Unifying Recommender Systems and Conversational User Interfaces

Alain D. Starke
Marketing and Consumer Behaviour Group, Wageningen University & Research
Wageningen, Netherlands
MediaFutures, University of Bergen
Bergen, Vestland, Norway
alain.starke@wur.nl

Minha Lee
Future Everyday Group
Eindhoven University of Technology
Eindhoven, Netherlands
m.lee@tue.nl

ABSTRACT

This paper considers unifying research on conversational user interfaces and recommender systems. Studies on conversational user interfaces (CUIs) typically examine how conversations can be facilitated (i.e., optimizing the *means*). Recommender systems research (RecSys) aims to retrieve and present recommendations in a user's session (i.e., optimizing the *ends*). Though these aims are overlapping across both areas, they can be better examined together to target the means and ends of what people can achieve with technology as *conversational recommender systems* (CRSs). We discuss the intersection of conversational user interfaces, recommender systems, and conversational recommender systems. We argue how conversations and recommendations can be designed holistically, in which recommendations can also be a means to foster engaging conversational interaction, while conversations as ends can better sustain curated, long-term recommendations.

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models**; *Interaction design theory, concepts and paradigms.*

KEYWORDS

conversations, recommender systems, conversational recommender systems, personalization

ACM Reference Format:

Alain D. Starke and Minha Lee. 2022. Unifying Recommender Systems and Conversational User Interfaces. In *4th Conference on Conversational User Interfaces (CUI 2022), July 26–28, 2022, Glasgow, United Kingdom.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3543829.3544524

1 INTRODUCTION

The growth of conversational user interfaces (CUIs) has led to efforts to deliver a better conversational user experience [55]. From a researcher's perspective, the conversations are treated as the *means* to achieve or explore different goals. Examples include keeping conversations moving [16], maintaining conversations with humans as long as possible (as in Alexa Prize competitions[1]), finding a common ground with Google Home when planning a trip [8], or playing interactive games with Alexa as a family [64]. An upcoming, related field is that of conversational recommender systems (CRSs) that build upon recommender systems (RecSys) research. RecSys presents personalized content, such as songs to listen to or movies to watch, (e.g., by services like Spotify and Netflix; cf. [19, 24]), often via a graphical user interface. *Conversational recommender systems* aim to achieve the end goal of presenting personalized content through conversations.

For users of today's smart speakers or assistants, *conversations* are often designed as a *means* to an end, to explore or accomplish a shared goal. In contrast, users of Netflix or Spotify *recommendations* are seen as an *end* result of the system's personalization efforts. For CUI researchers, how the *means* is carried out (e.g., conversational flow or repair) is a large concern, whereas RecSys researchers are pre-occupied with whether the *end outcome* (i.e., the recommendation) is fitting for their users [31]. Surely, CUI and RecSys researchers also do examine both means and ends of how people are affected by the end result (e.g., how well a conversational agent answered a query), as well as how systems get there (e.g., recommendations as ongoing attempts at personalization). Still, how both communities can benefit from each other's perspectives is a lacking effort.

This paper outlines the differences and commonalities across the CUI and RecSys communities. We discuss that they can find a common ground by not just considering conversations as a means and recommendations as an end. We point out how contributions in the shared subfield field of conversational recommender systems (CRSs) can add to the research performed on CUIs, while highlighting which aspects of CUI research could also add to recommender systems research.[2]

2 BACKGROUND

2.1 CUI Research

CUIs have been popularized in many products and applications, such as the Google Assistant[3] or Replika[4]. Thus, CUI is a broad

CUI 2022, July 26–28, 2022, Glasgow, United Kingdom

ACM ISBN 978-1-4503-9739-1/22/07…$15.00
https://doi.org/10.1145/3543829.3544524

[1]https://www.amazon.science/alexa-prize
[2]To be clear, we are concerned with comparing CRSs and CUIs for goal-directed dialogue, i.e., advice solicitation in a sense, so we do not consider more complex, open-ended dialogues here.
[3]https://assistant.google.com/
[4]https://replika.ai/

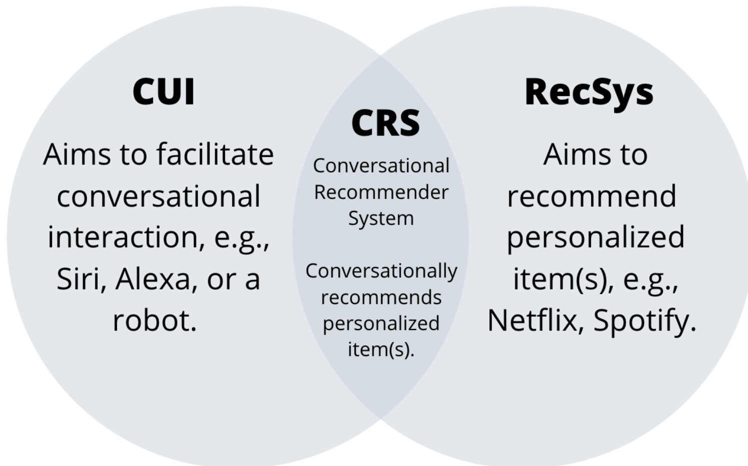

Figure 1: Diagram of how conversational recommender systems (CRSs) are positioned at the intersection of conversational user interfaces (CUIs) and recommender systems (RecSys).

term that can encompass easily accessible chatbots that are on communication platforms like Slack or Messenger, be it for work-related conversations or hobbies [11], or embodied as robots with multi-modal capabilities [63]. Increasingly, CUIs are collaborative partners in everyday rituals (e.g., cooking), in which the combination of modalities they use (e.g., speech and gaze), influence human conversational patterns. For example, CUIs can be equipped with eyes that are expected to gaze as a communication signal [33]. Most of today's CUIs are designed to react to our commands, in a task-oriented manner (cf. [44]).

There has been, however, increasing attention for research on CUIs that can have consequences on interpersonal relationships or well-being [36, 62, 69]. CUIs can be imbued with social traits by human interactants, intentionally or not, though people differentiate between functional vs. social CUIs [10]. For example, ELIZA was perceived to be a social chatbot, by acting as a therapist with only simple pattern-matching to responders' texts, which made people attribute intelligence to it [81]. Other examples of 'social' CUIs include customer service chatbots that should efficiently handle customer requests but are seen as human-like by some users [17], as well as a chatbot that was intended only to be functional as a news aggregator, but was expected to act in a social way by users nonetheless [38]. A common problem is the effective design of conversations, regardless of its purpose, be it functional, social, or both [55]. Hence, CUI research often deals with conversational *design* as the means to achieve or explore people's goals in various contexts, such as finding a song and playing it on a smart speaker. In other words, the goal lies in enhancing the conversational means to optimize any desired end.

2.2 Recommender System Research

There is also research on how conversations can be a means to an end of retrieving relevant content. Recommender systems (RecSys) are algorithms and interfaces that present content to users that match their preferences [31, 66]. This can be based on historical user interactions and ratings, or an active preference elicitation phase, in which a system inquires about what a user currently likes [65].

Three different RecSys approaches are dominant in terms of presenting personalized content. The most commonly used method is collaborative filtering (CF; [66]), which uses interaction data to determine which users and items are similar, i.e., which users are one's 'nearest neighbors'. A CF-based recommender would, for example, retrieve content that is popular among peers who have been identified to have preferences similar to the user. An important challenge is to mitigate the so-called cold-start problem, the phase in which the recommender cannot yet model the user's preferences due to a lack of historical data [14, 45]. Although this can be overcome by prompting users to rate or interact with a specific set of items to efficiently a map user's preferences [14], it is also possible to employ other approaches.

Two other approaches are content-based and knowledge-based recommendation [18]. Content-based methods typically determine similarity between items, based on different features: if a user likes an item (e.g., a recipe with specific ingredients), that user is recommended 'more of the same' (e.g., recipes with similar ingredients; cf. [71]). Knowledge-based recommenders are a third dominant approach, albeit at times regarded as a subset of content-based recommender, for they match items and users based on pre-defined rules between item features and user characteristics. For example, if a user of a recipe recommender system indicates to suffer from diabetes, that system may filter out recipes with high levels of sugar and fat [58].

Recommender systems are applied to domains where information and choice overload are likely [7, 22], such as e-commerce. They are also useful in domains where users like to explore different options for which they do not have clear item or outcome in mind [15, 35]. For example, the website AllRecipes.com employs food recommender system to retrieve and present users any of the 60,000 recipes from its website [77]. The main purpose of RecSys is, thus, to present a limited number of options that are most likely to be relevant to a user, either because of that user's past browsing history or because the item set is on average popular among all users [13, 31].

The overarching aim of RecSys design is to keep a user engaged to a service by providing relevant content [27, 31]. In other words, most research focuses on optimizing the *ends*. Although research also examines the 'means' of RecSys, i.e., *how* users interact with a recommender in terms of preference elicitation and effort [28, 72], most studies assess the quality of a recommender system through its predictive accuracy and how satisfied people are with the presented and chosen content [29, 32, 52]. This is also reflected in the employed methods, which also involve 'offline evaluation': assessing the predictive accuracy of different recommender models by splitting the dataset of interest into training and test sets [46]. In the recommender domain, studies that present improvements in algorithmic accuracy are also regarded as a significant research contribution.

Most recommender systems are designed as graphical user interfaces [31, 66]. They often combine text and images (cf. [80]), with which can be interacted through clicks or in a query-based fashion. These modalities are also related to the type of items recommended [23, 67]. For instance, recommenders that present a single set of options may be appropriate if the user already has strong preferences, while more interactions are required if a user wishes to explore (further). This volatility of user preferences plays a role in RecSys research [6], in the sense that they are formed *during* interaction [25, 79], as well as due to the recommender's choice architecture [25, 70, 73], i.e., how the options are presented.

2.3 Conversational Recommender Systems

Conversational recommender systems (CRSs) typically elicit user preferences through a conversation over multiple interactions. CRSs form a specific subgroup of RecSys and are touted to be appropriate to overcome cold-start problems [23], for they can allow users to refine an initial set of recommendations by critiquing specific attributes or characteristics of the search result(s) [24]. For example, a user may be refine an initial suggestion for a restaurant to visit by asking the recommender to change specific criteria (e.g., proximity, cuisine). Most CRSs are designed as one-off conversations, regardless of whether they aim to provide task-focused recommendations or open-domain recommendations. An assumption underlying some CRSs is that human interactants know what the conversation's goal is, whereas some may still be searching for a goal in various CUI contexts [49].

In contrast with general RecSys research, *collaborative filtering* is not dominant in CRSs [23]. As many CRSs do not maintain user histories, many recommenders use a *content-based*, "more like this" approach instead, retrieving items that are similar to the one a user is currently interacting with, but would like to refine. Consequently, most CRSs are interested in the 'short-term preferences' of users rather than the long-term ones [23], often constrained to the dialogue a user is in.

A key difference between CUI and RecSys are the assumptions about their users having exploratory intent [20]. Whereas one may engage in exploratory search with a conversational agent [50], a user turning to a recommender system often has a need for a recommendation but is typically unsure which one would suit her the best. Exploratory searches in CUIs are more likely to lead to breakdowns or 'hiccups' in a conversation [23, 59]; a CRS might propose a specific item or a set of items and then pause the interaction, even though a user might expect it to continue. Surveys on the state-of-the-art for CUIs and CRSs highlight that different themes are dominant. For CRSs, much work has gone into the type of modalities, whether this should include voice or a combination of voice and text, and how this leads to the most optimal system retrieval in terms of predictive accuracy [23]. Much less attention has been devoted to theories of conversation and speech and whether CRSs accurately reflect human conversations, such as in an inquiry to recommend a specific restaurant [9].

3 CONTRASTING THE TWO FIELDS OF RESEARCH

All conversational systems use some type of content retrieval to present a response that is most appropriate for a specific context [56, 76]. The fundamental functionalities of conversational agents across CUIs and CRS can be similar, whether this is based on a predictive model or a rule-based model, and whether this is based on natural language processing or fixed responses. However, the type of studies and their underlying goals vary extensively. CUIs tend to focus on studying evaluative outcomes with regard to the conversational *form*, while less attention is placed on what is presented, i.e., the advice, the items, etc. Lee et al. [38] illustrates the latter. In their study, they examine user trust in cryptocurrency through an RSS-feed based chatbot, which presents content that is not necessarily the most relevant for the user, but simply a snapshot of what fits within the user's category. The main construct in the study, trust, is analyzed without a connection to the presented news, and whether personalization would have affected this.

CUI studies also tend to vary the role that an agent has in a conversation and to examine how it is perceived by the user [30, 36]. For one, conversations can be designed more like a script that both the agent and user need to act out, with very few degrees of freedom. This may be due to specified roles, e.g., ELIZA's perceived role as a psycho-therapist [81], though recent examples like Replika are more advanced. For example, in a mental health context, how much self-disclosure is evoked, whether an agent cares for a user, or if it asks to be taken care of, affects the flow of conversation and how the agent is perceived by the user [36, 43]. Although conversational retrieval occurs to generate appropriate responses, as it would in a RecSys, there is less focus on utilitarian evaluation outcomes, such as whether specific conversational response items are liked by a user. In other words, CUI research often focuses on the value of the conversation itself and how to best carry it out in specific cases, rather than the system's predictive accuracy.

In contrast, CRSs are typically optimized towards people's discovery of new items, such as a movie to watch or recipe to try [3, 21]. CRSs often forgo how a conversation can flow and focus on *which* item or content should be presented [23, 51, 75]. Instead, they often focus on reducing the costs of interaction or information search [60], while little attention is devoted to conversational UX. Yet, recommender systems could use the conversation itself as a way to determine a user profile, (i.e., the user's role) to which the content should be personalized [23]. As of now, the recommender interface itself is rarely personalized towards the user or the conversation.

3.1 Evaluation Methods

How CUIs and CRSs are evaluated indicates where these two fields differ. Research questions in CUI papers often address an aspect of the conversation, and how this matches the user. For example, some studies examine how the interaction can become more efficient without sacrificing conversational aspects (e.g., the human-likeness of an agent [23, 30]). To this end, users are often asked to evaluate the conservation after a single interaction. Moreover, evaluation aspects are also more likely to involve attitudinal aspects, such as one's attitude towards a proposed mental health exercise or negative emotions in a conversation [30, 61, 68]. In contrast, CRSs, and RecSys in general rely on interaction data, such as ratings, likes, and choices [23]. Although inquiring on user perceptions is becoming more common in CRS research [60], this often does not involve conversational aspects, such as the perceived effort of using an CRS. What is more common are evaluations regarding the outcome, such as the perceived quality, choice satisfaction, and other forms of liking with regard to a chosen item [23]. In doing so, the conversation is treated as a potential 'hurdle' that should be overcome as efficiently as possible.

Based on the differences between CUIs and CRS, we also observe a need for expanded evaluation techniques. First, most CRSs are evaluated online by novel users, while long-term evaluations by experienced users should also be considered in research. Second, for CUI studies, a more refined approach can be applied to gauge the relevance of recommendations. To determine which algorithm yields the highest accuracy, CUIs could first evaluate their models 'offline', as is done in RecSys [66]. Some studies employ a dataset with user-item interaction data, e.g., ratings, to train multiple recommender models, by dividing the dataset into training and validation sets and using five-fold cross-validation. The models are typically evaluated on their predictive accuracy through an error metric, such as the root mean square error [23, 46, 75].

3.2 The role of crowdsourcing or collaborative filtering

Designing conversations and generating conversational content is a theme across both CUI and CRS studies, but the approaches differ. Recently, in CUI, it has been studied how the content of an agent's conversation can be derived from crowdsourced advice [1] or based on dialogue graph creation with crowd input [26]. For one, therapeutic advice [2] is an example area that can benefit from crowd-driven conversation design. While still emerging in CUI, crowdsourcing methods are widespread in RecSys. Most notably, collaborative filtering is used to identify users with similar interests [31, 66], based on their historical interaction data. This principle is loosely based on the homophily principle (i.e., similar people are more likely to be connected [53]), relating to how we are more partial to receiving advice from within our homogeneous social network than from people outside it. People with similar interests in a network can train algorithms for personalized recommendation (e.g., Spotify), which can be used for explaining how the recommendations are relevant [54]. More importantly, however, is that this is used to predict whether the recommendations are relevant. In studies with transparent, explanatory interfaces, RecSys literature has found social peer-based (vs. non-peer) and solicited (vs. unsolicited) recommendations to be more effective in motivating people (e.g., taking energy-saving measures [70, 74]). For example, to reinforce sustainable behavior by emphasizing the behavior of others, such as "75% of similar users do this" [73].

CUIs and CRSs have yet to fully grasp the role of crowdsourcing. This could be paired with advice solicitation, which caters to the user's perceived autonomy. For example, system-initiated recommendation is perceived as riskier than user-initiated recommendation [34]. In addition, users also feel less in control, particularly if the system indicates that the user's browsing history is used as the basis for its recommendation. What has not been looked into is how CUIs become actors in our networks that can also give solicited advice. They can have social roles of that are normally reserved for human-peers for norm establishment [37]. We thus believe that conversational recommender systems (CRS) [75] can be a research area on how CUIs can conversationally recommend to its human interactants on communication platforms (as chatbots) or as social actors (as robots or smart speakers). The latter would particularly be new to RecSys scholars.

3.3 Designing an identity

As recently noted by a growing body of work on identity [4, 5, 39], the design of a CUI (an agent's identity) is as important as user-oriented personalization (a user's identity). CUI research has focused on optimizing the agents or the conversational environment, but another touch point is whether the presented content affects how the conversation or the agent is perceived or evaluated by users. CUIs that talk to people are seen as entities in their own right, attributed with a mind of their own [40, 42]. Thus, how they recommend content can also affect how they are perceived. Even a simple chatbot can be seen as a separate identity, when it refers to itself in first person (i.e., "I"). Siri, for instance, states "go ahead, I'm listening", while waiting for a person to speak. RecSys, in contrast, are designed as tools without the first-person language and its users do not attribute a mind to an application, which foremost augments users' identity as an environment, not as an identity of its own. Nonetheless, while CRSs may be attributed with anthropomorphic traits like having a mind [34], people are more likely to anthropomorphize systems if their underlying algorithmic processes (e.g., in online behavioral targeting) are not transparently disclosed [48]. Given the focus on explainability in many RecSys studies (e.g., [57]), this may be at odds with CUIs that aim for social presence; future research can focus on how explainability and social presence can connect.

4 DISCUSSION AND CONCLUSION

We have described the key commonalities and differences between Conversation User Interface (CUI) and Recommender System (RecSys) research. The CUI community can consider opportunities to extend their research by incorporating principles from RecSys research, in particular conversational recommender systems (CRSs).

What stands out is a difference in how conversations are approached by the two fields. CUI research seems to focus on diverse contexts of use and facilitation of a conversation (cf. [47]), while CRSs seek to help people to find what they want more specifically. In other words, whereas CUIs seem to optimize the *means* of a conversation, CRSs aim to optimize the *end* of a conversation, following how RecSys optimizes for personalization accuracy. Some CUI papers have focused on a user's attitude towards agents and their evaluation of a conversation (e.g., attitude towards emotional interaction [30], engagement [30, 56], perceived personality or partner model of agents [12, 41, 78]), which partially overlaps with how CRSs are evaluated [23]. CRS papers mainly assess the algorithmic accuracy, by predicting and ranking conversational responses [76], while possibly also evaluating the quality of the presented or chosen recommendations [23]. There is less attention, however, for the perceived cognitive effort of engaging with such a system [51].

We can benefit from exploring how CUIs can adopt RecSys practices, as well as being attentive to how RecSys research can be more attuned to conversationally sensitive interactions. Hence, supporting dialogue facilitation and presenting the most relevant recommendations can go hand-in-hand. In fact, many people may be willing to 'forgive' mediocre conversation design if the outcome (e.g., the presented item) is relevant, and vice versa. CUI scholars could implement their expertise in building CRSs by applying conversation design principles. This aspect is rarely considered by recommender scholars [23], since they mainly optimize for behavioral outcomes (i.e., which item is chosen) and more rarely how the conversation is evaluated (cf. [51, 60]). As for the scope of research, conversationally handling complex or long-term topics, e.g., mental and self-care, is more difficult than that of simple or short-term goals, e.g., recipe recommendation (cf. [3, 36]). For a shared future research direction, we wholeheartedly 'recommend' scholars to optimize both the *means* and the *end* of a conversation.

REFERENCES

[1] Tahir Abbas, Vassilis-Javed Khan, Ujwal Gadiraju, Emilia Barakova, and Panos Markopoulos. 2020. Crowd of oz: A crowd-powered teleoperation system for enhanced human-robot conversations. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, New York, NY, USA, 81–83.

[2] Tahir Abbas, Vassilis-Javed Khan, Ujwal Gadiraju, and Panos Markopoulos. 2020. Trainbot: A Conversational Interface to Train Crowd Workers for Delivering On-Demand Therapy. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Vol. 8. AAAI, 3–12.

[3] Sabrina Barko-Sherif, David Elsweiler, and Morgan Harvey. 2020. Conversational agents for recipe recommendation. In *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*. Association for Computing Machinery, New York, NY, USA, 73–82.

[4] Alexandra Bejarano, Samantha Reig, Priyanka Senapati, and Tom Williams. 2022. You Had Me at Hello: The Impact of Robot Group Presentation Strategies on Mental Model Formation. In *Proceedings of the 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Association for Computing Machinery, New York, NY, USA, 363–371.

[5] Alexandra Bejarano and Tom Williams. 2022. Understanding and Influencing User Mental Models of Robot Identity. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, 1149–1151.

[6] James R Bettman, Mary Frances Luce, and John W Payne. 1998. Constructive consumer choice processes. *Journal of consumer research* 25, 3 (1998), 187–217.

[7] 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*. Association for Computing Machinery, New York, NY, USA, 63–70.

[8] Janghee Cho and Emilee Rader. 2020. The role of conversational grounding in supporting symbiosis between people and digital assistants. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (2020), 1–28.

[9] Konstantina Christakopoulou, Filip Radlinski, and Katja Hofmann. 2016. Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. Association for Computing Machinery, New York, NY, USA, 815–824.

[10] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, et al. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[11] Robert Dale. 2016. The return of the chatbots. *Natural Language Engineering* 22, 5 (2016), 811–817.

[12] Philip R Doyle, Leigh Clark, and Benjamin R Cowan. 2021. What do we see in them? identifying dimensions of partner models for speech interfaces using a psycholexical approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–14.

[13] Mehdi Elahi, Danial Khosh Kholgh, Mohammad Sina Kiarostami, Sorush Saghari, Shiva Parsa Rad, and Marko Tkalčič. 2021. Investigating the impact of recommender systems on user-based and item-based popularity bias. *Information Processing & Management* 58, 5 (2021), 102655.

[14] Mehdi Elahi, Francesco Ricci, and Neil Rubens. 2016. A survey of active learning in collaborative filtering recommender systems. *Computer Science Review* 20 (2016), 29–50.

[15] Cristhian Figueroa, Iacopo Vagliano, Oscar Rodríguez Rocha, and Maurizio Morisio. 2015. A systematic literature review of Linked Data-based recommender systems. *Concurrency and Computation: Practice and Experience* 27, 17 (2015), 4659–4684.

[16] Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for Voice Interface Design. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 26, 8 pages. https://doi.org/10.1145/3342775.3342788

[17] Asbjørn Følstad and Marita Skjuve. 2019. Chatbots for Customer Service: User Experience and Motivation. In *Proceedings of the 1st International Conference on Conversational User Interfaces (CUI '19)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3342775.3342784

[18] Marco de Gemmis, Pasquale Lops, Cataldo Musto, Fedelucio Narducci, and Giovanni Semeraro. 2015. Semantics-aware content-based recommender systems. In *Recommender systems handbook*. Springer, Boston, MA, USA, 119–159.

[19] Carlos A Gomez-Uribe and Neil Hunt. 2015. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)* 6, 4 (2015), 1–19.

[20] Ido Guy. 2018. The characteristics of voice search: Comparing spoken with typed-in mobile web search queries. *ACM Transactions on Information Systems (TOIS)* 36, 3 (2018), 1–28.

[21] Javeria Habib, Shuo Zhang, and Krisztian Balog. 2020. Iai MovieBot: A conversational movie recommender system. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. Association for Computing Machinery, New York, NY, USA, 3405–3408.

[22] Sheena S Iyengar and Mark R Lepper. 2000. When choice is demotivating: Can one desire too much of a good thing? *Journal of personality and social psychology* 79, 6 (2000), 995.

[23] Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)* 54, 5 (2021), 1–36.

[24] Yucheng Jin, Wanling Cai, Li Chen, Nyi Nyi Htun, and Katrien Verbert. 2019. MusicBot: Evaluating critiquing-based music recommenders with conversational interaction. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. Association for Computing Machinery, New York, NY, USA, 951–960.

[25] Eric J Johnson, Suzanne B Shu, Benedict GC Dellaert, Craig Fox, Daniel G Goldstein, Gerald Häubl, Richard P Larrick, John W Payne, Ellen Peters, David Schkade, et al. 2012. Beyond nudges: Tools of a choice architecture. *Marketing Letters* 23, 2 (2012), 487–504.

[26] Patrik Jonell, Per Fallgren, Fethiye Irmak Doğan, José Lopes, Ulme Wennberg, and Gabriel Skantze. 2019. Crowdsourcing a self-evolving dialog graph. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. Association for Computing Machinery, New York, NY, USA, 1–8.

[27] Raghav Pavan Karumur, Tien T Nguyen, and Joseph A Konstan. 2018. Personality, user preferences and behavior in recommender systems. *Information Systems Frontiers* 20, 6 (2018), 1241–1265.

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

[29] Bart P Knijnenburg and Martijn C Willemsen. 2015. Evaluating recommender systems with user experiments. In *Recommender systems handbook*. Springer,

Boston, MA, USA, 309–352.

[30] Rafal Kocielnik, Raina Langevin, James S George, Shota Akenaga, Amelia Wang, Darwin P Jones, Alexander Argyle, Callan Fockele, Layla Anderson, Dennis T Hsieh, et al. 2021. Can I Talk to You about Your Social Needs? Understanding Preference for Conversational User Interface in Health. In *CUI 2021-3rd Conference on Conversational User Interfaces*. Association for Computing Machinery, New York, NY, USA, 1–10.

[31] JA Konstan and J Riedl. 2012. Recommended for you: How online merchants predict your preferences and prod you to purchase. *IEEE Spectrum* 49, 10 (2012), 54–61.

[32] Joseph A Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction* 22, 1 (2012), 101–123.

[33] Dimosthenis Kontogiorgos. 2022. *Mutual Understanding in Situated Interactions with Conversational User Interfaces: Theory, Studies, and Computation*. Ph.D. Dissertation. KTH Royal Institute of Technology.

[34] Guy Laban and Theo Araujo. 2020. The effect of personalization techniques in users' perceptions of conversational recommender systems. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. Association for Computing Machinery, New York, NY, USA, 1–3.

[35] Béatrice Lamche, Uwe Trottmann, and Wolfgang Wörndl. 2014. Active learning strategies for exploratory mobile recommender systems. In *Proceedings of the 4th Workshop on Context-Awareness in Retrieval and Recommendation*. Association for Computing Machinery, New York, NY, USA, 10–17.

[36] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300932

[37] Minha Lee, Lily Frank, Femke Beute, Yvonne De Kort, and Wijnand IJsselsteijn. 2017. Bots mind the social-technical gap. In *Proceedings of 15th European conference on computer-supported cooperative work-exploratory papers*. European Society for Socially Embedded Technologies (EUSSET), 35–54.

[38] Minha Lee, Lily Frank, and Wijnand IJsselsteijn. 2021. Brokerbot: A cryptocurrency chatbot in the social-technical gap of trust. *Computer Supported Cooperative Work (CSCW)* 30, 1 (2021), 79–117.

[39] Minha Lee, Dimosthenis Kontogiorgos, Ilaria Torre, Michal Luria, Ravi Tejwani, Matthew J Dennis, and Andre Pereira. 2021. Robo-identity: Exploring artificial identity and multi-embodiment. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, New York, NY, USA, 718–720.

[40] Minha Lee, Gale Lucas, and Jonathan Gratch. 2021. Comparing mind perception in strategic exchanges: Human-agent negotiation, dictator and ultimatum games. *Journal on Multimodal User Interfaces* 15, 2 (2021), 201–214.

[41] Sunok Lee, Sungbae Kim, and Sangsu Lee. 2019. " What does your Agent look like?" A Drawing Study to Understand Users' Perceived Persona of Conversational Agent. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–6.

[42] Sangwon Lee, Naeun Lee, and Young June Sah. 2020. Perceiving a mind in a Chatbot: effect of mind perception and social cues on co-presence, closeness, and intention to use. *International Journal of Human–Computer Interaction* 36, 10 (2020), 930–940.

[43] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. " I Hear You, I Feel You": Encouraging Deep Self-disclosure through a Chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[44] Toby Jia-Jun Li, Jingya Chen, Haijun Xia, Tom M Mitchell, and Brad A Myers. 2020. Multi-modal repairs of conversational breakdowns in task-oriented dialogs. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. Association for Computing Machinery, New York, NY, USA, 1094–1107.

[45] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4 (2014), 2065–2073.

[46] Aldo Lipani, Ben Carterette, and Emine Yilmaz. 2021. How am I doing?: Evaluating conversational search systems offline. *ACM Transactions on Information Systems (TOIS)* 39, 4 (2021), 1–22.

[47] Kate Lister, Tim Coughlan, Francisco Iniesto, Nick Freear, and Peter Devine. 2020. Accessible conversational user interfaces: considerations for design. In *Proceedings of the 17th International Web for All Conference*. Association for Computing Machinery, New York, NY, USA, 1–11.

[48] Bingjie Liu and Lewen Wei. 2021. Machine gaze in online behavioral targeting: The effects of algorithmic human likeness on social presence and social influence. *Computers in Human Behavior* 124 (2021), 106926.

[49] Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. Towards Conversational Recommendation over Multi-Type Dialogs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 1036–1049. https://doi.org/10.18653/v1/2020.acl-main.98

[50] Xiao Ma and Ariel Liu. 2020. Challenges in Supporting Exploratory Search through Voice Assistants. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. Association for Computing Machinery, New York, NY, USA, 1–3.

[51] Tariq Mahmood and Francesco Ricci. 2009. Improving recommender systems with adaptive conversational strategies. In *Proceedings of the 20th ACM conference on Hypertext and hypermedia*. Association for Computing Machinery, New York, NY, USA, 73–82.

[52] Sean M McNee, Shyong K Lam, Joseph A Konstan, and John Riedl. 2003. Interfaces for eliciting new user preferences in recommender systems. In *International Conference on User Modeling*. Springer, Boston, MA, USA, 178–187.

[53] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 27, 1 (2001), 415–444.

[54] Martijn Millecamp, Cristina Conati, and Katrien Verbert. 2022. "Knowing me, knowing you": personalized explanations for a music recommender system. *User Modeling and User-Adapted Interaction* 32, 1 (2022), 215–252.

[55] Robert J Moore and Raphael Arar. 2019. *Conversational UX design: A practitioner's guide to the natural conversation framework*. Morgan & Claypool.

[56] Christine Murad, Cosmin Munteanu, Leigh Clark, and Benjamin R Cowan. 2018. Design guidelines for hands-free speech interaction. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. Association for Computing Machinery, New York, NY, USA, 269–276.

[57] Cataldo Musto, Alain D Starke, Christoph Trattner, Amon Rapp, and Giovanni Semeraro. 2021. Exploring the Effects of Natural Language Justifications in Food Recommender Systems. In *Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 147–157.

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

[59] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for how users overcome obstacles in voice user interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–7.

[60] Fedelucio Narducci, Marco de Gemmis, Pasquale Lops, and Giovanni Semeraro. 2018. Improving the user experience with a conversational recommender system. In *International conference of the italian association for artificial intelligence*. Springer, Boston, MA, USA, 528–538.

[61] Nicole Novielli, Fiorella de Rosis, and Irene Mazzotta. 2010. User attitude towards an embodied conversational agent: Effects of the interaction mode. *Journal of Pragmatics* 42, 9 (2010), 2385–2397.

[62] SoHyun Park, Anja Thieme, Jeongyun Han, Sungwoo Lee, Wonjong Rhee, and Bongwon Suh. 2021. "I wrote as if I were telling a story to someone I knew.": Designing Chatbot Interactions for Expressive Writing in Mental Health. In *Designing Interactive Systems Conference 2021*. Association for Computing Machinery, New York, NY, USA, 926–941.

[63] Hannah RM Pelikan and Mathias Broth. 2016. Why that Nao? how humans adapt to a conventional humanoid robot in taking turns-at-talk. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 4921–4932.

[64] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–12.

[65] Pearl Pu and Li Chen. 2008. User-involved preference elicitation for product search and recommender systems. *AI magazine* 29, 4 (2008), 93–93.

[66] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. Recommender systems: introduction and challenges. In *Recommender systems handbook*. Springer, Boston, MA, USA, 1–34.

[67] Aghiles Salah, Quoc-Tuan Truong, and Hady W Lauw. 2020. Cornac: A Comparative Framework for Multimodal Recommender Systems. *J. Mach. Learn. Res.* 21 (2020), 95–1.

[68] Daniel Schulman and Timothy Bickmore. 2009. Persuading users through counseling dialogue with a conversational agent. In *Proceedings of the 4th international conference on persuasive technology*. Association for Computing Machinery, New York, NY, USA, 1–8.

[69] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My chatbot companion-a study of human-chatbot relationships. *International Journal of Human-Computer Studies* 149 (2021), 102601.

[70] Alain Starke. 2019. The effectiveness of advice solicitation and social peers in an energy recommender system. In *IntRS 2019 Interfaces and Human Decision Making for Recommender Systems 2019 (CEUR Workshop Proceedings)*, Peter Brusilovsky, Marco de Gemmis, Alexander Felfernig, Pasquale Lops, John O'Donovan, Giovanni Semeraro, and Martijn C. Willemsen (Eds.). CEUR-WS.org, Aachen, DE,

65–71. 6th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, IntRS 2019 ; Conference date: 19-09-2019.

[71] Alain Starke, Edis Asotic, and Christoph Trattner. 2021. "Serving Each User": Supporting Different Eating Goals Through a Multi-List Recommender Interface. In *Fifteenth ACM Conference on Recommender Systems*. Association for Computing Machinery, New York, NY, USA, 124–132.

[72] Alain Starke, Martijn Willemsen, and Chris Snijders. 2017. Effective user interface designs to increase energy-efficient behavior in a rasch-based energy recommender system. In *Proceedings of the eleventh ACM conference on recommender systems*. Association for Computing Machinery, New York, NY, USA, 65–73.

[73] Alain Starke, Martijn Willemsen, and Chris Snijders. 2021. Promoting Energy-Efficient Behavior by Depicting Social Norms in a Recommender Interface. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 11, 3-4 (2021), 1–32.

[74] Alain D Starke. 2019. *Supporting energy-efficient choices using Rasch-based recommender interfaces*. Technische Universiteit Eindhoven, Eindhoven, NL.

[75] Yueming Sun and Yi Zhang. 2018. Conversational recommender system. In *The 41st international ACM SIGIR conference on Research & Development in Information Retrieval*. Association for Computing Machinery, New York, NY, USA, 235–244.

[76] Paul Thomas, Mary Czerwinksi, Daniel McDuff, and Nick Craswell. 2021. Theories of conversation for conversational IR. *ACM Transactions on Information Systems (TOIS)* 39, 4 (2021), 1–23.

[77] Christoph Trattner and David Elsweiler. 2017. Investigating the healthiness of internet-sourced recipes: implications for meal planning and recommender systems. In *Proceedings of the 26th international conference on world wide web*. Association for Computing Machinery, New York, NY, USA, 489–498.

[78] Sarah Theres Völkel, Samantha Meindl, and Heinrich Hussmann. 2021. Manipulating and Evaluating Levels of Personality Perceptions of Voice Assistants through Enactment-Based Dialogue Design. In *CUI 2021-3rd Conference on Conversational User Interfaces*. Association for Computing Machinery, New York, NY, USA, 1–12.

[79] Weiquan Wang and Izak Benbasat. 2013. Research Note—A contingency approach to investigating the effects of user-system interaction modes of online decision aids. *Information Systems Research* 24, 3 (2013), 861–876.

[80] Pontus Wärnestål. 2005. User evaluation of a conversational recommender system. In *Proceedings of the 4th Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. 32–39.

[81] Joseph Weizenbaum. 1983. ELIZA — a Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun. ACM* 26, 1 (Jan. 1983), 23–28. https://doi.org/10.1145/357980.357991