Exploring Mental Models for Transparent and Controllable Recommender Systems: A Qualitative Study

Thao Ngo*
thao.ngo@uni-due.de
University of Duisburg-Essen
Duisburg, Germany

Johannes Kunkel*
johannes.kunkel@uni-due.de
University of Duisburg-Essen
Duisburg, Germany

Jürgen Ziegler
juergen.ziegler@uni-due.de
University of Duisburg-Essen
Duisburg, Germany

ABSTRACT

While online content is personalized to an increasing degree, e.g. using recommender systems (RS), the rationale behind personalization and how users can adjust it typically remains opaque. This was often observed to have negative effects on the user experience and perceived quality of RS. As a result, research increasingly has taken user-centric aspects such as transparency and control of a RS into account, when assessing its quality. However, we argue that too little of this research has investigated the users' perception and understanding of RS in their entirety. In this paper, we explore the users' mental models of RS. More specifically, we followed the qualitative *grounded theory* methodology and conducted 10 semi-structured face-to-face interviews with typical and regular Netflix users. During interviews participants expressed high levels of uncertainty and confusion about the RS in Netflix. Consequently, we found a broad range of different mental models. Nevertheless, we also identified a general structure underlying all of these models, consisting of four steps: data acquisition, inference of user profile, comparison of user profiles or items, and generation of recommendations. Based on our findings, we discuss implications to design more transparent, controllable, and user friendly RS in the future.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Human-centered computing → User studies**.

KEYWORDS

Recommender Systems, Transparent AI, Mental Models, Grounded Theory, Think Aloud

ACM Reference Format:
Thao Ngo, Johannes Kunkel, and Jürgen Ziegler. 2020. Exploring Mental Models for Transparent and Controllable Recommender Systems: A Qualitative Study. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '20), July 14–17, 2020, Genoa, Italy.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3340631.3394841

1 INTRODUCTION

With the growing use of intelligent algorithms in current systems, such as recommender systems (RS), end-users find it increasingly hard to comprehend the rationale behind a certain recommendation. Thus, it is important for users to understand the relationship between user input and recommendation of RS [36]. In most cases, systems appear to users as black boxes, particularly in case of the increasingly used complex probabilistic techniques [12]. Previous research suggests that this opaqueness can lead to feelings of discomfort or even creepiness when a personalized recommendation matches a user's interest very accurately [41]. These feelings, in turn, may have negative consequences on users' trust in a RS and their intention to accept recommendations. Thus, recently, research efforts were made to increase transparency and control of a RS, e.g. through interactive explanatory interfaces [21, 42]. An important and understudied question in this context is what kind of *mental models* users form of RS. Based on in-depth knowledge about such mental models, designers of RS could make recommendations more transparent and controllable, thus mitigating the negative consequences.

Mental models can be defined as subjective knowledge representations of technological systems (e.g. computer programs) [26, 33]. Previous research indicates that users do construct mental models for RS. The soundness of these models influences satisfaction and effectiveness of interaction with the RS [9, 16]. As such, mental models focus on practical effectiveness and on making predictions about the outcome of the system. They are typically incomplete, inaccurate, and may contain areas of uncertainty [26, 33].

Due to this subjective nature of mental models, a qualitative approach seems to be most appropriate to investigate them. This approach allows us to investigate the users' unique perspectives in-depth and ask for *what* and *why* users hold certain mental models of a RS. Specifically, we chose the *Grounded Theory* (GT) methodology [5] due to its strong exploratory and data-driven nature. The participants' knowlegdge, experiences, and attitudes solely drive the data collection and analysis. Thus, the results from this methodology emerge from the data. In other words, they are *grounded* in them.

In GT, data sampling is performed purposefully, i.e. not randomly. Thus, to reveal what mental models users of RS, what assumptions these models entail, and what implications for future RS development can be derived from them, we focused on mental models of typical and regular RS users. In particular, we aim to answer four central research questions:

- RQ1: What are the mental models users hold of a RS?
- RQ2: To what extent is the RS perceived as transparent?
- RQ3: To what extent is the RS perceived as controllable?
- RQ4: What implications for RS design can be derived?

[*] Both authors contributed equally to this research.

UMAP '20, July 14–17, 2020, Genoa, Italy

ACM ISBN 978-1-4503-6861-2/20/07…$15.00
https://doi.org/10.1145/3340631.3394841

In this study, we chose Netflix as an example because it makes extensive and apparent use of recommendations [11]. Moreover, it is one of the most popular video-on-demand services in the U.S. and Germany [8, 38]. Thus, the sample of this study most likely has developed a mental model of Netflix.

We make two main contributions with this work: (1) A theoretical contribution in form of the exploration of mental models of RS. The mental models provide in-depth insights to the user assumptions of how a RS works internally. For example, we found that all mental models followed a basic structure, comprising four steps: data acquisition, inference of user profile, comparison of user profiles or items, and generation of recommendations. (2) A practical contribution in form of discussing how our theoretical results can be applied to the development of RS. For instance, we suggest to link recommendations and user preferences more explicitly than it is done to this date.

2 BACKGROUND AND RELATED WORK

RS have become widely adopted tools to pro-actively filter online content with respect to the current user's preferences. While recommendation algorithms are able to suggest items with high precision, quality criteria that go *beyond accuracy* [15, 24] were neglected for a long time. It has been argued that user-centric aspects, such as the system's perceived transparency or the degree of control users are able to exert, constitute important facets of a system's overall perceived quality [2, 29].

2.1 Transparency and Control in RS

Typically, RS appear as *black box* to their users as it remains opaque why items are recommended and how they relate to the users' preferences [13, 35]. Increasing the transparency of a RS constitutes a prominent issue in HCI design for RS [2, 9]. It can improve perceived quality of recommendations [18], their acceptance [6, 13], and users' confidence [36]. Therefore, many researchers have called for explainable RS, i.e. the increase of system transparency through (mostly textual) explanations (e.g. [40, 42]).

Another aspect that goes *beyond accuracy* is the extent to which users can exert control over the recommendation process. Allowing users to control what is recommended to them can increase user satisfaction [32] and the perceived accuracy of predictions [28]. While many RS rely on user ratings (e.g. implicitly by recording click-through streams, or explicitly by eliciting thumb up/down ratings) [31, 37], more advanced methods for controlling recommendations have been suggested. Examples include relating preferences and recommendations more directly [1, 19], or eliciting preferences for groups instead of single items [3, 22].

Transparency and control are not independent from each other. To exert control over their recommendations effectively, users need insights into the system's reasoning—at least to a certain degree [9, 40**?**]. Yet, the relation between transparency and control is not trivial to investigate and may lead to counter-intuitive observations. Tsai and Brusilovsky [42], for instance, found that, besides increasing transparency, explaining recommendations can also result in a *decrease* of the perceived degree of control. According to the authors, this might be due to information overload effects entailed by the explanatory interfaces.

Such observations underline that putting transparency and control into practice may not be straightforward. In this context, we add another aspect that might be responsible for this: a discrepancy between a user's *mental model* of a system and its actual behavior.

2.2 Mental models in RS

Mental models can be defined as knowledge representations of technological systems, which are generated through interaction with the respective system [26, 33]. Rumelhart and Norman [33] used the terms of *represented* and *representing world*. The mental model represents an object or a situation of the represented world inside the cognitive representing world. This points out that mental models are *constructed*, i.e. the representing world is incomplete as it only contains those properties of the represented world that were deemed necessary. Elsewhere, Norman [26] uses a slightly different terminology to which we adhere in this paper: Based on a *target system* (i.e. the represented world) the user invents a mental model (i.e. the representing world) to simulate system behavior and make assumptions about interaction outcomes. Norman underlines that the users' mental models are incomplete, contain areas of uncertainty and possibly superstition, and focus on practical effectiveness rather than technical accuracy. In contrast to the user's mental model, the *conceptual model* represents a *more appropriate* model of the target system in terms of accuracy, consistency and completeness. They are constructed by specialists regarding the target system (e.g. the system designers).

Yet, mental models need some degree of technical correctness to let users successfully predict system behavior and thus, use it effectively. If this is not the case, misaligned mental models can result in what Norman describes as "gulfs" between user and system [27]: The *gulf of execution* occurs when a user's mental model is erroneous in terms of how a specific task can be performed with the system. The *gulf of evaluation* occurs when the actual outcome of an action with the system diverges from what the user's mental model predicted. These gulfs are well-known in usability engineering and account for many problems and misconceptions arising in HCI. One reason for the occurrence of such gulfs may lie in the transfer of a mental model from one technical system to another. To save cognitive effort, users try to re-use mental models whenever it seems feasible [26, 27].

Shneiderman and Maes concluded that one important future challenge is to make users aware of how autonomous software agents (e.g. RS) came to decisions and thus, become predictable for users [34]. Even though they did not use the term of mental models explicitly, they described them implicitly as making practical predictions about the outcome of the system is the most central utility of mental models. Surprisingly, this aspect was not further investigated in the subsequent years. To this date, the literature on mental models of RS is relatively sparse.

Only few studies have examined mental models in the context of RS so far. In an initial online survey, Ghori et al. [10] presented scenarios of different RS platforms and asked for users' knowledge and beliefs about RS. While they did not explicitly elicit mental models of RS, they concluded that users hold a "cognitive model", understand that RS track user behavior, and have rudimentary ideas of filtering mechanisms. In an exploratory approach, Kodama et al.

[14] elicited different mental models that middle school students create of the Google search engine. They found, that these models were most often wrong and conclude that concepts behind algorithmic agents should be taught better. In line with this, Kulesza et al. [16] has shown in an experiment that users who increase soundness of their mental model during usage were more efficient in controlling their recommendations. This resulted in a higher satisfaction with the outcome. To operationalize the systematic consideration of users' mental models into actual software design, Eiband et al. [9] have proposed a stage-based, iterative prototyping approach that targets at making RS more transparent through offering explanations.

3 METHOD

The research goal of this study is to investigate the users' mental models. Since the structure of mental models is inherently subjective and individual, a quantitative approach would be insufficient for this goal as this approach aims at analyzing empirical data for predetermined hypotheses. Thus, to explore unknown and highly individual mental models, we deem a qualitative approach as more appropriate.

Our qualitative study followed the *Grounded Theory* (GT) methodology [5, 39]. GT is an established and well-defined methodology from social sciences for systematic data collection and analysis. This methodology has a strong exploratory focus, i.e. no clear theory about the topic at hand is presupposed. Concepts evolve from the data during conduct of the study and hence, are *grounded in* the data. Due to the lack of predetermined hypotheses, data sampling follows the approach of *theoretical sampling* [5]. This means that sampling is performed purposefully, not randomly. Furthermore, while sampling in quantitative research is typically randomized and person-wise, in qualitative research theoretical sampling is done iteratively and concept-wise. Data are collected, coded, and analyzed simultaneously. In this way newly occurring concepts determine the sampling during the study to explore them dynamically. For this, the differences in relevant concepts (also called contrasts) are deliberately varied until no further novel observations regarding the concept are made. Then, the state of so-called *theoretical saturation* [5, 25] is achieved. In this case, either another concept is explored or the study is concluded if the pursued theory is already well-developed.

In our study, the *theory* of GT are the different mental models users hold of a RS. We deliberately focused Netflix as an example for RS, since it 1) is well-known for its extensive and apparent use of recommendations [11], and 2) is wide-spread, increasing the likelihood for us to sample a broad variety of contrast in our concepts. In other words, this allows us to study different variations of one concept. As instruments we applied individual semi-structured face-to-face interviews, which we combined with a *Think Aloud* task and a drawing task to capture different facets of each mental model as broadly as possible. Following the approach of theoretical sampling, we deliberately recruited participants of whom we had information about their background and who fit to the current concept under consideration (e.g. the level of technical knowledge). Throughout the entire study, we only sampled participants with advanced Netflix experience (frequent use for at least one year), as we aimed to focus on the typical Netflix user. All in all, we recruited ten participants (six female) with an age range between 19 and 31 ($M = 24.70$, $SD = 4.57$). Hence, our sample represented the typical Netflix user group well [7]. The interviews were conducted in July and August 2019.

For our analysis, each interview was transcribed in a timely manner using easytranscript 2.50 and analyzed with MAXQDA 18.2.3. The transcribed interview of each participant was first coded by two independent raters. Subsequently, the two raters discussed and analyzed each interview jointly. During analysis, various analytic tools and mental strategies were used, including *microanalysis* of the data through open *line-by-line coding*, *constant comparison* and *axial coding* to summarize the open codings to categories, and *selective coding* to infer the mental model of Netflix for each participant. To ensure that codes and resulting categories emerge from the data, throughout the whole process *in-vivo codings* (i.e. verbatim codes from participants' statements) played a central role. In order to record impressions, evolving theoretical concepts and the relationships among them, raters made extensive use of memos, which constitutes a substantial aspect of the GT method.

This iterative process led to 10 distinct categories such as *evaluation strategy for items*, which pertains to how participants asses the quality of items (e.g. *content-based* vs. *non-content-based*), *search strategy for items*, which describes how participants decide whether to consume an item or not (e.g. through *internal* or *external* information acquisition), and *general model of RS*, which we focus within the scope of this work and which emerged from our data presented in Section 4.

3.1 Preparation

The study was approved by the local ethics committee of the University of Duisburg-Essen in Germany. Participants consented to the interview and audio recording. All personally identifiable information was anonymized.

3.1.1 Interview guide. We developed an interview guide with an interview duration of roughly one hour. All interview questions in the guide were open-ended. The interview started with a brief introduction of the interviewer and a short description of the purpose and motivation of the interview. It was emphasized that the study is concerned with the recommendation component of Netflix and that the main interest of the study lies in the exploration of the participants' experience with the personalized content of Netflix. The participants were then asked about their experience with Netflix: How often do they use Netflix? Since when do they use it? Which experience do they have with the recommendations in Netflix? Which parts of Netflix are subject to personalization? How confident do they feel with personalization in general?

The interview proceeded with the Think Aloud task: Participants were instructed to use their own Netflix account to find a comedy movie to watch in the evening that was in line with their preferences. After that, a hypothetical scenario was introduced. Participants had to imagine that the movie was not as good as expected. Thus, they should try to express negative feedback to Netflix for that movie. The purpose of this task was for the participants to reflect on the options to express their preferences to Netflix. Then, participants were asked about the functioning and data processing of Netflix: How does a RS like Netflix work? Which data are used by the system? What happens to the data in order to generate recommendations? Do they know about the thumb function of Netflix? What does it trigger?

Finally, participants received a sheet of paper and were asked to draw their very own image of Netflix. Participants were informed that they could perform this task freely, without any limitation. They were prompted to explain their drawing.

At the end of the interview, participants were debriefed and offered to ask questions and give general feedback on the interview. No incentives were given for participation, besides a certificate of taking part in the study[1].

3.2 Data sampling

Our data sampling was fundamentally influenced by the data-driven approach of GT and of theoretical sampling. Accordingly, we sampled participants not at random, but based on what concepts we decided to explore next. In general our data acquisition was organized in three phases with different foci. In the following section, we elaborate on our sampling decisions for this study.

3.2.1 First sampling phase: Typical Netflix users. As recommended by Corbin and Strauss [5], we first focused on a typical sample of the target population. According to Dahlgreen [7] 57% of Netflix users are female and roughly 50% of all Netflix users are between 18 and 34 years old. Our initial sample (P1, P2, P3, P4) were females with an age of 21, 24, 27, and 24. Thus, we were able to recruit a sample within the age range of typical Netflix users.

In this first sampling phase, we found the concepts of *centrality of self* and *item-based recommendation*. The first concept was especially salient in the interview of P3, while the item-based recommendation was most apparent in the drawing of P2. We found these concepts mostly through *comparison*. During further *axial coding*, we observed that all participants held rather *technical* mental models (see Section 4.3), i.e. they are close to the functioning of algorithms or procedures. This became mostly apparent through *microanalysis*, which revealed that P1, P3, and P4 generally used many technical terms (e.g. "*ip address*" (P1, P3), "*database*" (P1, P3, P4), and "*dynamic query*" (P3)).

Following the *flip-flop technique* [5], we turned this concept "upside down", asking ourselves questions such as: "*How are the mental models in case of lower technical knowledge?*", and "*How are the mental models in case of higher technical knowledge?*" In order to investigate these questions, we decided to sample low and high extremes on the dimension of technical knowledge next.

3.2.2 Second sampling phase: Low/high technical knowledge. Next, we purposefully sampled P5 and P6. Both participants were male and aged 31 and 30, respectively. While P5 had a very low technical background regarding RS (he held a bachelor's degree in arts and was currently unemployed), P6 had a high technical knowledge as he worked in computer science research and was currently engaged with decision support systems.

In this second sampling phase, we found mental models which differed in a *metaphorical* and *technical dimension* (see Section 4.3). P5 clearly held a metaphorical mental model: He drew Netflix as a monster serving recommendations with many arms (see Figure 2c). In contrast to this, P6 had a technical idea of Netflix.

In addition, P6 mentioned that he used the explicit rating function (thumbs up/down) occasionally to steer his Netflix account towards better recommendations. We found this aspect quite striking as we did not elicit any responses on what influence the explicit rating function may have until this point of the study. Rather tentative, we assumed a connection between usage of explicit ratings and decided to restrict our next sample to participants using explicit ratings frequently.

3.2.3 Third sampling phase: Use of explicit ratings. During this third sampling phase, we conducted interviews with participants P7, P8, P9, and P10 aged 19, 19, 22, and 30. P7 and P10 were male, while P8 and P9 were female. All declared using the thumbs function in Netflix frequently. During analysis of this sample, we observed the counterpart of *item-based recommendation*, namely *user-based recommendation* (see Section 4.2).

After analyzing the data, we found that the main concepts, which we found before, did not show further variation in these observations. Thus we deemed them as theoretically saturated. Especially, with regards to our main aspects of transparency and control, we did not see new insights in the interviews. Thus, we ended our data sampling at this point.

4 RESULTS

Overall, one general structure of users' mental models of RS emerged from our collected data. All participants followed the same pattern and divided the functioning of RS into four separate steps: (1) *data acquisition*, (2) *inference of a virtual user profile*, (3) *comparison of user profiles or items*, and (4) *generation of recommendations* (see Figure 1).

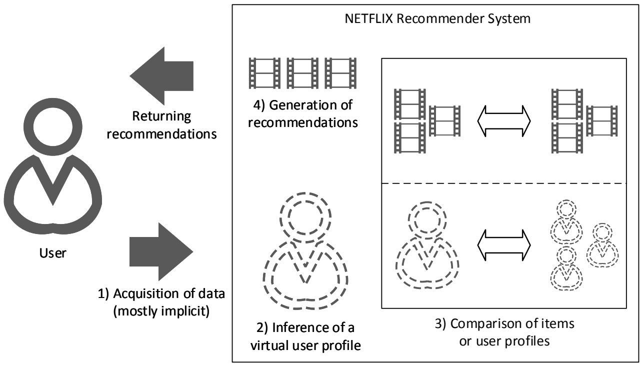

NETFLIX Recommender System
Returning recommendations
4) Generation of recommendations
User
1) Acquisition of data (mostly implicit)
2) Inference of a virtual user profile
3) Comparison of items or user profiles

Figure 1: Basic mental model found in all participants.

Regarding the acquisition of data in step (1), our analysis revealed that participants considered user characteristics, such as location, gender, and age, as well as user interaction behavior as relevant for Netflix. For the latter, we were able to form two categories: *implicit* user behavior, such as watching a movie, and *explicit* user behavior, such as pressing the thumbs-up button. From these data, participants assumed that Netflix derives a virtual user profile in step (2). This profile may contain *latent* item characteristics, which

[1] This was requested by 3 of the 10 participants since they needed to participate in empirical studies as part of their study program.

are not visible to the user. For example, P5 speculated: "*As far as I know, there are a lot of subcategories in the background which a user does not see on the interface.*" In step (3), participants assumed that comparisons between items or user profiles were made. These two general directions adhered to the concepts of *user-* and *item-based recommending* (see Section 4.2). Finally, step (4) corresponds to the actual selection of personal recommendations. Here, participants assumed that all process data cumulated into one recommendation. This assumption is for instance depicted in the drawing of P3 (see Figure 2b).

Regarding details of how the four steps are performed, participants made diverse assumptions. Nonetheless, across all interviews, participants expressed confusion and uncertainty when asked about the inner working of Netflix: "*I don't know which data they have of me.*" (P3), or "*It's a black box. I don't know how they do it. Maybe I should know it.*" (P7). Many of them also rejected the recommendations provided by Netflix, as P2 stated:

> "*Some [recommended] movies I find interesting, but there are also many things, I am not interested in. I feel that my preferences don't play a role, instead it's just [the movies] which people are currently talking about.*"

Furthermore, based on participants' drawings and statements, we derived the concept of "centrality of self" from our data as well as two dimensions that characterize the identified mental models. They are reported in the following sections.

4.1 Centrality of self

Some participants clearly viewed their own self as central component in their Netflix experience (P3, P5, and P6). This became particularly apparent in the drawing of P3 (see Figure 2b), as she confidently started the drawing task with the role of herself ("*Ok, I am still a little overburdened by what to begin with. Ok, in any case, first of all we need myself: the Netflix user.*"). Then, the entire drawing evolved around this central self. While many other participants used a content-based approach of explaining how recommendations are generated (see Section 4.2), content aspects of any kind were entirely absent in the drawing of P3. Instead, in different parts of the sketched functionality, users played a central role, for instance, when recommendations are generated based on what was watched before (box with dashed line in the southwest of Figure 2b). The *centrality of self* together with the importance of users per se and their social interrelation, was emphasized by the participant's estimation of an existing internal connection between Netflix and Facebook (northern arc in Figure 2b). P3 assumed that the history of what her friends watched in the past was also taken into account when recommendations for herself are generated, and vice versa. Note, that this drawing clearly depicted three of the four general steps discussed above: data is acquired (watched items and Facebook data), similarity is calculated between users and items (inside the box with the dashed line), and recommendations are generated (arrows inside the box on the right).

The concept *centrality of self* can also be found throughout the interview of P3. She, for instance, mentioned that her recommendations are sometimes inaccurate. This results in long searching sessions, which she described as tedious and confusing. Yet, the reason for this lack of decisiveness was sought at her own side:

> "*Because it takes an extreme amount of time to search, but I would not necessarily burden this on Netflix but on myself, since I am never satisfied with my choice.*"

Note, however, that when being asked, P3 did not assess herself as a person who has a hard time to decide in general ("*at the supermarket [. . .] I am very determined.*"). Even though, this person did not ascribe the problem of long searching sessions in Netflix to the RS, she fancied the idea of having better explanations for her recommendations. In particular, P3 formulated a wish to know more about the relation of recommendations and her own preferences. As a consequence, the category of "similar to . . . " recommendations was perceived as helpful, yet also as arbitrary. When confronted with the idea to be able to control to which preferences recommendations are generated, P3 expressed a strong affection for such a feature[2].

Other participants did not see the self as central as P3 but elaborated on the role of the self implicitly throughout the interview. P1, for instance, showed some aspects of *centrality of self*, when she was asked to clarify the difference between implicit and explicit ratings. She underlined that her explicit ratings have higher impact compared to her implicit interaction data because she used the thumb function seldom. In the same answer during the interview, P1 took over the role of Netflix talking about herself: "*Ok, now she clicked on something [i.e. rated an item], so we will give her more of that.*" Even though rather shallow, the *self* as a concept was mentioned in both statements. It constituted a counterpart to Netflix as a system making assumptions about the user. A similar effect could be observed in answers of P10. He emphasized his own responsibility for the influence on recommendations ("*If I dislike Adam Sandler but all the time [. . .] watch movies starring him, I do not have to wonder when a Adam Sandler comes out [of the RS].*").

4.2 User- vs. item-based recommendations

When participants were asked about the rationale they assumed behind items being recommended, we observed two major directions. While some participants assumed recommendations being generated with respect to similarity between items (P5, P8, P10), others followed a user-based approach (P1, P9). As P9 put it:

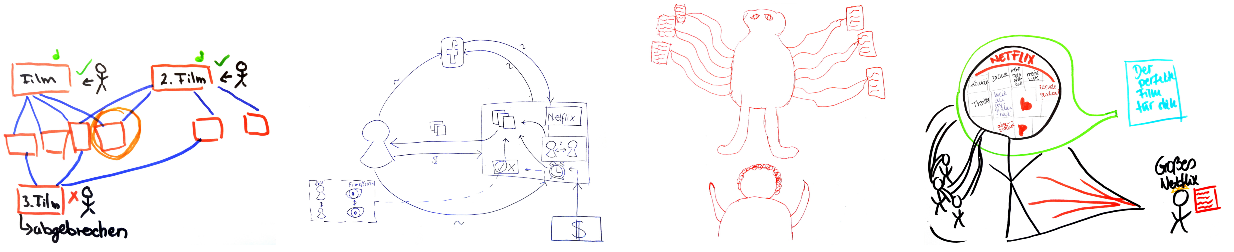
> "*[Recommendations base on] other users: what other users frequently watched, or gave a good rating for.*"

Strikingly, this style resembled closely the explanation model used by *Amazon* ("*Users who bought . . . also bought . . .* "). This resemblance was also explicitly mentioned by P9:

> "*I could imagine that it is like e.g. at Amazon. There it is also written that users bought something together.*"

P9 adhered to this form of thinking in her drawing as well (Figure 2d). Here, the entire process of generating recommendations was envisioned as inherently social. It depicted a crowd of users at the bottom left, which was connected to the personal Netflix agent (large stick figure in the middle). Through this connection the agent selects a movie as recommendation. Even Netflix in general was depicted as person, which instructs and overviews the entire process (at the bottom right of the drawing). When examining level

[2] Note, that such a function actually exists (next to the details for a movie or TV show). P3 also knew this function but, nonetheless, wished it to be more visible and that the "similar to . . . " category on the front page was replaced by an interactive version.



(a) Drawing of *P2*. A user watches three movies, from which two are liked. The recommendation (orange circle) is similar to the liked ones.

(b) Drawing of *P3*. The *centrality of self* is highly salient as the entire recommendation process evolves around this person's self.

(c) Drawing of *P5*. Netflix as a "tentacle monster", which can handle a huge range of recommendations with its many arms.

(d) Drawing of *P9*. The entire process of recommending is perceived as inherently social. This perspective highly emphasizes the role of users in Netflix.

Figure 2: Drawings of participants asked to illustrate their mental model of the inner workings of Netflix

of technical knowledge, P9 mentioned that she "*studies in that area*" and did have some knowledge about "*technical topics, AI, and such*".

Such social assumptions were juxtaposed by a model of Netflix that was based on content features of items. Examples for features being utilized for deriving similarity between items were "*actors*" (P8), "*buzz words*" (P10), and other content data such as "*movies set in the same time*" (P3). Another aspect for item-based comparison that was frequently mentioned were latent categories. Such categories were supposed to be only used "*in the background*" (P5) and had a finer granularity:

> "*there is not just action but also Asian action, German, and English... such things [...] for depicting more accurate [recommendations].*" (P6)

Apparently, this assumption originated from the RS in *Spotify* as P6 further explained: "*Once I saw a list somewhere containing Spotify genres. [...] They have somewhat over* 400 *genres*".

However, user- and item-based styles were not fully mutually exclusive. P2, P3, P4, and P6 showed aspects of both dimensions. P3, for instance, assumed a hybrid algorithm, which combines items watched by similar users and items that have a similar genre to the recently watched ones. Over all different styles, frequent use of verbs like *thinking*, *guessing*, and *believing* underlined the uncertainty about the inner workings of Netflix. The same applied to the *matching score*, which was shown for recommendations at Netflix. All participants agreed on being uncertain regarding what is actually *matched* when talking about the depicted score ("91 % *match – whatever that means.*" (P6)).

4.3 Technical vs. metaphorical

We found that the mental models can be characterized as either technical or metaphorical. Technical models were expressed by six participants (P2, P3, P4, P6, P7, P8). They used process diagrams and data flows to explain how Netflix arrives at its recommendations, which indicated a procedural understanding. For instance, P8 described:

> "*I am thinking about which data Netflix takes from me or already hold of me. [...] From this they know, what I like to watch. What else? Actors, producers... they take this from the movies I watched. Then, [Netflix] takes a look at the match and searches for [recommendations].*"

In the technical models, the general four steps were often made explicit by the participants. For example, P2 explained the steps *data acquisition*, *inference of user profile*, and *comparison of user profiles*:

> "*Probably everything is saved and collected for each user. And then, they compare users with similar profiles, in terms of the movies, to see whether these users have similar interests. Then, perhaps, one similar user has rated a movie positively the other one has not yet watched. Interviewer: And this is how they arrive at recommendations? P2: Yes, for instance.*"

The clear understanding of P2 was underlined by her drawing (see Figure 2a), which depicted the same process of recommending from an item-based perspective.

A different standpoint was taken by four participants (P1, P5, P9, P10), who used a metaphorical description of Netflix and thus, drew characters to illustrate how the RS works. P1, for instance, used a metaphor of a house:

> "*A huge complex house in which all the data and databases are somehow saved. [...] Of course, there are employees, but I think everything works with algorithms.*"

P1 focused on Netflix as a whole entity and in a more literal way than other participants. She expressed the four basic steps in the interview, however, for the drawing task, she chose the depiction of a house. This could be seen as a simplification of Netflix and a tangible understanding of Netflix which was based on the Netflix corporation building.

Additionally, some metaphorical mental models clearly entailed the participant's attitudes towards Netflix. P5 compared Netflix to a tentacle monster (see Figure 2c):

> "*It has all its tentacles and at each tentacle it offers its products, the movies it has. It's like a kraken monster. It has a huge range of offers, hence so many tentacles so that there is something for everybody.*"

This description expressed a negative view on Netflix. When browsing through the catalog of movies to find a matching one during the Think Aloud task, this participant expressed feelings of being lost and confused:

> "*Most things here mean nothing to me. These all are arbitrary and random images that do not catch me so that I think I don't want to look into [the details of the movie]. No. [. . .] everything seems absolutely random.*"

Such negative feelings were also quoted by other participants. Some, for instance, pointed out that personalization of Netflix engenders a loss of diversity in their movie consumption and therefore pose a risk of becoming trapped in a "filter bubble" (e.g. P1, P7, P9).

Apart from that, nearly all participants lacked trust in Netflix. Especially they doubted the system's integrity assuming that recommendations were biased towards in-house productions or third-parties, "*I have the feeling that in-house productions are mostly advertised and this is not necessarily good.*" (P6). When being asked about who influences the movie recommendations, P3 mentioned third-parties: "*The producers of the movies. [. . .] And perhaps record companies of movie soundtracks?! I don't know.*" Hence, at least some participants were aware of the economic interest of Netflix and third-parties. However, P9 justified their influence on the personalization process: "*I think it is good how it is because they do not exaggerate it and draw attention to it. It is also their production. [. . .] Therefore, it is good and also their right to advertise for themselves.*"

5 DISCUSSION

Regarding our first research question (" *What are the mental models users hold of the RS?*"), we found very diverse mental models, which, nonetheless, all adhered to a very basic structure—even among those participants with little technical knowledge. This structure consists of four steps: *data acquisition*, *inference of a user profile*, *comparison of items or users*, and *generation of recommendations* (see Figure 1). As this basic structure was held by all participants, we suspect that this structure might be prevalent in many typical and regular Netflix users. Our results extend the findings by Ghori et al. [10] substantially through the identification of this general model.

The subsequent sections are organized regarding our other research questions thus, asking for transparency and control in the identified mental models, and finally for possible implications for RS design.

5.1 As how transparent is Netflix perceived?

Across the four general steps, participants made various causal assumptions of how recommendations are derived and how their behavior as user affects them. We observed many discrepancies regarding these assumptions during single interviews, and especially between participants' drawings and their explanations throughout the rest of the interview. Assumptions were highly speculative and led to confusion—even superstition. This resulted in an effect we term *mystification* of the underlying RS: Participants invented various suppositions about the capabilities of the system, although they might be entirely unjustified and lack realistic evidence. One example illustrating this is P3's assumption that she receives recommendations, based on what her friends liked on Facebook. Thus, regarding RQ2, we conclude that users did not perceive the RS of Netflix as very transparent. We note that this was also not mitigated by the experience in using a RS, since we observed this in spite of the rather advanced experience with Netflix all participants had.

As a consequence of this lack of transparency, users encountered a *gulf of evaluation* (i.e. users did not understand what their recommendations were based upon) and were thus not able to exploit the full potential a personalized RS bears. We also found that mystified beliefs may harm the reputation of Netflix, which is shown by metaphorical mental models entailing negative attitudes towards the RS (e.g. P10 cynically drew Netflix as evil hungry black box eating user data and "pooping" recommendations). Reasons for this *mystification* and *gulf of evaluation* can be found in the dimensions we identified as concepts.

Participants, showing the *centrality of self* (Section 4.1), were clearly aware of the role of their own self, which we assume to be a general stance when encountering the surrounding world. Not surprisingly, this was also applied to the interpretation of recommendations. The users who expressed the centrality of self, wished to be more informed about which information about them is responsible for the recommendations. Participants were not able to understand this causality, which also resulted in a *gulf of evaluation* and, consequently, in a demand for a higher transparency regarding the influence of user preferences on recommendations.

In line with Norman [26], we observed that many of our participants tried to transfer their mental model of Amazon to Netflix. The RS of Amazon provides users with textual explanations for recommendations (i.e. products that were bought together with the currently inspected one). These explanations follow the algorithmic mechanism of item-based collaborative filtering, which we also found in the concept of *item-based recommending* (Section 4.2). While the prevalence of such algorithmic methods in the users' mental models, might be beneficial in some special cases (i.e. when source and target RS are algorithmically very similar), we assume such situations to be rather unlikely in practice. Our observations, for instance, show that the transfer of the mental model of Amazon to Netflix lead to false assumptions and misunderstandings. We ascribe this mainly to the different forms of how recommendations are presented. While Amazon follows an item-based approach, showing recommendations right next to the textual specification of single products, Netflix mainly presents recommendations in accordance to the entire user profile (i.e. "*top picks for you*"). Consequently, participants were highly unsure about how the list of recommendations was constructed.

5.2 As how controllable is Netflix perceived?

In our third research question, we asked ourselves to what degree the RS of Netflix is perceived as controllable by its users. As mentioned (e.g. in [9, 40**?**]), transparency and control are interdependent. We observed the same in our study: The lack of transparency led to a *gulf of execution* (i.e. participants were unable to figure out what interaction possibilities they had). Consequently, they also found it unclear how to steer the RS towards recommendations fitting their needs more adequately.

One reason we deem responsible, is again the transfer of mental models from Amazon to Netflix, mainly because the rather simplistic style of explanations provided by Amazon does not provide any direct entry points for interaction: Users might perceive that they cannot influence what "users who bought, also bought". As a consequence, many participants experienced no or little control over their recommendations, although they were aware of explicit interaction options (e.g. in form of expressing a like for a movie).

To make interaction with RS less confusing, more transparent, and controllable, we argue that mental models of RS need to be aligned with the conceptual model, which represents the actual algorithmic functioning. In other words, users need to be educated about how recommendations are derived and what possibilities for interactively controlling them they have. In such a way educated users understand recommendations and their causality better, are able to use the system more effectively, and thus, are more satisfied with it and the resulting recommendations.

5.3 Implications for RS development

Considering *RQ4* ("*What implications for RS design can be derived?*"), we derive four guidelines for the development of RS. While we are aware that these are based on one particular RS, we are confident that they pose valuable anchor points for general RS design.

5.3.1 Link components to existing mental models. To reduce confusion and cognitive complexity, RS developers might rely on our identified basic mental model (Figure 1) to be already present. In particular, we encourage developers of RS to increase transparency by relating components of their system to one or more of the model's four steps. This implies that it might not be necessary to explain each single step of the inner working of RS to users in detail.

5.3.2 Align UI components with recommendation algorithm. We suggest to align explanatory and interactive components with the underlying algorithmic pattern of recommending more precisely and explicitly. Here, especially item- and user-based recommending should be distinguished. Our results indicate that both pertain to diverse mental models and that they were transferred between RS, which caused many false expectations about system behavior. In this sense, prevalent mental models might need to be corrected regarding the system's actual functioning.

5.3.3 Heed the centrality of self. RS developers should emphasize the impact of the users' current preference profile on recommended items. We particularly suggest to link content features between consumed and recommended items, since we observed that the content of items is a paramount expected aspect in the process of recommending (see Section 4.2). This does not mean that the RS has to solely rely on content-based filtering though. There is some research on how to combine collaborative filtering with content data [20, 21, 23], which could be used to make systems based on collaborative filtering more transparent using the content of items. When communicating the relation of preferences and recommendations adequately, it can also be used to exert control over recommendations (see, e.g., [1, 19]).

5.3.4 Enlighten the mystification. A central challenge of making RS more transparent and controllable is to overcome the *mystification* of RS. While this is implicitly also addressed by the guidelines above, we observed that mystification was especially a result of metaphorical mental models. Hence we suggest to introduce standardized and accordingly aligned metaphors that correct or replace existing ones. This could, for instance, be achieved by personifying the RS, e.g. by depicting an anthropomorphic avatar. However, while the depiction of such avatars and the social presence they emit, were observed to improve trust and adoption of recommendations [17, 30], negative emotions may be triggered, e.g. due to *uncanny valley* effects [4]. Thus we deem the design of feasible metaphors for RS as distinctively challenging and emphasize that it requires further research in this topic.

5.4 Limitations

Despite the small size of $N = 10$, we consider our identified concepts as theoretically saturated because we noticed that the concepts of the mental models were very well developed early in the recruitment process. The main limitation of our work is the focus on a very specific sample of *one* single platform, namely regular and experienced Netflix users which most likely has contributed to the early theoretical saturation. Finally, due to the qualitative nature of this study, we cannot make assumptions about the prevalence of the identified mental models.

6 CONCLUSIONS AND FUTURE WORK

Applying a qualitative approach, we found a variety of mental models. Our participants expressed high degrees of uncertainty and confusion about the inner working of Netflix. Nonetheless, we elicited a general structure that all of these models adhered to which can be used for RS development in practice. Furthermore, the concepts of *centrality of the self* and *item- and user-based recommending* can serve as entry points for the design of transparent and controllable RS. Hence, this work contributes not only to the exploration of users' mental models of RS, but also provides insights for RS development in practice.

In future work, we plan to validate our findings through quantitative research. Especially, the general structure represents a solid baseline for hypotheses and confirmatory studies on a large user basis. Here, it might also be interesting to investigate a more diverse user group which differ in the frequency of use and experience with RS. We stress out that it is worthwhile to investigate other RS platforms as our study focused on one single platform. Finally, the aspect of transfer of mental models was a striking result of our study. Transfer of mental models can be important for RS developers as they could rely on this to build the RS. To further investigate the transfer of mental models, we suggest to conduct comparative studies with several examples of RS.

7 ACKNOWLEDGEMENT

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under Grant No. GRK 2167, Research Training Group "User-Centred Social Media".

REFERENCES

[1] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: A Visual Interactive Hybrid Recommender System. In *Proceedings of the Sixth ACM Conference on Recommender Systems (RecSys '12)*. ACM, New York, NY, USA, 35–42. https://doi.org/10.1145/2365952.2365964

[2] André Calero Valdez, Martina Ziefle, and Katrien Verbert. 2016. HCI for Recommender Systems: The Past, the Present and the Future. In *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16)*. ACM, New York, NY, USA, 123–126. https://doi.org/10.1145/2959100.2959158

[3] Shuo Chang, F. Maxwell Harper, and Loren Terveen. 2015. Using Groups of Items for Preference Elicitation in Recommender Systems. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*

(CSCW '15). ACM, New York, NY, USA, 1258–1269. https://doi.org/10.1145/2675133.2675210

[4] Leon Ciechanowski, Aleksandra Przegalinska, Mikolaj Magnuski, and Peter Gloor. 2019. In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems* 92 (March 2019), 539–548. https://doi.org/10.1016/j.future.2018.01.055

[5] Juliet Corbin and Anselm Strauss. 2008. *Basics of Qualitative Research (3rd ed.): Techniques and Procedures for Developing Grounded Theory.* SAGE Publications, Thousand Oaks, California. https://doi.org/10.4135/9781452230153

[6] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (Aug. 2008), 455. https://doi.org/10.1007/s11257-008-9051-3

[7] Will Dahlgreen. 2016. *Streaming wars: the actors Netflix and Amazon customers want to see.* Retrieved January, 15, 2020 from https://yougov.co.uk/topics/politics/articles-reports/2016/01/14/streaming-wars-actors-netflix-and-amazon-customers

[8] Deloitte. 2017. *Welchen Video-on-Demand-Anbieter nutzen Sie?* Retrieved April, 24, 2020 from https://de.statista.com/statistik/daten/studie/443820/umfrage/genutzte-video-on-demand-anbieter-in-deutschland/

[9] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces (IUI '18)*. ACM, New York, NY, USA, 211–223. https://doi.org/10.1145/3172944.3172961

[10] Muheeb Faizan Ghori, Arman Dehpanah, Jonathan Gemmell, Hamed Qahri-Saremi, and Bamshad Mobasher. 2019. Does the User Have A Theory of the Recommender? A Pilot Study. In *Proceedings of Joint Workshop on Interfaces and Human Decision Making for Recommender Systems*. CEUR-WS.org, 77–85.

[11] Carlos A. Gomez-Uribe and Neil Hunt. 2015. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Transactions on Management Information Systems* 6, 4 (Dec. 2015), 13:1–13:19. https://doi.org/10.1145/2843948

[12] Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Franco Turini, Fosca Giannotti, and Dino Pedreschi. 2018. A Survey of Methods for Explaining Black Box Models. *Comput. Surveys* 51, 5, Article Article 93 (Aug. 2018), 42 pages. https://doi.org/10.1145/3236009

[13] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining Collaborative Filtering Recommendations. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work (CSCW '00)*. ACM, 241–250. https://doi.org/10.1145/358916.358995

[14] Christie Kodama, Beth St. Jean, Mega Subramaniam, and Natalie Greene Taylor. 2017. There's a creepy guy on the other end at Google!: engaging middle school students in a drawing activity to elicit their mental models of Google. *Information Retrieval Journal* 20, 5 (Oct. 2017), 403–432. https://doi.org/10.1007/s10791-017-9306-x

[15] Joseph A. Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction* 22, 1-2 (March 2012), 101–123. https://doi.org/10.1007/s11257-011-9112-x

[16] Todd Kulesza, Simone Stumpf, Margaret Burnett, and Irwin Kwan. 2012. Tell Me More?: The Effects of Mental Model Soundness on Personalizing an Intelligent Agent. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1–10. https://doi.org/10.1145/2207676.2207678

[17] Johannes Kunkel, Tim Donkers, Catalin-Mihai Barbu, and Jürgen Ziegler. 2018. Trust-Related Effects of Expertise and Similarity Cues in Human-Generated Recommendations. In *Companion Proceedings of the 23rd International on Intelligent User Interfaces: 2nd Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces (HUMANIZE)*. http://ceur-ws.org/Vol-2068/humanize5.pdf

[18] Johannes Kunkel, Tim Donkers, Lisa Michael, Catalin-Mihai Barbu, and Jürgen Ziegler. 2019. Let Me Explain: Impact of Personal and Impersonal Explanations on Trust in Recommender Systems. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300717

[19] Johannes Kunkel, Benedikt Loepp, and Jürgen Ziegler. 2017. A 3D Item Space Visualization for Presenting and Manipulating User Preferences in Collaborative Filtering. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces (IUI '17)*. ACM, New York, NY, USA, 3–15. https://doi.org/10.1145/3025171.3025189

[20] Johannes Kunkel, Benedikt Loepp, and Jürgen Ziegler. 2018. Understanding Latent Factors Using a GWAP. In *Proceedings of the Late-Breaking Results track part of the Twelfth ACM Conference on Recommender Systems (RecSys'18)*. https://arxiv.org/pdf/1808.10260.pdf

[21] Benedikt Loepp, Tim Donkers, Timm Kleemann, and Jürgen Ziegler. 2019. Interactive recommending with Tag-Enhanced Matrix Factorization (TagMF). *International Journal of Human-Computer Studies* 121 (Jan. 2019), 21–41. https://doi.org/10.1016/j.ijhcs.2018.05.002

[22] Benedikt Loepp, Tim Hussein, and Jürgen Ziegler. 2014. Choice-based preference elicitation for collaborative filtering recommender systems. In *Proceedings of the 32nd International Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3085–3094. https://doi.org/10.1145/2556288.2557069

[23] Julian McAuley and Jure Leskovec. 2013. Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text. In *Proceedings of the 7th ACM Conference on Recommender Systems (RecSys '13)*. ACM, New York, NY, USA, 165–172. https://doi.org/10.1145/2507157.2507163

[24] Sean M. McNee, John Riedl, and Joseph A. Konstan. 2006. Being Accurate is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. ACM, New York, NY, USA, 1097–1101. https://doi.org/10.1145/1125451.1125659

[25] Janice M. Morse. 2015. "Data Were Saturated...". *Qualitative Health Research* 25, 5 (2015), 587–588. https://doi.org/10.1177/1049732315576699

[26] Donald A. Norman. 1983. Some Observations on Mental Models. In *Mental Models*, Dedre Gentner and Albert L. Stevens (Eds.). Psychology Press, New York, NY, USA, 7–14.

[27] Donald A. Norman. 1988. *The design of everyday things.* Basic Books, Inc., New York, NY, USA.

[28] John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, and Tobias Höllerer. 2008. PeerChooser: Visual Interactive Recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 1085–1088. https://doi.org/10.1145/1357054.1357222

[29] Pearl Pu, Li Chen, and Rong Hu. 2011. A User-centric Evaluation Framework for Recommender Systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems (RecSys '11)*. ACM, New York, NY, USA, 157–164. https://doi.org/10.1145/2043932.2043962

[30] Lingyun Qiu and Izak Benbasat. 2009. Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *Journal of Management Information Systems* 25, 4 (Dec. 2009), 145–182. https://doi.org/10.2753/MIS0742-1222250405

[31] Al Mamunur Rashid, Istvan Albert, Dan Cosley, Shyong K. Lam, Sean M. McNee, Joseph A. Konstan, and John Riedl. 2002. Getting to Know You: Learning New User Preferences in Recommender Systems. In *Proceedings of the 7th International Conference on Intelligent User Interfaces (IUI '02)*. ACM, New York, NY, USA, 127–134.

[32] Quentin Roy, Futian Zhang, and Daniel Vogel. 2019. Automation Accuracy Is Good, but High Controllability May Be Better. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 520:1–520:8. https://doi.org/10.1145/3290605.3300750

[33] David E. Rumelhart and Donald A. Norman. 1983. Representation in Memory.

[34] Ben Shneiderman and Pattie Maes. 1997. Direct Manipulation vs. Interface Agents. *interactions* 4, 6 (Nov. 1997), 42–61. https://doi.org/10.1145/267505.267514

[35] Itamar Simonson. 2005. Determinants of Customers' Responses to Customized Offers: Conceptual Framework and Research Propositions. *Journal of Marketing* 69, 1 (Jan. 2005), 32–45. https://doi.org/10.1509/jmkg.69.1.32.55512

[36] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 830–831. https://doi.org/10.1145/506443.506619

[37] E. Isaac Sparling and Shilad Sen. 2011. Rating: how difficult is it?. In *Proceedings of the Fifth ACM Conference on Recommender Systems (RecSys '11)*. ACM, New York, NY, USA, 149–156.

[38] Statista.com. 2018. *Most popular video streaming services in the United States as of July 2018, by monthly average users.* Retrieved January, 15, 2020 from https://www.statista.com/statistics/910875/us-most-popular-video-streaming-services-by-monthly-average-users/s

[39] Anselm Strauss and Juliet Corbin. 1994. Grounded theory methodology. In *Handbook of qualitative research*, Norman K. Denzin and Yvonna S. Lincoln (Eds.). SAGE Publications, Thousand Oaks, CA, USA, 273–285.

[40] Nava Tintarev and Judith Masthoff. 2015. Explaining Recommendations: Design and Evaluation. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, Boston, MA, USA, 353–382. https://doi.org/10.1007/978-1-4899-7637-6_10

[41] Helma Torkamaan, Catalin-Mihai Barbu, and Jürgen Ziegler. 2019. How Can They Know That? A Study of Factors Affecting the Creepiness of Recommendations. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)*. ACM, New York, NY, USA, 423–427. https://doi.org/10.1145/3298689.3346982

[42] Chun-Hua Tsai and Peter Brusilovsky. 2019. Explaining Recommendations in an Interactive Hybrid Social Recommender. In *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19)*. ACM, New York, NY, USA, 391–396. https://doi.org/10.1145/3301275.3302318