An Instrument for measuring users' meta-intents

Yuan Ma
Tim Donkers
Timm Kleemann
Jürgen Ziegler
a15205352825@gmail.com
tim.donkers@uni-due.de
timm.kleemann@uni-due.de
juergen.ziegler@uni-due.de
University of Duisburg-Essen
Duisburg, NRW, Germany

ABSTRACT

We propose the concept of meta-intents which represent high-level user preferences related to the interaction and decision-making in conversational recommender systems (CRS) and present a questionnaire instrument for measuring meta-intents. We conducted a two-stage user study, an exploratory study with 212 participants on Prolific, and a confirmatory study with 394 participants on Prolific. We obtained a reliable and stable meta-intents questionnaire with 22 question items, corresponding to seven latent factors (concepts). These seven factors cover important interaction preferences and are closely related to users' decision-making process. For example, the factor dialog-initiative reflects whether users prefer to follow the system's guidance or ask their own questions in a CRS. We conducted statistical analyses of meta-intents in two domains (smartphones and hotels), and a general chatbot scenario. We also investigated the influence of additional factors (demography, decision-making style) on meta-intents through Structural Equation Modeling (SEM). Our results provide preliminary evidence that the proposed meta-intents are domain and demography (gender, age) independent. They can be linked to the general decision-making style and can thus be instrumental in translating general decision-making factors into more concrete design guidance for CRS and their potential personalization. Meta-intents also provide a basis for future analyses of interaction behavior in CRS and the development of a cognitively founded theoretical framework.

CCS CONCEPTS

• **Human-centered computing → User models**.

KEYWORDS

Meta intents, Conversational UI design, Conversational recommender systems

CHIIR '23, March 19–23, 2023, Austin, TX, USA

ACM ISBN 979-8-4007-0035-4/23/03…$15.00
https://doi.org/10.1145/3576840.3578317

ACM Reference Format:
Yuan Ma, Tim Donkers, Timm Kleemann, and Jürgen Ziegler. 2023. An Instrument for measuring users' meta-intents. In *ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '23), March 19–23, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3576840.3578317

1 INTRODUCTION

Conversational recommender systems (CRS[36]) have been gaining increased attention in research and industry in recent years [11, 12]. Generally, conversational techniques can provide users with strong guidance to achieve their goals combined with a high level of flexibility in expressing their needs. Jannach et al. [17] define CRS as recommender systems supporting a task-oriented, multi-turn dialog and distinguishing between natural language-based, form-based, and critiquing approaches. Due to the advances in NLP techniques in recent years, natural language-based CRS have become a subject of particularly extensive research. In addition to the general style of interaction used, various general design-related properties can be distinguished. Fu et al. [10] classify NLP-based CRS based on dialog initiative. User-initiated CRS allows the user to lead a dialog rather than reacting to the system's questions. System-initiated CRS, on the contrary, guide the user through the dialog with a system-asks-user-responds pattern. Some systems use mixed-initiative dialog or combine natural language with a form-based style integrating GUI elements for entering responses [28]. A general question of high importance is how the CRS should formulate its questions. Questions need to be formulated at a level of abstraction that is appropriate for the user's knowledge and goals, asking, for instance, either about the intended use of the product or about some specific technical features. Dialog flow should follow the user's likely mental decision process, providing sufficient flexibility without becoming overly complex, and recommendations should be presented in appropriate numbers and with an appropriate level of detail. Many of these factors likely depend on the user's psychological characteristics and preferences, personalizing the conversation therefore seems a promising, yet challenging approach, to increase the effectiveness and acceptance of CRS.

To address these challenges a thorough understanding of user needs, interactive behaviors, and their decision-making style is needed. Little research, however, has investigated the influence of psychological user characteristics and general, dialog-related preferences in the context of CRS thus far [38]. Clearly, psychological factors are an important resource underutilized by current CRS. Recently, Ma et al. [29] proposed the concept of meta-intents, which are interaction-oriented characteristics that represent general user preferences when interacting with a CRS, such as obtaining detailed information about items or comparing products [29]. Different from *intents* which is a term in CRS field usually represents the low-level preferences refer to concrete properties of the desired item (specifically Add Details [7]), *meta-intents* represent a more abstract level, meta-level preferences that relate to the conversation and type of questions in a CRS [27]. Meta-intents can be considered a useful concept for understanding users' needs and for optimizing and personalizing CRS dialogs. However, in the previous work the empirical foundation for establishing valid meta-intents factors was limited and there was still a lack of a reliable instrument for measuring them.

In this paper, we further develop and validate the concept of meta-intents, and investigate the relation between established decision-making theory [15] and interaction preferences in CRS. We revisit the initial meta-intents definitions and better align them with the typical decision process in recommender scenarios. One of the first questions along this process is whether the user wishes to obtain system guidance and how strong the guidance should be, which is determined by a number of factors, such as the user's understanding of his or her own needs and the product features, as well as the degree of trust in the assistant. When system assistance is accepted, the question that comes into view is which form of assistance is preferred. When users are active and have a clear goal they may prefer to ask their own questions in an unrestricted manner (*user-initiative*), otherwise, the system will ask questions first (*system-initiative*). In the perspective of preference elicitation, we also need to consider the users' level of cognitive engagement in the decision process, for example, whether they like to examine all details about an item or only key features (*interest in details*) and whether they engage in performing detailed comparisons (*comparison-orientation*). Preferences may also differ with respect to the size (*scope of choice*) and diversity (*diversity-orientation*) of the resulting recommendation set and the possibility to critique recommended items (*critiquing-orientation*). More general individual preferences also need to be considered such as the degree of anthropomorphism of an assistant (*human-likeness*), the willingness to spend time in the dialog (*efficiency-orientation*), and the need for explaining the recommendations (*explanation-orientation*).

To refine and validate these concepts formulated with respect to the decision process supported by CRS, we performed two studies, following a classical two-stage approach comprising an exploratory and a confirmatory analysis step. This process resulted in a stable and validated set of meta-intents that we see as a starting point for analyzing CRS interactions more deeply, for predicting meta-intents from interaction behavior, and for personalizing CRS dialogs.

The studies presented in this paper make three major contributions:

- We propose and empirically validate a set of meta-intents that represent high-level user preferences with respect to interacting with a CRS.
- Based on exploratory and confirmatory studies we present a validated meta-intents questionnaire with 22 question items, corresponding to seven meta-intents factors (concepts).
- From an analysis of three recommendation scenarios (smartphones, hotels, and a neutral scenario), We provide evidence for the generalisability and domain-independence of the proposed meta-intents, and also show the impact of individual decision-making style on some of the meta-intents.

2 RELATED WORK

Conversational Recommender Systems (CRS) have become a rapidly growing and popular research area because they provide a flexible, human-like multi-turn dialog for preference elicitation, which is essential for generating personalized recommendations [25]. Jannach et al. [17] distinguish three types of CRS, differing in the style and structure of the interaction used: natural language-based, form-based, and critiquing-based.

NLP-based CRS have received considerable interest recently due to the advancements in natural language processing. They typically use a question-answer format [44]. NLP-based CRS provide users with a high level of freedom to express their preferences, but misrecognitions and misunderstandings still happen frequently, often leading to user frustration. NLP-based CRS clearly have a possibility for both *user-initiative* and *system-initiative* dialog. However, there is still a lack of community discussion on how to personalize them for users.

Form-based CRS present questions and answer in a GUI style, leading users through a predefined dialog structure. This type of CRS has many advantages as they provide guidance to the users, avoid errors, and can incorporate domain knowledge [21]. The disadvantage is that the question sequences/paths are hand-crafted, not enough freedom. An important problem for form-based CRS is that a number of users do not have enough patience to answer all the questions, so differentiated *efficiency-orientation* is important for personalizing question sequence length for users.

Critiquing-based CRS will first recommend an item and then capture users' feedback in the form of critiques [8]. It helps users to efficiently refine their preferences by providing more options. However, showing all features may not be suitable for all users, so knowing users' preference of *interest in details* will be helpful for the personalization of feature list length.

Currently, very limited research has as yet studied the influence of psychological user characteristics on the usage, perception, and acceptance of CRS, and the design implications of these characteristics. Papenmeier et al. [31] investigated human advisory dialogs, identifying some recurring strategies such as funneling to successively narrow down the space of potential items. Kleemann et al. [20] investigated user behavior and personal characteristics when using a form-based conversational advisor in combination with other decision aids (chatbot, faceted filtering, recommender), assessing users' component preferences, utilization, and switching behavior. [19]. Atas et al. [1] summarize that preferences are determined and adapted is influenced by various factors such as personality traits, emotional states, and cognitive biases. Ma et al. [29] proposed meta-intents, proved they can be linked with stable decision-making style characteristics, and can be beneficial with personalized CRS interaction [27, 29].

To provide design guidance for CRS and to potentially adapt them to the individual user, a deeper understanding of the psychological factors influencing users' decision-making and interaction behavior in CRS is required. For recommender systems in general, the influence of psychological characteristics on users' preference construction and decision-making has been shown repeatedly [1]. Lex et al. [24] distinguish between factors related to cognition, personality, and emotion. The influence of psychological characteristics such as the Big Five personality factors (e.g. [14, 37]), Need for Cognition [30], or cognitive biases [41] has been studied in several works. However, these studies mostly aim at better understanding user preferences with respect to the recommended items and at improving their accuracy. In contrast, the relationship between psychological factors and the design of advisory dialogs in CRS remains an underexplored area. Especially theories related to human decision-making styles appear to be promising points of departure for studying this relation. The distinction between rational and intuitive decision-making styles [15] or cognitive styles such as the need for cognition may influence users' assessment of CRS. More domain-specific theories such as Shopping Orientation [5, 6], distinguishing between task-focused and experiential shopping are also of interest. However, to our knowledge none of these approaches has yet been applied to CRS.

User goals and preferences when interacting with a CRS may be located on different levels of abstraction. Low-level preferences refer to concrete properties of the desired item. Jameson et al. [16] suggest high-level factors (such as economy and safety) but these factors are related to the product itself, not to the way users prefer to interact with a CRS. On a more abstract level, meta-level preferences that relate to the conversation and type of questions in a CRS have, to our knowledge, very limited study on it. Ma et al. [29] proposed meta-intents concepts and bring them into the CRS design [27, 28]. However, each meta-intents concept has only one question item. Due to the lack of multiple-question items and strict statistical tests, it can not be treated as a complete and reliable basis for CRS. In this paper, we further develop meta-intents theory and provide a detailed introduction and explanation of meta-intents concepts.

3 META-INTENTS

Meta-intents are constructs that we propose to describe users' interaction preferences with a CRS at an abstract level. While intent detection methods in CRS are usually directed at extracting user preferences at the level of specific item types and attributes, we see meta-intents as higher-level preferences that indicate in which way users prefer to conduct a conversation with the system. We assume that these preferences are affected by general psychological traits such as personality or decision-making style, yet they should be concrete enough to inform the design of a CRS and to possibly personalize the dialog in a number of aspects. The meta-intents concept thus aims at bridging the gap between general psychological traits that research has shown to be robust and valid but that are too abstract to directly guide the design of a dialog, and the item-specific goals that determine the user's search behavior.

While sufficiently concrete to provide guidance for dialog design and personalization, we aim at proposing constructs that are still at a level that is independent of the specific item domain, covering a broad range of recommendation scenarios. In fact, the concepts we propose may be applicable beyond the field of CRS, yet they should capture important aspects of dialog strategy, content, and wording in CRS. Based on these assumptions we define meta-intents as a set of abstract factors reflecting users' individual interaction preferences in CRS that are largely independent of the domain and specific item preferences but impacted by general personality traits.

With the studies presented in the following sections, we aim at establishing and validating a set of meta-intents that cover a reasonably large range of general CRS design questions. Our second goal is to develop a questionnaire instrument that can be applied to measure users' meta-intents for providing, for example, ground truth in future work aimed at automatically extracting meta-intents from user behavior. To this end, we empirically determine a number of relevant factors in two online surveys. In the first study, we apply an exploratory approach presenting participants with a large set of questionnaire items formulated with respect to an initial set of factors we derived from different pertinent CRS design options collected from the literature and own prior work. In the second study, a confirmatory factor analysis was applied to validate the factors. In this study, we furthermore analyzed the influence of decision-making style and demographic variables on the meta-intents through Structural Equation Modeling.

4 EXPLORATORY STAGE STUDY

4.1 Concept

We postulated the following initial set of meta-intents factors. Each factor stands for a construct and represents an interaction preference in CRS. The explanation of the construct and the source of the idea and the interaction problem to be solved are stated below. We see this list as an important step towards defining factors relevant to users' interaction preferences and decision-making process in CRS.

- **Task-focus**
  This factor indicates how well-defined the user's mental goal is and how stringently this goal is pursued in the search for a suitable item. It is a factor that is hypothetically strongly influenced by the user's decision-making style but may also vary most with the item domain addressed and the specific current needs of the user. However, in CRS the level of task-focus may also determine to what extent users are attracted to items they would not have considered initially ("unplanned items"). Users with a strong task-focus are less likely to consider unplanned products when shopping. In this case, recommending novel or advertised products may interfere with the user's goal. Users with low task-focus, in contrast, will react to seeing unplanned products when shopping, and presenting a more varied set of items or promoting new products may improve the user shopping experience. Task-focused versus experiential shopping behavior are also central personal characteristics in the Shopping Orientation model proposed by Büttner et al. [4, 6]. In terms of CRS design, task-focus may inform when to apply solid or soft

constraints [33, 40] in eliciting preferences, or whether to show novel and promotional products in the recommendation list.

- **Openness for guidance**
  A considerable number of users are reluctant to be strongly guided by a system in their search and decision process and may prefer to explore the available options themselves by using, for example, search or filtering techniques [19, 26]. Many CRS, however, provide strong guidance by leading users through a stepwise decision process, as is the case with GUI-based CRS or question-answer style chatbots. Some concerns may be related to design issues or the currently still limited recognition accuracy in NLP-based systems, but general personal variables such as sense of agency [9] and need for control are also likely to determine the acceptance of system guidance.
- **User-initiative**
  This meta-intent refers to the user's preference of initiating and leading a dialog rather than reacting to the system's questions. The choice between user-initiated or system-initiated interaction is a long-standing HCI theme ever since GUIs became available as an alternative to command languages, but its relevance is again amplified especially in chatbots which can apply either interaction style or a mix of both. Although related to the guidance factor discussed above, user-initiative can be seen as a factor specifically relevant for NLP-based dialogs. User-initiative allows users to express their preferences more freely but also puts more burden on memory where desired item features need to be remembered and recalled. Users with a preference for initiating dialog steps themselves will benefit from providing them with more flexibility and control [10, 12].
- **System-initiative**
  Users who prefer system-initiated conversation tend to interact with the CRS in a more passive and reactive manner. This style of interaction has less demands on the user's domain knowledge which is necessary to express preferences in a self-initiated manner and benefits from recognition instead of recall with respect to memory load. Preference for system-initiative is logically the opposite of user-initiative but we wanted to assess the two factors separately to see whether this is also the case in users' perception of a CRS.
- **Efficiency-orientation**
  This meta-intent refers both to the time spent in a recommendation dialog as well as the number of interaction steps that users find acceptable. CRS with their multi-turn dialog are typically less efficient than selecting from one-shot recommendations. On the other hand, the conversational model integrates preference elicitation in the dialog which needs to be achieved through other interactions delivering explicit or implicit preference signals in non-conversational techniques. Yet, time-consuming, multi-step dialogs often cause the user to abandon the dialog after the first few steps. We assume users differ in their willingness to engage in a longer conversation. It is therefore important to calibrate the length of the dialog, providing an acceptable trade-off between preference collection and the provision of personalized recommendations, ideally resolving the exploration-exploitation problem [12, 36, 42] in CRS subject to the user's efficiency orientation.
- **Human-likeness**
  The debate about the human-likeness of virtual agents has a long tradition [13, 22] which also constitutes an important design aspect for CRS. The factor refers to the question of whether users prefer an anthropomorphic or a more neutral, factual style of interaction. Human-likeness can be achieved by a number of means such as including in the dialog greetings and chitchat, using a personal style of language, or by embodying the agent in an animated character. Choosing an appropriate level of human-likeness (or unlikeness) has been shown to have a considerable impact on user experience [35].
- **Interest in detail**
  This factor describes the degree to which users pay attention to details in the search for and inspection of products. Some users may want to view all features of an item and to obtain a comprehensive understanding of the product. In contrast, other users may only be interested in the feature they care about most and are unwilling to spend time on examining other features. The factor also refers to the depth of understanding users want to get about an item which can be at the level of detailed (e. g. technical) features or at the level of product usage, match with the current mood, or other more general aspects. Concrete design implications relate, among others, to asking feature-level or usage-level questions in the dialog, and the extent of features presented for a recommended item, also whether unimportant features should be left out or folded away in the recommendation panel.
- **Comparison-orientation**
  Decision-making typically involves comparing the available options. Comparing items or features allows users to express their preferences in a relative rather than an absolute fashion. Consequently, comparison functions are a widely-used technique in online shopping. They are also applied, although to a lesser extent, in recommender systems. Comparison functions that are automatically shown or highlighted for access can also be integrated in CRS [39], while CRS-specific support for comparisons can be realized by letting users ask comparative questions [3] or by providing comparative explanations in the dialog flow.
- **Scope of choice**
  The number of recommendations shown to the user is a classical problem in recommender systems research. Longer lists provide users with a greater scope of choice which may increase the attractiveness of the recommendations but also increase choice difficulty [2], whereas shorter lists may reduce cognitive effort and facilitate decision-making. Depending on the design, recommendations in CRS can be shown in a recommendation list separate from the dialog, or inline as part of the conversation. Preference for longer or shorter recommendation lists may depend on general user traits such as decision-making style.

- **Diversity-orientation** The diversity of recommendations is a further factor that can influence recommendation attractiveness and effectiveness, and for which user preference differs [23]. Some users appreciate exploring unknown item types and coming across serendipitous finds [18]. For others, high diversity will increase the difficulty of decision-making and the effectiveness of the recommender.
- **Critiquing-orientation**
This meta-intent describes whether users are willing to use critiquing-oriented functions in a recommender system. Critique is a technique by which users can express preferences by modifying the attributes (or attribute sets) of a sample item shown to them [8]. Linking preference elicitation to a specific recommended product can serve as a cue for elaborating or remembering one's preferences, and may also play the role of an anchor based on which the suitability of other similar items is gauged.
- **Explanation-orientation**
The final proposed meta-intent refers to users' perception of explanations that are provided for recommendations. Explaining recommendations has received considerable attention in the research community recently, since it "helps to improve the transparency, persuasiveness, effectiveness, trustworthiness, and satisfaction of recommendation systems" [43]. However, there is also a debate about whether and to what extent users benefit from explanations in AI (see the meta-analysis by Schemmer et al. [34]). This indicates that the perception and acceptance of explanations may differ among users and it may be important to consider these preferences when designing the CRS dialog.

4.2 Method

To develop a reliable and valid questionnaire that can appropriately measure the 12 meta-intents concepts proposed, we initially compiled a total of 59 question items. In the initial proposal stage, we will not deliberately change the wording to make different concepts have different question items, but focus on each concept itself (as shown in Subsection 4.1) and construct question items from both positive and negative perspectives, attempting to represent this concept accurately. Most of the concepts were presented with six items which included positive questions and negative questions, except for task-focus (3 positive questions, 4 negative questions), user-initiative (2 positive questions), system-initiative factor (2 positive questions), interest in detail (5 positive questions and 2 negative questions), comparison-orientation (2 positive questions and 2 negative questions), explanation-orientation (2 positive questions and 2 negative questions), critiquing-orientation (3 positive questions). Due to the exploratory focus of this study, we defined a larger number of question items than was to be used in the final instrument. After Exploratory factor analysis (EFA), we filtered out items with low factor loadings and updated some question formulations that seemed to be less clearly understood. Since the initial item set contained some redundancy and less relevant questions, we do not show all 59 question items here but present the final version of the items in the section on the second study stage.

We set up two scenarios in the domains of smartphones and hotels, as well as a general, domain-independent scenario. Participants were instructed to imagine buying a new smartphone or booking a hotel online. The hotel and smartphone domains were chosen because they require a sufficiently complex decision process involving a variety of decision criteria. For most people, they are also well-known, real-life tasks that lets us assume they understand the smartphone/hotel features at least to a certain extent. Furthermore, smartphones have a large number of feature options.

In the general, domain-independent scenario, participants were supposed to buy an unspecified product online that they would currently need which we assume covers a broader range of product types. In a between-subjects design, we first randomly assigned participants to one of the three scenarios, obtained demographic data and then presented our questionnaire. All items were to be rated on a 5-point Likert scale.

4.3 Participants

We recruited participants from the crowdsourcing platform Prolific[1], a tool commonly used for academic surveys [32]. We preselected Prolific users based on the following criteria to maximize quality: **1**. participants should be fluent in English; **2**. their success rate should be greater than 95 %. We recruited 226 participants, and all of them finished the study. In our analysis, we only considered participants who passed three inner attention test questions (e.g. It's an attention test, please select strongly agree), leaving us with 212 participants. Each participant received compensation of 1.18£. 107 of the 212 participants were female. Their age ranged from 18 to 70 ($M = 38.50$, $SD = 13.30$). The average duration of the survey was 7.49 minutes ($SD = 2.68$).

4.4 Results

We first performed prerequisite tests for EFA. The Kaiser-Meyer-Olkin (KMO) value was .729 (> 0.7) and Bartlett's test was significant (< .001), which both indicate that our data meets the requirements for performing EFA. Next, we applied Principal Component Analysis (PCA) to extract factors, with Varimax rotation and Kaiser normalization, taking *eigenvalue > 1* as the threshold to determine the number of factors. As a result of the EFA, the 59 items were merged into 18 factors. Again, due to space limitations, we do not present the factor loadings for all items here but only describe the criteria that led us to delete, merge, or modify items. Complete details will be given for the second study.

Deleting an item is based on the criteria:

- the item's factor loadings are all < 0.5
- the item forms a factor individually after removing other items with low factor loadings
- the items' commonality extraction was < 0.5

We deleted two items we had associated with *openness for guidance* and *critiquing-orientation* because they would have each formed a factor individually. One question was removed for *human-likeness* and *diversity-orientation* respectively due to low factor loadings.

When several items simultaneously scored high on two different concepts, we merged the concepts. For example. we merged

[1] https://www.prolific.co

comparison-orientation and *interest in details* into one concept *interest in details*, because 3 of 4 *comparison-orientation* questions and 4 of 7 *interest in details* questions were automatically merged into the first factor. For the same reason, we also merged *diversity-orientation* and *scope of choice* into *scope of choice*, and merged *user-initiative* and *system-initiative* into *dialog-initiation*.

We reformulated an item if its factor loading was high but still below the threshold of 0.7 to make it more expressive.

We also noticed the most positive questions survived, so we decided to keep only the positive questions for each concept and make them uniform, except for *dialog-initiation*, as it has 2 opposing subconcepts: *user-initiative* and *system-initiative*. After the first exploratory stage of the study, we ended up with 8 concepts and 24 items and started the second confirmatory stage of the study to validate and confirm them.

5 CONFIRMATORY STAGE STUDY

5.1 Meta-intent questionnaires

The questionnaire resulting from the EFA phase comprised 24 items for 8 meta-intents concepts. In the following, we give an overview of the factors and items that we submitted to a confirmatory analysis:

- **Interest in details** (**1**. In most cases, I will compare different products in detail before making a decision. **2**. I usually inspect all product features before deciding. **3**. Even if a product meets my key needs, I still will check all product features to be sure I make the right choice.)
- **Explanation-orientation** (**1**. If a system recommends products to me, I would like to see an explanation of why it is recommended. **2**. The system should provide information about which data are used to produce a recommendation. **3**. For a recommended product, only showing the features is not enough for me. An explanation for the recommendation is also necessary.)
- **human-likeness** (**1**. I appreciate it if the system addresses me personally, like a human salesperson. **2**. When conversing with a digital assistant, I would like to have the feeling that I am talking to a real person. **3**. I like systems with which I can have a conversation in natural language.)
- **Scope of choice** (**1**. Seeing a larger range of suggested products at a glance helps me making my decision. **2**. When the system recommends products, I rather like to see a longer list than a short one. **3**. I find it helpful if a system shows a large number of recommendations.)
- **Task-focus** (**1**. I usually only buy products that serve my current needs. **2**. I tend to ignore products that do not match my initial goal. **3**. Usually, I only buy the things I had in mind originally.)
- **Efficiency-orientation** (**1**. I usually don't like spending much time searching for a product. **2**. Finding a product I like quickly is very important for me. **3**. The number of interaction steps needed for finding a product should be as low as possible.)
- **Dialog-initiation** (**1**. When conversing with a digital sales assistant (like a chatbot), I would prefer asking my own questions rather than answering the chatbot's questions. **2**. In a chatbot dialog, I believe my search will be more effective if I can ask my own questions. **3**. When interacting with a chatbot, I prefer answering the system's questions rather than asking my own questions. **4**. I prefer being guided by the system through a search process rather than formulating my own queries.)
- **Critiquing-orientation** (**1**. I can often only decide which product features I prefer when I see a sample product. **2**. Compared to searching, I prefer to see an example product first and then adjust its features according to my preference to find a suitable product.)

5.2 Method

To verify the quality of the questionnaire, we conducted a confirmatory study approximately one and a half months after the first-stage exploratory study. The method of the confirmatory study was basically the same as that of the exploratory study. We again used the Prolific platform, involving three domains (smartphone, hotel, neutral). All items were measured with a 5-point Likert scale. In addition to the meta-intents questions we included six items from the Decision Styles Scale (DSS [15]) questionnaire measuring users' degree of rationality and intuitiveness in their decision-making, and two demographic questions (gender, age). By including the DSS scales, we wanted to examine whether meta-intents are directly influenced by more general psychological traits, also the potential impact of age and gender was to be tested.

After performing the CFA and obtaining scores for the different meta-intents, we performed ANOVA with the *domain* as the independent variable to test whether there are significant differences between the domains and which meta-intents show domain-specific influences. We postulated the hypothesis that meta-intents are domain-independent.

5.3 Participants

We recruited 407 participants of whom 394 finished the study and passed inner attention test questions. Each participant received a compensation of 0.84£. The number of participants was much larger than ten times the number of question items (240), which is considered well in line with the standard for performing a CFA. 191 of the 394 participants were female. Their age ranged from 18 to 79 ($M = 39.41$, $SD = 14.10$). The average duration of the survey was 5.49 minutes ($SD = 2.63$). The reason why the average duration was shorter than in the exploratory phase is that the number of questions was significantly reduced in the confirmatory stage.

5.4 Results

5.4.1 Validity and reliability test.

First, we performed prerequisite tests for CFA, the Kaiser-Meyer-Olkin (KMO) value was .715 (> 0.7) and Bartlett's test was significant (< .001), which both indicate that our data meets the requirements for performing CFA. Next, we used Principal Component Analysis (PCA) to extract factors, with Varimax rotation and Kaiser normalization, assigning *8* as the number of factors (according to our 8 meta-intents concepts). The results of the CFA are shown in Table 1. The cumulative variance of the 8 factors is 66.56%. The minimum question item's commonality value is 0.46, and the others are greater than 0.5, which is acceptable. The minimum factor loading value (0.594) is greater than 0.5, which shows that the items reflect the concept expressed by the factor well, also passing the discriminant validity test. We furthermore conducted a reliability test. For each generated factor, we calculated the Cronbach's alpha value. We found that the values of the first 7 factors were greater than 0.5 and passed the reliability test, but for **critiquing-orientation** alpha was only 0.43 and failed the reliability test, so we finally obtained an meta-intents instrument with 7 concepts, including 22 specific question items.

5.4.2 *Distribution checking.*

After the validity and reliability test, we checked the statistical distributions of the seven factors and plotted them as histograms in Figure 1, with the value of each meta-intents factor represented by the mean of its items. The purpose of this illustration is to detect whether the distributions of the factors are balanced. We did not find very skewed distributions in meta-intents factors. For comparison, we also draw a histogram of stable DSS factors (rationality and intuitiveness). We observed that, except for **interest in details** and **human-likeness** which are slightly skewed, the other five factors are basically subject to normal distribution. Although the distributions of these two factors are slightly skewed, they have not yet reached a dominant level.

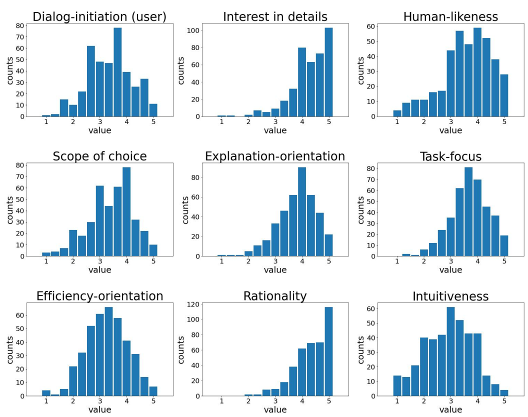

Figure 1: Distribution of scores for each factor

5.4.3 *Domain independence test.*

From the CFA we obtained values for the newly generated factors. This value is calculated by regressing the value of the question items following normalization (*M=0* and *SD=1*). These values were used in an ANOVA to test for differences between the three scenarios (smartphone, hotel and neutral scenario), using *domain* as a independent variable and 7 meta-intents factors (normalized regressing value) as dependent variables. The results are shown in the following Table 2. Because of multiple hypothesis testing, we performed the Benjamini-Hochberg (BH) procedure to control the false discovery rate (FDR). The results show that all meta-intents factors are not significantly different between the three domains. We want to emphasize that to strictly prove that the meta-intents are consistent in the three domains, we need to conduct 21 ($7 * 3$) pairs of equivalence tests (TOST procedure) which will be included in future work to carry out.

5.4.4 *Testing for gender influence.*

The gender question in the questionnaire contained four options, *male*, *female*, *others*, *don't want to say*. and in this part of the analysis on gender independence, we only use data from participants who chose male and female, which left 387 of 394 participants data. We used *gender* as grouping variable and 7 meta-intents factors (normalized regressing value) as test variables for independent-samples t-tests, and show the results in Table 3. Significance levels were again corrected by using Benjamini-Hochberg (BH) . The results show that in the *human-likeness* factor, the male group and the female group showed a significant difference, while no significant differences were observed in the other 6 meta-intents factors. Since the value of meta-intents factors is regressed and normalized, the mean (*M*) in the table can only reflect the difference between different groups of the same meta-intents factor, but cannot be used to compare different factors. An overview of the differences is provided in Figure 2. Note that the factor value in the Figure 2 is calculated from the mean value of the question items. In terms of *all* (male group + female group), we found that the *interest in details* factor has the highest mean, which is greater than 4 (scale is 1 to 5), and the mean of the other factors is relatively close, over around 3. In terms of *grouping* (male group vs. female group), significant differences in *human-likeness* between the two groups were also observed (whiskers have no overlap), but not for other factors.

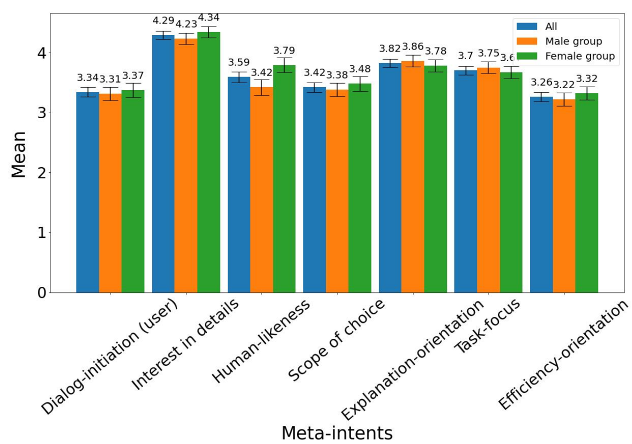

Figure 2: Mean scores obtained for the different meta-intents. Whiskers indicate the 95% confidence interval. The positive direction (larger value) of *dialog-initiated* represents *user-initiated*, and the negative direction (smaller value) represents *system-initiated*

5.4.5 *Influence of decision-making style.*

We assigned participants to two groups depending on their degree of intuitiveness or rationality in decision-making by applying k-means clustering. 174 participants could be assigned to a rationality group, while 220 participants were in an intuitiveness group as shown in Figure 3. To validate the clustering quality, besides visualization

Table 1: Final CFA results of total 24 question items (n=394). The first column represents meta-intents question items and *Cronbach's* α values of factors. The first row represents communities of question items and the generated latent factors. The factors in the experimental results are only digits, and we named them according to the problem items they have. The bold font indicates the values are greater than 0.5.

		Factors							
24 question items	*Commonalities*	dialog initiation	interest in details	human likeness	scope of choice	explanation orientation	task focus	efficiency orientation	critiquing orientation
dialog-initiation 3	0.74	**-0.85**	-0.01	-0.04	0.04	0.03	0.09	0.05	0.03
dialog-initiation 1	0.54	**0.79**	-0.05	0.16	0.14	0.02	0.14	0.13	0.08
dialog-initiation 2	0.72	**0.70**	0.02	0.28	0.10	0.02	0.14	0.08	0.11
dialog-initiation 4	0.62	**-0.66**	-0.05	0.11	0.17	0.01	0.04	0.07	0.23
interest in details 1	0.70	0.04	**0.83**	0.02	0.04	0.10	-0.05	-0.08	-0.04
interest in details 3	0.68	-0.04	**0.81**	0.03	0.13	0.12	0.12	0.01	0.00
interest in details 2	0.71	0.00	**0.78**	0.11	0.18	0.16	0.07	0.02	0.05
human-likeness 1	0.70	0.03	-0.02	**0.83**	0.08	0.09	-0.06	0.04	-0.01
human-likeness 2	0.70	0.07	0.00	**0.82**	0.06	0.01	-0.03	0.04	0.11
human-likeness 3	0.54	0.19	0.16	**0.67**	0.04	0.05	-0.01	0.09	0.11
scope of choice 2	0.62	-0.01	0.04	-0.11	**0.79**	0.20	-0.01	-0.07	-0.07
scope of choice 3	0.69	-0.01	0.18	0.16	**0.78**	0.06	-0.04	0.00	0.08
scope of choice 1	0.68	0.05	0.16	0.17	**0.72**	-0.07	-0.09	0.08	0.16
explanation-orientation 1	0.69	-0.01	0.10	0.08	0.01	**0.82**	0.03	0.02	-0.03
explanation-orientation 3	0.46	0.08	0.13	0.10	-0.01	**0.77**	0.02	-0.05	0.10
explanation-orientation 2	0.63	-0.07	0.10	-0.04	0.17	**0.63**	0.00	0.14	0.01
task-focus 1	0.72	-0.01	0.05	0.05	0.07	-0.02	**0.80**	0.02	-0.26
task-focus 3	0.46	0.05	-0.08	-0.13	-0.06	-0.01	**0.72**	0.15	0.09
task-focus 2	0.58	0.05	0.13	0.00	-0.14	0.08	**0.61**	0.02	0.20
efficiency-orientation 3	0.58	0.07	0.18	-0.04	-0.13	0.09	0.07	**0.78**	0.00
efficiency-orientation 2	0.54	0.05	-0.08	0.15	0.16	0.04	0.04	**0.69**	0.07
efficiency-orientation 1	0.67	-0.09	-0.43	0.08	-0.05	-0.01	0.15	**0.60**	0.06
critiquing-orientation 1	0.75	-0.03	0.03	0.03	0.00	0.07	-0.05	0.09	**0.86**
critiquing-orientation 2	0.55	-0.04	-0.04	0.34	0.24	0.00	0.17	0.00	**0.59**
Cronbach's α		0.79	0.76	0.72	0.73	0.63	0.56	0.52	0.43

we also calculate the Silhouette score (0.42) and Calinski-Harabasz score (366.30) which indicates the clustering into a rationality group and an intuitiveness group is acceptable. Here we want to explain that the rational cluster does not mean that the rational value of the individual cases in this cluster will be greater than the intuitive value, but the cluster center has a higher rational value and lower intuitive value than the other cluster. After observing the distribution of rational values and emotional values, we found that most people tend to depict themselves with a higher rational value (most are greater than 3), but the clustering results show that there are more intuitive people than rational people, which is in line with reality.

We used *rationality-intuitiveness cluster* as a grouping variable and 7 meta-intents factors (normalized regressing value) as test variables for the independent-samples t-test, and show the Benjamini-Hochberg-corrected results in Table 4. The results show that the rational group and the intuitive group have significant differences in four meta-intents factors, which are: *interest in details*, *human-likeness*, *explanation-orientation*, *efficiency-orientation.* This means that decision-making style has a significant effect on meta-intents. Here we also calculate the mean of the question items to represent the meta-intents factors and draw the Figure 4. It shows consistent results: significant differences in four factors across groups.

5.4.6 *Structural Equation Modeling.*

We noticed that multiple meta-intents factors showed significant differences between the decision style groups (rational vs. intuitive). In order to quantitatively analyze the impact of decision style on meta-intents, we conducted a structural equation modeling analysis and showed the results in Figure 5. In order to display the relationships between our main factors as clearly as possible, we left out the concrete question items and filtered out paths with insignificant impact and standardized regression coefficients less than 0.1. We found *rationality* has a significantly positive influence on *human-likeness* (0.27), *task focus* (0.16), and especially strong on *scope of choice* (0.37), *explanation orientation* (0.43), and *interest in details* (0.9). *Rationality* has a negative influence on *efficiency orientation* (-0.23). In contrast, *intuitiveness* have a strong positive influence on *efficiency oriented* (0.63) and *human-likeness* (0.35),

Table 2: Results of the one way ANOVA-test for meta-intents factors between smartphone, hotel and the neutral domains with Benjamini-Hochberg (BH) correction. Values marked with * are significant at a level of $p < .05$.

		Sum of Squares	*df*	Mean Square	*F*	*p*	BH adjusted *p*	Significant (FDR of 0.05)
Dialog-initiation	Between Groups	0.906	2	0.453	0.452	.637	.728	No
	Within Groups	392.094	391	1.003				
	Total	393	393					
Interest in details	Between Groups	5.658	2	2.829	2.856	.059	.236	No
	Within Groups	387.342	391	0.991				
	Total	393	393					
Human-likeness	Between Groups	0.577	2	0.289	0.288	.75	.75	No
	Within Groups	392.423	391	1.004				
	Total	393	393					
Scope of choice	Between Groups	4.473	2	2.237	2.251	.107	.214	No
	Within Groups	388.527	391	0.994				
	Total	393	393					
Explanation-orientation	Between Groups	9.31	2	4.655	4.744	**.009***	.072	No
	Within Groups	383.69	391	0.981				
	Total	393	393					
Task-focus	Between Groups	1.69	2	0.845	0.844	.431	.575	No
	Within Groups	391.31	391	1.001				
	Total	393	393					
Efficiency-orientation	Between Groups	5.517	2	2.759	2.784	.063	.168	No
	Within Groups	387.483	391	0.991				
	Total	393	393					

Table 3: Results from the independent samples *t*-test with Benjamini-Hochberg (BH) Procedure for meta-intents factors between the rational group and intuitive group. Values marked with * are significant at a level of $p < .05$. All 7 factors passed Levene's Test for equality of variances which is a prerequisite for the t-test.

	Male			Female			Levene's Test for Equality of Variances			t-test for Equality of Means				
	n	*M*	*SD*	*n*	*M*	*SD*		*F*	*p*	*t*	*df*	*p*	*d*	BH adjusted *p*
Dialog initiation	196.00	-0.01	0.97	191	0.01	1.04	Equal Var assumed	1.048	.307	0.192	385	.848	.020	.848
Interest in details	196.00	-0.09	1.02	191	0.09	0.99	Equal Var assumed	0.557	.456	1.758	385	.080	.179	.187
Human-likeness	196.00	-0.18	1.00	191	0.20	0.95	Equal Var assumed	0.788	.375	3.871	385	**<.001***	.394	**<.001***
Scope of choice	196.00	-0.02	0.93	191	0.05	1.07	Equal Var assumed	3.704	.055	0.650	385	.516	.066	.602
Explanation-orientation	196.00	0.09	0.95	191	-0.11	1.04	Equal Var assumed	0.470	.493	-2.014	385	.045*	-.205	.158
Task-focus	196.00	0.07	0.97	191	-0.06	1.03	Equal Var assumed	0.873	.351	-1.274	385	.203	-.130	.284
Efficiency-orientation	196.00	-0.06	1.01	191	0.08	0.98	Equal Var assumed	0.453	.501	1.459	385	.145	.148	.254

but no significant influence on *interest in details* and *explanation orientation*. This shows that the more rational persons are, compared to intuitive persons, the more they like details, comparison, scope of choice, and explanation when making purchase decisions. The more intuitive persons are, the more they care about efficiency and are unwilling to spend more time on finding and selecting the optimal item.

5.4.7 Age influence on meta-intents.

To verify the effect of age on meta-intents, we included this variable in the SEM, and the results are plotted in Figure 5 as well. We found that age has no significant influence on meta-intents factors except for *dialog-initiation*, but the standardized regression coefficient is only 0.17. We therefore claim that most meta-intents factors are independent of age. To verify the reliability of SEM, We performed a series of fit tests. The overall model fit is shown in Table 5 Since the proposed SEM model meets the evaluation criteria under multiple fitness indices (χ^2/df, GFI, AGFI, RMSEA) and is very close to the evaluations criteria in the other three fitness indices (TLI, NFI, CFI), it shows that our SEM is reliable.

6 DISCUSSION

6.1 Meta-intents instrument

With the aim of providing users with a better and eventually personalized interactive experience in CRS, we propose 12 meta-intents

Table 4: Results from the independent samples t-test for meta-intents factors between the rational group and intuitive group with Benjamini-Hochberg (BH) Procedure. Values marked with * are significant at a level of $p < .05$. A total of 5 factors passed Levene's Test for equality of variances which is a prerequisite for the t-test. The rest 2 factors reject the null alternative and accept the alternative hypothesis *equal variances not assumed*. We calculated for both conditions.

	Rationality			Intuitiveness			Levene's Test for Equality of Variances			t-test for Equality of Means				
	n	*M*	*SD*	*n*	*M*	*SD*		*F*	*p*	*t*	*df*	*p*	*d*	BH adjusted p
Dialog initiation	174	-0.01	1.06	220	0.01	0.95	Equal Var assumed	1.651	.200	-0.135	392.000	.892	-.014	.892
Interest in details	174	0.29	0.85	220	-0.23	1.05	Equal Var assumed	6.409	**.012**	5.240	392.000	.000		
							Equal Var not assumed			5.374	391.892	**.000**	.532	**.000**
Human-likeness	174	-0.14	1.09	220	0.11	0.91	Equal Var assumed	7.070	**.008**	-2.566	392.000	.011		
							Equal Var not assumed			-2.512	335.045	**.012**	-.260	**.021**
Scope of choice	174	-0.11	1.02	220	0.09	0.97	Equal Var assumed	0.737	.391	-1.904	392.000	.058	-.193	.081
Explanation-orientation	174	0.15	1.05	220	-0.12	0.95	Equal Var assumed	0.256	.613	2.701	392.000	**.007**	.274	**.016**
Task-focus	174	-0.10	1.03	220	0.08	0.97	Equal Var assumed	1.018	.314	-1.850	392.000	.065	-.188	.076
Efficiency-orientation	174	-0.31	0.98	220	0.24	0.95	Equal Var assumed	0.099	.753	-5.605	392.000	**.000**	-.569	**.000**

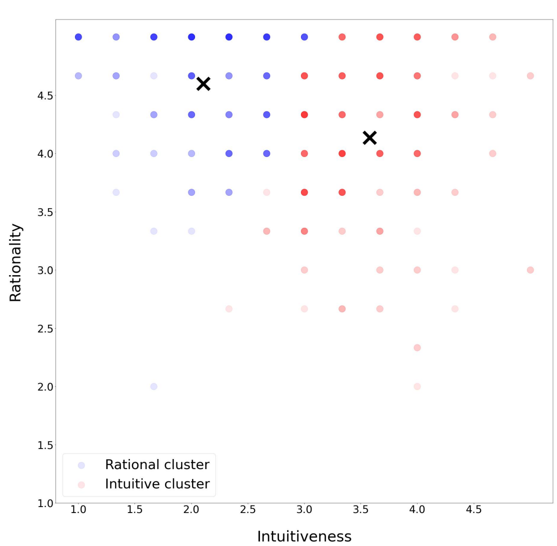

Figure 3: *K*-means clustering of rationality and intuitiveness factors. The black crosses are centroids of groups. Due to data overlapping, we present each user with a filled circle with 10% transparency, light colors mean sparse, and dark colors mean dense.

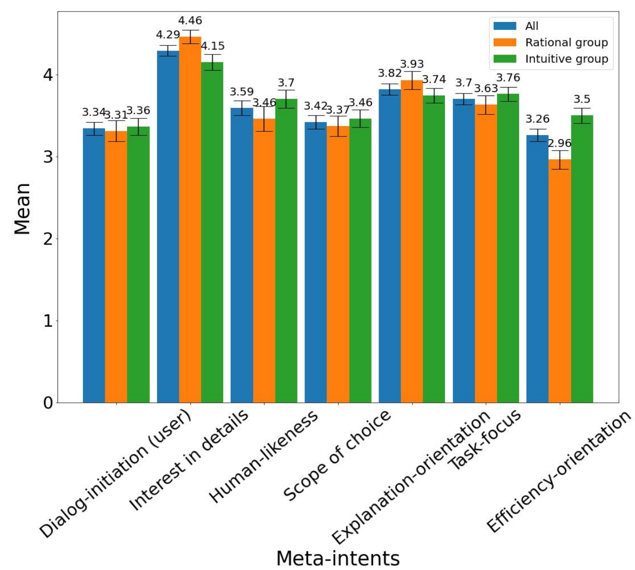

Figure 4: Mean scores obtained for the different meta-intents. Whiskers indicate the 95% confidence interval. The positive direction (larger value) of *dialog-initiated* represents *user-initiated*, and the negative direction (smaller value) represents *system-initiated.*

Table 5: The overall fitness indices of the proposed structural equation model.

	χ^2/df	GFI	AGFI	TLI	NFI	CFI	RMSEA
evaluation criteria	1< & <3	>0.8	>0.8	>0.9	>0.9	>0.9	<0.08
proposed SEM	2.019	.892	.868	.863	.790	.880	.051

concepts representing user interaction preferences, extending previous work [27, 29] that suggested the meta-intents concept. We suggest that the meta-intents concept can provide a theoretical basis for investigating user interaction with CRS in greater depth, and can, in particular, offer a framework for personalizing CRS dialogs. We followed a strict two-stage study approach including an exploratory and a confirmatory phase, resulting in a meta-intents instrument with 7 factors and 22 question items. Based on to the experimental results, we merged some concepts proposed in earlier

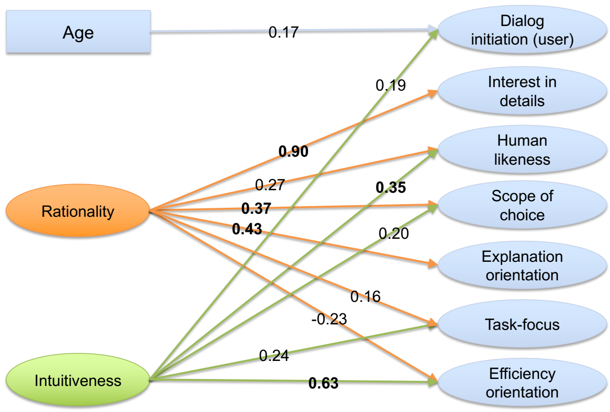

Figure 5: SEM of decision-making style, age and meta-intents. Rectangles represent specific questionnaire questions, ovals represent latent variables, arrows represent significant linear relationships, and values represent standardized regression coefficients. Standardized regression coefficients greater than 0.3 are highlighted in bold font.

work such as *comparison oriented* and *interest in details* because we could not find a sufficient discriminant validity for these factors. It should be noted that different meta-intents refer to different components and steps of a dialog. For example, *scope of choice* and *explanation orientation* refer to the presentation of recommendations while *interest in details* may refer both to the elicitation of preferences as well as the presentation of items. Therefore, these concepts may have different implications for dialog design and personalization.

Two initially proposed factors, *critiquing-orientation*, *diversity-orientation*, concepts that have received a lot of attention in the recommender systems community, had to be filtered out based on the statistical tests. One reason might be that users are unfamiliar with critiquing techniques and may also not fully understand the diversity concept. We do not suggest, however, that such design aspects are irrelevant, future work may find support for re-introducing them in the framework.

One of our major assumptions was the meta-intents represent interaction-oriented user preferences that are independent of the specific item domain. In fact, we did not find significant differences between the domains addressed in our scenarios. There was only a tendency in the factor explanation-orientation which showed a higher value for smartphones than for hotels. This might indicate that the nature of the product to be recommended, possibly its technical complexity or the financial risk involved, can alter the user's preferences in how the dialog should be designed. Since we have only investigated meta-intents in two specific product domains and in an open scenario, the findings still have to be taken with some caution.

A further limitation can be seen in the fact that participants in our studies only imagined using a CRS which was induced by the scenarios presented. We thus can not exclude the possibility that the preferences represented by the meta-intents may change when using a real system. This might influence the weights users put on the different factors and may also introduce additional factors. Nonetheless, our studies so far provide strong evidence for the relevance and stability of the proposed meta-intents.

The analysis of gender and age did not reveal a larger impact of these variables. It seems that females have some preference for anthropomorphic, more human-like dialogs in contrast to males who prefer a more neutral conversation. However, the difference is small and no other effects of gender were found. Also, age only had a small (regression coefficient of 0.17) but significant influence on dialog-initiation, that is, older persons seem to prefer dialogs that are initiated by the user.

A finding we see as very relevant is the evidence for a strong relation between the general decision-making style of a person and the different meta-intents. Our results consistently reflect that the rational group and the intuitive group have opposite preferences on the factors *interest in details*, *human-likeness*, *explanation-orientation* and *efficiency-orientation*. First, this indicates that meta-intents can actually serve as a bridge between general psychological traits and the concrete design of a CRS. Second, if a user profile would contain information about the user's decision style, there would be a number of concrete, rather evident options for personalizing the dialog.

Finally, one large open question is how to obtain information about a user's meta-intents in real-world settings. One way to achieve this would be to let users answer the questionnaire we developed which would lead to quite reliable results. However, this will not always be a practical option, especially in production e-commerce settings. As an alternative, we have some hope that meta-intents may be detected in the interaction behavior and patterns of users in a live CRS. We intend to investigate this question in future work.

7 CONCLUSION

In this paper, we have proposed the concept of meta-intents and a set of factors that represent different meta-intents in the context of CRS, and have validated these factors in a two-stage empirical approach. We define meta-intents as high-level user preferences for the conversation and information presentation in CRS that are more abstract than concrete user intents which refer to item and feature-related preferences. On the other hand, Meta-intents are more concrete than general personal characteristics such as decision-making style, need for cognition, or Big-5 personality traits. Meta-intents can thus be seen as bridging the gap between general user characteristics and the concrete design of CRS. Concerning the level of abstraction expressed in meta-intents, we expected them to be independent of the concrete item domain in which recommendations are provided. Our studies provide evidence that meta-intents are independent of the domain. The questionnaire we developed provides a stable and validated instrument for these future studies.

An outcome of our studies we see as particularly relevant both from a theoretical and a practical perspective is the finding that meta-intents are significantly influenced by the general decision-making style of a user. In fact, users who are either more rational or more intuitive in their decision-making behavior express opposite weights on a number of meta-intents such as interest in detail and the preference for efficiency in a dialog.

We see the findings of our studies as a promising point of departure for further investigating interaction-related user preferences in conversational recommenders, and there are still a number of open questions. Since our current studies were based on scenarios, not on using an actual CRS, user preferences need to be investigated in more realistic settings. Furthermore, there is the largely open question whether it will be possible to detect meta-intents in the interaction behavior of users in real CRS which would provide relevant input for personalizing the conversation on the fly.

ACKNOWLEDGMENTS

This work was partly funded by the German Federal Ministry of Education and Research in the project ACODA (funding no. 01QE1946C).

REFERENCES

[1] Müslüm Atas, Alexander Felfernig, Seda Polat-Erdeniz, Andrei Popescu, Thi Ngoc Trang Tran, and Mathias Uta. 2021. Towards psychology-aware preference construction in recommender systems: Overview and research issues. *Journal of Intelligent Information Systems* (2021), 1–23.

[2] Dirk Bollen, Bart P Knijnenburg, Martijn C Willemsen, and Mark Graus. 2010. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems*. 63–70.

[3] Alexander Bondarenko, Pavel Braslavski, Michael Völske, Rami Aly, Maik Fröbe, Alexander Panchenko, Chris Biemann, Benno Stein, and Matthias Hagen. 2020. Comparative web search questions. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 52–60.

[4] Oliver B Büttner, Arnd Florack, and Anja S Göritz. 2013. Shopping orientation and mindsets: How motivation influences consumer information processing during shopping. *Psychology & Marketing* 30, 9 (2013), 779–793.

[5] Oliver B Büttner, Arnd Florack, and Anja S Göritz. 2014. Shopping orientation as a stable consumer disposition and its influence on consumers' evaluations of retailer communication. *European Journal of Marketing* (2014).

[6] Oliver B Büttner, Arnd Florack, and Anja S Göritz. 2015. How shopping orientation influences the effectiveness of monetary and nonmonetary promotions. *European Journal of Marketing* (2015).

[7] Wanling Cai and Li Chen. 2020. Predicting User Intents and Satisfaction with Dialogue-Based Conversational Recommendations. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 33–42. https://doi.org/10.1145/3340631.3394856

[8] Li Chen and Pearl Pu. 2012. Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction* 22, 1 (2012), 125–150.

[9] Sarah Dean, Sarah Rich, and Benjamin Recht. 2020. Recommendations and user agency: the reachability of collaboratively-filtered information. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*. Association for Computing Machinery, New York, NY, USA, 436–445. https://doi.org/10.1145/3351095.3372866

[10] Zuohui Fu, Yikun Xian, Yongfeng Zhang, and Yi Zhang. 2020. Tutorial on Conversational Recommendation Systems. In *Fourteenth ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 751–753. https://doi.org/10.1145/3383313.3411548

[11] Zuohui Fu, Yikun Xian, Yongfeng Zhang, and Yi Zhang. 2021. WSDM 2021 Tutorial on Conversational Recommendation Systems *(WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 1134–1136. https://doi.org/10.1145/3437963.3441661

[12] Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2021. Advances and Challenges in Conversational Recommender Systems: A survey. *AI Open* 2 (2021), 100–126. https://doi.org/10.1016/j.aiopen.2021.06.002

[13] Moojan Ghafurian, Neil Budnarain, and Jesse Hoey. 2019. Role of emotions in perception of humanness of virtual agents. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 1979–1981.

[14] Tilahun Z Gizaw, H Dong Jun, and Ammar Oad. 2017. Solving cold-start problem by combining personality traits and demographic attributes in a user based recommender system. *International Journal of Advanced Research in Computer Science and Software Engineering* 7, 5 (2017).

[15] Katherine Hamilton, Shin-I Shih, and Susan Mohammed. 2016. The Development and Validation of the Rational and Intuitive Decision Styles Scale. *Journal of Personality Assessment* 98, 5 (2016), 523–535. https://doi.org/10.1080/00223891.2015.1132426

[16] Anthony Jameson, Martijn C Willemsen, Alexander Felfernig, Marco de Gemmis, Pasquale Lops, Giovanni Semeraro, and Li Chen. 2015. Human Decision Making and Recommender Systems. In *Recommender Systems Handbook*. Springer, Boston, MA, 611–648. https://doi.org/10.1007/978-1-4899-7637-6_18

[17] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A Survey on Conversational Recommender Systems. *ACM Comput. Surv.* 54, 5, Article 105 (may 2021), 36 pages. https://doi.org/10.1145/3453154

[18] Marius Kaminskas and Derek Bridge. 2016. Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 1 (2016), 1–42.

[19] Timm Kleemann, Benedikt Loepp, and Jürgen Ziegler. 2022. Towards Multi-Method Support for Product Search and Recommending. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 74–79. https://doi.org/10.1145/3511047.3536408

[20] Timm Kleemann, Magdalena Wagner, Benedikt Loepp, and Jürgen Ziegler. 2021. Modeling User Interaction at the Convergence of Filtering Mechanisms, Recommender Algorithms and Advisory Components. In *Mensch Und Computer 2021* (Ingolstadt, Germany) *(MuC '21)*. Association for Computing Machinery, New York, NY, USA, 499–511. https://doi.org/10.1145/3473856.3473859

[21] Ivica Kostric, Krisztian Balog, and Filip Radlinski. 2021. Soliciting User Preferences in Conversational Recommender Systems via Usage-related Questions. In *15th ACM Conference on Recommender Systems (RecSys '21)*. 724–729.

[22] Philipp Kulms and Stefan Kopp. 2019. More human-likeness, more trust? The effect of anthropomorphism on self-reported and behavioral trust in continued and interdependent human-agent cooperation. In *Proceedings of mensch und computer 2019*. 31–42.

[23] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems–A survey. *Knowledge-based systems* 123 (2017), 154–162.

[24] Elisabeth Lex, Dominik Kowald, Paul Seitlinger, Thi Ngoc Trang Tran, Alexander Felfernig, and Markus Schedl. 2021. Psychology-informed Recommender Systems. *Foundations and Trends® in Information Retrieval* 15, 2 (2021), 134–242. https://doi.org/10.1561/1500000090

[25] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the Cold Start Problem in Recommender Systems. *Expert Syst. Appl.* 41, 4 (March 2014), 2065–2073. https://doi.org/10.1016/j.eswa.2013.09.005

[26] Ewa Luger and Abigail Sellen. 2016. " Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 5286–5297.

[27] Yuan Ma, Tim Donkers, Timm Kleemann, and Jürgen Ziegler. 2022. Meta-Intents in Conversational Recommender Systems. In *Proceedings of the 4th Edition of Knowledge-aware and Conversational Recommender Systems Workshop co-located with 16th ACM Conference on Recommender Systems (RecSys 2022)*. 81–90.

[28] Yuan Ma, Timm Kleemann, and Jürgen Ziegler. 2021. Mixed-Modality Interaction in Conversational Recommender Systems. In *Interfaces and Human Decision Making for Recommender Systems 2021: Proceedings of the 8th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems*. 21–37.

[29] Yuan Ma, Timm Kleemann, and Jürgen Ziegler. 2022. Psychological User Characteristics and Meta-Intents in a Conversational Product Advisor. In *Proceedings of the 9th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems co-located with 16th ACM Conference on Recommender Systems (RecSys 2022)*. 18–32.

[30] Martijn Millecamp, Nyi Nyi Htun, Yucheng Jin, and Katrien Verbert. 2018. Controlling Spotify Recommendations: Effects of Personal Characteristics on Music Recommender User Interfaces. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (Singapore, Singapore) *(UMAP '18)*. Association for Computing Machinery, New York, NY, USA, 101–109. https://doi.org/10.1145/3209219.3209223

[31] Andrea Papenmeier, Alexander Frummet, and Dagmar Kern. 2022. "Mhm..." – Conversational Strategies For Product Search Assistants. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) *(CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 36–46. https://doi.org/10.1145/3498366.3505809

[32] Eyal Peer, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research. *Journal of Experimental Social Psychology* 70 (2017), 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

[33] Pearl Pu and Boi Faltings. 2004. Decision tradeoff using example-critiquing and constraint programming. *Constraints* 9, 4 (2004), 289–310.

[34] Max Schemmer, Patrick Hemmer, Maximilian Nitsche, Niklas Kühl, and Michael Vössing. 2022. A Meta-Analysis on the Utility of Explainable Artificial Intelligence in Human-AI Decision-Making. *arXiv preprint arXiv:2205.05126* (2022).

[35] Marita Skjuve, Ida Maria Haugstveit, Asbjørn Følstad, and Petter Brandtzaeg. 2019. Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human–chatbot interaction. *Human Technology* 15, 1 (2019), 30.

[36] Yueming Sun and Yi Zhang. 2018. Conversational Recommender System. In *The 41st International ACM SIGIR Conference on Research & Development in Information*

Retrieval (Ann Arbor, MI, USA) *(SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 235–244. https://doi.org/10.1145/3209978.3210002

[37] Marko Tkalcic and Li Chen. 2015. *Personality and Recommender Systems*. Springer, Boston, MA, 715–739. https://doi.org/10.1007/978-1-4899-7637-6_21

[38] Bo Xiao and Izak Benbasat. 2007. E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. *MIS Q.* 31, 1 (mar 2007), 137–209.

[39] Zhihui Xie, Tong Yu, Canzhe Zhao, and Shuai Li. 2021. Comparison-based Conversational Recommender System with Relative Bandit Feedback. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1400–1409.

[40] Markus Zanker, Markus Jessenitschnig, and Wolfgang Schmid. 2010. Preference reasoning with soft constraints in constraint-based recommender systems. *Constraints* 15, 4 (2010), 574–595.

[41] Jingjing Zhang. 2011. Anchoring Effects of Recommender Systems. In *Proceedings of the 5th ACM Conference on Recommender Systems* (Chicago, Illinois, USA) *(RecSys '11)*. Association for Computing Machinery, New York, NY, USA, 375–378. https://doi.org/10.1145/2043932.2044010

[42] Xiaoying Zhang, Hong Xie, Hang Li, and John CS Lui. 2020. Conversational contextual bandit: Algorithm and application. In *Proceedings of the web conference 2020*. 662–672.

[43] Yongfeng Zhang and Xu Chen. 2020. Explainable Recommendation: A Survey and New Perspectives. *Foundations and Trends in Information Retrieval* 14, 1 (2020), 1–101.

[44] Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. 2018. Towards Conversational Search and Recommendation: System Ask, User Respond. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) *(CIKM '18)*. Association for Computing Machinery, New York, NY, USA, 177–186. https://doi.org/10.1145/3269206.3271776