"I Really Don't Know What 'Thumbs Up' Means": Algorithmic Experience in Movie Recommender Algorithms

Oscar Alvarado[1,2(✉)], Vero Vanden Abeele[1(✉)], David Geerts[2(✉)], and Katrien Verbert[1(✉)]

[1] Department of Computer Science, KU Leuven, Leuven, Belgium
{oscarluis.alvaradorodriguez, vero.vandenabeele,
katrien.verbert}@kuleuven.be
[2] Mintlab, KU Leuven, Leuven, Belgium
david.geerts@kuleuven.be

Abstract. Many of our daily activities and decisions are driven by algorithms. This is particularly evident in our interactions with contemporary cultural content, where recommender algorithms deal with most of their access, production, and distribution. In this context, the Algorithmic Experience (AX) design framework emerged to guide the design of users' experiences with algorithms for social media platforms. However, thus far, a framework to design specifically for AX within the context of movie recommender algorithms was lacking. To this end, the present study combines a semiotic inspection analysis of the Netflix interface with sensitized design workshops and semi-structured interviews to explore AX requirements for movie recommender algorithms. Linking the analysis with design opportunities, we shed light on AX suited for movie recommender systems. Moreover, we extend the current AX design framework with two new design areas: algorithmic usefulness and algorithmic social practices.

Keywords: Algorithms · Movie recommendations · Algorithmic experience · Recommender systems

1 Introduction

Algorithms are involved in most of our daily activities and decisions [41], becoming publicly relevant [17], enacting power [4, 32] or governance [24, 42]. Additionally, algorithms act as cultural gatekeepers when they sort, rank, manage, distribute and produce existing music [27, 31, 34], movies or tv shows [18, 20], videos [35] and other kinds of cultural expressions. They are even seen as relevant cultural objects on their own [16], promoting the formulation of the algorithmic culture concept [38].

Unfortunately, most algorithms and their decisions possess an inscrutable nature [24], emanated from complex processes and being influenced by uses and constant changes in their inner workings or interfaces [24]. In addition, general low algorithmic awareness among users [14] may also produce negative experiences, as invisibility,

anxiety, and inequalities [6], bias in the personalization processes [7] and possible human interventions [10].

Different researchers are trying to overcome these issues and enable more positive interactions with algorithms. For instance, Hamilton et al. [21] invite research on the design of algorithmic interfaces, balancing user needs for transparency with the advantages of automatic implementations. Diakopoulos presents a call for algorithmic accountability and propose an algorithmic transparency standard [11]. Other academics highlight the importance of a human-centered design of algorithmic systems [3, 28] or adhering to a design framework for *algorithmic experience* (AX) [1, 33] in the area of social media platforms.

Simultaneously, there are efforts to increase transparency and trust of recommender systems and, specifically, to support better user control with such recommender algorithms [2, 5, 9, 15, 26, 36, 40]. Although several researchers have presented different user interfaces to address the black-box nature of recommender systems, this research is to date still rather ad hoc [23]. A specific visualization is presented and shown to improve user trust and acceptance, but the generalizability of the results is limited. In this paper, we present a framework to support such generalizability with a framework for Algorithmic Experience (AX) of movie recommender systems.

Centering our study on the Netflix movie recommender algorithm, we applied three main methods to develop this framework. First, we analyzed the Netflix user interface to unearth the intentions of the designer towards the algorithm. Second, we performed sensitizing workshops to elicit AX requirements among Netflix users. Finally, we conducted follow-up semi-structured interviews to expand the AX requirements elicitation.

Building on the AX framework [1] for social media, we adapt the framework for movie recommender algorithms by expanding it with two new design areas: algorithmic usefulness and algorithmic social practices. This specialized framework enriches the present debate on AX and recommender algorithms, enables refined design guidelines, and promotes positive user experiences with movie recommender algorithms.

2 Background

Different academic approaches in the study of algorithms provide inspiration and insights for an AX definition of movie recommender algorithms.

2.1 Algorithms, Audiences and Cultural Content

Academics portray the relevance of algorithms in the cultural context. For instance, Striphas defines algorithmic culture as "the unfolding of human thought, conduct, organization and expression into the logic of big data and large-scale computation" [38]. Additionally, Morris states that recommender algorithms frame the interaction between cultural goods and those who encounter them [31], impacting culture management [31].

Furthermore, algorithms define cultural audiences. Gillespie argues that "trending" algorithms produce specific algorithmically identified audiences based on profiles [16], even becoming sources of cultural concern. Similarly, Prey explains how personalized media enact a sense of looking for distinct predilections of users, but "there are in fact no individuals, but only ways of seeing people as individuals" [34]. According to the author, these platforms represent individuals only by their data, defining a constantly modulated and never conclusive algorithmic identity [34]. In the end, these technologies are reducing the individual to their behavioral feedback cues on the platform [34].

Other researchers study different methods to measure algorithmic decisions on cultural products. For example, Rieder, Matamoros-Fernández, and Coromina proposed *ranking cultures* to determine the algorithm's intentions towards the cultural content [35]. They observed how YouTube's results are not only based on popularity, but also on vernaculars such as the video issue date and its own definition of novel videos.

Academics have also described the Netflix recommender algorithm and its cultural implications. Gomez-Uribe and Hunt describe the Netflix recommender engine as the key pillar [18] for its movie service. By data gathering and personalization techniques, Netflix allows the existence of niche audiences that are too small and almost impossible to exist in other impersonalized contexts [18]. They also express that personalization promotes better results from the recommender system, and increases overall engagement with the platform [18]. In a different context, Hallinan and Striphas emphasize the importance of studying the context that influences the design, development and social consequences of movie recommender algorithms [20], by analyzing results of a contest called the "Netflix Prize" proposed by the same company.

In general, these studies reflect on algorithmic culture, algorithmic effects on cultural products, audience creation, algorithms development, and how algorithms transform our current consumption and production of cultural products.

2.2 The Relevance of, and Bad Experiences with, Recommender Algorithms

Different researchers have aimed at unpacking the many algorithmic implications for users and societies. Even if this work is not specifically addressing recommender engines, it portrays various similarities with movie recommender algorithms.

Willson and Beer suggest particular attention for those algorithms in which we are delegating everyday activities, working semi-autonomously with no supervision from human counterparts [4, 41]. Additionally, Gillespie defines *public relevance algorithms* as those delimited by six provisional functions: selecting or excluding information products, inferring or anticipating information about their users, defining what is relevant or legitimating knowledge, flaunting impartiality with no human mediation, and provoking behavioral changes in users practices [17].

Research has also documented negative user experiences with recommender algorithms. Bozdag describes the layers of bias in algorithmic filtering and personalization [7]. Additionally, Bishop reports YouTubers' anxiety and inequalities with the

platform's algorithm [6]. These examples highlight the relevance of addressing AX with recommender systems.

2.3   Transparency, Human-Centered Algorithms and Algorithmic Experience

Answers to these previous challenges come from different perspectives. For example, Diakopoulos presents an *algorithmic transparency standard* for media-related algorithms based on five categories that might be considered for disclosing [11]. First, human involvement in the decisions of algorithms should be explained including the purpose of the algorithm and possible automated or human editorial goals. Second, the collected data must be described in terms of its quality, accuracy, uncertainty, timeliness, representativeness, including its definition, collection, and edition. Third, the algorithmic input data should be transparent, including its model and modeling process. Fourth, the inferences made by the algorithm must be clear, including their potential for errors. Fifth, the algorithmic presence should be clear, whether it is used at all, and if personalization is in use, promoting user awareness.

Furthermore, researchers highlight the importance of including users in algorithmic development. Baumer proposes "human-centered algorithm design" to bring together "algorithmic systems and the social interpretations thereof" [3]. Also, Lee, Kim, and Lizarondo describe a human-centered implementation of an algorithmic service [28].

Other academics turn towards the experience with these systems. Alvarado and Waern propose Algorithmic Experience (AX) as a conceptualization of "the ways in which users experience systems and interfaces that are heavily influenced by algorithmic behavior" [1]. They identify five design areas to improve AX in social media platforms [1]. First, *algorithmic profiling transparency* is described as a design opportunity to promote user perception on what the algorithm is tracking to create personalized results. Second, *algorithmic profiling management* is described as the design opportunity to manage the user's algorithmic profiling. Third, *selective algorithmic remembering* is identified as the design opportunity to allow the user to avoid future algorithmic results based on previous and no longer relevant algorithmic profiling. Fourth, *algorithmic user control* describes the design opportunity to regulate how and when the algorithm is going to produce and show its results. Finally, *algorithmic awareness* is described as the design opportunity to promote understanding of how the algorithm works and measures user behavior. Similarly, Oh et al. picture a new way on HCI research, based on Algorithmic Experience "as a new stream of research on user experience" [33] that considers constant relationships with algorithms.

2.4   Interaction Design for Recommender Systems

Extensive work has focused on designing the interaction experience with recommender systems. For example, Knijnenburg et al. present a framework to evaluate recommender systems with a user-centric approach [26]. Additionally, Jugovac and Jannach review the state of the art on user interaction with these systems [25], presenting strategies for preference elicitation and alternatives for interactive recommendations.

The "black box" nature of recommenders has also been studied with different tactics. For instance, He, Parra and Verbert survey interaction strategies in recommender systems and group them in six groups [23]: transparency, justification, controllability, diversity, cold start phase, and context. Also, Gedikli, Jannach, and Ge compare different types of explanations for recommenders algorithms [15], while Tintarev and Masthoff evaluate seven different goals for explanations in recommender systems [39, 40]: transparency, scrutability, trust, effectiveness, persuasiveness, efficiency and satisfaction.

Bakalov et al. advise five aspects to evaluate user models and personalization effects in recommender systems [2]: usefulness, ease of use and learning, satisfaction, trust, and user modeling. Also, Cramer et al. explore eight aspects to evaluate an art recommender system [9]: perceived transparency, competence, usefulness and need for explanations in the system, understanding, intent to use, acceptance, and ease of use.

In general, previous proposals bring efforts to evaluate recommender algorithms in terms of transparency and explanations after their implementation. However, there is still no framework that could provide suggestions for the human-centered design of movie recommender algorithms based on user experiences, or a specialized AX framework for movie recommender systems.

3   Methods and Results

Studying algorithmic experience (AX) with human-centered approaches constitutes a challenging endeavor, due to low algorithmic awareness among users [14]. Therefore, this study uses a mixed-method approach to understand the AX of movie recommendations in Netflix. First, we applied a self-sensitizing technique to understand the intentions of the interface designer towards the recommender algorithm interface using Semiotic Engineering [37]. Second, we held sensitizing workshops to elicit AX requirements from Netflix users. Finally, we conducted individual follow-up semi-structured interviews to explore complementary aspects of Netflix AX, based on recommender systems interface design research.

3.1   Study 1: Semiotic Inspection

The Semiotic Engineering Process (SEP) is a scientific HCI methodology derived from semiotics and communication theory [37]. It offers a method for analyzing interfaces and designers goals called the Semiotic Inspection Method (SIM) [37]. SIM recognizes interfaces as a communication process between designers and users, exposing the former's intentions behind the design. In contrast with other heuristic methods, SIM is not directed by strict usability principles and is not centered on the user's experience.

SIM consists of five stages which were applied for this paper by the main author. The first three stages allow to iteratively analyze the static, dynamic and metalinguistic signs [37] embedded in the interface of the system. While *static signs* can be interpreted at a single moment in time without temporal and causal relations, *dynamic signs* emerge only through the interaction with the interface, containing both a temporal and causal context [37]. The intrinsic relation between dynamic and static signs also

produce meanings in *metalinguistic signs* to communicate a specific message to the user [37]. The fourth stage compares all the signs collected in the previous steps, to find the designer's meta-communication message or the designer's final goal with the interface [37]. Finally, the fifth stage evaluates the system's communicability, revealing relationships and (in)consistencies between the designer goals and the interface [37].

The outcome of the method is an analysis of the communication strategy of the system and the proposed message to the user. In this paper, we applied the method to the analysis of the user interface in the Netflix recommendation system.

However, because SIM is limited in identifying absences (i.e. signs that should be present but are not) in an interface, we complemented the method by adding specific requirements for algorithms from two studies. First, we included the five categories from the *algorithmic transparency standard* [11] which serves as a design framework for accountability and transparency in media-related algorithms: explaining human involvement, describing data collection, providing limits to the algorithmic model, clarifying made inferences and promoting awareness of the algorithm. Second, we complemented this with two purposes for interactive visualizations in recommender systems proposed by He, Parra and Verbert: diversity and cold start [23]. The other four goals for interactive visualization in recommender algorithms were excluded from the analysis either because they overlapped with the categories from Diakopoulos [11], or because they were not relevant for the Netflix case.

SIM was applied using the Netflix desktop platform in English, using a 27″ screen, browsing with Mozilla Firefox explorer version 60.0.2 during June 25th and 26th, 2018.

Semiotic Inspection Results. Static signs were mostly content containers. Nowadays, static signs in web platforms are mostly wireframes or dedicated spaces with mutable contents. Also, certain areas were identified as static signs but contained dynamic contents. For example, when logging into Netflix, a prominently featured show is initially shown, as pictured in Fig. 1. This space presents a background video with other signs, such as two buttons for playing the show or adding it to the user's list, respectively.

Fig. 1. Prominent featured content on Netflix's landing page.

The top area of the interface contains two features: sorting content by category and the user list with manually saved shows. Other static signs in this area are a "search icon" and the user's image (avatar). Scrolling down shows another static title "Netflix

Originals" and a horizontal list composed of different images of shows. When scrolling down, smaller horizontal lists are found, composed of images with their correspondent titles. After the lists "Trending now", "Continue Watching for [user]" and "Watch It Again", the names of these lists relate to specific reasons for why these items are proposed such as "Because you liked [content]".

Dynamic signs are mostly encountered in the constantly changing movies inside the platform every time the user logs in, usually in the recommendation lists and in the initially prominently featured show. For example, the featured show changes its title and background video according to the show or movie being promoted.

The smaller horizontal lists also change their movie background images or white titles dynamically, depending on the content category. Additionally, as shown in Fig. 2, these small images possess a dynamic feature: hovering on them presents varied texts and buttons such as a "play" button, the show's title and description, a "Match" green text next to a percentage scoring, the show's age classification, the available number of seasons for that show, the "thumbs up/down" buttons, a "+" button to add that show to the user's list, and a "down arrow". As Fig. 3 shows, a click on this "down arrow" expands the show to cover the entire screen width and provides further details about the show. Clicking on "more like this" opens a new horizontal list of recommendations containing similar shows. There is no sign or indication that explains the selection of these recommendations or the inner logic for them, except for the initial name of the show from which these recommendations are generated.
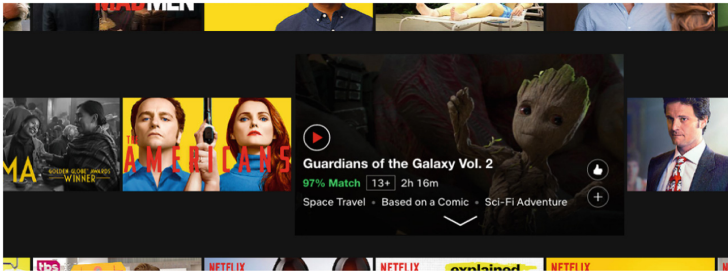

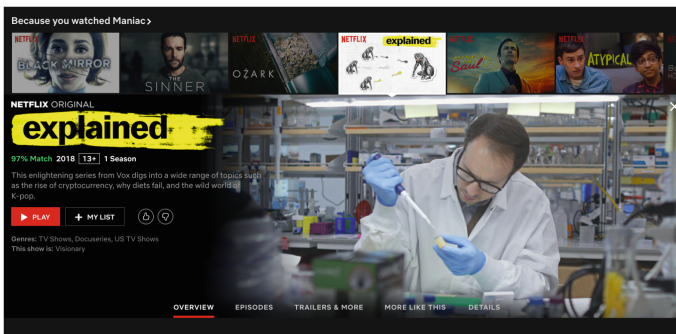**Fig. 2.** Dynamic feature while hovering on a movie.

Fig. 3. The Netflix interface, dynamically expanding for details about a show.

Metalinguistic signs in the Netflix interface are mostly directed towards the voting system. There is a clear design intention in the voting system to define "thumbs up/thumbs down" icons as positive and negative feedback, respectively. Furthermore, voting seems to mean that a user has already watched that show from the perspective of the designer since all the scored shows appear later in a "Watch It Again" list. This is not confirmed by any other signs, leading to confusion because voting could also be based on past consumption outside the platform or by following peers' suggestions.

Similarly, it is not clear if there is a meta-linguistic sign in the user's list. Adding elements in this list possibly influences the recommender algorithm or it just organizes user's content, but again this is not confirmed by any sign.

The fourth phase of the method showed the general intention of the interface. After iterating and comparing static, dynamic, and metalinguistic signs, it is possible to determine the meta-communication in relation with the recommender algorithm: to promote movie watching in a fast and easy way, guided mostly by the recommendations. There is no design intention to give the user control of the recommender system besides the "thumbs up/thumbs down" buttons. Finally, the fifth phase does not picture any inconsistencies or relations with the defined meta-communication strategy.

Additionally, the complementary requirements were used to analyze Netflix's interface. When using Diakopoulos' framework for algorithmic transparency [11], no indication of possible direct human involvement on the recommendations can be found. Moreover, besides explicit signs such as recommendation lists with texts like: "Because you liked [content]", there are no signs about *the data collection process, data transformation processes, the algorithm's model, the inferences made by the system*, or any reference about the user categorization. Finally, there is no clear sign to delimit where the algorithm is presenting its results, or a space "free" from the algorithm influence.

When using the diversity and cold start goals for interactive recommenders defined by He, Parra and Verbert [23], there is no sign for diversity in the recommender system, echoing He et al.'s finding: only one surveyed recommender system included features for representing recommendation *diversity* [23]. However, Netflix's cold start solution does show signs in relation to the recommender system. In this case, the system shows an interface as in Fig. 4 which consist of several signs: (1) a metalinguistic sign that consists of "thumbs up" icons to refer to those contents being selected, (2) static signs such as texts inviting the user to select three contents he/she likes or a text explaining how this selection will help the system to find better recommendations as well as a button to proceed and a layer of pictures, and (3) a dynamic number indicating the amount of selections ("4" in Fig. 4).
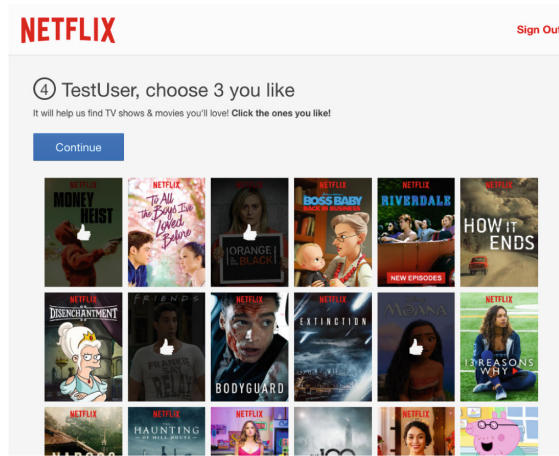
Fig. 4. The Netflix interface after registration, addressing the "cold start" phase.

3.2   Study 2: Sensitizing Workshops

This method gathers algorithmic experience (AX) requirements for the recommender algorithm in Netflix, based on the experiences of the users with the platform. Active Netflix user recruitment was done using the university departmental mailing lists and Facebook/WhatsApp groups around the city. A Netflix gift card was offered during recruitment and raffled in every workshop to encourage user participation.

Since algorithmic awareness is generally low among users [14], every workshop started with a priming tutorial about AX and algorithms in known platforms. Afterwards, participants contributed to a group discussion with their perceptions about AX of Netflix recommendations. Participants were also invited to log in with their Netflix accounts using laptops, being able to use and browse their accounts during the entire workshop. The discussion was guided by semi-structured questions derived from Alvarado and Waern groups of design opportunities for AX [1: 6], but other perspectives were also welcomed. Every workshop was recorded for further analysis.

Five different sessions were organized in total, with 15 active Netflix users between 18 and 35 years old, all undertaking at least a master's program and tech-savvy. To maintain user's anonymity, they will be referred with a number as an identifier. More information about the workshop (represented by letters from A to E) they participated in and their gender (represented by F or M) is detailed in Table 1.

Table 1. Participants for sensitizing workshops and follow up interviews.

Participant	Netflix experience (E) (M = months; Y = years)	Workshop	Sex
P1	6M < E < 1Y	A	M
P2	1Y < E < 2Y	A	M
P3	1M < E < 6M	A	M
P4	1Y < E < 2Y	A	F
P5	3Y < E < 5Y	B	M
P6	1Y < E < 2Y	B	F
P7	3Y < E < 5Y	C	M
P8	5Y < E < 10Y	C	M
P9	1Y < E < 2Y	D	M
P10	1Y < E < 2Y	D	M
P11	5Y < E < 10Y	D	M
P12	2Y < E < 3Y	E	F
P13	1M < E < 6M	E	M
P14	5Y < E < 10Y	E	M
P15	3Y < E < 5Y	E	F

Sensitizing Workshop Results. Coding was based on the five categories from the AX framework for social media including, profiling transparency and management, algorithmic awareness and control, and selective algorithmic memory. This was complemented with new codes for results that did not fit inside the previous categories.

Algorithmic Profiling Transparency. Netflix uses a green matching score to support transparency. However, this was not properly perceived by many participants. P5, P6, P8, and P14 reported not noticing it before the workshop. Similarly, P10 did not comprehend the meaning of the matching score. In contrast, P3 and P13 noted how obvious Netflix' profiling activity is: "I can see my past in these [recommendations]".

Specific information was detailed by users for algorithmic profiling transparency. For example, P4 said she would like to know how Netflix justifies a specific match score. Likewise, P3 wanted to check past behavior to understand his recommendations.

Participants also expressed trust issues. P4 did not trust the algorithm and wondered whether popularity is a reason for recommendations. Similarly, P5 and P6 desired to understand the recommendation reasoning. Likewise, P8, P13, and P15 reported obscurity in the recommender algorithm due to the lack of explanations in the data collection.

Again, some suggestions were made in this area. For instance, P14 and P15 requested to know their preferences according to their algorithmic user profile. Moreover, P12 said that phrases like "Because you have watched this…" are approximations to explain the algorithmic profiling, but she expected more detailed information.

Algorithmic Profiling Management. This area was exemplified by P4, P6, and P10 who expressed a need to "tune up" the algorithm: "to say that you like specific actors or a genre". Furthermore, P1 desired an option to help him practice a language: an option

to delimit dubbed content characteristics for his profile. Similarly, P7, P8, P13, and P14 wanted options to avoid some contents: "I do not like it, I do not dislike it, just ignore this". Also, P7 and P9 preferred options to avoid already watched movies.

In relation to algorithmic profiling management, the profile interaction also seemed unclear to the participants. For example, P11 expressed: "You really don't know what 'thumbs up' means, or if adding content to the list changes anything". Likewise, P11 and P13 said the "thumbs up/down" options are very limited to manage their profiling. Similarly, P13 and P14 declared they did not know when to properly use the "thumbs up/down". Interestingly, P5, P6, P12, and P14 reported not using this feature.

Algorithmic Awareness. Users expressed that they were not aware of an "algorithm-free" space. For P6, the entire platform was the recommender. Likewise, P4 said it would increase her trust in the system if she knew where exactly the algorithmic influence was present inside the platform. Following AX interrelation of its design areas [1:6–8], P13 expressed that knowing where the algorithm has its influence should be part of the transparency "package" for any of these platforms.

Algorithmic User Control. P8 expressed to have an "explorative mode" and the interface did not help with it: "you cannot skip the categories they predefined for you". This encouraged "hacking" or tricks to discover new content. This was echoed by P15, who exposed that certain web sites offer category IDs to find content: "I do not understand why they do not allow you to reach all the content they have", describing a need for a space free from the algorithmic filtering. Also, P12, P13, and P15 expressed the need of "turning off" the algorithm for being able to "choose something different".

In relation, users reported that manual searching was their most common way to "bypass" the algorithm. Despite this practice, P4 described this manual search feature as limited for the user because she had to know the name of the desired show beforehand, which made her feel manipulated. "I usually want to have freedom of choice, but here it feels they want to say which way to go". A solution expressed by P6 and P15 to avoid the algorithm was an alphabetical or chronological sorting feature for the movies.

Some users expressed that there is no way to "turn off" the algorithm. P3 and P5 said that even sorting by genres showed again recommendations. P3 and P4 suggested a blank landing screen to promote content exploration and reduce the "imposed" feeling.

Other users expressed a need to avoid the algorithm only at specific moments. For instance, P11 and P12 wanted to "stop" the algorithm for precise periods of time, to "ignore" shows that were not truly what they liked and avoid them in the future.

Algorithmic user control was also expressed in the need for a way to define what the recommender algorithm should consider. This is illustrated by P3, P4, and P5 who wanted a dashboard to "turn off" certain algorithm inputs.

Similarly, P8, P9, P10, and P14 wanted to tell the system their current "watching mood" to adjust the recommendations. While P8 felt that Netflix saw him as static: "but I change constantly", P7 said the platform "pushed" him to an inert profile.

Selective Algorithmic Remembering. P3, P5, P6, P11, P13, and P15 agreed on wanting a feature to "erase" previously watched contents. Also, they desired to delete other people activities when they share their accounts to improve the recommendations.

Other Results. Other results did not fit in the current AX framework. P4 described unhelpfully how the system was "making guesses" during the *cold start phase*, offering contents that she did not know before or that would never be part of her preferences. Similarly, P3 declared that his recommendations were terribly limited during this stage.

Available content seemed also relevant for the experience. For instance, P14 felt that recommendations worked better in the US, where the platform offers more content.

It seems also that the interests of Netflix as a content producer influence the experience. P3, P13, P14, and P15 reported *impressions of dishonesty* since the system mostly offered Netflix's original content and "kept pushing those titles". Also, P6 and P11 noticed the interface tended to locate Netflix Originals with high matching scores.

Interestingly, AX seems to be affected by a common practice among users: *sharing their own accounts* with other people even though it is possible to create separate profiles in the platform. Most users like P2, P6, P11, P12, and P15 agreed that this sharing "messed up" the recommendation algorithm and its results. Similarly, users reported following peers or "real-life" recommendations rather than the recommender algorithm. P2, P6, P8, and P14 mentioned they usually do not use the recommendation engine but recommendations of friends instead. In relation, P5 and P6 desired to add a "*social perspective*" for the recommender, with features such as sending recommendations to friends or following trusted users. Also, P5, P6, P7, P10, P13, P14, and P15 agreed on using third-party recommendations and scorings such as Rotten Tomatoes, IMDB, or similar sources. For instance, P5 would like a comparison between these sources and the "Match" score in the interface.

3.3   Study 3: Follow-Up Semi-structured Interviews

We complemented the previous AX elicitation with semi-structured follow-up interviews to reinforce users' impressions about Netflix AX and to gather more results, relying on theories related to recommender systems experience.

Three theories were used for this purpose. Firstly, we used Bakalov et al.'s five aspects to evaluate user models and personalization effects in recommender systems [2]: usefulness, ease of use and learning, satisfaction, and trust. Secondly, the interviews were also inspired by three concepts from Cramer et al. [10: 473], for evaluating trust and acceptance in a content-based art recommender system: understanding of the system, acceptance of the system, and perceived need of explanations. Thirdly, Tintarev and Masthoff [40] four explanatory aims for recommenders were included for the interviews: effectiveness, persuasiveness, efficiency, and satisfaction.

Every interview was recorded for further analysis. An iterative process was implemented to code and later organize the results, described in the following section.

Follow-Up Semi-structured Interviews Results. Ten participants from all previous workshop were recruited again for the follow-up interviews, two of them female. The same participant identifiers previously used are reapplied in this section to maintain anonymity and to show their relationship with previous results.

Usefulness. Users described recommender usefulness as closely related to other key concepts such as satisfactory results, better and fast decisions, enjoying the algorithm and the system's knowledge about user preference. For instance, P3 expressed that the

algorithm knew what he wanted to watch but was not currently giving satisfactory results: "right now it prolongs the decision and makes me go somewhere else". On the other hand, when he tried to look for a show manually, he enjoyed the offering of "similar" recommendations. Similarly, the algorithm was only occasionally useful for P10: "…sometimes they surprise you, they give you happy accidents and that is a good enjoyable feeling". He expressed that the recommender did not know what he wanted to watch: "but they try to give the best guess". Likewise, P11 and P14 reported that their recommendations were arbitrary and not truly what they preferred: "It is just offering popular stuff". P11 reported spending too much time browsing around with no useful decision. Also, P15 said she did not get satisfactory results with the recommendation system and reported not noticing the matching score at all before the workshop.

Trust. This area was mostly related to Netflix' commercial interests and previous transparency results. P2 described the recommendations as "an honest guess", sometimes better than recommendations of friends. In contrast, most participants gave negative comments. For instance, P5 expressed that he did not trust or enjoy the algorithm and preferred to remove it. He did not trust the algorithm because most recommendations had a high matching score, they were not related to his preferences or the quality of the show, and because they promoted too much their own content. Similarly, P10 did not understand completely what the percentage meant, which affected his trust negatively. P11 as well preferred to know what data Netflix was collecting and to have a way to avoid recommendations to improve his trust in the system. P15, P6, and P8 agreed on not trusting the system because Netflix own contents had "more weight" in their interfaces. In a similar vein, P3 expressed that they "push" too much Netflix Originals.

Ease of Use and Learning. This category was ambiguous among users. Encouraging comments were expressed by P3 who said that his recommendations were easy to browse. Similarly, P2 said that sometimes people do not even know they are using the recommender system because of its ease of use. Likewise, P10 described the "thumbs up/down" as easy to use, but he did not know whether the user's list influenced the system. On the other hand, negative comments were voiced by P6 who described the "thumbs up/down" options as too binary to properly manage the recommender. Moreover, P15 and P13 reported the match percentage as hard to perceive. P15 also mentioned that the specific recommendations titles such as "Because you watched [show]" were hard to understand according to the inner logic for those recommendations.

User Control. P2 needed more structure in the recommendation's organization, while P15 preferred a distribution based on self-defined categories: "I'm not that committed to spend hours in the platform to improve the recommender system". Similarly, P8 desired to select between the shows' time duration and their number of seasons. Moreover, P10, P15, P6, and P8 agreed on suggesting a feature to ask the user his or her current mood and filter the recommendations based on what they felt like.

Content. P2 and P3 believed that having more content would improve the chances for satisfactory recommendations and faster decisions. Additionally, P15 and P6 said reducing the number of recommendations in the interface could speed up user decisions.

Transparency. Some contradictions were voiced the transparency of the recommender system. A positive comment was expressed by P2 when he described the system as self-explanatory with texts as "because you watched this…". On the other hand, P3, P14, P8, and P5 agreed on not having a clue on what parameters the recommender used and the weight it gave to them. Similarly, P6 and P5 preferred to know what metrics were used for the suggestions. Furthermore, P16 said that it was not helpful to know how the recommender works without being able to control it. In relation, P14 said increasing transparency would promote accountability of the recommender system and understanding on how to "guide" the algorithm for better results. Also, P13 said comprehending the match percentage would improve the ease of use and enjoyment. Also, P11 said that more transparency would rise the usefulness of the recommender and produce faster decisions.

Explanations. P3 and P13 felt they needed a feature that appears only when requested: "it needs to be there just in case I want to know the information". Likewise, P3 did not want to know how the mathematics work, but "two or three sentences" expressing a general answer. In relation, P11 desired a quantifiable explanation: for example, how many people have watched the show or how many people liked it. Finally, P15 said: "At least say something about what is the input for the algorithm".

External Sources. P15 and P6 want to include a "social" variable to the algorithm. In relation, P6 relied more on friends' inputs because the suggestions of the system were not effective. Moreover, P11, P13, and P5 suggested adding scores from Rotten Tomatoes or similar services. Additionally, P13 believed that reviews from other people could provide valuable information and increase his trust in the recommender system.

4   Algorithmic Experience for Movie Recommender Algorithms

The five groups of design opportunities for Algorithmic Experience (AX) [1] in social media, were echoed in our study case. Hence, this AX framework is also found valid for movie recommender algorithms. However, we extend the framework to include two new AX opportunities not considered before: *algorithmic social practices* and *algorithmic usefulness*, as shown in Fig. 5.
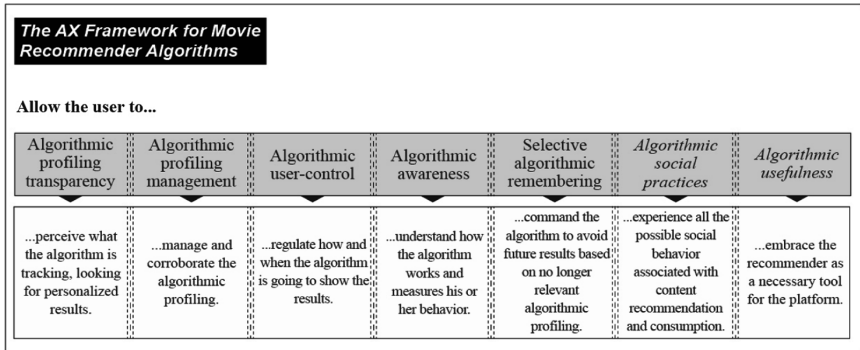
Fig. 5. Design areas for algorithmic experience in movie recommender algorithms.

This framework can also guide the requirement elicitation for other movie recommendation systems (e.g. Amazon Prime, Hulu, others). Approximations could be done outside the movie application domain (e.g. YouTube or Twitch), but they require specific studies to determine whether their streaming context possesses similar requirements.

4.1   Algorithmic Profiling Transparency

Transparency remains a relevant requirement for AX in movie recommender algorithms. This design area for AX is associated with showing clearly the profile created by the algorithm to achieve personalized recommendations. Possible improvements in this category could be defined by the algorithmic transparency standard [12]. Improvements include showing explicitly possible human involvement both in the recommendations or the user profile, explaining clearly the data collection process, model and inferences done by the algorithm. Additionally, it can be useful to explain where exactly the algorithm influence is included in the interface, both related to collecting data from the user or providing recommendations. Finally, features to check viewing history or inspect preferred user categories according to the algorithm could also improve this profiling transparency.

4.2   Algorithmic Profiling Management

During workshops and interviews, users expressed a need to manage and corroborate the preferences gathered by the recommender to promote or avoid specific types of content at specific moments. This management can be included in the user's profile [1].

A common related feature associated with the algorithmic profile management is the "cold start" phase. This phase defines the initial stage in which the recommender engine does not know enough about a user to provide effective recommendations. This concern is addressed by recommender systems designers through different strategies [23] and was also articulated during both the workshops and the SIM analysis as a new aspect that could be included in the area of *algorithmic profiling management.*

Therefore, the "cold start" phase seems to be a first opportunity in the interaction with a recommender system to offer appropriate algorithmic profiling management. Besides a simple selection of movies, opportunities could be to allow users to choose between predefined profiles for specific categories or their mixture or let them pick friends, influencers or groups with common preferences.

Algorithmic profile management is also related to negative privacy experiences. Ambiguous impressions from users suggest issues with profile management that can negatively affect the AX, depending on the user's attitude towards data collection and processing. Therefore, opportunities should be available to control/erase behavioral information when desired, in line with legislation [19].

Finally, algorithmic profile management was also discussed in relation to the options of the interface to interact with the profiling mechanisms. For example, the strategy of Netflix is delimited by the "thumbs up/down" buttons and possibly by adding movies to the user's list, features that are misunderstood and too limited according to the workshop results and the SIM analysis. Possible solutions in this area could be developing more detailed user controls for the recommendations.

4.3   Algorithmic Awareness

Algorithmic awareness requirements were also voiced during the workshops. First, there should be a clear distinction between algorithmically generated recommendations and recommendations that are simply self-promotional, an issue constantly expressed and related to Netflix's commercial interests. Second, users also reported a low understanding of the "match percentage" and the "thumbs up/down" buttons.

Third, ambiguous opinions about the platform's ease of use and learning were also articulated by the participants. These comments were mostly based upon having so many recommendations with high scores, which were found untrustworthy and negatively evaluated. Moreover, interaction with the recommender should promote algorithmic awareness. For example, implications for future recommendations using features as "Thumbs up/down" or the "User list" should be clearly stated in the interface, including when they are supposed to be used (before or after watching a show).

Understanding how the algorithm works improves algorithmic awareness, trust, and transparency and could be improved via direct explanations. Users asked for these explanations as a "second layer" option or not always directly visible.

4.4   Algorithmic User Control

Opportunities for user control were also mentioned by users. First, users need a way to communicate their current "mood" to the recommender, such as "explorative" to avoid algorithmically defined recommendations or to promote recommendations not directly related with their personal profile or preferences. Second, this explorative mode could be also related to the user's desire for an "algorithm free" space, or a way to "turn off" the recommender engine. Opportunities in this direction could be sorting the content alphabetically, by categories, by year of production, and other similar features. This functionality was described by users as an opportunity to reduce algorithm hacking or

"looking for tricks" to find alternative content in the platform. Third, users also mentioned a need to "turn off" the recommender data collection during specific periods of time or just to "stop" it so a specific movie could be avoided for future recommendations. This feature could be particularly useful when users share their account with relatives or friends and do not desire that activity to influence their own recommendations.

In relation to this area of AX, users also expressed the need to indicate faulty algorithmic recommendations with the "thumbs down" option. Also, users desired to provide explanations for the "downvoting" to detail why the recommender algorithm should avoid similar content in the future.

4.5 Selective Algorithmic Remembering

Regarding algorithmic remembering, users said they would like to make the algorithm "forget" previous specific activities to avoid related future recommendations. Again, this feature could be helpful when users share their accounts with other people to avoid future irrelevant content or just to curate their viewing history.

It may be helpful to make a distinction between algorithmic profiling management and selective algorithmic remembering. While the latter opens an opportunity to manage recommendations based on complete movie categories such as comedies, westerns, and others, the former aims to delete specific shows or movies that have been watched in the past. In this case, the user looks for a need to refine the recommendations with specific contents, rather than entire categories.

4.6 Algorithmic Usefulness

This category allows the user to embrace the recommender as a necessary tool for the platform and "enjoy using it". An initial and obvious opportunity in this area is to provide an effective recommender algorithm that could predict users' preferences as accurately as possible. Bad recommendations will affect negatively the AX. Closely related, users also expressed that the algorithm should not only recommend "guaranteed bets" or popular shows. Instead, it could take "certain calculated risks" to offer alternative/diverse contents outside of the mainstream general consumption or usual preference of the user. In this context, a recommender algorithm that only promotes own content is considered bad AX, untrustworthy, turning the feature into a "disposable" tool.

Another finding in this area was discovered during the workshops and interviews: the amount and diversity of the available content for the recommendations affects the AX. Less movies will not only create a negative experience with the recommendations but will also reduce the probabilities for appropriate algorithmically generated suggestions. In relation, when a user manually searches for a specific name of a show or movie, it seems to be a good idea to at least recommend "similar" contents when that specific content is not available in the platform, as Netflix does already.

Finally, other results were closely related with previous studies that portrayed some opportunities in this area [2, 39, 40]. For example, a positive AX in movie recommender algorithms is promoted by appropriate knowledge about the user preferences,

satisfactory results, better and faster decisions, produce enjoyment, express appropriate knowledge about the user preferences, producing enjoyment and persuading the user to choose the algorithm recommendations because of their usefulness.

4.7   Algorithmic Social Practices

This design opportunity allows the user to experience the possible social behaviors associated with content consumption, such as the recurrent habit of sharing a personal account. Even though Netflix offers many accounts in a single subscription, users continuously expressed that they share the same account with different people, which in the end negatively influences their AX. In relation, a solution supported by the users is that the algorithm should offer to "stop tracking" momentarily when a movie does not correspond to their personal preference. Moreover, users expressed a need to be able to erase previous viewing activity to "tune up" their recommendations after sharing their accounts. Another solution is to consider multiple users in front of the screen and offer a "mixed accounts" mode, providing reciprocal recommendations to improve AX.

Other design opportunities in this area were based on external recommendation sources. Users continuously mentioned the habit of following specialized platforms and friends (or "influencers") to guide their viewing preferences. Therefore, they desire to add a "social factor" to the recommender algorithm in which friends could share movie suggestions or follow trust-worthy accounts to mimic watching preferences. Furthermore, it was mentioned to add "third party" grading systems such as Rotten Tomatoes or IMDB to compare and complement their grades with the recommender matching score. Finally, users want to check other users' reviews on movies and shows.

5   Examples to Improve AX of Movie Recommender Systems

A way to improve *algorithmic profiling transparency* is generating a "profile view" showing which inputs the algorithm is currently considering for recommendations as exemplified by a previous study [30]. Likewise, a solution for *algorithmic profiling management* is proposed by an interface that offers to adjust recommendations related to current preferences and user models [2]. Solutions for the "cold start" are described by a previous study that showed significant improvements by offering groups of movies rather than single movie selection [8]. Also, this phase could use representations of movie communities, watching trends, or other "social" options [23]. Additionally, a previous study showed improvements in *user control* by letting users choose movies based on recency and popularity [22] as a way to "turn off" the algorithm. Finally, improving accuracy for *algorithmic usefulness* is not enough [29]. For instance, a significant portion of users prefers an option that allows to choose between different algorithmic strategies for movie recommendations [13].

6   Limitations

The study was performed in Belgium, inferring implications on user perceptions on their recommendations due to regionally available content. Also, the user group was dominated by males, tech-savvy and highly educated users, which could imply possible bias among the results and uncertainty of the same results among other groups.

7   Conclusion

This study explored the algorithmic experience (AX) of recommender algorithms using Netflix. It applied a mixed approach for the analysis: first, a semiotic inspection method to study the designer's intention towards the recommender system; second, a sensitized workshop to elicit AX-based requirements; and third, follow-up interviews to collect more AX requirements based in recommender systems design theories.

From the analysis, we propose a specialized AX framework for movie recommender algorithms with seven design opportunities: algorithmic profiling transparency, algorithmic management, algorithmic awareness, algorithmic user-control and selective algorithmic remembering, algorithmic usefulness and algorithmic social practices.

This new specialized AX framework for movie recommender algorithms contributes to a focused approach for designing these systems. Future research could measure if the framework already includes all the necessary aspects for a positive AX in this context and the objective implications on user experience when they are implemented.

References

1. Alvarado, O., Waern, A.: Towards algorithmic experience. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI 2018, pp. 1–9. ACM Press, Montreal (2018)
2. Bakalov, F., et al.: An approach to controlling user models and personalization effects in recommender systems. In: Proceedings of 2013 International Conference on Intelligent User Interfaces - IUI 2013, pp. 49–56 (2013)
3. Baumer, E.P.: Toward human-centered algorithm design. Big Data Soc. **4**, 1–12 (2017)
4. Beer, D.: The social power of algorithms. Inf. Commun. Soc. **20**(1), 1–13 (2017)
5. Berkovsky, S., Taib, R., Conway, D.: How to recommend ? User trust factors in movie recommender systems. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces, IUI 2017, pp. 287–300 (2017)
6. Bishop, S.: Anxiety, panic and self-optimization. Converg. Int. J. Res. New Media Technol. **24**(1), 69–84 (2018)
7. Bozdag, E.: Bias in algorithmic filtering and personalization. Ethics Inf. Technol. **15**(3), 209–227 (2013)

8. Chang, S., Harper, F.M., Terveen, L.: Using groups of items to bootstrap new users in recommender systems. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW 2015, pp. 1258–1269. ACM Press, New York (2015)
9. Cramer, H., et al.: The effects of transparency on trust in and acceptance of a content-based art recommender. User Model. User-Adapt. Interact. **18**(5), 455–496 (2008)
10. Cramer, H., Thom, J.: Not-so-autonomous, very human decisions in machine learning: questions when designing for ML the human side of machine learning. In: AAAI Spring Symposium on Designing the User Experience of Machine Learning Systems, pp. 412–414 (2017)
11. Diakopoulos, N.: Accountability in algorithmic decision making. Commun. ACM **59**(2), 56–62 (2016)
12. Diakopoulos, N., Koliska, M.: Algorithmic transparency in the news media. Digit. J. **5**(7), 809–828 (2017)
13. Ekstrand, M.D., Kluver, D., Harper, F.M., Konstan, J.A.: Letting users choose recommender algorithms. In: Proceedings of the 9th ACM Conference on Recommender Systems - RecSys 2015, pp. 11–18. ACM Press, New York (2015)
14. Eslami, M., et al.: I always assumed that I wasn't really that close to [her]. In: Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems - CHI 2015, Seoul, pp. 153–162 (2015)
15. Gedikli, F., Jannach, D., Ge, M.: How should i explain? A comparison of different explanation types for recommender systems. Int. J. Hum Comput Stud. **72**(4), 367–382 (2014)
16. Gillespie, T.: #trendingistrending: when algorithms become culture. In: Algorithmic Cultures Essays Meaning, Performance New Technologies, vol. 189, pp. 52–75 (2016)
17. Gillespie, T.: The relevance of algorithms. In: Media Technologies Essays on Communication, Materiality, and Society, pp. 167–194 (2014)
18. Gomez-Uribe, C.A., Hunt, N.: The Netflix recommender system. ACM Trans. Manag. Inf. Syst. **6**(4), 1–19 (2015)
19. Goodman, B., Flaxman, S.: EU regulations on algorithmic decision-making and a "right to explanation". In: 2016 ICML Workshop on Human Interpretability in Machine Learning, WHI 2016, pp. 1–9 (2016)
20. Hallinan, B., Striphas, T.: Recommended for you: the Netflix prize and the production of algorithmic culture. New Media Soc. **18**(1), 117–137 (2016)
21. Hamilton, K., Karahalios, K., Sandvig, C., Eslami, M.: A path to understanding the effects of algorithm awareness. In: Proceedings of the Extended Abstracts of ACM Conference on Human Factors in Computing Systems - CHI EA 2014, pp. 631–642 (2014)
22. Harper, F.M., Xu, F., Kaur, H., Condiff, K., Chang, S., Terveen, L.: Putting users in control of their recommendations. In: Proceedings of the 9th ACM Conference on Recommender Systems - RecSys 2015, pp. 3–10. ACM Press, New York (2015)
23. He, C., Parra, D., Verbert, K.: Interactive recommender systems: a survey of the state of the art and future research challenges and opportunities. Expert Syst. Appl. An Int. J. **56**(C), 9–27 (2016)
24. Introna, L.D.: Algorithms, governance, and governmentality: on governing academic writing. Sci. Technol. Hum. Values **41**, 1 (2016)
25. Jugovac, M., Jannach, D.: Interacting with recommenders – overview and research directions. ACM Trans. Interact. Intell. Syst. **7**(3), 1–46 (2017)
26. Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. User Model. User-Adapt. Interact. **22**(4–5), 441–504 (2012)

27. Kulesza, T., Stumpf, S., Burnett, M., Kwan, I.: Tell me more? The effects of mental model soundness on personalizing an intelligent agent. In: Proceedings of 2012 ACM Annual Conference on Human Factors in Computing Systems - CHI 2012 (2012)

28. Lee, M.K., Kim, J.T., Lizarondo, L.: A human-centered approach to algorithmic services: considerations for fair and motivating smart community service management that allocates donations to non-profit organizations. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI 2017, Denver, pp. 3365–3376 (2017)

29. McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems - CHI EA 2006, pp. 1097–1101. ACM Press, New York (2006)

30. Millecamp, M., Htun, N.N., Jin, Y., Verbert, K.: Controlling spotify recommendations: effects of personal characteristics on music recommender user interfaces. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization - UMAP 2018, pp. 101–109 (2018)

31. Morris, J.W.: Curation by code: infomediaries and the data mining of taste. Eur. J. Cult. Stud. **18**(4–5), 446–463 (2015)

32. Neyland, D., Möllers, N.: Algorithmic IF … THEN rules and the conditions and consequences of power. Inf. Commun. Soc. **20**(1), 45–62 (2017)

33. Oh, C., Lee, T., Kim, Y., Park, S., Kwon, S., Suh, B.: Us vs. them: understanding artificial intelligence technophobia over the Google deepmind challenge match. In: Conference on Human Factors in Computing Systems - CHI 2017, pp. 2523–2534 (2017)

34. Prey, R.: Nothing personal: algorithmic individuation on music streaming platforms. Media Cult. Soc. **40**, 1086–1100 (2017). https://doi.org/10.1177/0163443717745147

35. Rieder, B., Matamoros-Fernández, A., Coromina, Ò.: From ranking algorithms to 'ranking cultures': investigating the modulation of visibility in YouTube search results. Convergence **24**(1), 50–68 (2018)

36. Sinha, R., Swearingen, K.: The role of transparency in recommender systems. In: CHI 2002 Extended Abstracts on Human Factors in Computing Systems - CHI 2002, pp. 830–831 (2002)

37. De Souza, C.S., Leitão, C.F.: Semiotic Engineering Methods for Scientific Research in HCI (2009)

38. Striphas, T.: Algorithmic culture. Eur. J. Cult. Stud. **18**(4–5), 395–412 (2015)

39. Tintarev, N., Masthoff, J.: A survey of explanations in recommender systems. In: Proceedings - International Conference on Data Engineering, pp. 801–810 (2007)

40. Tintarev, N., Masthoff, J.: Evaluating the effectiveness of explanations for recommender systems: methodological issues and empirical studies on the impact of personalization. User Model. User-Adapt. Interact. **22**(4–5), 399–439 (2012)

41. Willson, M.: Algorithms (and the) everyday. Inf. Commun. Soc. **20**(1), 137–150 (2017)

42. Ziewitz, M.: Governing algorithms: myth, mess, and methods. Sci. Technol. Hum. Values **41**(1), 3–16 (2016)