Do psychological traits influence the perceived usefulness of rule recommendations in configuration tasks?

Federica Cena, Cristina Gena, Claudio Mattutino, Michele Mioli, Barbara Treccani, Fabiana Vernero & Massimo Zancanaro

Preprint version of the article: Federica Cena, Cristina Gena, Claudio Mattutino, Michele Mioli, Barbara Treccani, Fabiana Vernero & Massimo Zancanaro (28 Aug 2024): Do psychological traits influence the perceived usefulness of rule recommendations in configuration tasks?, Behaviour & Information Technology, DOI: 10.1080/0144929X.2024.2396478

Do psychological traits influence the perceived usefulness of rule recommendations in configuration tasks?

Federica Cena[a], Cristina Gena[a], Claudio Mattutino[a], Michele Mioli[a], Barbara Treccani[b], Fabiana Vernero[a] and Massimo Zancanaro[b]

[a]Department of Computer Science, University of Turin, Torino, Italy; [b]Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

ABSTRACT

In this paper, we describe an empirical evaluation of the user's perceived usefulness of recommendations for configuration tasks in a smart home scenario. Our results suggest that while overall recommendations help improve the performance in the task, the psychological traits of `Self-efficacy` and `Need for Cognition` play an important role in determining the perceived usefulness of the recommendations. These effects are somehow different from those reported in other studies where recommendations are provided for a choice task rather than a constructive task. Thus, our results offer evidence regarding the importance of incorporating the evaluation of personality traits into the design of configuration constructive tasks in order to make informed decisions on whether and how to provide users with recommendations. Additionally, they show that it is crucial to take into account the nature of the task and, where possible, individual competencies.

1. Introduction

Smart environments, such as smart homes and smart manufacturing, integrate a large number of heterogeneous *Internet-of-Things (IoT)* components, namely sensors and actuator devices, in a single system aimed at supporting users in pursuing their goals (Bansal and Kumar 2020), for example automating routine procedures or preventing emergency situations. Although the specificity of each installation of such environments (that is, the idiosyncratic routines of each family or the peculiar requirements of a specific factory) would require individual adaptation of the environment, dealing with the complexity of those infrastructures is difficult for non-expert users. Indeed, the research field of End-User Development (EUD) investigates how to provide end-users with adequate support to this purpose (Lieberman et al. 2006).

We call a *configuration task* in a smart environment the specification of the set of operations that a user needs to perform in order to activate the desired set of actions by a subset of the actuator devices present the environment, given a specific set of values detected by a subset of sensors. For example, the windows in a smart home should be automatically closed if the weather sensors detect that it starts raining; or, in a smart factory, the engine should be stopped as soon as a person is detected entering into the danger area of machinery.

In the EUD domain, it has long been recognised that a useful approach to deal with configuration tasks by non-expert users is by means of *trigger-action rules* (Ghiani et al. 2017). *Trigger-action rules* are similar to *if* $<$*condition*$>$ *then* $<$*actions*$>$ statements in which the conditional part contains an expression on the values detected by a sensor and the $<$*actions*$>$ part specifies a sequence of actions to be performed. For example, the first configuration task presented above may be realised with the following *trigger-action* rule: '*if a weather station* [trigger object] *detects that it is starting to rain* [trigger event] *the windows* [action object] *are closed* [action event]'. *Trigger-action* rules are a simplified version of *Event-Condition-Action* rules use in software engineering (Casati et al. 2000; Paton 2012), and a large literature investigates how they may fit the user mental model (see for example (Ghiani et al. 2017; Huang and Cakmak 2015; Zancanaro et al. 2022)).

Therefore, a *configuration task* can be more formally defined as the list of the following *preferential choices* (Jameson et al. 2011):

(1) What devices (sensors and actuators) present in the environment have to be considered

(2) How to connect several sensors and several actuators to achieve the desired behaviour;
(3) What event (*trigger*) should activate an interaction pattern;
(4) What behaviour (*action*) should be carried out after a trigger event has occurred.

In this respect, it is clear that people may need support in getting through such a process. Indeed, helping people make better decisions is an important function of recommender systems (Jameson et al. 2015). Yet, resolving the preferential choices of a configurational task is different from the typical cases in which recommender systems have been used, such as movie recommendations. In this case, there is a constructive task, where the users need to take a number of decisions, from a relatively small number of alternatives but with several options to consider, aimed at developing a solution to a problem. The major indicator of success is the performance on the task (i.e. the configuration of sensors and actuators works as planned). On the contrary, recommenders are usually employed in decision tasks where the users have to opt for a relatively simple choice but among a very large number of alternatives (such as deciding which movie to rent or which book to buy). The indicator of success is satisfaction with the decision taken.

Our first research question concerns the usefulness of recommendations of either trigger conditions or actions in the process of defining *trigger action* rules. We consider both perceived usefulness (RQ1a) and the ability of recommendations to improve users' performance (RQ1b):

Research Question 1a (RQ1a): Do users find it useful to receive suggestions on either trigger conditions or actions to carry out the configuration task?

Research Question 1b (RQ1b): Are suggestions on trigger conditions and actions actually useful (i.e. they improve users' performance) to carry out the configuration task?

From research in Psychology, we understand that personality of people plays an important role in decision-making (Deniz 2011). Indeed, in the last decade there has been an increasing research interest in more user-oriented approaches to recommender systems (Dhelim et al. 2022; Tkalcic and Chen 2015) to understand the impact that personality can play in accepting and using recommendations, specifically in decision tasks (Dhelim et al. 2020; Koren, Bell, and Volinsky 2009) and how they are exposed to unfair recommendations (Yalcin and Bilge 2023).

Several investigations focus on the the Big-Five model (McCrae and John 1992) of personality (see the related work section), that is best suited in relational contexts.

Since we are dealing with an effortful cognitive task, we aimed at investigating different perspectives on the concept of personality. We decided to focus on three largely well know constructs: `Locus of Control`, `Self-efficacy`, and `Need for Cognition`.

`Locus of Control` (Rotter 1966) is a construct that is used to categorise people's beliefs about how much control they have over the events of their lives. People with an external `Locus of Control` tend to perceive their life outcomes as arising from factors out of their control. People with an internal `Locus of Control` tend to attribute the cause of events in their lives to their own actions, motivations, or competencies. `Self-efficacy` (Bandura 1986) is a closely related construct. It can be defined as people's belief in their capacity to exercise control over their own functioning (i.e. their motivation, cognition, behaviour), over their social environment, and, in general, over events that may affect their lives. The self-evaluation that `Locus of Control` and `Self-efficacy` are supposed to measure (i.e. the beliefs people hold regarding their power to affect situations in their life) may influence the goals towards which people strive and the amount of energy they spend to achieve these goals, but also their inclination towards challenging cognitive activities and their wish to be helped when they face these problems (i.e. their appreciation of possible aid and assistance). In this respect, these constructs are related to `Need for Cognition` (Cacioppo and Petty 1982), a personality construct reflecting the extent to which people are inclined towards demanding cognitive activities. More specifically, `Need for Cognition` is 'a need to structure relevant situations in a meaningful, integrated way […] a need to understand and make reasonable the experiential world' (Cohen 1957) and can be briefly defined as 'the tendency for an individual to engage in and enjoy thinking' (Cacioppo and Petty 1982).

We posit the following research question:

Research Question 2 (RQ2): Do `Locus of Control`, `Self-efficacy`, and `Need for Cognition` influence the perception of usefulness of the recommendations provided to a user in a given context?

We investigate the above research questions by means of an empirical study on a configuration task in a smart home scenario carried out using a prototype developed in the context of the project Empathy [1]

In order to better clarify the effect of personality on the perception of usefulness of recommendation, we compare two different approaches to recommendation: (a) as a means towards an easier solution, by providing a recommendation of possible rules while the user is working on it, and (b) after the rule has been defined, as a way to improve the understanding (and possibly the learning) of the system.

Our final goal was to determine whether it might be useful to infer and include the personality traits of `Locus of Control`, `Self-efficacy` and `Need for Cognition` in an extended user model which uses them to provide recommendations only to those users who can possibly appreciate them.

The paper is structured as follows. Section 2 presents the main relevant work regarding personality-based recommenders and rule recommendations for End-User Development. Section 3 introduces the prototype we used for our empirical evaluation. Section 4 describes the experiment and its results and in Section 5 we discuss the implications of our findings. Finally, Section 6 concludes the paper and presents future work.

2. Related work

2.1. Personality in recommender systems

Personality explains individual differences in long-lasting emotional, interpersonal, experiential, attitudinal, and motivational styles (McCrae and John 1992). Several studies have shown that personality plays an important role in decision-making (Deniz 2011) and can help understand user preferences (Dhelim, Aung, and Ning 2020) in various domains, such as music (Bansal, Flannery, and Woolhouse 2021), movies (Kim 2020), websites and online behaviours (Kosinski et al. 2014). Starting from the pioneering work of Dunn et al. (2009), Nunes (2009) and Hu and Pu (2010), the possibility of incorporating personality-related information into recommenders has attracted considerable attention. In recommender system terms, in fact, personality can be understood as a user profile that is both context- and domain-independent (Tkalcic and Chen 2015).

The Big-Five model. Among the many examples which can be mentioned for this purpose, several studies aimed at using personality characteristics as complementary information to overcome the *cold start* problem which hinders traditional recommendation algorithms. For example, in Hu and Pu (2011), personality-based neighbourhoods were built for the target users, and various personality-aware approaches were compared to the traditional user-based Collaborative Filtering (CF) method with promising results. Starting from the idea that personality information learned in one domain could be transferred to a different one, in a cross-domain recommendation perspective (Cantador, Fernández-Tobías, and Bellogín 2013), Fernández-Tobías et al. (2016) also compared different methods to alleviate the cold start problem, and provided guidelines for their application to different scenarios. More recently, this problem was addressed in Moscato, Picariello, and Sperlí (2021), with a specific focus on personality-based music recommendation.

Relevant work in the *group recommendation* area referred to the Big-Five to investigate the relationship between choice satisfaction and characteristics of the individuals and the groups (Delic et al. 2018), while Wu, Chen, and He (2013) studied how personality can impact users' preferences for *recommendation diversity* and demonstrated a personality-based diversity adjusting approach in the movie recommendation domain. Based on the results of a comprehensive study with 1840 users of the MovieLens recommender, Karumur, Nguyen, and Konstan (2018) discussed how personality information could be used to deal with a number of the traditional recommender systems issues, such as newcomer retention, cold start problem, novelty, diversity, and popularity in recommendations, group and cross-domain recommendation generation.

Other personality models. The RIASEC model (Holland 1997), whose main goal is to explain vocational behaviour and thus help people attain professional satisfaction, distinguishes six personalities: *realistic*, *investigative*, *artistic*, *social*, *enterprising* and *conventional*. It was applied in an e-commerce prototype to increase customers' likelihood to purchase (Bologna et al. 2013) and in a hybrid recommender to help adolescents define their future career plans (Ochirbat et al. 2018). The Bartle model (Bartle 1996), which identifies four player types based on their personality (*achiever*, *explorer*, *socialiser*, *killer*) was exploited for the in-game recommendation of new activities to undertake (De Simone et al. 2021). The Thomas-Kilman Conflict Mode Instrument (TKI) (Thomas 1992) describes an individual's behaviour in conflict situations along two basic dimensions, i.e. *assertiveness* and *cooperativeness*. In the context of group recommendation, the TKI was used to generate recommendations that take into account the group personality composition, so as to maximise satisfaction for the group as a whole (Quijano-Sánchez, Recio-García, and Díaz-Agudo 2010). More recently, it was exploited to explore how group members' conflict resolution style can affect recommendation quality (Nguyen et al. 2019).

Need for Cognition, Locus of Control, and Self-efficacy. Experimenting with a music recommender system, Millecamp et al. (2019) found that participants with a high *Need for Cognition* were more confident about the playlists they had created when the system provided no explanations for its recommendations. In fact, since they spontaneously reasoned about possible grounds for their recommendations, system-provided explanations were redundant and might even lead to distrust of the system when recommendations were not appropriate. In an online dating scenario (Tong et al. 2018), daters with a high `Need for Cognition` proved to be more willing to rely on a recommender system when they faced very large choice sets (i.e. more than 800 potential mate profiles), as a consequence of their tendency towards adaptive behaviour, while daters lower in `Need for Cognition` were always likely to use the recommender, independently of the number of available options. In the related area of decision-making, the possible influence of `Need for Cognition` on nudging techniques (defaults and social influence) was explored in an online shopping scenario (Ingendahl et al. 2021): results show that nudges are effective across different levels of consumer's `Need for Cognition`.

Locus of Control is also linked to decision-making and the use of information. For example, a recent study (Sharan and Romano 2020) investigated whether players trust advice generated from either an AI-based algorithm or other humans, in a decision-making card game. Participants with a low (external) `Locus of Control` showed patterns of concordance with recommendations, independently of their source, while the opposite holds for people with a high (internal) `Locus of Control`.

Finally, partially related work in the area of technology acceptance showed that general *Self-efficacy* mediates the effects of Big-Five traits on perceived ease of use and usefulness of audiovisual technology (Manolika et al. 2022). Furthermore, computer `Self-efficacy` significantly influences the overall acceptance of different technologies, such as video conferencing software (Alfadda and Mahdi 2021) and digital learning environments (Reddy et al. 2021).

2.2. Recommendation in the EUD domain

End-User Development (EUD) is 'a set of methods, techniques, and tools that allow users of software systems, who are acting as non-professional software developers, at some point to create, modify, or extend a software artifact' (Lieberman et al. 2006). Since the beginning, EUD systems have shown that recommendations can improve usability and user experience (Haines et al. 2010). Haines et al. (2010) explored the different classes of recommendations made in EUD systems.

One of the earliest approaches to EUD was Programming By Example (PBE). Some systems in this category continuously observe the user's actions to find repetitions over which they can learn a looping program to complete the user's task. Examples include EAGER, Dynamic Macro, and APE (Cypher 1991; Masui and Nakayama 1994; Ruvini and Dony 2000). By recommending automation directly within the user's workflow, these systems achieved EUD transparently without the user's awareness of having programmed the system. A more general approach to automation within the user's workflow relies on activity recognition to observe what the user is doing and infer what that user is trying to accomplish. For example, Lumière system used Bayesian user models to offer context-dependent assistance (Horvitz et al. 1998). While Lumière could offer assistance on various tasks, it was limited to assisting the tasks encoded by the developers. Some systems combine aspects of both approaches. WARP, like Lumière, utilised probabilistic models for activity recognition on a wide variety of tasks (Yorke-Smith et al. 2009), but, like PBE systems, was also able to continuously extend its knowledge base to handle new tasks. Task Assistant was another system that made recommendations over an extensible knowledge base (Peintner et al. 2009). It used manual procedures produced by EUD to inform its recommendations for further automated procedures of future tasks.

Haines et al. (2010) proposed that the following kinds of recommendations should be supported in EUD: (1) recommending shared procedures; (2) improving activity recognition; (3) suggesting preferred defaults; (4) suggesting more likely generalisations; (5) suggesting potential problems based on similar programs; (6) suggesting examples; (7) suggesting solutions to programming problems; (8) suggesting reusable code elements; (9) assisting novices and suggesting best practices. In the latter case, Haines et al. (2010) suggested collaborative or social recommender techniques. According to these authors, however, few systems today use recommender technologies to support decision-making within EUD. They identify four main classes of recommendations that could be improved by using recommender technology: (1) inserting automation into the user's workflow; (2) helping the user make the right decisions; (3) handling errors; (4) supporting unplanned sharing.

To the authors' knowledge, in the EUD field, no study investigates the relationship between users' psychological traits and recommendations.

3. The prototype

Before introducing the evaluation, we describe the rule recommender prototype[2] and its main components, which we have realised in the context of the Empathy project (for more details, see Cena et al. 2021).

We have implemented a *rule-based recommender service*, developed starting from an *If This Then That* (IFTTT)[3] dataset of 98,744 rules. First, we have applied the association rules algorithm Apriori on the dataset to extract the most relevant associations between smart objects. To find stronger associations, we have then related the individual objects to the classes of an extended version of EUPont ontology (Corno, De Russis, and Roffarello 2017), representing IoT end-user programming environments. We ran the Apriori algorithm on the classes and more significant results emerged.

Based on this data, we have built recommendation services that suggest object categories (namely the ontology classes) to compose a trigger-action rule: when the input category is a trigger category, a set of action categories are recommended; conversely, when the input is an action category, a set of trigger categories are recommended. We have built a web interface showing these services, see Figures 1 and 2.

As we can see from Figure 1, when choosing the trigger (or the action on the other page), the associated action (or trigger) categories are shown with a number that represents the confidence originating from the Apriori algorithm for association rule learning. The categories with a higher confidence value represent the most used pairings in the IFTTT dataset. Once trigger or action categories have been chosen, three examples of rules from the IFTTT dataset are suggested (see Figure 2). Assuming that users can be described through *user models* to receive personalised recommendations, this recommender service will be extended in the future with a user modelling component that will consider user psychological traits according to the findings of the experiment described in this paper.

Implementation details. As far as the prototype's implementation is concerned, the recommender is a module built in Java that contains all the logic required to provide the different recommendation techniques. The data are stored in a mongoDB instance and are organised in different collections and databases. We used the same architecture to realise the web pages hosting the experiment which will be described in Section 4. Finally, the web interface has been developed using Vue.js and Vuetify as a material design framework. The application server is implemented with Spring Boot and exposes a set of REST APIs.

4. The experiment

The goal of the experiment was to assess how helpful (in terms of performance) and useful *trigger*, *action*, and *rule recommendations* are to users who are given a (simulated) configuration task.

Participants were asked to write three trigger-action rules, selecting appropriate trigger and action objects from a list of options and then specifying how they should interact. Before carrying out the configuration task, participants in the experimental group could access the recommender prototype (see Section 3), where they could retrieve suggested object categories to include in their rules, as well as example rules themselves. On the contrary, participants in the control group had to carry out the configuration task without receiving any suggestion, and they could access the recommender prototype only after they had completed the tasks. Although participants in the control group could not benefit from recommendations during the configuration task, they were nevertheless asked to explore the prototype for three reasons: firstly, we wanted to provide them with exactly the same information as participants in the experimental group, for fairness reasons; secondly, they were also asked to assess recommendation usefulness, imagining that -in the future- they would have to carry out a configuration task similar to the one they had just completed; thirdly, this allowed us to test two different approaches to recommendation, one where suggestions provided before/during the task can be used as a means to more easily reach a solution, and one where suggestions provided after the task completion can help participants to improve their understanding of the system.

Performance, perceived usefulness, and other metrics for user experience have been investigated in connection with personality constructs such as `Locus of Control`, `Self-efficacy`, and `Need for Cognition` in order to determine whether these could be taken into account in the future user modelling component and in the resulting recommendation generation process.

The experiment was conducted online. Ethical approval for this study was obtained from the bioethical committee of the University of Turin, with approval number: 0675432

Design. We used a *between-subjects* design, where the main independent variable is *access to recommendations*, with two possible values, *before* and *after*: the experimental group had access to the recommendations on trigger-action rules *before* the configuration task, while the control group could explore possible recommendations *after* completing the task only. The dependent variables are participants' performance, the perceived easiness and the perceived enjoyment of the configuration task, participants' satisfaction

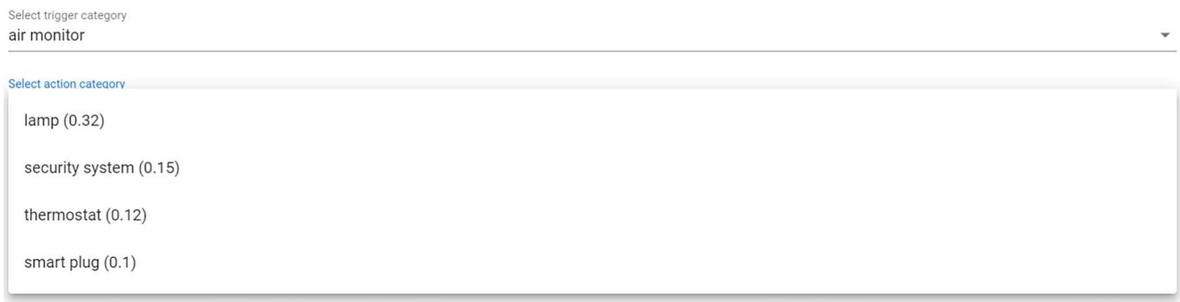

Figure 1. Web interface for the recommendation of action object categories.

with their tasks, and the perceived usefulness of recommendations.

Hypothesis. Regarding RQ1a, we hypothesised that all users would judge recommendations useful to carry out the proposed configuration task. Even more so, we expect that the accessibility of recommendations, before of after the task, impacts participants' performance (RQ1b), with the experimental group being able to compose more effective rules and trigger/action associations than users in the control group. We also expect participants in the experimental group to find the simulated configuration tasks easier, more enjoyable, and more satisfying. Regarding RQ2, we hypothesised that personality traits have an impact on the perceived usefulness of the recommendations. More specifically, we hypothesised that users with more external `Locus of Control` would find it more useful to receive recommendations before the task than users with internal `Locus of Control`. Concerning the `Need for Cognition` and `Self-efficacy` we hypothesised that people with high scores in these traits would find it more useful to look at the recommendations because they can appreciate the effort required to understand the recommendations in order to improve their performance.

Participants. We recruited 64 participants (68.4% in the 18–25 age range, 26.2% in the 26–35 age range, and 3.2% in the 36–44 age range), among students attending HCI and interaction design master courses, with both a technical and humanistic background. They declared an average familiarity with smart objects and IoT technology (2.8 out of 5), while most of them declared to possess at least one smart object (mostly voice assistants and location services).

Apparatus and materials. We created a series of web pages to deliver the experiment online. They contained

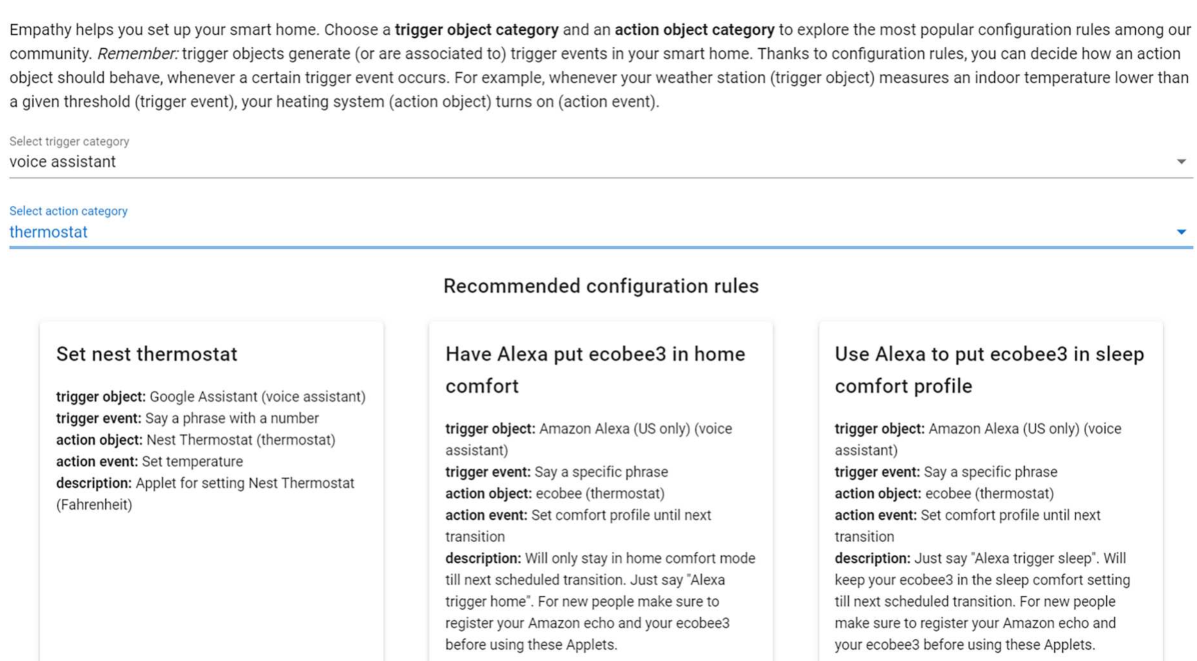

Figure 2. Web interface for the recommendation of configuration rules.

open and closed questions for the pre- and post-tests, forms to carry out the experimental tasks (rule writing), as well as the link to our recommendation prototype. The exact order and content of the web pages was adapted according to the group participants belonged to, either experimental or control.

The test could be performed through any browser connected to the Internet. All the user data were anonymously collected and stored.

Procedure. For an overview of the experiment procedure, see Figure 3. The first steps were the same for participants in both the experimental and the control group. After having provided their informed consent to participate in the experiment, participants were allowed to access a welcome page offering an overview of the study. Then, they were asked to fill in a pre-test, where data on their behaviours and personality traits (`Locus of Control`, `Self-efficacy`, and `Need for Cognition`), on socio-demographic aspects (i.e. age, education), as well as on familiarity with technology and IoT, smart objects' possession, and potential smart home goals[4] were collected. After that, the smart home configuration task was introduced through the following scenario:

> You took home on your own, and, given your passion for technology, you decided to set up your home with a few smart objects. Imagine having to go through the phase of configuring your devices, and having to decide how to make them interact to make your environment truly intelligent. Try writing three rules by which you would configure your smart home. To do this, you need to choose a trigger object and an action object and write how you would like them to interact. A trigger object generates (or is associated with) events that cause a rule to run. Through a rule, you can specify which action object should be activated and how it should behave in correspondence with a trigger event. Here is an example of a rule: if a weather station (trigger object) measures an internal temperature below a certain threshold (trigger event), the heating system (action object) turns on (action event).

At this point, the experiment procedure was adapted according to the group participants belonged to. After having read the scenario, in fact, only the participants in the experimental group were given access to the website described in Section 3, which offers the users help in configuring their smart home by providing rule recommendations they could explore before doing the configuration task introduced in the scenario. Then, all participants were asked to write three configuration rules: for each one, they had to choose a trigger object and an action object, then they had to write the rule's logic in natural language (see Figure 4). After that, participants had to evaluate their experience by answering a few questions regarding ease, fun, and satisfaction with what they had done in the task.

At this point, the participants in the control group were given access to the recommendation prototype and, finally, all participants were asked to assess the recommendations: in particular, they evaluated how useful it was (for the control group: *it would be*) to discover recommended action objects from a trigger object and vice-versa, as well as how useful it was (for the control group: *it would be*) to discover example rules with the chosen trigger and action objects.

Measures. Participants' familiarity with technology and with IoT was self-assessed through two 5-point Likert scales ranging from 'no familiarity at all' to 'extreme familiarity'.

`Locus of Control`, `Self-efficacy` and `Need for Cognition` have been assessed using validated questionnaires from the International Personality Item Pool (IPIP) web site[5]. The IPIP items are statements presenting possible behaviours[6]: respondents are required to assess how accurate such items are in describing their own behaviour, rating them on a 5-point Likert scale with points labeled as follows: very accurate, moderately accurate, neither accurate nor inaccurate, moderately inaccurate, and very inaccurate. In particular, to measure `Locus of Control`, we used the 5-item IPIP rational scale[7] based on Levenson's `Locus of Control` scale (Levenson 1981). For `Self-efficacy`, we used the 10-item IPIP scale[8] that contains the same items as those used to measure 'Independence' in Gough's California Psychological Inventory (Gough 1996). To measure `Need for Cognition`, we used the 10-item IPIP scale[9] based on the original scale by Cacioppo and Petty (Cacioppo and Petty 1982).

Participants' performance with rule writing was assessed by three domain experts who first worked separately, and then discussed their evaluations in order to achieve consensus. For each rule, the experts separately assessed participants' choice of the trigger and action objects (assigning 1 point if the objects matched their roles, 0 points otherwise) and their ability in composing a configuration rule (assigning 1 point if the rule was syntactically correct, but either non-sensical or incoherent with the chosen trigger and action objects, 2 points if the rule was both formally correct and logical, 0 points otherwise).

The easiness and enjoyment of the configuration task, participants' satisfaction with the rules they had created, as well as their evaluation of the usefulness of

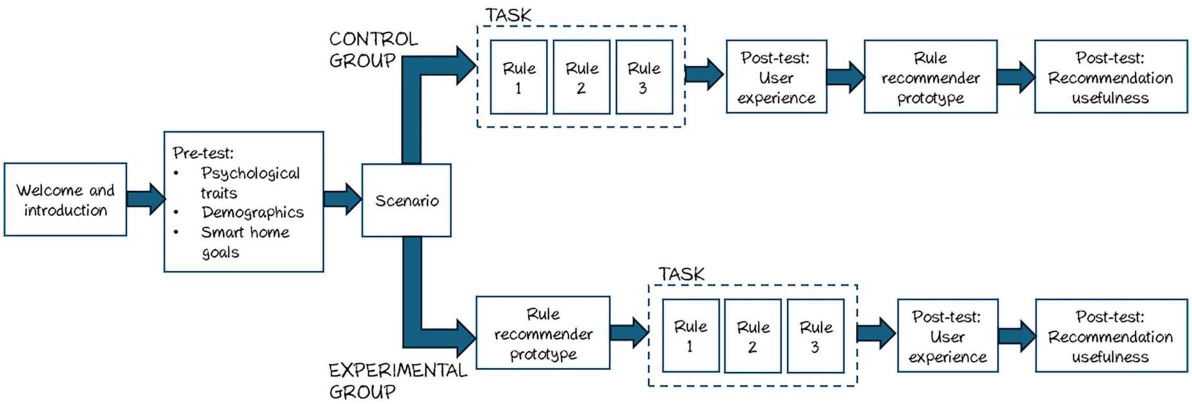

Figure 3. Experiment procedure.

trigger, action, and rules' recommendations were collected using dedicated 5-point Likert scales, with points ranging from 'not at all' to 'extremely'.

Results (descriptive statistics). The self-declared familiarity with technology is quite high, with mean 4.15 (std= 0.54) on a 5-point scale. The minimum

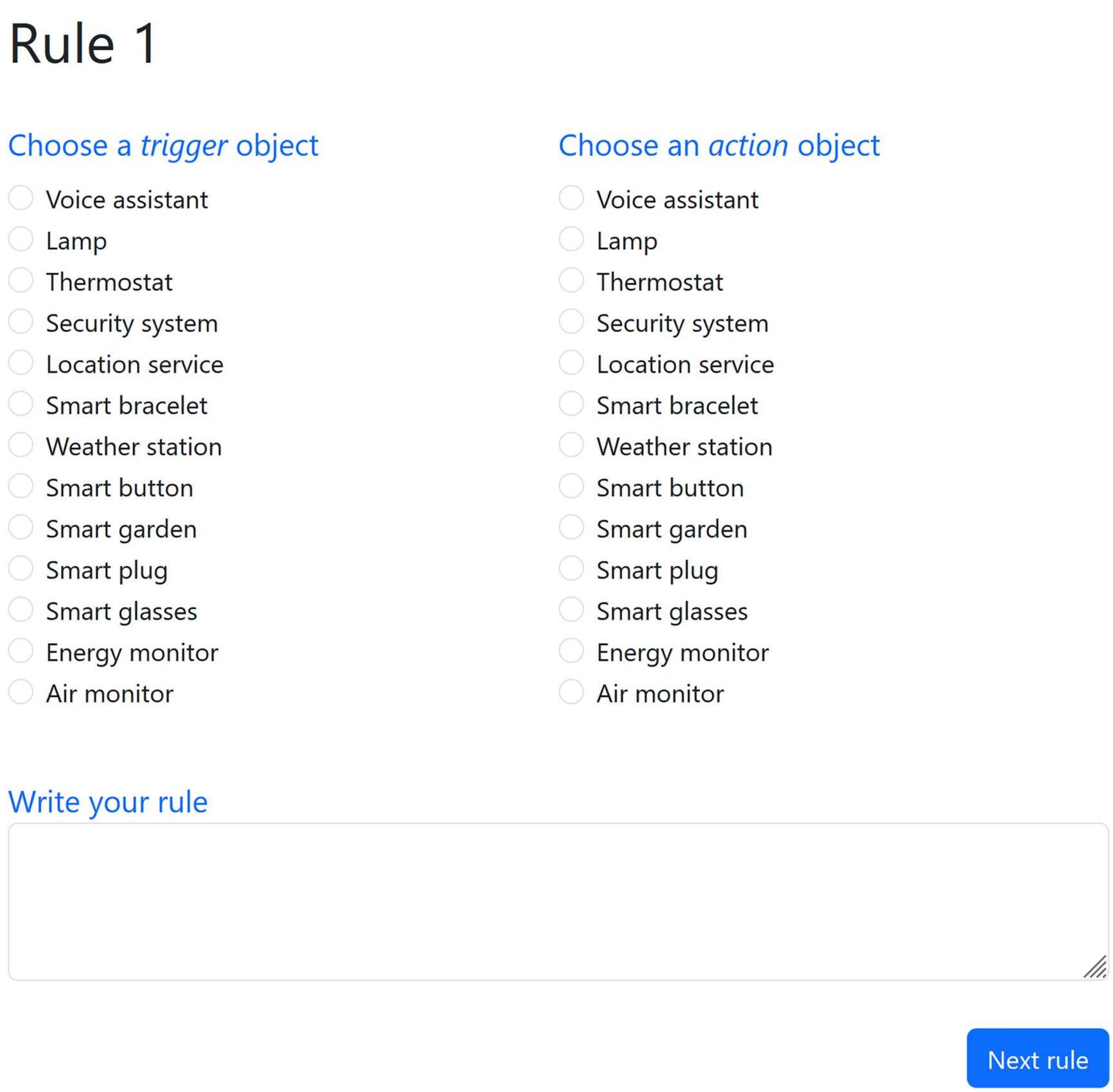

Figure 4. Experimental task: for each rule, users must choose the trigger and action objects and write the rule's logic in natural language.

value is 3, and it is not normally distributed. The familiarity with IoT devices is lower (mean= 2.81; std= 1.09), and the distribution, although not normal, is wider. The two variables have a significant correlation (Pearson's $r = 0.53$, $p<0.01$). For both these measures, there is a small but significant difference between the two conditions (for familiarity with technology, Mann Whitney $w = 314.000$, $p<0.05$ and for familiarity with IoT, Mann Whitney $w = 334.500$, $p<0.05$). The values of the means for the two conditions are reported in Table 1.

For what concerns the personality traits, for `Locus of Control`, the items show a relatively good internal coherence (Cronbach alpha= 0.67, comparable to 0.61 reported in the International Personality Item Pool). We will therefore consider the scale as composed of the average scores of the items. The variable is normally distributed (Shapiro–Wilk $w = 0.971$, $p = 0.154$) with mean 3.25 (std= 0.73). There is no significant difference between the two conditions (t test, $t = -0.469$, $p = 0.641$).

For `Self-efficacy`, the items show a relatively good internal coherence (Cronbach alpha= 0.61), although lower than the 0.81 reported in the International Personality Item Pool. We consider the scale as composed by the average scores of the items. The variable is normally distributed (Shapiro–Wilk $w = 0.981$, $p = 0.465$) with mean 3.74 (std= 0.44). There is no significant difference between the two conditions (t test, $t = -0.092$, $p = 0.927$).

For Need for Cognition, the items show a good internal coherence (Cronbach alpha= 0.722), but lower than the 0.84 reported in the International Personality Item Pool. We consider the scale as composed by the average scores of the items. The variable is normally distributed (Shapiro–Wilk $w = 0.977$, $p = 0.286$) with mean: 3.72 (std= 0.519). There is no significant difference between the two conditions (t test $t = -0.013$, $p = 0.990$).

Regarding the performance on the task, we took into consideration the scores provided by the experts on the selection of the trigger/action and on the composition of the rule for each one of the three exercises, as explained above. The scores taken together have a good internal consistency (Cronbach alpha= 0.805), hence, we aggregate them together by averaging the individual values. The resulting scale has a value between 0 (totally wrong) and 1 (completely correct). The variable has a mean of 0.80 (std= 0.29), and it is not normally distributed.

Regarding the subjective experience on the task, it was measured with 3 single items for task easiness (mean= 2.84, std= 0.79), enjoyment (mean= 2.98 std= 0.86), and satisfaction (mean= 3.016, std= 0.735). None of them is normally distributed (Shapiro-Wilk, respectively $w = 0.835$, $p = 0.000$; $w = 0.881$, $p = 0.000$; $w = 0.846$, $p = 0.000$).

Finally, for the reported usefulness of the recommendations, the participants assessed separately the usefulness of recommendations about the triggers, the action to use, and the rules that could be defined. Since the items have a good internal consistency (Cronbach alpha= 0.844), we consider the average score as a single scale. The variable is normally distributed (Shapiro-Wilk $w = 0.969$, $p = 0.124$) with mean 3.583, std= 0.768).

Results (inferential statistics). For the performance measure, there is a significant difference between the two conditions (Mann Whitney $w = 344.0$, $p < 0.05$) with the participants in the experimental group performing over 10% better than the participants in the control group (see Table 1).

For the subjective experience, there are no statistical differences between the two conditions for either task easiness, enjoyment and satisfaction (Mann Whitney, respectively, $w = 535.000$, $p = 0.384$; $w = 496.000$, $p = 0.817$; $w = 564.000$, $p = 0.190$).

For the usefulness of recommendations, we found a statistically significant difference between the two conditions (t test t= 3.193, $p = 0.002$) with the participants in the control condition valuing around 15% more useful the recommendations with respect to the experimental condition (see Table 1).

In order to control for the possibility that the differences in performance and perceived usefulness are due to the difference between familiarity with technology and with IoT, the correlations among these variables have been checked (Pearson's correlation test). None of the pairwise correlations are significantly different from zero (correlation values are reported in Table 2).

In order to investigate the impact of personality traits on the perceived usefulness of the recommendations, we analysed participants' usefulness ratings with linear mixed effects models implemented in R with lme4 (Bates et al. 2014). These models allow to estimate the genuine effects of the variables under investigation by separating these effects from those of other confounding variables, that is, differences between participants and/or items (i.e. the specific questions about the usefulness of the recommendations) due to factors that we were not interested in Baayen (2008).

In order to find the model that best fits the value, we built a succession of models starting from a model including only participants and items as random effects (*Model 1*) and progressively adding all the terms of the previous one with an additional term. The fit of each model was compared with that of the

Table 1. Descriptive statistics for the main variables (*: there is a significant difference between the two conditions, $p<0.05$; ns: the difference is non-significant).

	Overall		Control con.		Experimental con.		
	mean	std.	mean	std.	mean	std.	Sig.
Pre-task measures:							
Familiarity with technology	4.15	0.54	3.97	0.49	4.31	0.54	*
Familiarity with IoT	2.81	1.08	2.47	0.97	3.13	1.10	*
Locus of Control	3.25	0.73	3.20	0.69	3.29	0.77	ns
Self-efficacy	3.74	0.44	3.73	0.46	3.74	0.43	ns
Need for Cognition	3.72	0.52	3.72	0.46	3.73	0.58	ns
Post-task measures:							
Performance	0.78	0.29	0.74	0.27	0.82	0.31	*
Easiness	2.84	0.79	2.93	0.91	2.75	0.67	ns
Enjoyment	2.98	0.86	3.00	0.91	2.75	0.67	ns
Satisfaction	3.02	0.74	3.13	0.82	2.91	0.64	ns
Usefulness of suggestions	3.58	0.77	3.88	0.73	3.30	0.70	*

previous one, using the log-likelihood test. For a discussion on this procedure, see (Baayen, Davidson, and Bates 2008) and for an example, albeit in a different context, see also the work of Zorzi et al. (2012)).

Before entering personality variables, we entered as fixed effects other factors (and their interactions) that could influence usefulness ratings. Therefore, the second model (*Model 2*) was defined by adding the group (experimental vs. control) to which participants belonged. The comparison between *Model 1* and *Model 2* showed a significant improvement in the model's fit, $\chi^2(1) = 8.949$, $p<0.001$[10]. Overall, participants who did not see recommendations before the execution of the task, but only afterwards, judged the recommendations as more useful than participants who read the recommendations before performing the task (i.e. participants in the control vs. experimental groups, respectively; 3.883 vs. 3.302).

In a subsequent model, *Model 3*, we added the participant's performance in the configuration task. The comparison between *Model 2* and *Model 3* did not reveal a significant improvement in the model's fit, $\chi^2(1) = 2.325$, $p = .127$. *Model 3* was subsequently updated by adding the interaction between group and performance. The comparison of the resulting model, *Model 4*, with either *Model 3* or *Model 2* showed that not even this addition provides a significant improvement to the model's fit, $\chi^2(1) = 0.071$, $p = .790$ and $\chi^2(2) = 2.396$, $p = 0.301$, respectively. Consequently, participants' performance was discarded from subsequent models.

In the next steps, we updated *Model 2* by adding in succession (one after the other) each of the personality variables, both as the main fixed effect and in interaction with the group. The addition of the `Need for Cognition` (*Model 5*) did not yield to a significantly better fit, $\chi^2(1) = 0.017$, $p = .895$, but the addition of the interaction between `Need for Cognition` and group (*Model 6*) resulted in a model that fitted the data significantly better than both *Model 5* and *Model 2*, $\chi^2(1) = 13.091$, $p<.001$ and $\chi^2(1) = 13.109$, $p<0.01$, respectively.

None of the subsequent comparisons between *Model 6* and each of the models including the other personality scores or their interactions with the group showed a significant improvement of the fit, all $\chi^2 s \leqslant 2.969$, $ps > .05$[11], that is, the models that provide for `Locus of Control`, `Self-efficacy` or their interaction with the group, besides providing for the interaction between `Need for Cognition` and group (*Models 7,8,9,10,11,12*; $Log-Likelihood \leqslant -214.60$), did not fit the data significantly better than the model providing for the interaction between `Need for Cognition` (here: NC) and group only (*Model 6*: $usefulness{\sim}(1|subject) + (1|item) + group{*}NC$; Log-Likelihood$= -216.47$). Nevertheless, it is worth noting that, by repeating the procedure starting from the `Self-efficacy` and adding the other variables later, the results are similar albeit with slightly lower log-likelihood scores. Indeed, the correlation between `Self-efficacy` and `Need for Cognition` is pretty high (Pearson's correlation r$= 0.716$, $p<0.001$) while the correlation of those traits with `Locus of Control` is significantly different from zero but lower (respectively, r$= 0.357$, $p<0.05$ with `Self-efficacy`, and r$= .484$, $p<0.01$ with `Need for Cognition`).

According to the simplest and best-fit model, therefore, participants' `Need for Cognition` had an

Table 2. Pearson's correlation coefficients between the familiarity variables and the two main factors that differentiate the experimental group from the control group (in bold, the coefficients that are statistically significant).

	Fam. Tech	Fam. IoT	Performance	Usefulness sugg.
Fam. Tech	–	**0.526**	0.012	−0.168
Fam. IoT	–	–	−0.058	−0.143
Performance	–	–	–	0.125
Usefulness sugg.	–	–	–	–

impact on their judgments of the usefulness of the recommendations. However, this impact was different in the two groups. As shown in Figure 5, for participants in the control group (i.e. those participants who saw the recommendations after having performed the task), the higher the `Need for Cognition` score, the lower the usefulness rating score. In contrast, for participants in the experimental group (i.e. those participants who saw the recommendations before having performed the task), the usefulness rating score increased as the `Need for Cognition` score increased. Similar results are obtained by replacing `Need for Cognition` with `Self-efficacy`.

In order to analyse separately the relationship between `Need for Cognition` and usefulness in the two groups, we conducted two separate analyses for data from participants of the control and experimental groups.

We defined two models (one for the experimental group data and the other for the control group data), including only participants and items as random effects (*Model 1exp* and *Model 1ctrl*). We updated both models by adding the `Need for Cognition` score (*Model 2exp* and *Model 2ctrl*). Both the comparisons of *Model 1exp* with *Model 2ex*p and that of *Model 1ctrl* with *Model 2ctrl* revealed that the addition of `Need for Cognition` led to a significant improvement of the fit, both $\chi^2(1) \geqslant 4.605$, $p < .05$; In both groups, therefore, `Need for Cognition` had a significant, albeit different, impact on performance.

In the successive models, we added, at first, participant's performance (*Model 3exp* and *Model3ctrl*) and then the interaction between this factor and `Need for Cognition` *(Model 4exp* and *Model4ctrl*). While the addition of performance did not provide a significant improvement in the models' fit, both $\chi^2(1) \leqslant 2.224$, $ps = .136$, the addition of the interaction between this factor and `Need for Cognition` resulted in models that fit the data better than all the previous ones, all $\chi^2(2) \geqslant 5.500$, all $ps < .05$.

As shown in Figure 6, the usefulness rating score increased as the performance score increased for participants with a high `Need for Cognition` while the usefulness data of participants with a low `Need for Cognition` show a less clear pattern. Similar results are obtained by replacing `Need for Cognition` with `Self-efficacy`.

5. Discussion

For what concerns our first research question, the results seem to confirm that exposure to examples and recommendations of possible rules improves performance in a configuration task (RQ1b). Yet, there seems to be no difference in the perceived easiness, enjoyment, or satisfaction with the task when the recommendations are provided before (and therefore potentially useful to improve the performance) or after the task. On the contrary, the perceived usefulness of the recommendations (RQ1a) is overall greater when the recommendations are provided after the task (therefore, just useful to get a better understanding and learning of the system).

This latter aspect can be explained by considering the impact that the `Need for Cognition` (and to a little lesser extent also `Self-efficacy`) has on the perception of the usefulness of the recommendations. Indeed, our data suggest that individuals with either high or low `Need for Cognition` tend to use differently (or to a different extent) the recommendations when performing the task, and the resulting performance in the task might, in turn, have an impact on the perceived usefulness of the recommendations. People's inclination to engage in demanding cognitive activities and challenging problems (`Need for Cognition`) can affect how much they appreciate the recommendations about how to solve these problems, but the `Need for Cognition` actually impact their judgments about usefulness which depends on whether these people have had the opportunity to use the recommendations in solving the task. Those who enjoy demanding cognitive tasks very likely try to use recommendations when these recommendations are presented before performing the task, and they may appreciate the aid.

Therefore, our second research questions (RQ2) can be answered positively for what concerns `Need for Cognition` (and possibly also `Self-efficacy`) by noting how the influence of those traits may explain an otherwise unexpected result.

Nevertheless, `Locus of Control` does not contribute to the model in our study. Its lack of impact on the perceived usefulness is somehow surprising (but it was also recognised in the work of Millecamp et al. (2019)). In our case, it might be because the recommendations are not perceived as scaffolding for compensating for lack of skills but as an integral part of the task. Yet, this aspect should be better investigated in further work.

Our findings seem coherent with Tong et al.'s (2018) observation that people with a high `Need for Cognition` were more willing to profit from recommendations when the task they had to carry out was perceived as more challenging, as a consequence of their ability to employ more sophisticated strategies to make an optimal decision. Indeed, this insight is coherent with the definition of `Need for Cognition`,

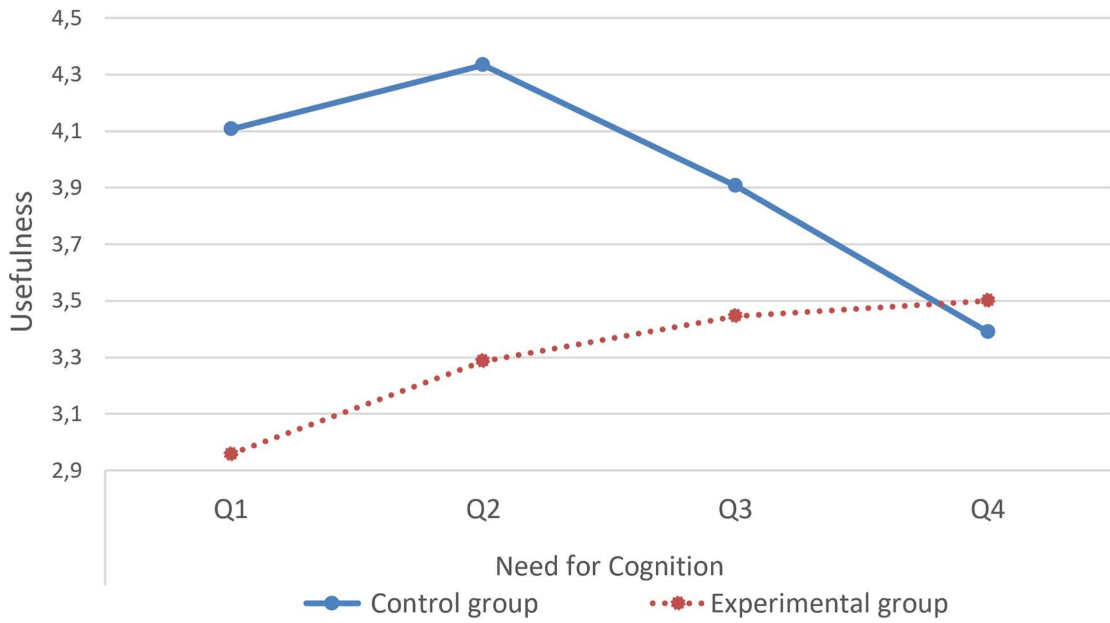

Figure 5. Mean usefulness scores as a function of the `Need for Cognition` (NC) scores (divided into quartiles, Q1-Q4) in the experimental and control groups.

especially if we consider the fact that individuals with high values for this trait tend to and enjoy to integrate multiple sources of information to make sense of the surrounding world (Cacioppo and Petty 1982; Cacioppo et al. 1996).

On the other hand, our results seem in contrast with those of Millecamp et al. (2019), who found that people with a low `Need for Cognition` tend to benefit more from explanations about a system's recommendation of a musical choice. Yet, this difference might be explained by the different contexts and tasks in which the users were engaged: Millecamp et al. assessed a choice process while we proposed a configuration task. In this respect, the higher or lower inclination to engage in effortful cognitive activities might be directly related to the specific activity in our case (that is, the willingness to perform the task well) while it might be just indirectly linked to the acceptance of the proposed choice. Indeed, Millecamp et al. measured the confidence of a choice as related to the personality trait.

The different effect on the specific task is worth to be better explored in further studies since it may have a stronger influence on the users of recommender systems in different contexts: for example, in education, for which constructive tasks are more frequent than decisions on recommendations about different options (which are more common in entertaining domains).

Again on RQ1, it is interesting to note how the perception of usefulness is related in a non-trivial way to performance. The interaction between performance and high scores on `Need for Cognition` may be interpreted as evidence that the provision of recommendations before or during the task may benefit more people with a higher `Need for Cognition` because

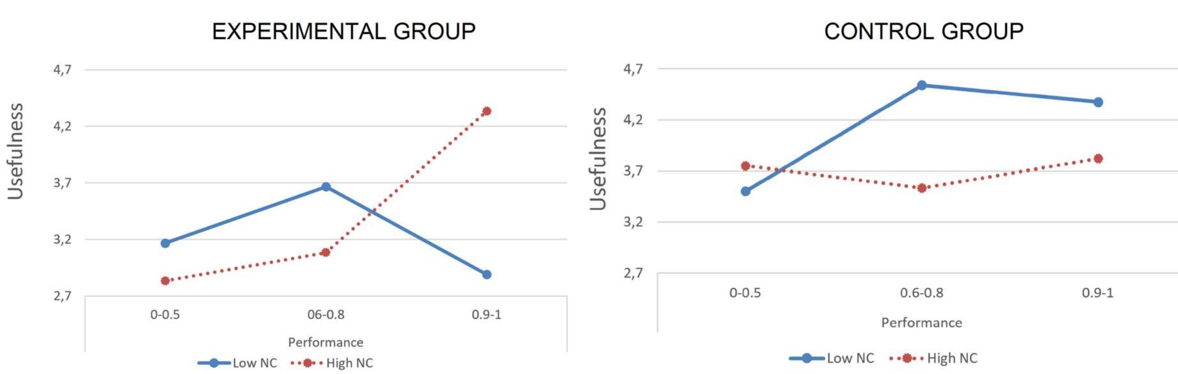

Figure 6. Mean usefulness scores of experimental (left) and control (right) participants with either high (>3.5) or low `Need for Cognition` as a function of performance.

they can effectively put them into practice to improve their performance. Of course, a simple correlation may not demonstrate this link of causality, and further studies are required.

In our sample, `Self-efficacy` and `Need for Cognition` are highly correlated, and therefore `Self-efficacy` seems to play a similar role to `Need for Cognition`, albeit in a less strong way. Indeed in literature, the two personality traits are often considered very close to each other. Nevertheless, `Need for Cognition` has a stronger emphasis on the willingness to commit to a specific task, while `Self-efficacy` concerns a more general sense of capability. Our study does not allow us to speculate further on this difference, but we believe this might be an interesting follow-up.

Although several aspects may need to be confirmed by further studies, we may derive some implications for the design of configuration tasks in the emerging domain of smart environments.

Implication 1 – recommendations as a learning tool: possibly differently than in other contexts, recommendations for configuration tasks may have a learning purpose; it is therefore important to integrate recommendations both at the initial phases of use to foster understanding and acceptance but perhaps in the subsequent stages too to guide the users towards a more advanced use.

Implication 2 – it is important to model the `Need for Cognition` of users to provide a better support: our results suggest that users with higher `Need for Cognition` appreciate (and possibly use) the recommendations in a different way than those with lower `Need for Cognition`; if the system were able to estimate the actual `Need for Cognition` of a specific users, it would possibly provide recommendations at the right moment (that is, before the task for the users with lower `Need for Cognition` and after for users with higher `Need for Cognition`).

Implication 3 – modelling personality traits for improving support: in general, it might be important to adapt the interaction with the users depending on personality; further and more specific studies are still needed but our investigation provides support to the idea that human-computer interaction can be improved (either in terms of effectiveness or user experience) by adapting to specific personality traits; indeed, our study also demonstrates that the impact of personality is not easy to understand: in our case, the most promising trait, `Locus of Control`, seems not to be relevant while `Need for Cognition` and Self Efficacy had a stronger impact but in a complicated way.

6. Conclusion

In this paper, we have described an empirical study that aimed to assess how helpful trigger, action, and rule recommendations are to users who are given a configuration task in a smart home scenario. Perceived usefulness has been investigated in connection with the psychological constructs of `Locus of Control`, `Self-efficacy`, and `Need for Cognition`.

Our results suggest that overall recommendations actually help improve users' performance in the task. In addition, our research also provides evidence that the personality traits of `Need for Cognition` and `Self-efficacy` may play an important role in assessing the perceived usefulness of recommendations, and they have some relation in conjunction with the performance of the task. These effects are somehow different from those reported in other studies where recommendations are provided for a choice task rather than a constructive task, as it happens in the End-User Development domain, and where the focus is on explanations rather than recommendations themselves.

Although these findings need to be better evaluated and generalised, they seem to suggest that it is important to model these personality traits as part of a *user model* to effectively decide if and how it is effective to provide users with recommendations. Furthermore, this choice should also depend on the type of task and the actual level of performance or the expected one given the users' competence.

As a limitation of our work, we can mention the convenience sampling strategy and the homogeneity of the sample, which is composed mainly of students of HCI and interaction design university courses, with similar demographic characteristics. By its nature, the study targeted potential users with a minimum level of familiarity with and interest in technology, as well as an adequate level of understanding of the trigger-action condition concept. Without this minimum level of interest and knowledge, there was a risk of very poor task performance, low participation levels, and/or excessive frustration for the participants, which would have compromised the study's results. It is also worth noting that users potentially interested in configuring smart environments typically possess this minimum level of motivation, familiarity, and knowledge. From this perspective, the students tested in the present study were ideal participants. The fact that, overall, participants exhibited an average level of familiarity with IoT devices and the relatively wide distribution of their scores on the IoT familiarity scale (see the Results section) provides some indication that our results might be generalisable to a broader population. However, further studies involving other types of potential users are needed to investigate to what extent our findings can be generalised to people of different ages and with different educational and professional backgrounds.

A further limitation is that the artificial setting of the experiment, which requires people to imagine a usage scenario, may have impacted the attitude towards it and, consequently, the performance by inducing some bias in the measures. Somehow connected to this point is the fact that, especially in a real life context, there may be several other factors which impact users' perception and adoption of recommendations, possibly mediating the effects of psychological traits. In our study, we have controlled that no such impact can be ascribed to differences in familiarity with technology and IoT, as self-assessed by participants, but other facets of these concepts, as well as other factors (think, for example, of specific user needs in configuration tasks), can be expected to play a role. While investigating all such factors is out of the scope of this study, it can represent an interesting direction for future work.

Addressing the aforementioned limitations, in future work we also plan to replicate the study with a wider and more heterogeneous sample of users, both in the same context and in different domains, to see if our results are valid also in different environments and for different types of tasks (e.g. constructive vs. choice tasks), thus aiming at generalising our current results beyond End-User Development. At the same time, we are developing a rule recommender system to test the effectiveness of the provision of recommendations using the scores of personality traits stored in the user model as decision thresholds.

Notes

1. http://www.empathy-project.eu/
2. https://app.empathy.di.unito.it
3. IFTTT is a private commercial company that runs services that allow a user to program a response to events (https://ifttt.com/).
4. For example, we asked questions like *Would you be interested in accessorising your home with smart objects?*, *What goals would you like your smart objects to help you achieve?*, *How much do you think you would be willing to spend to buy smart objects for your home?*
5. https://ipip.ori.org/
6. An example from Levenson's `Locus of Control` scale (Levenson 1981) is '(I) believe that my success depends on ability rather than luck'.
7. https://ipip.ori.org/newSingleConstructsKey.htm\#\#Locus-of-Control
8. https://ipip.ori.org/newCPIKey.htm\#\#Self-Efficacy
9. https://ipip.ori.org/newSingleConstructsKey.htm\#\#Need-for-Cognition
10. The likelihood-ratio test compares the goodness of fit of two competing statistical models. A relatively more complex model is compared to a simpler model to see if it fits a particular dataset significantly better. The likelihood ratio test statistic is expressed by a chi-square value (χ^2) with an associated *p*-value. *P*-values lower than the conventional threshold of .05 indicate that there is a significant difference in the goodness of fit between the two models
11. Notice that χ^2s (chi-squares) is the plural form of χ^2 and *ps* (*p* values) is the plural form of *p*.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work has been supported by the Italian Ministry of University and Research (MIUR) under grant PRIN 2017 'EMPATHY: EMpowering People in deAling with internet of THings ecosYstems'. This research received partial support from the PNRR project FAIR - Future AI Research (PE00000013), under the PNRR MUR program funded by the NextGenerationEU

ORCID

Massimo Zancanaro http://orcid.org/0000-0002-1554-5703

References

Alfadda, H. A., and H. S. Mahdi. 2021. "Measuring Students' Use of Zoom Application in Language Course Based on the Technology Acceptance Model (tam)." *Journal of Psycholinguistic Research* 50 (4): 883–900. https://doi.org/10.1007/s10936-020-09752-1.

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics*. Cambridge: Cambridge University Press.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. "Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items." *Journal of Memory and Language* 59 (4): 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall.

Bansal, J., M. B. Flannery, and M. H. Woolhouse. 2021. "Influence of Personality on Music-Genre Exclusivity." *Psychology of Music* 49 (5): 1356–1371. https://doi.org/10.1177/0305735620953611.

Bansal, S., and D. Kumar. 2020. "IoT Ecosystem: A Survey on Devices, Gateways, Operating Systems, Middleware and Communication." *International Journal of Wireless Information Networks* 27 (3): 340–364. https://doi.org/10.1007/s10776-020-00483-7.

Bartle, R. 1996. "Hearts, Clubs, Diamonds, Spades: Players Who Suit Muds." *Journal of MUD Research*1 (1): 19.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2014. “Fitting Linear Mixed-Effects Models Using lme4.” arXiv preprint arXiv:1406.5823.

Bologna, C., A. C. De Rosa, A. De Vivo, M. Gaeta, G. Sansonetti, and V. Viserta. 2013 . “Personality-Based Recommendation in E-Commerce.” In *EMPIRE 2013: Emotions and Personality in Personalized Services*. CEUR Workshop Proceedings, edited by S. Berkovsky, E. Herder, P. Lops, and O. C. Santos, Vol. 997. CEUR-WS.org.

Cacioppo, J., and R. Petty. 1982. “The Need for Cognition.” *Journal of Personality and Social Psychology*42 (1): 116–131. https://doi.org/10.1037/0022-3514.42.1.116.

Cacioppo, J. T., R. E. Petty, J. A. Feinstein, and W. B. G. Jarvis. 1996. “Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition.” *Psychological Bulletin* 119 (2): 197–253. https://doi.org/10.1037/0033-2909.119.2.197.

Cantador, I., I. Fernández-Tobías, and A. Bellogín. 2013. “Relating Personality Types with User Preferences in Multiple Entertainment Domains.” In *EMPIRE 2013: Emotions and Personality in Personalized Services*. CEUR Workshop Proceedings, edited by S. Berkovsky, E. Herder, P. Lops, and O. C. Santos, Vol. 997. CEUR-WS.org.

Casati, F., S. Castano, M. Fugini, I. Mirbel, and B. Pernici. 2000. “Using Patterns to Design Rules in Workflows.” *IEEE Transactions on Software Engineering* 26 (8): 760–785. https://doi.org/10.1109/32.879813.

Cena, F., C. Gena, C. Mattutino, M. Mioli, A. Moreno, and F. Vernero. 2021. “Supporting Configuration Choices in Smart Environments through Personalized Recommendations.” In *Proceedings of the 2nd International Workshop on Empowering People in Dealing with Internet of Things Ecosystems Co-Located with INTERACT 2021*, Bari, Italy, Online / Bari, Italy, September 30, 2021. CEUR Workshop Proceedings, edited by G. Desolda, V. Deufemia, M. Matera, F. Paternò, M. Zancanaro, and F. Vernero, Vol. 3053, 23–27. CEUR-WS.org.

Cohen, A. 1957. “Need for Cognition and Order of Communication as Determinants of Opinion Change.” In *Research with the Locus of Control Construct*, edited by C. Hovland, 1–15. Yale University Press.

Corno, F., L. De Russis, and A. M. Roffarello. 2017. “A High-Level Approach towards End User Development in the Iot.” In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 1546–1552. CHI EA ’17, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3027063.3053157.

Cypher, A. 1991. “Eager: Programming Repetitive Tasks by Example.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY: ACM

Delic, A., J. Neidhardt, T. N. Nguyen, and F. Ricci. 2018. “An Observational User Study for Group Recommender Systems in the Tourism Domain.” *Journal of Information Technology & Tourism* 19 (1-4): 87–116. https://doi.org/10.1007/s40558-018-0106-y.

Deniz, M. 2011. “An Investigation of Decision Making Styles and the Five-Factor Personality Traits with Respect to Attachment Styles.” *Educational Sciences: Theory and Practice* 11 (1): 105–113.

De Simone, L., D. Gadia, D. Maggiorini, and L. A. Ripamonti. 2021. “Design of a Recommender System for Video Games Based on in-Game Player Profiling and Activities.” In *CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter. CHItaly ’21*, Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3464385.3464742.

Dhelim, S., N. Aung, M. A. Bouras, H. Ning, and E. Cambria. 2022. “A Survey on Personality-Aware Recommendation Systems.” *Artificial Intelligence Review* 55 (3): 2409–2454. https://doi.org/10.1007/s10462-021-10063-7.

Dhelim, S., N. Aung, and H. Ning. 2020. “Mining User Interest Based on Personality-Aware Hybrid Filtering in Social Networks.” *Knowledge-Based Systems*206:106227. https://doi.org/10.1016/j.knosys.2020.106227.

Dhelim, S., H. Ning, N. Aung, R. Huang, and J. Ma. 2020. “Personality-Aware Product Recommendation System Based on User Interests Mining and Metapath Discovery.” *IEEE Transactions on Computational Social Systems* 8 (1): 86–98. https://doi.org/10.1109/TCSS.6570650.

Dunn, G., J. Wiersema, J. Ham, and L. Aroyo. 2009. “Evaluating Interface Variants on Personality Acquisition for Recommender Systems.” In *User Modeling, Adaptation, and Personalization*, edited by G. J. Houben, G. McCalla, F. Pianesi, and M. Zancanaro, 259–270. Berlin, Heidelberg: Springer Berlin Heidelberg.

Fernández-Tobías, I., M. Braunhofer, M. Elahi, F. Ricci, and I. Cantador. 2016. “Alleviating the New User Problem in Collaborative Filtering by Exploiting Personality Information.” *User Modeling and User-Adapted Interaction* 26 (2-3): 221–255. https://doi.org/10.1007/s11257-016-9172-z.

Ghiani, G., M. Manca, F. Paternò, and C. Santoro. 2017. “Personalization of Context-Dependent Applications Through Trigger-Action Rules.” *ACM Transactions on Computer-Human Interaction (TOCHI)* 24 (2): 1–33. https://doi.org/10.1145/3057861.

Gough, H. 1996. *CPI Manual*. 3rd ed. Palo Alto, CA: Consulting Psychologists Press.

Haines, W., M. T. Gervasio, A. Spaulding, and B. Peintner. 2010. “Recommendations for End-User Development.” In *Proceedings of the ACM RecSys Workshop on User-Centric Evaluation of Recommender Systems and their Interfaces*. CEUR Workshop Proceedings, edited by B. P. Knijnenburg, L. Schmidt-Thieme, and D. Bollen, Vol. 612. CEUR-WS.org.

Holland, J. L. 1997. *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments*. 3rd ed. Odessa, FL: Psychological Assessment Resources.

Horvitz, E., J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. 1998. “The Lumière Project: Bayesian User Modeling for Inferring the Goals and Needs of Software Users.” In *UAI’98*, 256–265. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Hu, R., and P. Pu. 2010. “A Study on User Perception of Personality-Based Recommender Systems.” In *International Conference on User Modeling, Adaptation, and Personalization*, 291–302. Springer.

Hu, R., and P. Pu. 2011. “Enhancing Collaborative Filtering Systems with Personality Information.” In *Proceedings of the Fifth ACM Conference on Recommender Systems.*

RecSys '11, 197–204. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2043932.2043969.

Huang, J., and M. Cakmak. 2015. "Supporting Mental Model Accuracy in Trigger-Action Programming." In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 215–225. New York, NY: ACM

Ingendahl, M., D. Hummel, A. Maedche, and T. Vogel. 2021. "Who Can Be Nudged? Examining Nudging Effectiveness in the Context of Need for Cognition and Need for Uniqueness." *Journal of Consumer Behaviour* 20 (2): 324–336. https://doi.org/10.1002/cb.1861.

Jameson, A., S. Gabrielli, P. O. Kristensson, K. Reinecke, F. Cena, C. Gena, and F. Vernero. 2011. "How Can We Support Users' Preferential Choice?." In *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Extended Abstracts Volume, May 7-12, 2011*, edited by D. S. Tan, S. Amershi, B. Begole, W. A. Kellogg, and M. Tungare, 409–418. Vancouver, BC, Canada: ACM. https://doi.org/10.1145/1979742.1979620.

Jameson, A., M. C. Willemsen, A. Felfernig, M. De Gemmis, P. Lops, G. Semeraro, and L. Chen. 2015. "Human Decision Making and Recommender Systems." In *Recommender Systems Handbook*, 611–648. Springer.

Karumur, R. P., T. T. Nguyen, and J. A. Konstan. 2018. "Personality, User Preferences and Behavior in Recommender Systems." *Information Systems Frontiers* 20 (6): 1241–1265. https://doi.org/10.1007/s10796-017-9800-0.

Kim, D. D. E. 2020. "Drawn to the Screen by Who We are and Who We Aspire to Be: Brand-Self Congruence Differences in Movie Preferences." *International Journal on Media Management* 22 (3-4): 144–167. https://doi.org/10.1080/14241277.2021.1920022.

Koren, Y., R. Bell, and C. Volinsky. 2009. "Matrix Factorization Techniques for Recommender Systems." *Computer* 42 (8): 30–37. https://doi.org/10.1109/MC.2009.263.

Kosinski, M., Y. Bachrach, P. Kohli, D. Stillwell, and T. Graepel. 2014. "Manifestations of User Personality in Website Choice and Behaviour on Online Social Networks." *Machine Learning* 95 (3): 357–380. https://doi.org/10.1007/s10994-013-5415-y.

Levenson, H. 1981. "Differentiating among Internality, Powerful Others, and Chance." In *Research with the Locus of Control Construct*, edited by H. M. Lefcourt, 1–15. Academic Press.

Lieberman, H., F. Paternò, M. Klann, and V. Wulf. 2006. *End-User Development: An Emerging Paradigm*, 1–8. Dordrecht: Springer Netherlands. https://doi.org/10.1007/1-4020-5386-X_1.

Manolika, M., R. Kotsakis, M. Matsiola, and G. Kalliris. 2022. "Direct and Indirect Associations of Personality with Audiovisual Technology Acceptance Through General Self-Efficacy." *Psychological Reports* 125 (2): 1165–1185. https://doi.org/10.1177/0033294121997784.

Masui, T., and K. Nakayama. 1994. "Repeat and Predict–Two Keys to Efficient Text Editing." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '94*, 118–130. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/191666.191722.

McCrae, R., and O. P. John. 1992. "An Introduction to the Five-Factor Model and Its Applications." *Journal of Personality* 60 (2): 175–215. https://doi.org/10.1111/jopy.1992.60.issue-2.

Millecamp, M., N. N. Htun, C. Conati, and K. Verbert. 2019. "To Explain or Not to Explain: The Effects of Personal Characteristics when Explaining Music Recommendations." In *Proceedings of the 24th International Conference on Intelligent User Interfaces, IUI '19*, 397–407. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3301275.3302313.

Moscato, V., A. Picariello, and G. Sperlí. 2021. "An Emotional Recommender System for Music." *IEEE Intelligent Systems* 36 (5): 57–68. https://doi.org/10.1109/MIS.2020.3026000.

Nguyen, T. N., F. Ricci, A. Delic, and D. G. Bridge. 2019. "Conflict Resolution in Group Decision Making: Insights From a Simulation Study." *User Modeling and User-Adapted Interaction* 29 (5): 895–941. https://doi.org/10.1007/s11257-019-09240-9.

Nunes, M. A. S. N. 2009. *Recommender Systems Based on Personality Traits: Could Human Psychological Aspects Influence the Computer Decision-Making Process?*. Berlin: VDM Verlag.

Ochirbat, A., T. K. Shih, C. Chootong, W. Sommool, W. Gunarathne, H. H. Wang, and Z. H. Ma. 2018. "Hybrid Occupation Recommendation for Adolescents on Interest, Profile, and Behavior." *Telematics and Informatics* 35 (3): 534–550. https://doi.org/10.1016/j.tele.2017.02.002.

Paton, N. W. 2012. *Active Rules in Database Systems*. New York, NY: Springer Science & Business Media.

Peintner, B., J. Dinger, A. C. Rodriguez, and K. L. Myers. 2009. "Task Assistant: Personalized Task Management for Military Environments." In *Proceedings of the Twenty-First Conference on Innovative Applications of Artificial Intelligence*, edited by K. Z. Haigh, and N. Rychtyckyj, July 14-16, 2009. Pasadena, California, USA: AAAI.

Quijano-Sánchez, L., J. A. Recio-García, and B. Díaz-Agudo. 2010. "Personality and Social Trust in Group Recommendations." In *2010 22nd IEEE International Conference on Tools with Artificial Intelligence*, Vol. 2, 121–126. Washington, DC: IEEE Computer Society. https://doi.org/10.1109/ICTAI.2010.92.

Reddy, P., K. Chaudhary, B. Sharma, and R. Chand. 2021. "The Two Perfect Scorers for Technology Acceptance." *Education and Information Technologies* 26 (2): 1505–1526. https://doi.org/10.1007/s10639-020-10320-2.

Rotter, J. B. 1966. "Generalized Expectancies for Internal Versus External Control of Reinforcement." *Psychological Monographs* 80 (1): 1–28. https://doi.org/10.1037/h0092976.

Ruvini, J. D., and C. Dony. 2000. "Ape: Learning User's Habits to Automate Repetitive Tasks." In *Proceedings of the 5th International Conference on Intelligent User Interfaces, IUI '00*, 229–232. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/325737.325854.

Sharan, N. N., and D. M. Romano. 2020. "The Effects of Personality and Locus of Control on Trust in Humans

Versus Artificial Intelligence." *Heliyon* 6 (8): e04572. https://doi.org/10.1016/j.heliyon.2020.e04572.

Thomas, K. W. 1992. "Conflict and Conflict Management: Reflections and Update." *Journal of Organizational Behavior* 13 (3): 265–274. https://doi.org/10.1002/job.4030130307.

Tkalcic, M., and L. Chen. 2015. "Personality and Recommender Systems." In *Recommender Systems Handbook*, 715–739. Springer.

Tong, S. T., E. F. Corriero, R. G. Matheny, and J. T. Hancock. 2018. "Online Daters' Willingness to Use Recommender Technology for Mate Selection Decisions." In *IntRS@RecSys*.

Wu, W., L. Chen, and L. He. 2013. "Using Personality to Adjust Diversity in Recommender Systems." In *Proceedings of the 24th ACM Conference on Hypertext and Social Media, HT '13*, 225–229. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2481492.2481521.

Yalcin, E., and A. Bilge. 2023. "Popularity Bias in Personality Perspective: An Analysis of How Personality Traits Expose Individuals to the Unfair Recommendation." *Concurrency and Computation: Practice and Experience* 35 (9): e7647. https://doi.org/10.1002/cpe.v35.9.

Yorke-Smith, N., S. Saadati, K. L. Myers, and D. N. Morley. 2009. "Like an Intuitive and Courteous Butler: A Proactive Personal Agent for Task Management." In *AAMAS '09, International Foundation for Autonomous Agents and Multiagent Systems*, 337–344. Richland, SC.

Zancanaro, M., G. Gallitto, D. Yem, and B. Treccani. 2022. "Improving Mental Models in IoT End-User Development." *Human-Centric Computing and Information Sciences* 12:633–659. https://doi.org/10.22967/HCIS.2022.12.048.

Zorzi, M., M. Bonato, B. Treccani, G. Scalambrin, R. Marenzi, and K. Priftis. 2012. "Neglect Impairs Explicit Processing of the Mental Number Line." *Frontiers in Human Neuroscience* 6:125. https://doi.org/10.3389/fnhum.2012.00125.