Psychologically Informed Design of Energy Recommender Systems: Are Nudges Still Effective in Tailored Choice Environments?

Alain D. Starke and Martijn C. Willemsen

Abstract Studies in psychology have shown various ways how humans can be influenced in their choices and behavior. Many of these persuasive strategies and nudges are now also used online, affecting how digital choice environments are designed. In the sustainability domain, these strategies have been used to promote specific pro-environmental behaviors, such as through green energy defaults and social norms (e.g., '75% of people re-use their towel'). Most of these nudges are, however, evaluated in one-size-fits-all interventions, not reflecting to an extent to which today's digital environments are personalized. Not only does this call for smarter, personalized nudges, it also overlooks the fact that various nudges would be applied in tailored choice environments. In particular, recommender systems have become ubiquitous, directly tailoring advice to end users, which might deem nudges to become superfluous. Hence, it remains an open question whether nudging is still effective if the advice is also tailored. This chapter explores the effectiveness of different (smart) nudges in the context of tailored choice systems for household energy conservation. We have developed an approach for a psychology-informed recommender system that presents personalized, attitude-tailored energy-saving advice to end users. Our approach comprises an algorithm and interface nudges that are both personalized and operationalized through smart default and social norm interventions. We present the

A. D. Starke (✉)
Amsterdam School of Communication Research, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands
e-mail: a.d.starke@uva.nl

MediaFutures: Research Centre for Responsible Media Technology & Innovation, University of Bergen, Lars Hilles gate 30, Bergen 5008, Norway

M. C. Willemsen
Human-Technology Interaction Group, Eindhoven University of Technology, Den Dolech 2, Eindhoven 5612 AZ, The Netherlands
e-mail: m.c.willemsen@tue.nl

Jheronimus Academy of Data Science (JADS), Sint Janssingel 92, 's-Hertogenbosch 5211 DA, The Netherlands

B. Ferwerda et al. (eds.), *A Human-Centered Perspective of Intelligent Personalized Environments and Systems*, Human–Computer Interaction Series,
https://doi.org/10.1007/978-3-031-55109-3_9

results of multiple studies performed with our energy recommender systems, providing evidence for the limited effectiveness of interface nudges in a personalized advice context. We discuss the design implications and what nudging and persuasion mean in a world in which most decisions are digitized and content is personalized.

Keywords Recommender systems · Nudging · User modeling · Tailoring · Psychology · Energy conservation

1 Introduction

Digital technologies have the capacity to steer human behavior. Some of these technologies have an explicit persuasive intent [1], trying to promote a set of behaviors. In some cases, these are in the best interest of the user or simply try to help users to make 'better' decisions [2]. To improve their effectiveness, such systems can tap into psychological theories about how people form preferences, make decisions, and act accordingly.

Most studies in human-computer interaction, however, take a technology-centered perspective, prioritizing the optimization of algorithms and interfaces. Predicting what should be presented next is typically the main goal, for the which the underlying models are typically the result of a machine learning approach with opaque factors and interrelations. In that sense, there is a still a world to learn from social science domains, such as psychology, to improve technologies that are used in human-computer interaction, both, for users and researchers, to understand them better.

This chapter focuses on using psychological principles and theories to design recommender system technologies. We present how algorithmic approaches and interfaces can take a psychology theory as a starting point. We do so by addressing the problem of household energy conservation. This is a domain in which various studies have demonstrated the benefits of tailoring and nudging strategies [3, 4], but where technologies do not focus as on information retrieval and deep learning as found in computer science domains. Instead, tailoring often involves tailored feedback, while personalization is typically name-based [5, 6].

One of the main questions is whether nudges and persuasive strategies are still effective in addition to a tailoring strategy (cf. [1]). Tailoring or personalization of content is central to recommender systems [7], but is not the only strategy that can be used to steer user choices or support behavioral change. Various field and lab studies in the context of energy conservation have shown the effectiveness of different behavioral change strategies [4, 8], such as feedback and commitment [3, 9], goal-setting [3, 4], social comparison [9–13], defaults [14, 15], and many more, which also include digital contexts [12, 16]. The examples presented in this chapter, however, examine these effects in the context of a tailored or personalized list of options, which appear to mitigate the effectiveness of such nudges. Most studies that present personalized approaches do not necessarily tailor the content, but, for instance, present feedback that applies specifically to a user [17].

It is important to point out our conceptual conventions here, which are at odds with the views of some other authors (e.g., [18]). Regarding tailoring, we refer to presenting content to a user that fits their preferences, based on a user model of past choices or current needs. When referring to 'nudges', we *only* refer to specific changes in a choice environment that lead to predictable behavior [19], without changing *which items* are presented. The same applies to persuasive strategies that affect how content is presented in an interface. This contrasts with some studies which use 'nudge' as a term to refer to an intervention that seeks to affect behavior by changing which items are presented. For example, a study that used tailoring 'to nudge' people to save energy does not fall within our scope of nudges [20], even though some studies deem tailoring to be a persuasive strategy or a (smart) nudge [21].

This chapter first introduces our tailoring and personalization technology: the recommender system. While there are a few examples of psychologically informed recommenders across domains [22], with some of them closely mimicking psychological theories [23], we focus on the energy conservation domain. We discuss challenges for energy recommender systems, how this domain specifically calls for goal-directed design of recommenders, and how they should be evaluated to assess their effectiveness of psychologically informed interventions. In doing so, we outline different approaches in nudging and persuasion and how their effectiveness may be affected in the context of a tailored recommender system. To illustrate the arguments made, we present multiple examples that are adapted from studies on energy recommender systems [24–28].

2 Energy Recommender Systems

One of the main technologies studied in the context of algorithmic personalization are recommender systems. Recommender systems are defined as information filtering systems that present the most relevant content to users, based on their past preferences [7, 29]. Recommender systems comprise both algorithms, which are based on user models and interfaces, for which the latter may or may not be personalized [30]. In this case, personalization refers to system-initiated changes to an interface or system that facilitates a user's needs and interests [31, 32].

This chapter focuses on energy recommender systems [33]. This type of recommender aims to support its users to engage in pro-environmental behavior, typically targeting electricity use or environmental behaviors at home [34]. Initial recommender research for households has focused on capturing the trade-offs involved in energy-saving behavior [35], such between frequency of behavior and investment costs. For example, turning off lights in rooms is a frequent behavior without monetary costs, while the installation of solar PV is usually a one-time investment. More contemporary research tends to focus on the use of machine learning methods to detect which appliances are used in a household, which can be used to generate energy-efficient recommendations [34].

The challenges faced in the energy domain are distinct from various other domains, because behavioral change is also involved, going beyond one-time interactions and engagement. For instance, some recommenders are used to drive sales in e-commerce [36], which involves one-time transactions, while recommenders in other domains present content that cannot necessarily be bought, such as tourism-related recommenders [37–39], as well as recipe recommenders [40–42]. For the latter, 'taste' still plays a big role for the initial selection, but other behavioral costs that are more contextual play a big role [43], such as available time (for traveling and cooking) and skills (how difficult a recipe is to prepare).

2.1 Goal-Directed Recommender Systems for Energy Conservation

Designers of some recommender systems have a goal or intent [44]. Such systems are not 'neutral' decision-support systems [45], but they convey a message to the user that is in line with a specific goal that the recommender's designer has in mind. This steps away from a recommender optimizing for *any* item to be chosen, but instead optimizes for specific items or item features to be chosen.

This resonates with a number of variations on recommender systems such as those developed for self-development. Recommenders can help users to develop and understand their own taste and preferences [46, 47], so-called self-actualization. Whereas many applications apply a strategy in which items are presented with the highest predictive accuracy, users are supported in their discovery of a domain if recommendation strategies also optimize for diversity [46].

In this chapter, we specifically examine the goal-based or goal-directed behavior of energy conservation [48]. While some recommenders have used goals in the terminology (e.g., [49, 50]), this is often in the context of a user eliciting a short-term desire, to which an interface is adapted. Recommenders that help users to attain a new lifestyle or to adopt new behaviors are sparse [51], as most studies remain focused on the short term through session-based evaluations (e.g., in food [40]).

In the context of energy conservation, not all behaviors are equally likely to be performed [52]. While it is rather straightforward to turn off a light [27], and most people are willing to do so [53], it is perceived as more difficult to buy energy-efficient appliances and to learn habitual behaviors [54]. From a recommender point of view, easy energy-saving measures are considered popular as most people do them [55], which would make them seemingly good candidate items. However, if a system continues to 'dwell on the past', a user will never be presented new content [56], which could additionally contribute to their goal of energy conservation.

Household energy-saving behavior faces a 'difficulty' ordering [48, 57, 58]. Some behaviors come with fewer behavioral costs, which involves any type of costs related to executing a behavior, including cognitive and financial costs [52]. Psychological theories describing behavior and behavioral intent, such as the Theory of Planned

Behavior and Campbell's Paradigm [52, 59], postulate that individuals with stronger energy-saving attitudes are more likely to engage in any behavior. Moreover, only those with a sufficiently strong attitude are likely to engage in a rather difficult behavior [52]. Hence, a person who is careless about energy conservation is far less likely to spend time and costs to buy the most energy-efficient washing machine possible.

Most recommender approaches are not compatible with such an ordering in the database of items [51]. That is, user preferences are considered to move around freely in a multidimensional preference space, instead of being subject to a person's motivational disposition and capabilities. This has implications for both the algorithm and the interface of a recommender system. On the one hand, algorithms should be sensitive to the self-actualization and goal-setting that may underpin user preferences. This should also be considered in approaches that are not necessarily knowledge-based, where explicit goals can be asked [43, 60]. On the other hand, interfaces can support specific behavioral goals that a user may have, or a designer of the recommender for that matter. By changing *how* items are presented or explained, user preferences may be steered toward a goal more easily.

For such approaches to be successful, we echo the sentiment that recommenders need to go beyond behaviorism [51]. To be able to apply the principles of environmental psychology to an energy recommender system, a broader discussion of psychologically informed methods is necessary [22]. In the next subsection, we describe methods of persuasion and digital nudging and how they have been and can be applied in recommender systems. Subsequently, we also outline the methods for our examples: Because not only changes in behavior are relevant but also how different approaches are perceived and how systems are experienced, we describe a framework for mediated effects. This comprises a discussion of the user experience framework of Knijnenburg et al. [61].

2.2 Digital Nudging and Persuasion

The goal of a recommender system is typically not to just filter information neutrally [45]. Instead, developers of such systems may have various (secondary) goals. Most notably, companies that embed a recommender as part of their services usually aim to sell something and would like to improve customer retention [62]. In other cases, recommender systems may seek to promote a specific type of behavior that is in line with the values of the system's designer. This way, a recommender system may go beyond being a simple decision-support system and also become a persuasive tool [45, 63].

By definition, recommender systems differentiate between the algorithm and the interface [7, 29]. The effects are typically studied separately [30], as many studies specifically examine algorithmic improvements using prediction models (i.e., through 'offline evaluation' [29]), while other studies focus on comparing the user evaluation of different interfaces [64]. This means that the effectiveness of changes in the interface is always examined in conjunction with a personalized algorithm.

In this chapter, we examine changes in the recommender interface through the lenses of two fields: Persuasive technology and behavioral economics (i.e., nudging). Persuasive technology is defined as technology that aims to change a user's attitude or a specific behavior or a set of behaviors [1]. Nudges are changes in a choice architecture that lead to predictable behavior [19].

2.2.1 Persuasive Technology

Persuasion can be described as part of the communication-persuasion paradigm [65]. A persuasion attempt involves a source (i.e., a person or system sending a message), the message itself (e.g., advice), the target (e.g., a system user), and the effect (e.g., attitudinal change). The recommender interface or system itself can be considered the source, for which trust is important. The message is, however, arguably more important because it entails the recommendations provided by the system.

Persuasive technologies use persuasion and social influence to affect human behavior [1]. Most persuasion approaches aim to affect attitudes. In turn, this is expected to affect behavior, following various attitude-behavior paradigms proposed in psychology [52, 59, 66]. Strategies that are typically exploited tap into psychological theories [67], such as instruction styles (e.g., authority arguments), social influence, and motivational frames (e.g., intrinsic, gain vs. loss framing) [1]. Besides attitudes, another goal is to raise the self-efficacy levels of users, for example by helping them to take the first step toward a new behavior or habit. In the energy domain, persuasive technologies have helped users to manage their electrical appliances [68].

Algorithmic tailoring is just one of many persuasive strategies to affect user attitudes and behavior. Even though recommender systems can do more than tailoring, their influence is often collapsed into single concepts. For example, in the context of persuasive communication, tailoring and the effects of recommender systems are often referred to as 'algorithmic persuasion' [69]. Other taxonomies differentiate between adaptation (i.e., using personal characteristics) and personalization (i.e., using a person's name) regarding tailoring strategies [70]. What is also often overlooked in many of these approaches is the extent to which an approach is tailored, as expressed by the predictive accuracy in recommender system research. For example, the effectiveness of an approach tailored toward a user's gender and location may be far less effective than one that also considers a user's past behavior. In various studies, however, any form of user-based targeting 'counts' as adaptation or personalization [70], without taking heed of accuracy or classification metrics as is common in recommender system research. Moreover, the role of interface design is often overlooked or considered as not being part of the recommender.

2.2.2 Nudging

A nudge is defined as a change in a choice architecture that leads to predictable changes in human behavior, without forbidding any options or changing their economic incentives [19]. This definition stems from behavioral economics, where deviations from economically Pareto-optimal behavior are considered ‘cognitive biases’ or irrational [71]. The ‘predictable’ in the definition taps into these cognitive biases. For example, humans tend to place more weight on possible losses when making decisions under risk or uncertainty [72], which is expected to affect each decision over time. Overall, changing how choice options are presented (i.e., framing) or organized (i.e., changing prominence, salience or number of options) tends to lead to an average change in user choices.

Not forbidding options is an important aspect to consider for recommender systems. They typically determine what is feasibly shown to end users and therefore may hide other content. For example, a healthy eating intervention in a supermarket where products with a positive nutrition label are presented at eye level counts as a nudge, whereas a supermarket that bans unhealthy products from the shelf would not [19, 73]. In the context of recommender systems, a choice architecture thus mostly refers to *how* choices are presented, such as through the number of options, the ranking of the list, or informational framing. Incorporating such nudges in the design of recommender systems has become more common in recent years [18, 22], but there is little research specifically on ‘what happens’ to persuasive elements and nudges if they are added to a tailored recommender system, compared to when they are examined in isolation or in one-size-fits-all choice contexts.

2.2.3 Why Tailoring Is Not a Nudge

Discerning nudges and tailoring methods conceptually is one of the key challenges in this chapter. We argue that tailoring is not compatible with the nudging definition for recommender systems, because in most domains (with a large number of items) content filtering leads systems to hide a lot of options that are not shown.

Before delving deeper into our argument, we first discuss the opposite viewpoint, namely that changes in the recommended content (i.e., tailoring) would be a nudge (cf., [20]). Jesse and Jannach [18] argue that recommendations can also be considered nudges, for they reduce the physical effort of searching for alternatives, make a complex choice more structured, and apply filters on the content (i.e., ordering and hiding). They conceptually differentiate between four types of nudges: decision information (e.g., salience), decision structure (e.g., partitioning of choices), decision assistance (e.g., commitment), and social decision appeal (e.g., social influence). While decision information and social decision appeals both relate to information presented on screen, Jesse and Jannach [18] only consider information related to the recommended items for ‘decision information’. Decision structure is conceptually closest to the nudging definition, for it includes the composition of options, choice defaults, and the reduction of choice-related effort. The latter is, for instance, related to a reduction in the number of options shown or simply tailoring, which lies at the core of a recommender algorithm.

Contrary to Jesse and Jannach [18], we argue that tailoring is *not* a nudge when considering the design of a recommender system. We argue to conceptually separate the *what*, i.e., the items and content, from the *how*, i.e., the interface aspects. The latter also includes, for example, what information from the content is shown, such as framing effects. The selection of the best options for a user, constraint on the available information and computational capacity, is an inherent part of a recommender system. As some options need to be shown to a user while other options would be hidden, the tailoring strategy employed by a recommender system promotes the presentation of the most relevant items in a specific order. This is a step away from the definition of nudging: not all options in the choice architecture are still freely accessible. For example, on video-streaming platform YouTube, the search function could overcome front-page recommendations, but some of the near billion videos are most likely not found that way. In this case, the recommender system's tailoring strategy is profoundly different from any of the standard nudging strategies: If it does not show certain videos, users are practically 'forbidden' from choosing these options. Or, to phrase it in terms of behavioral economics, the incentives of non-shown items have significantly changed.

Moreover, some past studies have examined tailoring in recommender systems and other nudges simultaneously [18, 28]. Tailoring strategies still allow for further persuasion and nudging, for tailoring merely determines what content best fits a user. Designers of a recommender system could still decide to further re-rank a given list of tailored items, to explain them in terms of specific attributes or social norms, or to re-organize them visually. Such more 'classical' nudges or forms of persuasion are compatible with a tailoring strategy.

Recommender systems, in the past and the future, present content that is tailored. The application areas deviate from 'defaults' or 're-ranking' because some items are hidden or very difficult to find for users. We consider this to be a ground truth, and a base on which the effectiveness of choice architecture design needs to be further tested. Not only has there been a transformation of various 'offline' nudges to the digital realm, they have usually been tested in one-size-fits-all contexts. The application of these to a tailored environment is bound to lead to different outcomes: For one, because the fit between the content and the user might deem the user to not be 'distracted' by 'presentation effects'. If the content already fits the user as it is adequately tailored, nudges like defaults or ordering within the tailored content might have smaller effects than if the content is not tailored toward the user.

The examples described later in this chapter (from Sect. 3 onwards) combine tailoring and additional nudging strategies in a single recommender system. Most of the results point toward limited effects of the additional nudging 'on top of' the tailoring algorithm. All of these recommender systems are evaluated through the evaluation framework of Knijnenburg and Willemsen [74]. This approach discerns between different types of system aspects to form a path model that examines mediated effects of a recommender system. To support the understandability of this chapter, we include a primer that outlines the details of this evaluation framework. If this is already familiar to you, please skip to Sect. 3.

2.3 A Primer on Explaining the User Experience of Recommender Systems

This chapter includes various examples of the evaluation of recommender systems. All evaluations are performed in accordance with the evaluation framework of Knijnenburg et al. [61], a method to explain the user experience that is gaining in popularity, but is not applied in all online evaluations of recommender systems.

For computer scientists, the main topic of investigation is to optimize the accuracy of personalization algorithms. To this end, data sets in different domains are evaluated by performing a so-called 'offline' evaluation [7]. The accuracy of different recommendation models are examined by performing a train-test split on the data and assessing the predictive accuracy across all items in a data set [36]. For example, if the actual rating predicting for a movie is 4 out of 5 stars, a prediction model may output 3.9 out of 4, leading to an error of 0.1. Such differences are often taken at face value, in the sense that the smaller the error rate, the better. Moreover, while tailored approaches are rarely quantified in social science studies, the accuracy can also represent how tailored an approach is.

This algorithmic accuracy is typically only the first step of recommender system evaluation. As soon as the best performing algorithm(s) is/are selected, an online evaluation will follow suit. There, it is examined to what extent users prefer to interact with these different algorithms and how they are perceived. Since the offline evaluation only considers the *predictive* accuracy, it could be that a user's perception is completely perpendicular. For example, a news recommender system that leverages visual feature of photos in news article may enhance their predictive accuracy, even though this does not corroborate with a user's idea of accuracy or similarity [75].

To make sense of these different types of metrics, their interrelations can be examined in a path model. To this end, several recommender system evaluation frameworks have been proposed in the past decades [61, 76, 77]. In this chapter, we focus on the evaluation framework proposed by Knijnenburg and Willemsen [74], as described in the Recommender Systems handbook in 2015. We illustrate the different concepts based on the interventions and measures from our own studies, which we will discuss in more detail in Sect. 3.

The evaluation framework discerns between six different types of aspects [61, 74]. Five of these are used in this chapter, for which an overview is presented in Fig. 1. At the center are the *objective system aspects* of a recommender interface, of which the effects are measured across different groups of participants, such as in an A/B-test or between-subject design. Objective aspects comprise different recommender algorithms (e.g., different approaches or different candidates from a single algorithm

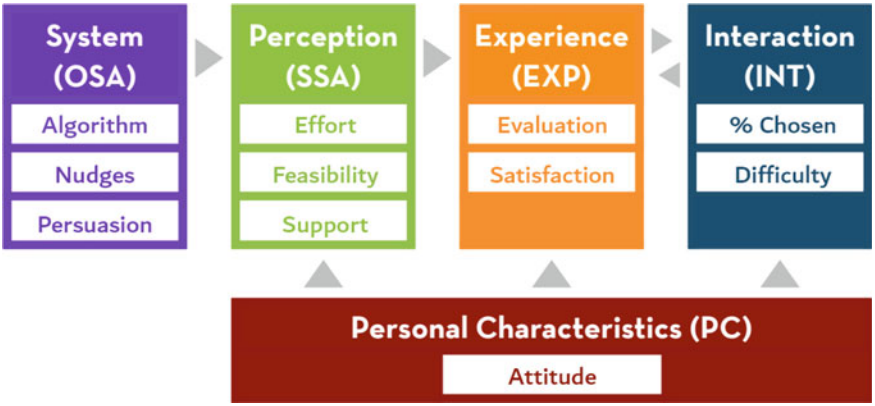

Fig. 1 The user-centric evaluation framework for recommender systems, adapted from Knijnenburg and Willemsen [74]. The examples depicted are used later in this chapter

[78]), changes in the interface design (e.g., different slider defaults [79], or list configurations [80]), and other adaptations that can be assessed 'objectively'. This is central to the evaluation, as research questions typically examine whether changes in the objective aspects lead to changes in other aspects. Objective aspects can lead to changes in three categories of other aspects. These are typically mediators or dependents in such a path model analysis.

The objective system aspects under investigation are not only different algorithmic approaches. In fact, we seek to test a single algorithm for which we take different recommendation strategies. For example, we compare attitude-tailored advice to an approach where the most popular energy-saving measures are presented. Moreover, we examine the addition of different list configurations (e.g., partitioning) and the relative effectiveness of different nudges in an interface.

Subjective System Aspects are central to most online evaluation approaches. They typically act as a mediator between the objective and interaction aspects [74]. These stem from evaluation questionnaire that aim to isolate how a user perceived a specific aspect of the entire system. In this chapter, we focus on how effortful an interface is perceived to use [25], the understandability of the interface [28], and other perception aspects that either concern the algorithm, the nudges in the interface, or the interface as a whole.

Beyond subjective aspects are *interaction aspects* and *experience aspects*. Interaction aspects include a user's behavioral traces of using a system, as well as behavioral outcomes, such as the characteristics of the items chosen (e.g., kWh savings) or self-reported behavior. These can act as a mediator in the path model, where changes in an objective aspect affect what does behaviorally, which may in turn affect experience aspects. The user experience is the overall evaluation of the user, in relation to the system, process, or the outcome. An important outcome variable in this chapter is choice satisfaction (i.e., level of satisfaction with the items that are chosen), for it may predict user retention for a given system, as well as positively affect actual energy-saving behavior [25]. Finally, *personal characteristics* may moderate the relations between the other aspects, or act as a control or covariate. For example, adding the user's attitudinal strength to a model may help to explain part of the variance in the amount of kWh savings chosen.

Using these aspects, the evaluation should consider whether an approach or interface helps a user to attain an energy-saving goal. Changes to the algorithm or interface should not only lead to different choices, but also lead to more positive perceptions, an acceptable level of system satisfaction or choice satisfaction, which can have a positive effect on behavioral intention or retention indicators. Since many evaluation methods are still session-based, which in principle only allows for cross-sectional conclusions to be drawn [81], the evaluation should be extensive to understand whether a user is likely to engage in more goal-based behavior, such as energy conservation, in the future.

The examples presented later in this chapter focus mainly on short-term changes, but also examine self-reported behavior four weeks later. For the path analyses, the effects are tested in Structural Equation Models, which can be considered as a combination of confirmatory factor analyses for latent aspects and simultaneous regression analyses. The main strength of this method is to examine mediated effects. For example, a certain algorithm may trigger changes in a user's quality perception, which only in turn leads to changes in that user's experienced level of choice satisfaction. For further details, please refer to the chapter by Knijnenburg and Willemsen [74].

As mentioned before, an important assumption for many recommender approaches is that past preferences are representative for future choices [7]. This way, human preferences would be represented by historical data. This stability of preferences would not only mean that 'similar' recommendations can be successful over time, but also that they would not be subject to the decision environment.

This is, however, not consistent with how human preferences are formed. Preferences are typically constructed when making a decision [82], being strongly dependent on a user's underlying attitude as well as the decision-making interface at hand [73]. An important aspect is the choice architecture of the interface [83], which refers to how options are presented and how this leads to predictable changes in behavior: a nudge [19]. A large body of literature from psychology and behavioral economics has demonstrated that humans are subject to systematic biases [67, 71, 72], which can affect their decision-making and possibly, in turn, their short- and long-term preferences. Incorporating these principles into a recommender system design should thus lead to more representative output and user decisions.

3 Psychologically Informed Energy Recommender Systems

This chapter examines the household energy conservation domain for its recommender technology. This involves individual actions to mitigate one's environmental impact [4]. Most measures that can be taken aim to reduce one's energy consumption at home, reducing electricity and gas consumption [57, 58].

It is clear that recommender systems aim to predict what a user will like and, consequently, will choose, buy, and/or use. It is, however, rather domain-dependent what can be concluded from different user interactions and dependent variables. For example, it is reasonable to expect that a user choosing a movie will watch it in the short term, but much less is known about whether and when people selecting recipes in a food recommender system will actually cook them [40].

The domain of energy conservation involves the aspect of behavioral change [84]. This can be costly, in the sense that an individual needs to change their habits and to investment in energy-efficient appliances. This also involves a lot of time and cognitive effort, which cannot materialize from a motivational vacuum [8].

For a psychologically informed recommender system to be feasible, a clear dimensionality of the domain is required. The next subsection discusses what attributes and features play a role in the energy-saving domain, after which we provide examples of psychology-based user modeling and interface nudging in an energy recommender system. We discuss studies from Starke et al. [27, 28], where examine whether psychologically informed approaches are effective, regarding chosen energy savings and the user evaluation.

3.1 Conceptualizing and Predicting Energy-Saving Behavior

The dimensionality of energy conservation determines how energy-saving behavior should be measured. The scientific literature, however, does not provide an unambiguous answer [58]. Whereas questions of dimensionality in recommender systems tend to involve data-driven solutions [7], typically through a reduction of dimensions (e.g., with Singular Value Decomposition), such questions are approached more theoretically in psychology. Some studies examine whether specific energy-saving attributes represent a defining characteristic, with which behaviors can be categorized [57, 85]. One popular distinction is two-dimensional [86], based on whether energy-saving measures involve investment costs. This differentiates between curtailment behaviors (i.e., reductions of existing behaviors or habits, such as turning off lights after leaving a room) and efficiency (i.e., one-time investments to make appliances more energy-efficient, such as buying a freezer with an A+ EU energy label) [55].

Various studies have proposed different representations for energy conservation [52, 57, 58]. These vary from a single dimension [48], all the way up to and beyond four dimensions [57]. For example, Dietz et al. [85] differentiate between behaviors that are for daily use, maintenance, adjustment, or weatherization. Defining attributes include the frequency of the behavior (e.g., daily or once every three years), financial costs (i.e., investment costs and maintenance), and the length or effort of performing the behavior. For example, the dimensionality of Boudet et al. [57] includes 'weekend project' to refer to energy-efficient investments that also require a lot of time.

Such multidimensional representations can be used by recommender systems if they are also included or operationalized as features in a data set. Some of the energy-saving features, however, involve trade-offs [87] that need to be considered in decision-making models, such as in a critiquing-based recommender [88]. For example, if a user prefers to perform behaviors that score low on behavioral frequency, a likely consequence is it will score higher on investment costs [33, 35]. Another issue is that some of these representations are not formally linked to user characteristics. For example, the four-dimensional approach of Boudet et al. [57] and its nine attributes are not formally linked to user characteristics [54].

More generally, an important behavioral determinant is one's environmental or energy-saving attitude [87, 89]. An attitude is a psychological construct that encapsulates an individual's general feeling or opinion about something or someone. It represents a mental and emotional state that is an acquired by an individual, which characterizes that individual's feelings toward an attitudinal object [90]. An attitude is not tangible, but can typically be measured when asking for evaluative statements [87]. For example, environmental attitudes can be assessed by inquiring whether an individual is concerned about the environment and whether humans or nature should be prioritized [91]. In most domains, attitudes are typically not as flexible as emotions, for they are formed through longer term exposure and experience.

Attitudes are used in theoretical frameworks to predict behavioral intention and behavior. A popular approach is the Theory of Planned Behavior [59], which predicts whether an energy-saving behavior will be performed, as determined by a person's attitude, the presence of subjective norms, and perceived behavioral control. This aims to explain, in part, why an attitude-behavior gap may exist, which refers to the situation where a person has a positive esteem of a behavior, but cannot or does not wish to act due to other reasons [52, 92]. Attitude-behavior gaps can occur if the contextual behavioral costs are too high. Consider, for example, a European citizen with a strong pro-environmental attitude who uses the plane to visit relatives in North America, due to a lack of alternatives. Another example are climate activists, who are accused of hypocrisy if they appeal for pro-environmental behavior while not fully acting environmentally friendly according to outsiders [93, 94].

Such an attitude-behavior gap is undesirable from the point of view of a recommender [56]. For example, in collaborative filtering, historical user data is assumed to reflect user preferences now and over time [7], without making substantial inferences about specific characteristics. Although such 'preference inertia' is rather naive, for user preferences tend to be strongly decision context-dependent [82, 95], it is more effective to use an algorithmic approach with which users can be matched to items more directly, thus to design an algorithm in which there is a formal relation between a user attitude and item characteristics [56].

The studies we report in this chapter rely on a one-dimensional representation of energy-saving behavior [48, 58]. It follows the rationale of Campbell's Paradigm [52], an attitude theory that is named after psychologist Donald Campbell [66]. It postulates that one's disposition or attitude toward energy conservation becomes apparent through the increasingly difficult behavioral steps an individual is willing to take to save energy [27, 52]. Indeed, *"actions speak louder than words"*, which is formalized by differences in execution difficulty between behaviors: while *saying* that protecting the environment is relatively easy, it requires cognitive costs, money, and time to actually take an energy-saving measure [58, 96], such as installing solar

PV on one's rooftop [27]. In Campbell's Paradigm, attitudes are formally *behavior-based*: they become stronger if a user engages in more or more difficult behaviors [24, 27, 52]. This approach addresses the gap between *evaluative* attitudes and behaviors [96].

To operationalize Campbell's Paradigm, a formal relation between users and behaviors is required. This way, a user model can be created that predicts which behaviors are appropriate for users to recommend, which can be implemented in an energy recommender system. The next subsection shows how this form of *psychologically informed user modeling* has been implemented in studies on energy recommender systems [27].

3.2 Examples of Attitude-Based User Modeling and Tailoring

The first example concerns the development of so-called Rasch-based energy recommender systems [27]. Hence, the psychometric Rasch model was used to formalize the relation between a user's attitude and a measure's behavioral costs are formalized [52]. This is an item response theory model that predicts whether a user would perform a particular behavior [97]. It accomplishes this by representing an underlying characteristic (i.e., a latent factor) as a function of the challenges faced by a group of characteristic-related items [98].

3.2.1 The Rasch Scale

The Rasch model is formalized as follows. It is a logistic function that predicts whether an individual n performs a measure i, through the arithmetic difference between a user's attitude θ and a measure's behavioral costs δ. Equation 1 shows that an increase in one's attitude increases the probability that one performs an energy-saving measure [52, 97, 98]:

$$P(X_{ni} = 1) = \frac{e^{\theta_n - \delta_i}}{1 + e^{\theta_n - \delta_i}} \tag{1}$$

It is assumed that each measure's probability distribution, which is defined as an item-characteristic curve [98], is shaped as a sigmoid function. This is depicted in Fig. 2. The three measures depicted vary in terms of how difficult they are to perform. Their behavioral cost levels can be inferred from the x-axis at the 50% probability points, as the engagement probability is 50% if user attitude and behavioral costs are equal. For example, Fig. 2 shows that an individual with an attitude of 0.9 logits has a 50% probability of using a water-saving shower head.

In the context of energy conservation, Rasch maps energy-saving actions and individuals onto a single measurement scale [24, 27, 48, 52]. The extent of effort (i.e., the behavioral difficulty or behavioral costs) required to perform a behavior is

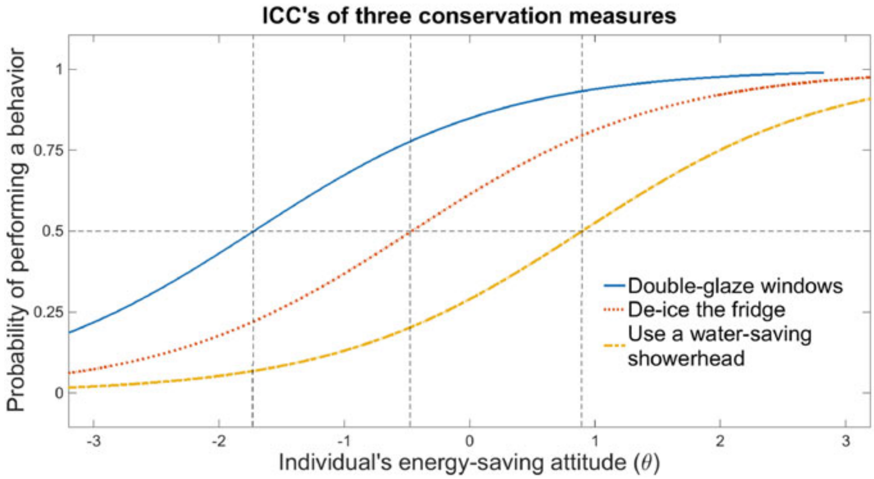

Fig. 2 Item Characteristics Curve of three energy-saving measures. The behavioral cost levels (δ) are defined as the point where an individual with attitude θ has a 50% probability of performing it. Image adapted from [27]

directly related to the number of people in a sample who have reported to perform it. Actions that are commonly taken result in lower effort costs compared to those undertaken by fewer individuals [54, 96]. Similarly, a person's energy-saving attitude increases in line with the number of actions they have undertaken from the scale [48].

This rationale is depicted in Fig. 3, showing five example measures. Some energy-saving measures, such as turning off lights after leaving a room, are performed frequently and are therefore deemed to have low behavioral costs. Consequently, a person with only a weak energy-saving attitude, who performs few measures overall, would be likely to perform that measure. Further up the scale, measures such as 'install solar PV' and 'erect a small windmill' are depicted as having high behavioral costs, because fewer people perform them. As a result, those measures are only feasible for users with strong energy-saving attitudes.

A Rasch scale can be formed by collecting engagement data from a group of people for a set of measures. For the dichotomous Rasch model in Eq. 1, user-item engagement data can either be 'yes' or 'no', regarding whether a measure is typically performed or is installed at home, but partial credit modeling with additional response options is possible [98]. The fit of the persons and, most importantly, the measures on the one-dimensional scale are through mean-square fit statistics [58, 98]. This compares the observed response values with those that are predicted by the Rasch model, computing the mean of sums of squared standardized residuals. This is the difference between observed and predicted response score standardized by its variance. In doing so, so-called *infit* statistics are used to assess item fit, where infit is the mean sum of squares weighted by variance of the residuals, and 1 indicates perfectly fitting data [58].

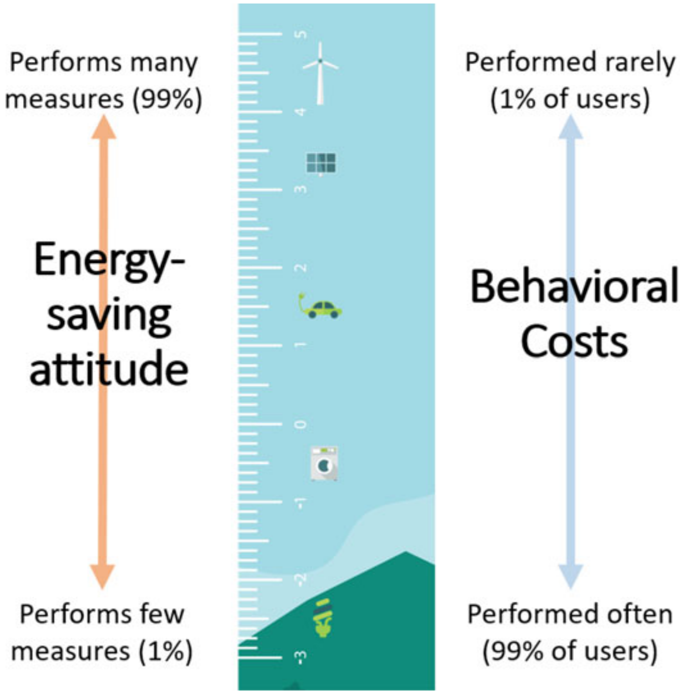

Fig. 3 Schematic overview of the Rasch scale used in studies of Starke et al. [27]. It comprises a one-dimensional construct on which both persons and energy-saving measures can be scaled, in terms of their energy-saving attitude and behavioral costs, respectively. Persons whose attitude equals a measure's behavioral costs have a 50% probability of performing that behavior. Depicted are relatively low-cost behaviors, such as turning off lights after leaving a room, and high-cost behaviors, such as installing Solar PV on one's rooftop or erecting a small windmill

An initial Rasch scale was formed in a preliminary study [24, 27]. Through a sample with 263 Dutch-speaking participants (57% male), recruited through convenience sampling, a set of 88 items was rated on their engagement levels. Participants were asked to indicate their engagement levels for a subset of measures, answering the question whether they already performed a measure with either 'yes', 'no', or 'not applicable', whether the latter was reserved for measures incompatible with one's housing situation.

The column labeled δ_{Study1} in Table 1 reports the order of difficulty of the measures. These could all be considered behavioral steps toward the goal of saving energy [48]. Our analysis in Winsteps software [99] showed that a one-dimensional construct could be inferred reliably [27]. The first dimension explained 36.9% of the variance, while what remained was residual quantification variance that was caused by the Rasch model's estimated probabilities for discrete, dichotomous events (0 or 1) [27]. A principal component analysis (similar to SVD) on the residual variance would only lead to a trivial increase of 1.8% in the explained variance, providing further evidence for a unidimensional model.

Referring to the earlier mentioned dichotomy [55, 86], it was found that curtailment and efficiency measures were distributed across the scale. However, similar to previous findings [58, 100], curtailment measures were easier to perform and more likely to be performed, compared to efficiency measures. Curtailment measures ($n = 46$) had on average a lower behavioral cost level (−0.67 logistic units) than efficiency measures ($n = 33$; 1.08). Nonetheless, Table 1 shows that some curtailment measures had high behavioral costs (e.g., keeping the rear of a refrigerator dust-free, $\delta = 2.43$), while some efficiency measures were found to have low behavioral costs (e.g., using a laptop instead of a PC, $\delta = -3.45$) [27].

3.2.2 Evaluation Study

Subsequently, Starke et al. [27] examined user preferences in a tailored advice scenario. Although the Rasch scale was expected to be informative about what measures would be appropriate for which user, it was not clear yet whether this difficulty order resonated with users of a recommender system.

We investigated how users perceived advice from a psychologically informed energy recommender system. Each user would be presented with two lists of nine energy-saving measures, which were tailored around their attitude: 3 measures that were −1 logit below their attitude, 3 measures equal to their attitude, and 3 measures +1 logit above it. For each list, users had to remove measures they already performed, after which they were asked to rank-order them in terms of their own preferences.

A group of 196 participants from the JFS participant database and Eindhoven University of Technology completed this task. To determine each participant's energy-saving attitude, they were first presented 13 semi-randomly sampled energy-saving measures. The strategy divided the scale in Table 1 into 13 subsets based on the δ level, randomly selecting one measure from each subset, ensuring that users would be presented both easy and difficult measures. For each measure, they were asked to indicate whether they performed it, either responding 'yes', 'no', or 'not applicable'. The resulting attitude was determined through the total number of 'yes' responses of each participant (with a small correction for 'not applicable' responses), taking the average behavioral cost level of the corresponding set in Table 1.

A total of 279 ordered lists were analyzed through rank-ordered multilevel logistic regression analyses [27]. A measure's final rank was predicted through the difference between a user's attitude and a measure's behavioral costs, i.e., the *attitude-cost difference*. In addition, a number of control variables were used: a measure's starting position in the list, whether a measure is of the curtailment type, a measure's estimated kWh savings, and a measure's investment costs.

The evaluation study showed a couple of things [27]. First, when analyzing the ranking of measures that a user did not perform yet, we found that measures that fell below a user's attitude were more likely to be ranked higher. This indicated that among tailored advice, users were more likely to prefer relatively easy measures.

Table 1 List of energy-saving measures fitted onto the one-dimensional Rasch scale. Measures were used in research conducted in the Netherlands in 2014, which involves gas for heat. Reported are each measure's behavioral cost levels (δ), based on the pre-study ('Study 1' [24, 27]) and the evaluation study ('Study 2' [27])

#	Name of energy-saving measure	δ_{Study1}	δ_{Study2}	Set
1	Save up laundry	−5.73	−3.23	1
2	Take a shower instead of a bath	−4.82	−4.41	1
3	Wash laundry at low temperatures	−3.95	−1.64	1
4	Air-dry laundry	−3.69	−2.93	1
5	Use a laptop instead of a desktop PC	−3.45	−3.62	1
6	Turn off the lights after leaving a room	−2.97	−2.78	1
7	Use public transportation instead of a car	−2.90	−2.52	1
8	Use a woolen blanket instead of an electric blanket	−2.51	−3.03	2
9	Use properly sized cooking equipment	−2.51	−2.69	2
10	Lower the thermostat while away from home	−2.49	−1.92	2
11	Do not put warm things in the fridge	−2.45	−2.40	2
12	Turn off the PC screen after use	−2.20	−0.71	2
13	Close the curtains/shutters in the evening	−2.09	−1.57	2
14	Shift gears at low speeds	−1.89	−2.07	3
15	Cook with a lid on the pan	−1.81	−1.81	3
16	Use energy-saving bulbs (CFL's)	−1.75	−1.31	3
17	Double-glaze windows	−1.72	−1.24	3
18	Air rooms for 20 min daily	−1.51	−1.21	3
19	Cook on gas stove instead of electric	−1.36	−2.23	3
20	Lower the thermostat one degree	−1.25	−0.47	4
21	Set thermostat to 14 °C before going to bed	−1.20	−0.90	4
22	Do not defrost food using a microwave	−1.18	−0.52	4
23	Turn off the TV instead of stand-by	−1.06	−0.74	4
24	Maintain correct tire pressure	−0.94	−0.27	4
25	Stir-fry food	−0.91	−0.62	4
26	Turn off the PC at the main switch	−0.81	−0.21	5
27	Turn off the coffee machine completely	−0.57	−1.30	5
28	Turn off the dishwasher after use	−0.57	−0.37	5
29	Insulate the cavity wall	−0.51	0.50	5
30	Turn off the washing machine completely	−0.49	−0.22	5
31	De-ice the fridge	−0.46	0.59	5
32	Unplug chargers	−0.32	−0.44	6
33	Take short showers	−0.29	0.52	6
34	Hand-wash dishes (no dish washer)	−0.22	−0.72	6
35	Configure PC power management	0.00	0.29	6
36	Shorten PC/laptop stand-by time	0.01	−0.43	6

(continued)

Table 1 (continued)

#	Name of energy-saving measure	δ_{Study1}	δ_{Study2}	Set
37	Air clothes instead of washing them	0.07	−0.39	6
38	Clean the cooker hood suction filters	0.14	0.29	7
39	Place fridge in a suitable position	0.18	−0.14	7
40	Use LED lighting	0.37	0.38	7
41	Decalcify your coffee machine and/or kettle	0.40	0.35	7
42	Sweep instead of using a vacuum cleaner	0.43	0.40	7
43	Use a smart thermostat	0.47	−0.11	7
44	Put a weather strip on the door	0.47	0.19	8
45	Use a HE boiler or CHP	0.47	0.98	8
46	Use household devices without displays	0.48	2.67	8
47	Use an ‘A+’ energy-class fridge	0.51	−0.09	8
48	Install motion sensors	0.51	−0.02	8
49	Insulate floors	0.56	0.29	8
50	Use a mini PC instead of desktop computer	0.59	3.89	9
51	Make coffee without using a heating plate	0.74	−0.84	9
52	Decalcify the washing machine	0.75	0.59	9
53	Use green power	0.85	0.22	9
54	Turn off the fridge while on holiday	0.87	1.84	9
55	Turn off the PC when away from keyboard	0.88	−0.74	9
56	Use a water-saving showerhead	0.90	0.34	10
57	Put your shirts briefly in the laundry dryer instead of ironing them	0.96	1.15	10
58	Cover the windscreen of your car	0.96	0.81	10
59	Replace dimmer switches	0.99	1.04	10
60	Use an ‘A-label’ energy-saving laundry dryer with a heat pump	1.17	1.36	10
61	Use day and night tariffs	1.21	0.75	10
62	Set boiler temperature to 65 °C	1.24	1.04	11
63	Set the mixing valve at a lower temperature	1.31	1.17	11
64	Put weather strips on the windows	1.46	0.59	11
65	Insulate hot water pipes	1.53	0.79	11
66	Clean the water heater	1.63	1.26	11
67	Install a door closer	2.34	1.39	11
68	Turn off the oven before the end of cooking time	2.40	2.27	12
69	Keep the rear of the fridge dust-free	2.43	1.25	12
70	Apply heat reflection foil to radiators	2.77	3.46	12

(continued)

Table 1 (continued)

#	Name of energy-saving measure	δ_{Study1}	δ_{Study2}	Set
71	Replace a radio alarm with a ‘classic’, unplugged alarm clock	2.88	3.51	12
72	Use a cabled telephone instead of a handheld phone	2.91	2.49	12
73	Install solar PV	3.17	1.60	12
74	Install a solar boiler	3.26	2.53	13
75	Slow down the PC processor	3.70	3.47	13
76	Use a pull bell instead of an electrical bell	3.82	3.20	13
77	Wash using a ‘hot-fill’ washing machine	4.04	2.36	13
78	Use software for dynamic energy use in a laptop or PC	5.18	0.52	13
79	Erect a small wind mill to produce electric energy	5.49	4.42	13

This effect is also depicted in Fig. 4, where negative differences indicated relatively difficult measures. These were, thus, ranked much lower. Second, we observed a *negative* effect of kWh savings on the predicted ranking, indicating that users were actually less likely to adopt measures that saved a lot of energy. This suggested that kWh savings were not a factor in user decision-making, but that expected effort or difficulty seemed to be a much more important predictor. Third, we observed small ranking bias, as measures that were ranked higher at the start were also more likely to be ranked higher in the end. In contrast, whether a measure was of the curtailment type did not affect the final ranking.

Finally, the data collected in the evaluation study are reported in Table 1. It also includes an updated scale using the data obtained in Study 2, which led to changes in the δ values, but only a few values were changed significantly. For those measures, such as “Use a mini PC instead of desktop computer”, it was advisable to exclude it from the scale in further studies.

3.2.3 Implications

It has emerged that users prefer relatively easy-to-perform measures, within the attitude-tailored advice context. Based on the Rasch scale used, there is a trade-off between feasibility and novelty when generating advice. On the one hand, the behavioral costs of measures should not exceed a user’s attitude too much, because this might prevent a user being able or willing to perform a measure at all. On the other hand, measures should be novel, in the sense that the user does not already perform them, which would make the advice redundant. In Starke et al. [27], further analysis reveals that tailored advice is likely to be more effective than simply presenting ‘popular’ or ‘middle-of-the-scale’ items, but this needs to be validated further in

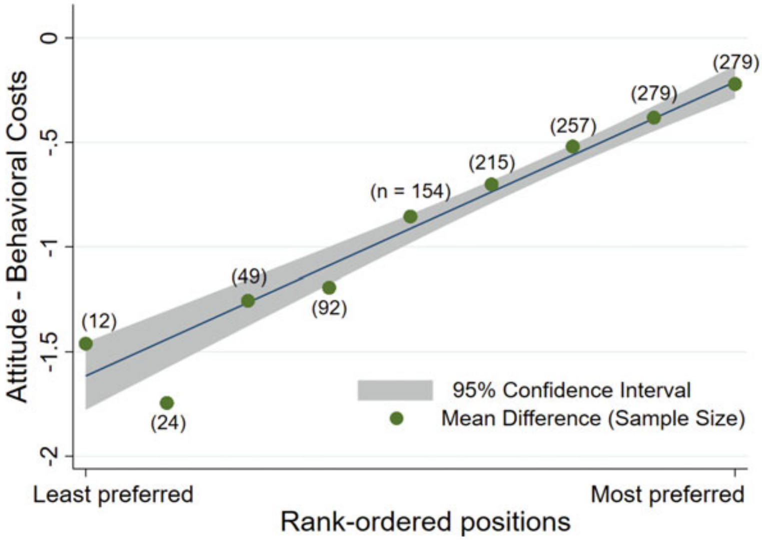

Fig. 4 Linear regression fit for the 'attitude-cost difference' on the rank-ordered positions of different measures. A negative difference indicated that a measure's behavioral costs were higher than a user's attitude. Users were asked to remove measures they already performed from lists, which led to varying sample sizes. Lists considered for analysis were at least 2 and at most 9 measures long

A/B testing. This is reported in the next section, reporting on the findings of a 2017 study by Starke et al. [25].

A possible confounding factor had a surprising effect. The amount of kWh savings of an energy-saving measure had a negative effect on the rank-ordered position of a measure [27]. While kWh savings were not depicted in the study's interface, one would expect that when inspecting behaviors or measures related to the goal of saving energy, users would look for higher gains. However, it seems that feasibility and the expected effort or difficulty is a more important predictor of user preferences, which is also observed in another study by Starke et al. [101]. Moreover, a possible correlation between efficiency measures leading to higher savings might also explain this negative impact [86].

Nonetheless, energy recommender systems are not *only* decision-support tools. There is also a clear persuasive intent [45], for there is an attainable goal of saving energy. Since the algorithm itself does not seem to persuade users to select measures with higher energy savings, Starke et al. [102] propose to make changes to the decision environment. They tap into literature on explainability [103], creating a fit score for a recommender interface, and examine literature from social psychology [104, 105], introducing a social norm nudge to support the adoption of energy-saving measures with higher savings. This presents another example of where psychology informs interface nudges that can affect potential behavioral change.

3.3 Combining Psychology-Informed Tailored User Modeling and Other Nudging Strategies

The tailoring algorithm is limited in what it can achieve in terms of persuasion. Whereas it may lead to more central processing of information [70], more elaborate strategies to affect user attitudes and choices may be required. As pointed out earlier, psychological theory can also be used to manipulate various aspects of the decision-making environment to lead to predictable behavior of the user [19]. This adheres to the definition of a nudge [83].

The predictability aspect of this definition encapsulates the insights from psychology. It uses mechanisms that have been found to determine attitudes and behavior. In this chapter, we use a number of mechanisms that play a strong role in affecting user preferences. This includes the use of defaults, partitioning, persuasive scores [25], and social norms [28, 102].

3.3.1 Defaults

A first example of psychologically informed interface nudging is the use of defaults. They represent the first choice or option configuration that a user encounters and can be a powerful tool. For instance, they have shown to increase the proportion of organ donors if an ‘opt-out’ system is used for registration, instead of an ‘opt-in’ system [106]. A smaller, but more common example is software installation, which allowed users to make custom changes, but the ‘standard setting’ without customization was typically set as the default, simply because this was the most appropriate for most users.

When used in the context of nudging, defaults are used to improve user decision-making [19, 73]. There are, however, more manipulative examples, such as airline Ryanair that presents additional options (e.g., fast lane boarding, insurances, mailings) as opt-out rather than opt-in, often leading to higher costs for the consumer. Using defaults for profit rather than better decision-making is nowadays referred to as a ‘dark pattern’ [107], a term used to describe phenomena from behavioral economics where a choice architect taps into predictive behavior for the benefit of the choice architect rather than the decision maker. While this term has also been used in relation to unfair algorithms [108], its use where psychology is applied as a manipulative tool is more common.

In their first study, Starke et al. [25] consider their Rasch scale algorithm as a method to generate ‘smart defaults’. Much like a recommender algorithm, smart defaults are based on a user model [21, 109], changing what options are presented first. Starke et al. [25] developed the ‘Saving Aid’ energy recommender system, on which users could received attitude-tailored energy-saving advice. The Dutch interface is depicted in Fig. 5, which presents five measures at a time. The attitude-tailored condition would designate which three measures best fit the user (i.e., ‘Aanbevolen’ is ‘Recommended’). Users could navigate the possible measures by clicking any of the

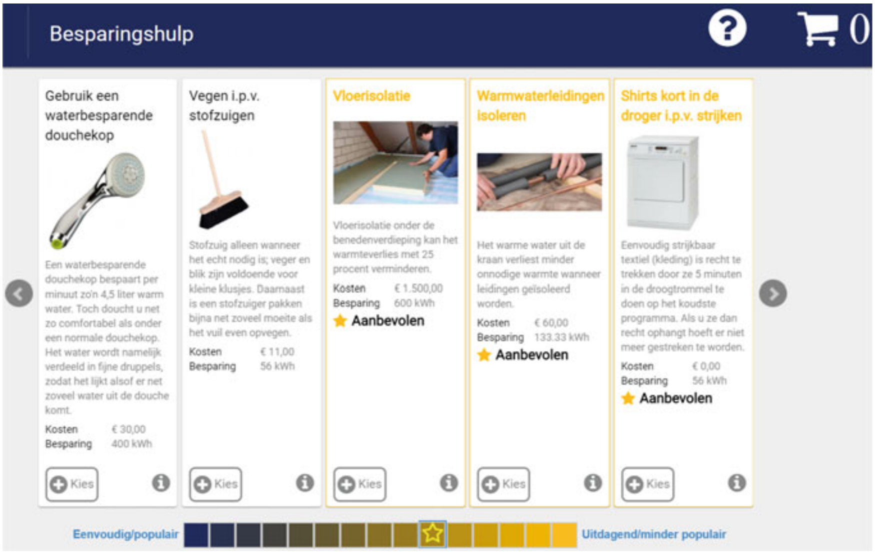

Fig. 5 Overview of the Saving Aid interface, as it was used for the energy recommender system study on defaults by Starke et al. [25]. The interface allowed users to explore the energy-saving measures on the Rasch scale by scrolling through them horizontally. The tailoring involved showing different options as smart defaults, based on the user's attitude

squares at the bottom, which were labeled from 'easy/popular' to 'challenging/less popular'. They could choose any number of energy-saving measures they pleased, which would be put in a digital shopping cart and sent to the user via email after the study was over, disclosing additional explanations.

Note that we refer to smart defaults for this research design. This is because the tailoring strategy and interface combined do not necessarily forbid any options, as with some effort all options can still be accessed. Nonetheless, the tailoring algorithm of Rasch does significantly change how much effort is required to find other items if a user is misplaced because of a specific default.

The Saving Aid interface used a 2x2 between-subject design [25]. On the one hand, users were either presented measures from the start (i.e., left-hand side) of the Rasch scale or attitude-tailored advice. On the other hand, measures were ordered in ascending or descending order of behavioral costs. Effectively, this meant that users were either presented attitude-tailored advice, based on an initial preference elicitation phase, or the most popular/easiest (i.e., ascending difficulty) or the most novel/difficult (i.e., descending difficulty).

The key of this study lies in its evaluation, which followed the user experience framework of Knijnenburg and Willemsen [74]. If a smart default would be further away from a user's attitude, based on the tailoring algorithm, it was expected that the interaction would be perceived as more effortful, simply because a user would need to click more to reach measures that fit that user's attitude [25]. This was also

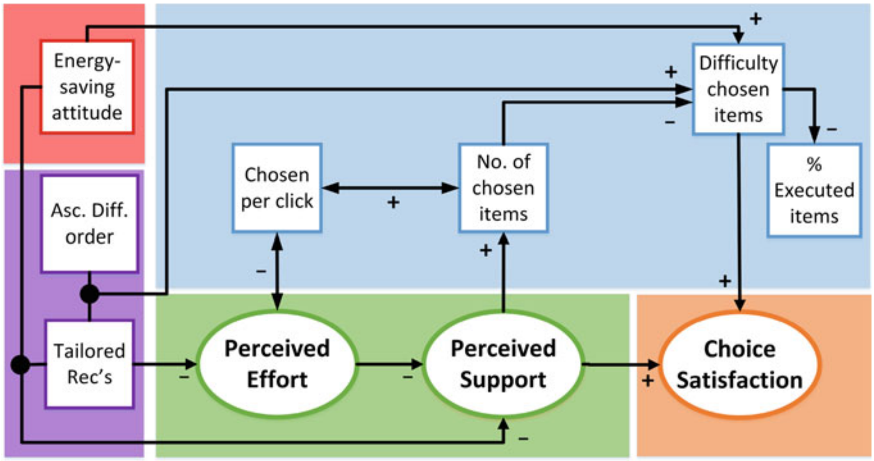

Fig. 6 Structural Equation Model (SEM) for the smart default study [25]. Relations between aspects resemble correlations. Coefficients and standard errors are omitted for simplicity. Details are deported in Starke et al. [25]. Colors follow the guidelines of Knijnenburg et al. [61]: Objective aspects are purple, observed variables are in blue, personal characteristics are in red, perception aspects in green, and experience aspects in orange. Difficulty is considered synonymous for behavioral costs

found in our analysis with $N = 209$ users, as tailored recommendations led a reduced effort perceptions, which in turn led to higher levels of perceived support and choice satisfaction. This is also depicted in a simplified Structural Equation Model, which is reported in Fig. 6, while full mathematical details can be found in Starke et al. [25]. These findings showed that a psychologically informed algorithm can lead to an improved user experience, as well as affect user choices.

In a similar vein, the 'most popular' and 'most novel' baselines were expected to be appropriate for users with a weak or strong attitude, respectively. These expectations did not fully materialize, but multiple attitude-related findings were presented [25]: (1) Users with stronger attitude chose more difficult measures (cf. Fig. 6), (2) users with stronger attitudes were less positive about tailored advice (possibly because these were so challenging), and (3) there was an interaction between the difficulty ordering and whether advice was tailored, showing non-tailored descending and tailored ascending led users to choose more difficult measures. This showed a relation between two psychologically informed recommender components, namely a user's behavior-based energy-saving attitude and the choice architecture design based on defaults and difficulty ordering.

In relation to this chapter, we examine the two main interventions as two different nudges that lead to a smart default. The tailoring strategy led to positive changes in user perceptions, also when a user's ability or attitudinal strength is considered. In contrast, the ordering had little effect on the user evaluation and only affected behavioral outcomes when combined with tailoring. This shows that additional nudging (i.e., re-ranking a list) is not very effective when the content is already tailored, a conclusion also found in our second example.

3.3.2 Fit Scores and Partition Nudges

Whereas Study 1 of Starke et al. [25] focuses on defaults, Study 2 uses two different mechanisms. An updated platform of the Saving Aid tool was developed, where instead of a 'horizontal' recommendation list, a 'vertical' interface was developed. This interface allowed users to explore a filtered selection of items, and thus applied a tailoring strategy. The interface was subject to a 2×3 between-subject research design: the measures in the recommended tab were either optimized for an attitude-cost difference of +1, 0, or −1 logit, while there was either a explanatory fit score present or not. The fit scores were based on the Rasch model. Depending on whether the condition was +1, 0, or −1 logit, items with a predicted adoption probability of 75%, 50%, or 25%, respectively, would yield a fit score of 100%. The probability distribution followed the item-characteristic curve as also depicted in Fig. 2.

The Dutch interface, with fit scores, is depicted in Fig. 7. It divided the best fitting energy-saving measures across three different sub-lists, based on their behavioral costs and the fit with the user: 'Basic', 'Recommended', and 'Challenging'. These were considered partition nudges [110]: a framing of a relevant item attribute

Fig. 7 Overview of the Study 2 in Starke et al. [25]. Depicted is the top of the Dutch Saving Aid energy recommender interface. Measures were divided across three 'tabs' that could be considered difficulty signposts: 'Basic' (NL: Basis), 'Recommended' (NL: Aanbevolen), and 'Challenging' (NL: Uitdagend). Within signposts, measures were ranked on their fit with the user's attitude, which is explained through the 'Match' fit score. For each measure, the estimated annual savings (in kWh and Euro) and investment costs are depicted

that might affect user decision-making, enabling users to look for items based on a prescribed item attribute. In related energy recommender research on signposting [111], attributes were changed from kWh to euro, examining whether this would affect which values a user taps into when making a decision. In Starke et al. [25], we tried to encourage the adoption of more difficult measures by making this salient, and framing it as a challenge. This tapped into the manipulation of fit scores and difficulty levels.

A sample of 288 users participated in a study with this interface, among which only 46 completed a follow-up questionnaire on their energy use four weeks later [25]. The attitudinal level of each user was assessed in the preference elicitation phase, after which they were presented attitude-tailored advice and could choose any number of measures they liked. Afterwards, they were inquired on their perceived feasibility of the advice, the perceived support from the interface, and their level of choice satisfaction. It was expected that fit scores would boost feasibility, while the recommended difficulty would have an impact based on whether a fit score was present or not.

A structural equation model (SEM) analysis was performed [25], based on the evaluation framework by Knijnenburg and Willemsen [74]. Multiple interaction effects between the fit score and either the recommended difficulty or user attitude were found (cf. Fig. 8). First, users with a strong attitude benefited less from the fit

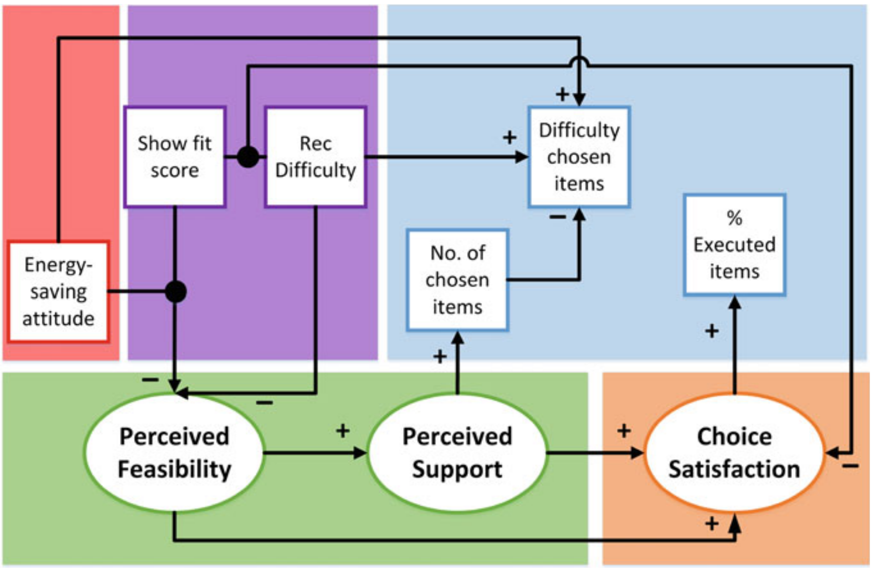

Fig. 8 Structural Equation Model (SEM) for the Fit Score study [25]. Relations between aspects resemble correlations. Coefficients and standard errors are omitted for simplicity. Details are deported in Starke et al. [25]. Colors follow the guidelines of Knijnenburg et al. [61]: Objective aspects are purple, observed variables are in blue, personal characteristics are in red, perception aspects in green, and experience aspects in orange. Difficulty is synonymous for behavioral costs

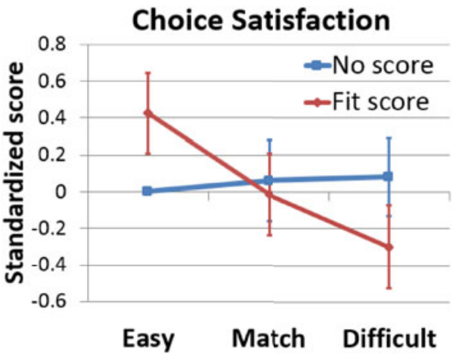

Fig. 9 Standardized scores for choice satisfaction in the fit score study [25], as a function of the research design: the attitude-cost difference of the recommended tab (easy = +1, match = +0, and difficult = −1 logit), and the presence of a fit score. Image adapted from Starke et al. [25]

score, as it decreased their perceived feasibility of the recommended measures. Second, higher behavioral cost or difficulty levels of advice also decreased feasibility, which was not mitigated by the fit scores. Third, there was a negative interaction effect between fit score and recommendation difficulty on choice satisfaction. Taken together, it seemed that the fit score was particularly unproductive for users with high ability levels. Measures that were relatively challenging in an absolute manner were presented with high fit scores, which high-ability users probably did not expect or found feasible. In contrast, fit scores were beneficial for users with a low attitude.

To contextualize the effects depicted in the SEM, we examined the effects on choice satisfaction in Fig. 9. It shows a flat line for the no-fit score condition, but a clear ‘easiness’ preference in the fit score condition that is in line with the 2020 study from Starke et al. [27]. The salience of how well a measure matched a user, in combination with the difficulty signposts, seemed to have backfired, as it did not match a user’s feasibility perception. This is an example of how a psychologically informed design of the interface can undermine the algorithmic rationale.

This study has made clear that it is challenging to overcome the tendency of users to prefer relatively easy energy-saving measures. Simply explaining the algorithm to the user, as is done in many recommender studies [103], seems to backfire in this domain. In our interface, the persuasive strategy to use a fit score was not able to overcome preferences triggered by the tailoring strategy. Hence, it seemed that the tailoring approach of the recommender system (i.e., ‘does the item really fit me?’) was a more important determinant of user choice and evaluation outcomes than the persuasive elements presented in the interface (e.g., ‘is the item promoted?’). In lieu of these findings, we examined the possibilities of using a different type of persuasive strategy or interface nudge that is still related to the algorithm, but uses ‘external’ information: the presentation of a descriptive social norm.

3.3.3 Social Norms

The final example provided in this chapter is the inclusion of a social norm nudge 'on top of' the Rasch model algorithm [28, 102]. This means that in addition to a tailoring strategy that presented only the best fitting measures, we used a social nudge. Social norms leverage the tendency of humans to compare socially to others, because of a lack of information about how to behave in general or what 'good' behavior is [105, 112]. A convincing message that affects what people like is one that describes a majority norm [112]. Indicating that a moderately large proportion of relevant peers engages in a certain behavior [105, 113] can persuade users of an HCI system to behave more socially desirable [114, 115]. While also 'injunctive norms' exist [11, 116], that state what people *should* do, we focus on 'descriptive norms', which simply state what others are doing [13].

Two main mechanisms explain why people act in line with social norms [117]. The first one is compliance, which is the propensity to act consistent with presented norms, and can be considered as a direct follow-up to a request. The second mechanism is conformity, which is the act of adapting one's behavior to match an apparent majority. Both mechanisms can fulfill a person's need for accuracy or appropriateness regarding their behavior or decision-making, because it reduces the uncertainty surrounding a certain behavior [105]. It can also be used to gain social approval; a notable example was the trend to drive hybrid cars (i.e., a Prius) in parts of the USA, which was touted 'green to be seen' or moral licensing [118].

Our example is a representative study for psychologically informed recommender design [102], for it mimics the rationale and in part the setup of a classic psychological study. It is based on a 2008 study from Goldstein et al. [119], which can be described as the 'hotel room towel study'. As part of the experiment, bathroom doors in hotel rooms were equipped with different door hangers that aimed to persuade the guest(s) to re-use their towels. The control group door hanger made an environmental appeal, much like many door hangers in hotels do [119], asking guests to "help save the environment". The treatment door hangers emphasized that a majority of people actually re-use their towels, stating: "Join your fellow guests in helping to save the environment", after which the door hanger states that 'almost 75%' of people comply with this request. Combined, this is a descriptive social norm [105], which is shown to be more effective than an environmental appeal to promote towel re-use [119].

Furthermore, the social norm message in Goldstein et al. [119] differentiates between different sources. Whereas one the variants refers to hotel guests in general, Goldstein et al. [119] also referred to other social peers. They vary to what extent individuals identify with the reference group, proposing abstract identities such as 'citizens' and 'men' or 'women', as well as very specific identities, such as 'hotel guests in room XXX'. Although people rate their citizen and gender identities as most important, it emerged that mentioning the specific hotel room led to the highest compliance rate.

Where the fit score was also a specific explanation of the algorithm, the social norm message can also be explained in terms of the Rasch algorithm. For example, if we would take a measure of the middle of the scale (i.e., $\delta \approx 0$), we could state that

50% of people or Dutch citizens do this (cf. [98]). This might already sound more convincing for measures that are not necessarily visible. While this rate increases for easier measures, going up to close to 100% based on the data in [27], it decreases sharply for more difficult measures. For example, considering Table 1, we would have to report that close of 0% have erected a windmill in their backyard.

To overcome these low numbers, we designed normative messages that were more selective in the source of their social norms. If we consider an attitude-tailored scenario as in the previous examples [25, 27], then reporting on the behavior of similar people in terms of attitude would lead to adoption rates around 50%. If you would, however, report on people with a stronger attitude, the adoption rates go up.

To craft convincing norm messages, we considered advice sources that have shown in the advice-taking literature that they can be effective. We propose the following three normative messages [28, 102]:

- 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."

Global norms simply report on the proportion of people that perform a measure, tapping into data used to form the Rasch scale. Similar norms take the engagement probabilities of the current user, following Eq. 1, while the experienced norms use the same equation, but take a stronger attitude than the user's. We took $\theta + 1$, which led to a probability of 75% for attitude-tailored advice. This rationale is also described in Table 2, which illustrates the norm percentages for two recommendation scenarios.

This study used an updated version of the Saving Aid interfaces from previous studies [25]. This time round, it only included a single list of twenty recommended measures, which were tailored toward the user's attitude and ordered on their estimated kWh savings. The interface is shown in Fig. 10, depicting similar norm scores on the right-hand side. Besides that, the interface depicts information on various energy-saving attributes if the user would hover images or click 'more information'.

An experiment was performed, with four different between-subject conditions [102]. The research design was in line with Goldstein et al. [119]: Users were either

Table 2 Scenario of attitude-tailored advice ($\delta = \theta$) to illustrate what norm percentages are presented for each norm source, depending on the user's attitudinal strength (on the left-hand side). For this table, imagine there are two users: User 1 has a relatively weak energy-saving attitude, User 2 has a relatively strong attitude

	Norm percentages for attitude-tailored advice: $\delta = \theta$		
	Global norms (%)	Similar norms (%)	Experienced norms (%)
User 1: $\theta_1 = -1$	72	50	75
User 2: $\theta_2 = 1$	30	50	75

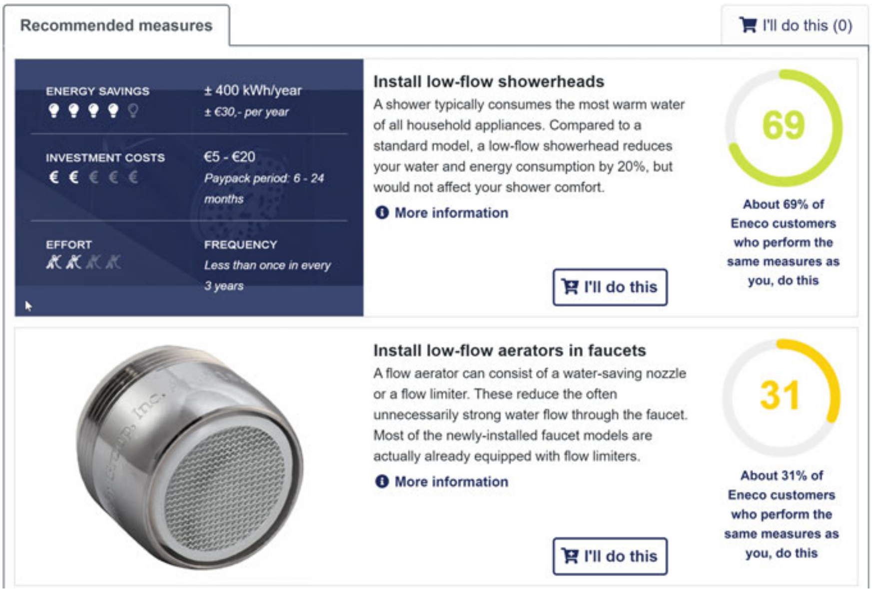

Fig. 10 Snapshot of the 'Saving Aid' interface used in Starke et al. [28, 102]. Depicted here is the Similar Norm condition, reporting the Rasch model probabilities as a proxy for the compliance among peers with a similar attitude as the user

assigned to one of the three norm conditions, Global, Similar, or Experienced, or to the environmental baseline. The environmental baseline involved a 'Savings Score', attributing a score of 100 to the top-item, which had the highest kWh savings, and lower scores as the savings decreased. The main difference with Goldstein et al. [119] was, obviously, that not a single behavior was promoted, but a list of 20 tailored energy-saving measures.

Analysis of Choice Data. The interface was distributed among an 'innovation panel' of a large Dutch electricity supplier (i.e., Eneco). A sample of 207 participants ($M = 53.5 years,\ SD = 14.0$) was analyzed, assessing the kWh savings chosen compared to other measures within the list, as well as the total savings per condition. The procedure was similar to other studies of Starke et al. [25, 26, 101]: Participants were invited to use the Saving Aid interface (cf. Fig. 10) and to choose any number of measures they would like to perform at home, which would be sent to them by email. Afterwards, they were asked to evaluate their interaction. In this study [28, 102], users were inquired on their perceived feasibility of the recommended measures and their level of choice satisfaction.

The findings can be divided into 'between' and 'within' effects. Overall, no between-list effects were observed, as measures were not more likely to be chosen, nor did another analysis reveal any changes in the chosen kWh savings. In contrast, we did observe so-called within-list effects. Measures with a relatively higher score in the Global and Experienced conditions were more likely to be chosen than a mea-

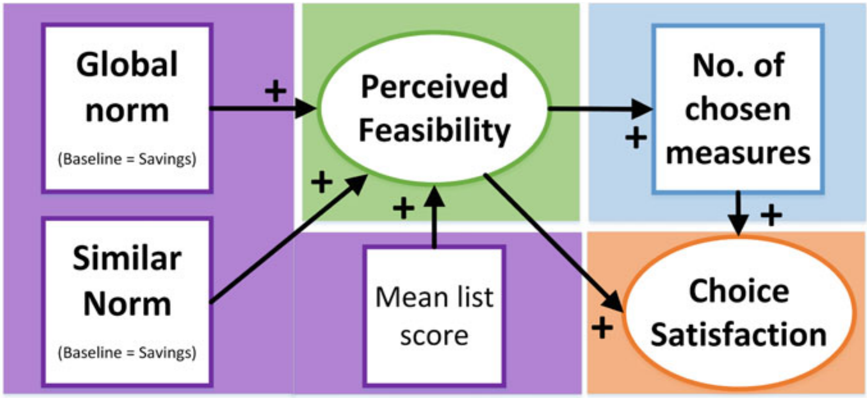

Fig. 11 Structural Equation Model (SEM). Relations between aspects resemble correlations; β-coefficients and standard errors are omitted for simplicity. Details are deported in Starke et al. [28]. Colors follow the guidelines of Knijnenburg et al. [61]: Objective aspects are purple, observed variables are in blue, perception aspects in green, and experience aspects in orange

sure with higher kWh saving scores. In fact, in line with the earlier examples in this chapter [25], higher kWh savings seemed to slightly decrease the probability that a measure would be chosen. This suggested that higher norm scores could have a positive effect for a measure to be chosen, but that the presence of norms did not lead to more choices.

Structural Equation Model Analysis. We further analyzed how the normative interventions affected the user evaluation. While this was not part of the original Goldstein et al. [119] study, they did report a discrepancy between participants identifying more citizen and gender-based norms, but not acting upon them as much as rather specific norms.

Figure 11 depicts a part of the original Structural Equation Model (SEM) analysis. It focuses on the direction of effects of the different normative messages, excluding personal characteristics. We found that similar and global norms increased the perceived feasibility of recommendations, compared to seeing an interface with Savings Scores. This suggested that the normative messages seemed to lower perceived thresholds toward performing energy-saving measures, effectively reducing the perceived or expected behavioral costs, even if users did not make more choices. All this was found while controlling for the mean list score, which was higher for some users and in some conditions.

Important to the evaluation of psychologically informed intervention are the mediated relations. Depicted in Fig. 11 is that perceived feasibility was affected by different norms. In turn, however, it also affected how many measures were chosen and how users evaluated their chosen measures. Starke et al. [28] assessed the relations from the two norms to choice satisfaction, showing that this relation was fully mediated by perceived feasibility and, in part, by the number of chosen measures. This meant that the study did have an impact on the user's perception and only, in turn, had other effects on choice and user experience.

Taken together, this study showed the limited effectiveness of social norm persuasion or nudging 'on top of' a tailored advice list. When presenting alongside a filtered list of best-fitting measures, social norm nudges affect what users choose, but in a reduced way. Whereas its application in a one-size-fits-all context led to more pro-environmental behavior (i.e., the hotel room towel study [119]), here it only shifted preferences among the presented options but not on how many options were chosen. Once again, an interface nudge that did not affect *what* measures was presented but only *how* had a reduced effectiveness in a tailored list.

3.3.4 Implications

The different examples all show how psychological user modeling can be used for algorithmic tailoring. The approach with the psychometric Rasch model is shown to be more effective than 'best guesses', even though it is a simple one-parameter item response theory algorithm [98]. On top of such a tailoring strategy, however, the additional benefit of interface nudges and persuasion is small, even though these were also mostly psychologically informed. This is despite that some studies are partial replications from psychological studies, in which an effect was observed in a one-size-fits-all context.

Translating psychological findings to the recommender context can, thus, come with unexpected results. The fact that a recommender already tailors its content has a non-trivial influence on any other 'nudging' or persuasive interface aspect. Whereas using social norms to promote any behavior that may or may not align with a user's preferences has shown to be effective, using it to promote a list of tailored measures seems to be much less effective. That is not to say there were no effects, as it was clear that it affected within-list preferences in Starke et al. [28, 102]. However, there seems to be diminishing marginal returns on additional persuasive strategies in recommender system research, to the extent that a statistically significant contribution is difficult to observe. This could be specific to the recommender, for there is some evidence that combining persuasive strategies in artificial agents or social robots is effective [120, 121].

It is important to emphasize that tailoring in recommender systems typically hides options or makes it significantly more difficult to find them. Even though the current studies dealt with relatively small and accessible data sets (i.e., between 79 and 135 measures), we observed a mitigated 'nudging effectiveness'.

4 Discussion

Affecting what people like and do is challenging in various scientific domains. Achieving longitudinal behavioral change is not addressed extensively in many human-computer interaction studies. This is despite the existence of various 'behavioral change technologies': From the introduction of persuasive technology [1], all the way toward the newly introduced ‘Digital Twins’ [122], the number of domains where human-computer interaction can be central to behavioral change is expanding. Recommender systems have the potential to be among those technologies, but many approaches would need to start incorporating behavioral change techniques from the persuasion and nudging fields beyond tailoring. However, it remains an open question whether the benefits of the whole of tailored content and nudging strategies can be larger than the sum of its individual parts.

This chapter has highlighted the design of psychologically informed recommender systems and the extent to which they can be effective in a behavioral change domain. Energy conservation specifically taps into various theories and concepts related to environmental psychology that, for the purpose of our studies, have been translated to recommender system aspects. In general, we have found that attitude-tailored methods are more effective than one-size-fits-all approaches, but that the benefit of additional nudging and persuasion is small.

The examples in this chapter show that the effectiveness of various nudges and persuasive strategies is significantly reduced when used in a tailored advice context. Even though these techniques, such as social influence, have proven their merit in one-size-fits-all contexts, both ‘offline’ and ‘online’ [17, 105, 119], something changes when promoting multiple behaviors that fit a specific user. A possible explanation that uses a different taxonomy for tailoring [70] is that tailoring might lead to more conscious processing of information processing, therefore mitigating the effectiveness of ‘peripheral’ nudging and persuasion techniques. However, many of the techniques used in this chapter are also cognitively oriented [123], mostly relying on a central form of information processing.

Based on the findings outlined in this chapter, we argue that tailored advice is a category or concept distinct from ‘nudges’ in the context of recommender systems. When dealing with a large number of items, tailoring leads to significant changes in the costs of looking for ‘hidden’ items, or it might even be impossible to find them at all. This is at odds with the definition of nudging, there being changes in a choice architecture that do not affect any of the user’s incentives. When considering joint work of recommender systems and digital nudging (cf. [18]), we propose to not consider tailoring to be a nudge. Also, we would suggest to avoid loose use of the term ‘nudging’ when only using algorithmic tailoring in a recommender system for behavioral change.

We propose to perform more conceptual work in this area, to tease apart different psychological aspects in a recommender system. We argue that some of the core concepts are not always completely understood. For example, while it makes sense that tailored content is ‘better’ or ‘more useful’ than a random guess, it is less clear how it affects an individual specifically at a cognitive and decision-making level. And, how does this compare to the effects instilled by other types of nudges and persuasion? We aim to work on this problem further, but also encourage to join this endeavor and to go beyond a 1-on-1 application of psychological theories to recommender systems.

We also highlight the benefits of our psychologically informed algorithm. We argue that the Rasch model is specifically effective in recommender domains where behavioral change is involved. Its inherent trade-off between feasibility and novelty is formalized, and is more informative than the simplistic 'accuracy = 1—diversity' rule of thumb. Rasch can help users to explore a trajectory on what to do next in domains of goal-directed behavior.

The psychological methods outlined in this chapter are far from complete. For one, most interface manipulations tap into theory from psychology similar to nudging. Instead, more methods from the field of persuasive technology could also be applied, particularly for cases that are more focused on 'personalization' [44]: tailored toward user characteristics rather than user preferences only. As authors with a track record in the research field of recommender systems but also in persuasion, we feel that these are two parallel worlds, even though there is much to learn from one another. Perhaps that technology can be humanized even further, when these two worlds meet.

Acknowledgements Alain Starke's contribution to this chapter received support from the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the centers for Research-based Innovation scheme, project number 309339.

References

1. Fogg BJ, Tseng H (1999) The elements of computer credibility. In: CHI, pp 80–87
2. IJsselsteijn W, De Kort Y, Midden C, Eggen B, Van Den Hoven E (2006) Persuasive technology for human well-being: setting the scene. In: Persuasive technology: first international conference on persuasive technology for human well-being, PERSUASIVE 2006, Eindhoven, The Netherlands, 18–19 May 2006. Proceedings 1. Springer, pp 1–5
3. Abrahamse W, Steg L, Vlek C, Rothengatter T (2007) The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents. J Environ Psychol 27(4):265–276
4. Steg L (2008) Promoting household energy conservation. Energy Policy 36(12):4449–4453
5. Demarque C, Charalambides L, Hilton DJ, Waroquier L (2015) Nudging sustainable consumption: the use of descriptive norms to promote a minority behavior in a realistic online shopping environment. J Environ Psychol 43:166–174
6. Midden CJ, Kaiser FG, Teddy McCalley L (2007) Technology's four roles in understanding individuals' conservation of natural resources. J Soc Issues 63(1):155–174
7. Jannach D, Zanker M, Felfernig A, Friedrich G (2010) Recommender systems: an introduction. Cambridge University Press, Cambridge, UK
8. Abrahamse W, Steg L, Vlek C, Rothengatter T (2005) A review of intervention studies aimed at household energy conservation. J Environ Psychol 25(3):273–291
9. Nolan JM (2021) Social norm interventions as a tool for pro-climate change. Curr Opin Psychol 42:120–125
10. Abrahamse W, Steg L (2013) Social influence approaches to encourage resource conservation: a meta-analysis. Glob Environ Chang 23(6):1773–1785
11. Allcott H (2011) Social norms and energy conservation. J Public Econ 95(9–10):1082–1095
12. Petkov P, Köbler F, Foth M, 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, pp 21–30
13. Schultz PW, Nolan JM, Cialdini RB, Goldstein NJ, Griskevicius V (2007) The constructive, destructive, and reconstructive power of social norms. Psychol Sci 18(5):429–434

14. Frederiks ER, Stenner K, Hobman EV (2015) Household energy use: applying behavioural economics to understand consumer decision-making and behaviour. Renew Sustain Energy Rev 41:1385–1394
15. Liebe U, Gewinner J, Diekmann A (2021) Large and persistent effects of green energy defaults in the household and business sectors. Nat Hum Behav 5(5):576–585
16. Benders RM, Kok R, Moll HC, Wiersma G, Noorman KJ (2006) New approaches for household energy conservation-in search of personal household energy budgets and energy reduction options. Energy Policy 34(18):3612–3622
17. Petkov P, Goswami S, Köbler F, Krcmar H (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, pp 587–596
18. Jesse M, Jannach D (2021) Digital nudging with recommender systems: survey and future directions. Comput Hum Behav Rep 3:100052
19. Thaler RH, Sunstein CR (2008) Nudge: Improving decisions about health, wealth, and happiness. Yale University Press, New Haven, CT, USA
20. Bothos E, Apostolou D, Mentzas G (2015) Recommender systems for nudging commuters towards eco-friendly decisions. Intell Decis Technol 9(3):295–306
21. Smith NC, Goldstein DG, Johnson EJ (2013) Smart defaults: from hidden persuaders to adaptive helpers. SSRN, INSEAD
22. Lex E, Schedl M (2022) Psychology-informed recommender systems: A human-centric perspective on recommender systems. In: Proceedings of the 2022 conference on human information interaction and retrieval, pp 367–368
23. Winecoff AA, Brasoveanu F, Casavant B, Washabaugh P, Graham M (2019) Users in the loop: a psychologically-informed approach to similar item retrieval. In: Proceedings of the 13th ACM conference on recommender systems, pp 52–59
24. Starke A, Willemsen MC, Snijders C (2015) Saving energy in 1-d: Tailoring energy-saving advice using a rasch-based energy recommender system. In: DMRS, pp 5–8. Citeseer
25. Starke A, Willemsen M, Snijders C (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, pp 65–73
26. Starke A (2019) The effectiveness of advice solicitation and social peers in an energy recommender system. In: IntRS@ RecSys, pp 65–71
27. Starke AD, Willemsen MC, Snijders CC (2020) Beyond "one-size-fits-all" platforms: applying campbell's paradigm to test personalized energy advice in the Netherlands. Energy Res Soc Sci 59:101311
28. Starke A, Willemsen M, Snijders C (2021) Promoting energy-efficient behavior by depicting social norms in a recommender interface. ACM Trans Interact Intell Syst (TiiS) 11(3–4):1–32
29. Ricci F, Rokach L, Shapira B, Kantor PB (2010) Recommender Systems Handbook, 1st edn. Springer, New York, NY, USA
30. Konstan JA, Riedl J (2012) Recommender systems: from algorithms to user experience. User Model User-Adap Inter 22:101–123
31. Hutmacher F, Appel M (2023) The psychology of personalization in digital environments: from motivation to well-being-a theoretical integration. Rev Gen Psychol 27(1):26–40
32. Serino CM, Furner CP, Smatt C (2005) Making it personal: how personalization affects trust over time. In: Proceedings of the 38th annual hawaii international conference on system sciences, pp 170–170. IEEE
33. Knijnenburg B, Willemsen M, Broeders R (2014) Smart sustainability through system satisfaction: tailored preference elicitation for energy-saving recommenders. In: 20th Americas conference on information systems (AMCIS 2014), 7–9 Aug 2014, Savannah, Georgia, USA. AIS/ICIS, pp 1–15
34. Himeur Y, Alsalemi A, Al-Kababji A, Bensaali F, Amira A, Sardianos C, Dimitrakopoulos G, Varlamis I (2021) A survey of recommender systems for energy efficiency in buildings: principles, challenges and prospects. Inf Fusion 72:1–21

35. Knijnenburg BP, Willemsen MC (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, pp 381–384
36. Konstan J, Riedl J (2012) Recommended for you: how online merchants predict your preferences and prod you to purchase. IEEE Spectr 49(10):54–61
37. Braunhofer M, Elahi M, Ricci F (2014) Techniques for cold-starting context-aware mobile recommender systems for tourism. Intelligenza Artificiale 8(2):129–143
38. Cena F, Console L, Gena C, Goy A, Levi G, Modeo S, Torre I (2006) Integrating heterogeneous adaptation techniques to build a flexible and usable mobile tourist guide. AI Commun 19(4):369–384
39. Van Setten M, Pokraev S, Koolwaaij J (2004) Context-aware recommendations in the mobile tourist application compass. In: Nejdl WDBPE (ed) Proceedings of adaptive hypermedia conference, pp 235–244
40. Elsweiler D, Hauptmann H, Trattner C (2022) Food recommender systems. Recomm Syst Handbook 11:871
41. Starke A, Asotic E, Trattner C (2021) "serving each user": supporting different eating goals through a multi-list recommender interface. In: Proceedings of the 15th ACM conference on recommender systems, pp 124–132 (2021)
42. Starke AD, Asotic E, Trattner C, Van Loo EJ (2023) Examining the user evaluation of multi-list recommender interfaces in the context of healthy recipe choices. ACM Trans Recomm Syst (2023)
43. Musto C, Starke AD, Trattner C, Rapp A, Semeraro G (2021) Exploring the effects of natural language justifications in food recommender systems. In: Proceedings of the 29th ACM conference on user modeling, adaptation and personalization, pp 147–157
44. Yoo K-H, Gretzel U, Zanker M (2021) Persuasive recommender systems: conceptual background and implications. Springer, New York, NY, USA
45. Jameson A, Willemsen MC, Felfernig A, De Gemmis M, Lops P, Semeraro G, Chen L (2015) Human decision making and recommender systems. Recomm syst Handbook 611–648
46. Knijnenburg BP, Sivakumar S, Wilkinson D (2016) Recommender systems for self-actualization. In: Proceedings of the 10th ACM conference on recommender systems, pp 11–14
47. Wilkinson D (2018) Testing a recommender system for self-actualization. In: Proceedings of the 12th ACM conference on recommender systems, pp 543–547
48. Kaiser FG, Wilson M (2004) Goal-directed conservation behavior: the specific composition of a general performance. Personality Individ Differ 36(7):1531–1544
49. Jiang W, Pardos ZA (2019) Time slice imputation for personalized goal-based recommendation in higher education. In: Proceedings of the 13th ACM conference on recommender systems, pp 506–510
50. Setten M, Veenstra M, Nijholt A, Dijk B (2006) Goal-based structuring in recommender systems. Interact Comput 18(3):432–456
51. Ekstrand MD, Willemsen MC (2016) Behaviorism is not enough: better recommendations through listening to users. In: Proceedings of the 10th ACM conference on recommender systems, pp 221–224
52. Kaiser FG, Byrka K, Hartig T (2010) Reviving campbell's paradigm for attitude research. Personal Soc Psychol Rev 14(4):351–367. https://doi.org/10.1177/1088868310366452
53. Lundberg DC, Tang JA, Attari SZ (2019) Easy but not effective: why "turning off the lights" remains a salient energy conserving behaviour in the United States. Energy Res Soc Sci 58:101257. https://doi.org/10.1016/j.erss.2019.101257
54. Attari SZ, DeKay ML, Davidson CI, Bruin WB (2011) Changing household behaviors to curb climate change: how hard can it be? Sustain: J Rec 4(1), 9–11
55. Gardner GT, Stern PC (2008) The short list: The most effective actions us households can take to curb climate change. Environ: Sci Policy Sustain Dev 50(5):12–25
56. Starke, A (2019). RecSys challenges in achieving sustainable eating habits. In: HealthRecSys'19: workshop in conjunction with ACM RecSys, Copenhagen, Denmark

57. Boudet HS, Flora JA, Armel KC (2016) Clustering household energy-saving behaviours by behavioural attribute. Energy Policy 92:444–454
58. Urban J, Ščasnỳ M (2016) Structure of domestic energy saving: how many dimensions? Environ Behav 48(3):454–481
59. Ajzen I (1991) The theory of planned behavior. Organ Behav Hum Decis Process 50(2):179–211
60. Hauptmann H, Leipold N, Madenach M, Wintergerst M, Lurz M, Groh G, Böhm M, Gedrich K, Krcmar H, Effects and challenges of using a nutrition assistance system: results of a long-term mixed-method study. User Model User-Adap Interact 1–53
61. Knijnenburg BP, Willemsen MC, Gantner Z, Soncu H, Newell C (2012) Explaining the user experience of recommender systems. User Model User-Adap Interact 22:441–504
62. Falk K (2019) Practical recommender systems. Simon and Schuster, New York, NY, USA
63. Felfernig A, Jeran M, Ninaus G, Reinfrank F, Reiterer S (2013)Toward the next generation of recommender systems: applications and research challenges. Multimed Serv Intell Environ: Adv Recomm Syst 81–98
64. Brusilovsky P, Schwarz E (1997) User as student: Towards an adaptive interface for advanced web-based applications. In: User modeling: proceedings of the sixth international conference UM97 Chia Laguna, Sardinia, Italy, 2–5 June 1997. Springer, pp 177–188
65. O'keefe DJ (2018) Persuasion. In: The handbook of communication skills. Routledge, Milton Park, UK, pp 319–335
66. Campbell DT (1963) Social attitudes and other acquired behavioral dispositions
67. Kahneman D (2011) Thinking, fast and slow. Macmillan, London, UK
68. Ham J, Midden C, Beute F (2009) Can ambient persuasive technology persuade unconsciously? using subliminal feedback to influence energy consumption ratings of household appliances. In: Proceedings of the 4th international conference on persuasive technology, pp 1–6
69. Zarouali B, Boerman SC, Voorveld HA, Noort G (2022) The algorithmic persuasion framework in online communication: conceptualization and a future research agenda. Int Res
70. Dijkstra A (2008) The psychology of tailoring-ingredients in computer-tailored persuasion. Soc Pers Psychol Compass 2(2):765–784
71. Dobelli R (2013) The art of thinking clearly: better thinking, better decisions. Hachette UK, London UK
72. Kahneman D (2003) Maps of bounded rationality: psychology for behavioral economics. Am Econ Rev 93(5):1449–1475
73. Thaler RH, Sunstein CR (2003) Libertarian paternalism. Am Econ Rev 93(2):175–179
74. Knijnenburg BP, Willemsen MC (2015) Evaluating recommender systems with user experiments. Recomm Syst Handbook 309–352
75. Larsen SØ, Trattner C (2021) Predicting feature-based similarity in the news domain using human judgments. In: Proceedings of the 9th international workshop on news recommendation and analytics (INRA 2021) co-located with 15th ACM conference on recommender systems (RecSys 2021)
76. Pu P, Chen L, Hu R (2011) A user-centric evaluation framework for recommender systems. In: Proceedings of the fifth ACM conference on recommender systems, pp 157–164
77. Zangerle E, Bauer C (2022) Evaluating recommender systems: survey and framework. ACM Comput Surv 55(8):1–38
78. Bollen D, Knijnenburg BP, Willemsen MC, Graus M (2010) Understanding choice overload in recommender systems. In: Proceedings of the fourth ACM conference on recommender systems, pp 63–70
79. Liang Y, Willemsen MC (2021) The role of preference consistency, defaults and musical expertise in users' exploration behavior in a genre exploration recommender. In: Proceedings of the 15th ACM conference on recommender systems, pp 230–240
80. Starke AD, Sedkowska J, Chouhan M, Ferwerda B (2022) Examining choice overload across single-list and multi-list user interfaces. In: 9th joint workshop on interfaces and human decision making for recommender systems (IntRS'22)

81. Jannach D, Quadrana M, Cremonesi P (2022) Session-based recommender systems. In: Recommender systems handbook. Springer, New York, NY, USA, pp 301–334
82. Bettman JR, Luce MF, Payne JW (1998) Constructive consumer choice processes. J Consum Res 25(3):187–217
83. Johnson EJ, Shu SB, Dellaert BG, Fox C, Goldstein DG, Häubl G, Larrick RP, Payne JW, Peters E, Schkade D (2012) Beyond nudges: tools of a choice architecture. Mark Lett 23:487–504
84. Asensio OI, Delmas MA (2016) The dynamics of behavior change: evidence from energy conservation. J Econ Behav Org 126:196–212
85. Dietz T, Gardner GT, Gilligan J, Stern PC, Vandenbergh MP (2009) Household actions can provide a behavioral wedge to rapidly reduce us carbon emissions. Proc Natl Acad Sci 106(44):18452–18456
86. Karlin B, Davis N, Sanguinetti A, Gamble K, Kirkby D, Stokols D (2014) Dimensions of conservation: exploring differences among energy behaviors. Environ Behav 46(4):423–452
87. Kaiser FG, Wölfing S, Fuhrer U (1999) Environmental attitude and ecological behaviour. J Environ Psychol 19(1):1–19
88. Viappiani P, Faltings B, Pu P (2006) Preference-based search using example-critiquing with suggestions. J Artif Intell Res 27:465–503
89. Eagly AH, Chaiken S (1993) The psychology of attitudes. Harcourt Brace Jovanovich College Publishers, San Diego, CA, USA
90. Wilson TD, Lindsey S, Schooler TY (2000) A model of dual attitudes. Psychol Rev 107(1):101
91. Dunlap RE, Van Liere KD, Mertig AG, Jones RE (2000) New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised nep scale. J Soc Issues 56(3):425–442
92. Kollmuss A, Agyeman J (2002) Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior? Environ Educ Res 8(3):239–260
93. Gunster S, Fleet D, Paterson M, Saurette P (2018) “Why don’t you act like you believe it?": competing visions of climate hypocrisy. Front Commun 49
94. Stuart D (2022) Tensions between individual and system change in the climate movement: an analysis of extinction rebellion. New Polit Econ 27(5):806–819
95. Coupey E, Irwin JR, Payne JW (1998) Product category familiarity and preference construction. J Consum Res 24(4):459–468
96. Greve W (2001) Traps and gaps in action explanation: theoretical problems of a psychology of human action. Psychol Rev 108(2):435
97. Rasch G (1961) On general laws and the meaning of measurement in psychology. In: Proceedings of the fourth berkeley symposium on mathematical statistics and probability, vol 4, pp 321–333
98. Bond TG, Fox CM (2006) Applying the rasch model: fundamental measurement in the human sciences, 2nd ed. Psychology Press, London, UK
99. Linacre JM (2016) WinstepsR Rasch measurement computer program. Winsteps.com, Beaverton, Oregon. http://www.winsteps.com Accessed 01 Jan 2016
100. Kempton W, Harris CK, Keith JG, Weihl JS (1985) Chapter 6: do consumers know “what works” in energy conservation? Marr Family Rev 9(1–2):115–133
101. Starke AD, Willemsen MC, Snijders C (2021) Using explanations as energy-saving frames: A user-centric recommender study. In: Adjunct Proceedings of the 29th ACM conference on user modeling, adaptation and personalization, pp 229–237
102. Starke AD, Willemsen MC, Snijders C (2020) With a little help from my peers: Depicting social norms in a recommender interface to promote energy conservation. In: Proceedings of the 25th international conference on intelligent user interfaces, pp 568–578
103. Tintarev N, Masthoff J (2010) Designing and evaluating explanations for recommender systems. In: Recommender systems handbook. Springer, New York, NY, USA, pp 479–510
104. Festinger L (1954) A theory of social comparison processes. Hum Relat 7(2):117–140
105. Cialdini RB, Goldstein NJ (2004) Social influence: compliance and conformity. Annu Rev Psychol 55:591–621

106. Johnson EJ, Goldstein D (2003) Do defaults save lives? Am Assoc Adv Sci
107. Johnson EJ (2021) The elements of choice: why the way we decide matters. Penguin, London, UK
108. O'neil C (2017) Weapons of math destruction: how big data increases inequality and threatens democracy. Crown, New York, NY, USA
109. Butt A, Donald MS, Foster FD, Thorp S, Warren GJ (2018) One size fits all? Tailoring retirement plan defaults. J Econ Behav Org 145:546–566
110. Tannenbaum D, Fox CR, Goldstein NJ (2013) Partitioning menu items to nudge single-item choice. In: 34th annual meeting for the society for judgment and decision making, pp 1–12. Citeseer
111. Ungemach C, Camilleri AR, Johnson EJ, Larrick RP, Weber EU (2018) Translated attributes as choice architecture: aligning objectives and choices through decision signposts. Manage Sci 64(5):2445–2459
112. Cialdini RB (1996) Activating and aligning two kinds of norms in persuasive communications. J Interpret Res 1(1):3–10
113. Handgraaf MJ, De Jeude MAVL, Appelt KC (2013) Public praise versus private pay: effects of rewards on energy conservation in the workplace. Ecol Econ 86:86–92
114. Chen CC, Shih S-Y, Lee M (2016) Who should you follow? Combining learning to rank with social influence for informative friend recommendation. Decis Supp Syst 90:33–45
115. Zhu H, Huberman B, Luon Y (2012) To switch or not to switch: understanding social influence in online choices. In: Proceedings of the SIGCHI conference on human factors in computing systems, pp 2257–2266
116. Chen C-F, Knight K (2014) Energy at work: social psychological factors affecting energy conservation intentions within chinese electric power companies. Energy Res Soc Sci 4:23–31
117. Cialdini RB, Trost MR (1998) Social influence: social norms, conformity and compliance
118. Griskevicius V, Tybur JM, Bergh B (2010) Going green to be seen: status, reputation, and conspicuous conservation. J Person Soc Psychol 98(3):392
119. Goldstein NJ, Cialdini RB, Griskevicius V (2008) A room with a viewpoint: using social norms to motivate environmental conservation in hotels. J Consum Res 35(3):472–482
120. Ham J, Cuijpers RH, Cabibihan J-J (2015) Combining robotic persuasive strategies: the persuasive power of a storytelling robot that uses gazing and gestures. Int J Soc Robot 7:479–487
121. Mayer RE, Sobko K, Mautone PD (2003) Social cues in multimedia learning: role of speaker's voice. J Educ Psychol 95(2):419
122. Reinders MJ, Starke AD, Fischer AR, Verain MC, Doets EL, Van Loo EJ (2023) Determinants of consumer acceptance and use of personalized dietary advice: a systematic review. Trends Food Sci Technol 131:277–294
123. Cadario R, Chandon P (2020) Which healthy eating nudges work best? A meta-analysis of field experiments. Mark Sci 39(3):465–486