boston, MA, USA, pp. 201–210.

- Chen, L., and Pu, P. 2009. "Interaction design guidelines on critiquing-based recommender systems," *User Modeling and User-Adapted Interaction* (19:3), pp. 167–206.
- Chen, L., and Pu, P. 2011. "Critiquing-based recommenders: survey and emerging trends," *User Modeling and User-Adapted Interaction* (22:1-2), pp. 125–150.
- Chin, D. N. 2001. "Empirical Evaluation of User Models and User-Adapted Systems," *User Modeling and User-Adapted Interaction* (11:1-2), pp. 181–194.
- Darby, S. 2010. "Smart metering: what potential for householder engagement?," *Building Research & Information* (38:5), pp. 442–457.
- Davis, F. D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), p. 319.
- Diehl, K., Kornish, L. J., and John G. Lynch, J. 2003. "Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity," *Journal of Consumer Research* (30:1), pp. 56–71.
- Dooms, S., De Pessemier, T., and Martens, L. 2011. "An online evaluation of explicit feedback mechanisms for recommender systems," in *7th International Conference on Web Information Systems and Technologies (WEBIST-2011)*, Noordwijkerhout, The Netherlands, pp. 391–394.
- Edwards, W., and Fasolo, B. 2001. "Decision Technology," *Annual Review of Psychology* (52:1), pp. 581–606.
- Felfernig, A. 2007. "Knowledge-Based Recommender Technologies for Marketing and Sales," *International Journal of Pattern Recognition and Artificial Intelligence* (21:2), pp. 333–354.
- Fischer, C. 2008. "Feedback on household electricity consumption: a tool for saving energy?," *Energy Efficiency* (1:1), pp. 79–104.
- Gardner, G. T., and Stern, P. C. 2008. "The Short List: The Most Effective Actions U.S. Households Can Take to Curb Climate Change," *Environment: Science and Policy for Sustainable Development* (50:5), pp. 12–25.
- Gena, C., Brogi, R., Cena, F., and Vernerio, F. 2011. "The Impact of Rating Scales on User's Rating Behavior," in *User Modeling, Adaption and Personalization*, J. A. Konstan, R. Conejo, J. L. Marzo, and N. Oliver (eds.), (Vol. 6787) Berlin, Heidelberg: Springer, pp. 123–134.
- Guttman, R. H., and Maes, P. 1999. "Agent-Mediated Integrative Negotiation for Retail Electronic Commerce," in *Agent Mediated Electronic Commerce*, Lecture Notes in Computer Science, P. Noriega and C. Sierra (eds.), Springer Berlin Heidelberg, pp. 70–90.
- Guttman, R. H., Moukas, A. G., and Maes, P. 1998. "Agent-mediated Electronic Commerce: A Survey," *Knowledge Engineering Review* (13:2), pp. 147–159.
- Hassenzahl, M. 2005. "The thing and I: understanding the relationship between user and product," in *Funology, From Usability to Enjoyment*, M. A. Blythe, K. Overbeeke, A. F. Monk, and P. C. Wright (eds.), Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 31–42.
- Häubl, G., Dellaert, B. G. C., Murray, K. B., and Trifts, V. 2004. "Buyer Behavior in Personalized Shopping Environments," in *Designing Personalized User Experiences in eCommerce*, Human-Computer Interaction Series, C.-M. Karat, J. O. Blom, and J. Karat (eds.), Springer Netherlands, pp. 207–229.
- Häubl, G., and Trifts, V. 2000. "Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids," *Marketing Science* (19:1), pp. 4–21.
- Hostler, R. E., Yoon, V. Y., and Guimaraes, T. 2005. "Assessing the impact of internet agent on end users' performance," *Decision Support Systems* (41:1), pp. 313–323.
- Hutton, R. J. B., and Klein, G. 1999. "Expert decision making," *Systems Engineering* (2:1), pp. 32–45.
- Johnson, D., and Grayson, K. 2005. "Cognitive and affective trust in service relationships," *Journal of Business research* (58:4), pp. 500–507.
- Kline, R. B. 2004. *Beyond significance testing: reforming data analysis methods in behavioral research*, Washington, DC: American Psychological Association.
- Knijnenburg, B. P., Reijmer, N. J. M., and Willemsen, M. C. 2011. "Each to his own: how different users call for different interaction methods in recommender systems," in *Proceedings of the fifth ACM conference on Recommender systems*, Chicago, IL, pp. 141–148.
- Knijnenburg, B. P., and Willemsen, M. C. 2009. "Understanding the effect of adaptive preference elicitation methods on user satisfaction of a recommender system," in *Proceedings of the third ACM conference on Recommender systems*, New York, NY, pp. 381–384.
- Knijnenburg, B. P., and Willemsen, M. C. 2010. "The effect of preference elicitation methods on the user experience of a recommender system," in *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, Atlanta, GA, pp. 3457–3462.

- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., and Newell, C. 2012. "Explaining the user experience of recommender systems," *User Modeling and User-Adapted Interaction* (22:4-5), pp. 441–504.
- Knijnenburg, B. P., Willemsen, M. C., and Hirtbach, S. 2010. "Receiving Recommendations and Providing Feedback: The User-Experience of a Recommender System," in *E-Commerce and Web Technologies*, F. Buccafurri and G. Semeraro (eds.), (Vol. 61) Berlin, Heidelberg: Springer, pp. 207–216.
- Konstan, J. A., and Riedl, J. 2012. "Recommender systems: from algorithms to user experience," *User Modeling and User-Adapted Interaction* (22:1-2), pp. 101–123.
- Koren, Y., Bell, R., and Volinsky, C. 2009. "Matrix Factorization Techniques for Recommender Systems," *Computer* (42:8), pp. 30–37.
- Kramer, T. 2007. "The Effect of Measurement Task Transparency on Preference Construction and Evaluations of Personalized Recommendations," *Journal of Marketing Research* (44:2), pp. 224–233.
- Krishnan, V., Narayanashetty, P. K., Nathan, M., Davies, R. T., and Konstan, J. A. 2008. "Who Predicts Better?: Results from an Online Study Comparing Humans and an Online Recommender System," in *Proceedings of the 2008 ACM Conference on Recommender Systems*, Zurich, Switzerland, pp. 211–218.
- Lee, Y. E., and Benbasat, I. 2011. "The Influence of Trade-off Difficulty Caused by Preference Elicitation Methods on User Acceptance of Recommendation Agents Across Loss and Gain Conditions," *Information Systems Research* (22:4), pp. 867–884.
- Lehman, P. K., and Geller, E. S. 2005. "Behavior Analysis and Environmental Protection: Accomplishments and Potential for More," *Behavior and Social Issues* (13:1), pp. 13–32.
- McGinty, L., and Smyth, B. 2006. "Adaptive Selection: An Analysis of Critiquing and Preference-Based Feedback in Conversational Recommender Systems," *International Journal of Electronic Commerce* (11:2), pp. 35–57.
- McKenzie-Mohr, D. 2000. "Fostering sustainable behavior through community-based social marketing," *The American psychologist* (55:5), pp. 531–537.
- McNee, S. M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S. K., Rashid, A. M., Konstan, J. A., and Riedl, J. 2002. "On the recommending of citations for research papers," in *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, New Orleans, LA, pp. 116–125.
- McNee, S. M., Lam, S. K., Konstan, J. A., and Riedl, J. 2003. "Interfaces for Eliciting New User Preferences in Recommender Systems," in *User Modeling 2003*, P. Brusilovsky, A. Corbett, and F. Rosis (eds.), (Vol. 2702) Berlin: Springer Heidelberg, pp. 178–187.
- Al-Natour, S., Benbasat, I., and Cenfetelli, R. T. 2006. "The role of design characteristics in shaping perceptions of similarity: The case of online shopping assistants," *Journal of the Association for Information Systems* (7:12), p. 34.
- Olson, E. L., and Widing, R. E. 2002. "Are interactive decision aids better than passive decision aids? A comparison with implications for information providers on the internet," *Journal of Interactive Marketing* (16:2), pp. 22–33.
- Pazzani, M. J., and Billsus, D. 2002. "Adaptive Web Site Agents," *Autonomous Agents and Multi-Agent Systems* (5:2), pp. 205–218.
- Pedersen, P. E. 2000. "Behavioral Effects of Using Software Agents for Product and Merchant Brokering: An Experimental Study of Consumer Decision-Making," *International Journal of Electronic Commerce* (5:1), pp. 125–141.
- Petkov, P., Köbler, F., Foth, M., and Krcmar, H. 2011. "Motivating Domestic Energy Conservation Through Comparative, Community-based Feedback in Mobile and Social Media," in *Proceedings of the 5th International Conference on Communities and Technologies*, Brisbane, Australia, pp. 21–30.
- Pu, P., Chen, L., and Hu, R. 2011. "A user-centric evaluation framework for recommender systems," in *Proceedings of the fifth ACM conference on Recommender systems*, Chicago, IL, October, pp. 157–164.
- Randall, T., Terwiesch, C., and Ulrich, K. T. 2007. "User Design of Customized Products," *Marketing Science* (26:2), pp. 268–280.
- Reijmer, N. J. M. 2011. "Improving energy saving decisions by matching recommender type with domain knowledge and mindset," Master Thesis, Eindhoven, Netherlands: Eindhoven University of Technology.
- Shanteau, J. 1988. "Psychological characteristics and strategies of expert decision makers," *Acta Psychologica* (68:1–3), pp. 203–215.

- Smyth, B. 2007. "Case-Based Recommendation," in *The Adaptive Web: Methods and Strategies of Web Personalization*, Lecture Notes in Computer Science, P. Brusilovsky, A. Kobsa, and W. Nejdl (eds.), (Vol. 4321) Berlin: Springer Verlag, pp. 342–376.
- Smyth, B., and McClave, P. 2001. "Similarity vs. Diversity," in *Case-Based Reasoning Research and Development*, Lecture Notes in Computer Science, D. W. Aha and I. Watson (eds.), Springer Berlin Heidelberg, pp. 347–361.
- Sparling, E. I., and Sen, S. 2011. "Rating: How Difficult is It?," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, Chicago, IL, pp. 149–156.
- Spiekermann, S., and Paraschiv, C. 2002. "Motivating human–agent interaction: Transferring insights from behavioral marketing to interface design," *Electronic Commerce Research* (2:3), pp. 255–285.
- Steg, L. 2008. "Promoting household energy conservation," *Energy Policy* (36:12), pp. 4449–4453.
- Venkatesh, V., Morris, M. G., Gordon B. Davis, and Davis, F. D. 2003. "User Acceptance of Information Technology: Toward a Unified View," *MIS Quarterly* (27:3), pp. 425–478.
- Vijayasarathy, L. R., and Jones, J. M. 2001. "Do Internet Shopping Aids Make a Difference? An Empirical Investigation," *Electronic Markets* (11:1), pp. 75–83.
- Wilson, C., and Dowlatabadi, H. 2007. "Models of Decision Making and Residential Energy Use," *Annual Review of Environment and Resources* (32:1), pp. 169–203.
- Xiao, B., and Benbasat, I. 2007. "E-commerce Product Recommendation Agents: Use, Characteristics, and Impact," *Mis Quarterly* (31:1), pp. 137–209.
- Xiao, B., and Benbasat, I. 2014. "Research on the Use, Characteristics, and Impact of e-Commerce Product Recommendation Agents: A Review and Update for 2007–2012," in *Handbook of Strategic e-Business Management*, F. J. Martínez-López (ed.), Springer Berlin Heidelberg, pp. 403–431.