Designing for the Better by Taking Users into Account: A Qualitative Evaluation of User Control Mechanisms in (News) Recommender Systems

Jaron Harambam
j.harambam@uva.nl
Institute for Information Law, University of Amsterdam
Amsterdam, The Netherlands

Dimitrios Bountouridis
d.bountouridis@tudelft.nl
Delft University of Technology
Delft, The Netherlands

Mykola Makhortykh
m.makhortykh@uva.nl
Amsterdam School of Communications Research, University of Amsterdam
Amsterdam, The Netherlands

Joris van Hoboken
j.v.j.vanhoboken@uva.nl
Institute for Information Law, University of Amsterdam / Law, Science, Technology, and Society (LSTS), Vrije Universiteit Brussels
Amsterdam / Brussels, The Netherlands / Belgium

ABSTRACT

Recommender systems (RS) are on the rise in many domains. While they offer great promises, they also raise concerns: lack of transparency, reduction of diversity, little to no user control. In this paper, we align with the normative turn in computer science which scrutinizes the ethical and societal implications of RS. We focus and elaborate on the concept of user control because that mitigates multiple problems at once. Taking the news industry as our domain, we conducted four focus groups, or moderated think-aloud sessions, with Dutch news readers (N=21) to systematically study how people evaluate different control mechanisms (at the input, process, and output phase) in a News Recommender Prototype (NRP). While these mechanisms are sometimes met with distrust about the actual control they offer, we found that an intelligible user profile (including reading history and flexible preferences settings), coupled with possibilities to influence the recommendation algorithms is highly valued, especially when these control mechanisms can be operated in relation to achieving personal goals. By bringing (future) users' perspectives to the fore, this paper contributes to a richer understanding of why and how to design for user control in recommender systems.

CCS CONCEPTS

• **Information systems** → Recommender Systems; Personalization; User Interfaces;User Centered Design; • **User/Machine Systems** → Human factors; • **Human Centered Computing** → Interaction paradigms.

KEYWORDS

recommender systems; filter bubble, transparency, diversity, accuracy, user control, ethics, qualitative research, user study, self-actualization; user centered design

ACM Reference Format:
Jaron Harambam, Dimitrios Bountouridis, Mykola Makhortykh, and Joris van Hoboken. 2019. Designing for the Better by Taking Users into Account: A Qualitative Evaluation of User Control Mechanisms in (News) Recommender Systems. In *Thirteenth ACM Conference on Recommender Systems (RecSys '19), September 16–20, 2019, Copenhagen, Denmark.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3298689.3347014

1 INTRODUCTION

Algorithmic recommender systems (RS) play an important role in contemporary societies. In domains as diverse as commerce, leisure and news, RS are deployed to help consumers deal with an overload of content by providing personalized suggestions. Also, RS help content providers to increase user engagement and thus boost revenues [41]. RS have thus great promise, but also raise concerns. The most prevalent one is that RS create "filter bubbles" [37], isolating people from a diversity of contents and viewpoints [13, 19]. While this concern plays a pivotal role in public debates, current research is much more ambivalent about the restrictive effects of RS [14], and all obviously depends on how they are optimized. Filter bubble concerns notwithstanding, there is a clear lack of transparency about the workings and objectives of RS [4, 45, 47], they generally allow for little accountability [9], and offer users few ways to direct or correct RS [15].

These concerns do not stand alone, but are situated in broader public discussions about the ethical and societal implications of the rise of algorithms and Big Data, and appropriate legal and policy frameworks to govern them [3, 27, 36, 49]. Increasingly, issues of fairness, accountability, transparency, and ethics in socio-technical systems become part of computer science research questions and communities (FAT*/FATREC conferences). In this paper, we align with that normative turn by exploring the concept of user control in RS, defined here as having *direct effects* on recommendations.

RecSys '19, September 16–20, 2019, Copenhagen, Denmark

ACM ISBN 978-1-4503-6243-6/19/09…$15.00
https://doi.org/10.1145/3298689.3347014

We contend that enhancing user control mitigates multiple problems at once, and adds to the individual and societal value of RS. Besides empowering users to make RS more responsive to their interests, needs and aspirations [11, 29], it simultaneously requires RS to be (more) transparent and explainable [17, 21]. This makes users not only more satisfied [16, 38], and trustful of RS [11], they are also activated to explore beyond their known interests [5], which increases the diversity of recommended contents [19], and lessens algorithmically induced blind spots, or filter bubbles [15]. It could even be that the mere possibility to control RS will already take away most concerns. In short, enhancing user control can help improve many issues associated with RS.

Although existing empirical studies of user control in RS [17, 21] are a good base to build on, they generally do not offer a systematic analysis of a wider range of control mechanisms. Moreover, since they are quantitative in design, they focus less on the meanings, interpretations and complexities of users. However, when thinking about why and how to design for more user control in RS, it is important to better understand such concerns, motivations and aspirations in all their nuance [11]. In this study, we therefore set out to *qualitatively* answer the following research question: **how do people evaluate different control mechanisms in news RS, which ones do they prefer, and why?** We focus on the news context because the ethical and societal ramifications of RS are most salient here: being (unknowingly) excluded from some types of shoes is different from some forms of information, as having a diverse and broad range of contents and viewpoints is considered necessary to participate in democracy [18].

We conducted four focus groups, or moderated, think-aloud sessions, with Dutch news readers ($N = 21$) in which we collectively reviewed and interacted with different control mechanisms in a News Recommender Prototype (NRP). We distinguish three phases of user control (input; recommender algorithms; the output)[17, 21], and built three distinct control mechanisms accordingly. We systematically asked users about their opinions of these mechanisms. Such qualitative focus groups are geared towards bringing diverse positions to the foreground in much empirical detail, which future quantitative studies can test across broader populations [35]. By showing contemporary concerns and desires, this approach makes it possible for scholars, designers and policy makers to delve into the lifeworlds of users/citizens and incorporate that in future products and regulations [11]. This paper contributes as such to a richer understanding of why and how to design for user control in RS and offers concrete ways to mitigate the concerns associated with them.

2 RELATED WORK

In the development of RS, most (academic) attention has been given to improving their technical performance [24, 41]. By now it has become accepted that the quality of RS go beyond accuracy [25, 34] and that designers should include metrics like diversity [5], novelty [6], context [1], and serendipity [31]. Successful RS should furthermore take into consideration factors like transparency and explainability in order to secure societal value and trust [47]. Integral to this shifting emphasis from "algorithms to experience" [30] is the central role users should have in (the design of) RS [28, 40]

Three comprehensive reviews on the current state of the art of interactive RS show how scholars and practitioners have tried to give users more control possibilities in research and real-life settings [17, 21, 24](For the SOTA of news recommenders, see [26]). They identify three different phases in the recommendation process where interaction/control could happen: **input** (or preference elicitation), **process** (algorithmic computation), **output** (presentation of the results). While these are not exhaustive of all control mechanisms, they do form a good analytical start for this study.

Input Most RS use implicit user feedback to infer people's preferences, but one could also let people themselves indicate preferences on (static/relative) user profile forms [21]. In the context of news, that could take the form of (un)following different news categories/topics. While this approach can be less accurate and prone to indicating desirable instead of actual preferences, relying on behavioral data *only* has drawbacks as well [11]. The most important ones being that online user behavior is heavily determined by the design choices made, and that it is unable to capture future or aspirational interests. Moreover, people's preferences are are highly complex, contextual, and sometimes internally contradictory [10, 20, 48]. Giving people the (extra) option to indicate their preferences satisfies therefore multiple goals, beyond assumed accuracy.

Process Users can control RS at the process level either by adjusting the algorithm parameters/weights [2, 10, 16] and/or by choosing between different types of algorithms [10, 15]. Instead of the common imperative to build one best-fit algorithm, this form of user control assumes that certain algorithms are better suited to certain tasks. A good RS would thus "have a family of algorithms at its disposal to select from" [34, p.6] so that users can deploy RS to their own (continuously changing) needs and aspirations.

Output This control mechanism offers ways to order the recommended items in ways that befit the user and her interest/needs at each specific moment [2, 42]. The benefit is that the effects are immediately noticeable: with one click users have a different list of recommendations. Some post-filtering functionalities are simple and straightforward, but others can become more visually complex and offer advanced ordering options [21, 23]

A known challenge is to "finding the sweet spot" between offering "rich user control while ensuring acceptable cognitive load"[23, p.13]. As observed in other domains, it will only be some tech-savvy few who might have the patience, capabilities and curiosity to adjust settings, but history has shown that most will not spend much time tinkering with them. That is the paradox of choice: the more to choose from, the less we actually seem to do. However, much depends on how such control mechanisms are designed and aligned with user expectations [30, 34]. This insight informs our study to qualitatively explore which control mechanisms are desired by (future) users and why.

3 METHODOLOGY

3.1 Research Design: Focus Groups

Following a user-centric approach, we held focus groups with news readers to understand how they evaluate (different) control mechanisms in RS [35]. Although user studies are increasingly used in RS research, qualitative designs are still limited [11, 33]. With our qualitative study, we add complexity and nuance to the literature,

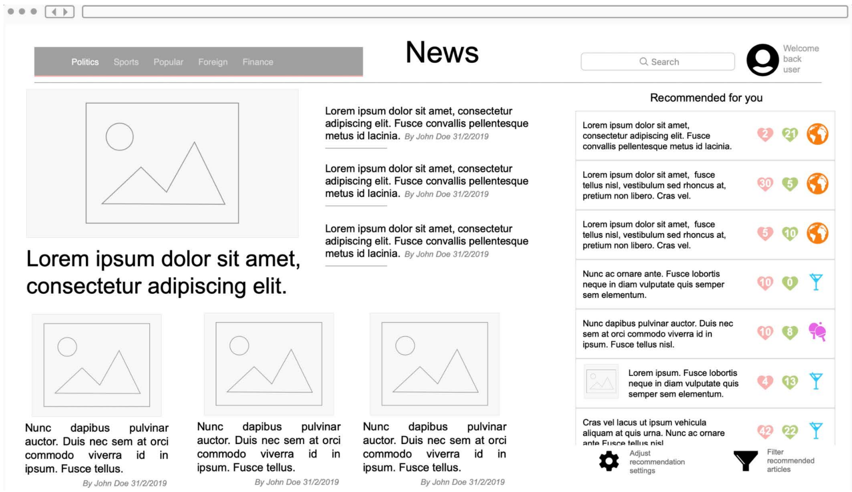

Figure 1: A screenshot of the news website prototype used for our study. The left section contains fixed headlines, while the right one contains a list-view of personalized recommendations and a serties of related control mechanisms.

and incorporate insightful ideas given by participants [43]. Our goal is, as a first step, to foreground distinct opinions and concerns. Quantitative studies can build forth on this work and measure their distribution across relevant populations.

While RS may mostly be used solitary, we decided to hold focus groups to let participants be confronted with different positions, reflect about that, and spur as such the discussion. We recruited participants through the news reader panel of the media organization with whom we collaborate (De Persgroep) and through our university student panel. There may be self-selection at work: some participants clearly had affinity with the topic, either in positive or negative terms. This is, however, not problematic. Following qualitative research strategies, generalization is not the goal, instead we aim for diversity to make distinct positions most clear [44]. Participants (N = 21; 4-6 per group) were equally divided in gender, age ranged from 20 through 65. We conducted the focus groups in February 2019, participants received a 10 euro gift card.

During the focus groups, we collectively reviewed and interacted with our NRP. We explained and showed on a big screen the workings of each control mechanism and systematically asked them about their opinions regarding *understandability*, *ease of use*, *usefulness*, *sense of control*, and *willingness to use*. This means that the focus groups were conducted in a semi-structured way: each control mechanism was discussed along the same script, however, there was enough opportunity for us and the participants to diverge from it and go deeper or into completely new issues. The questions asked were simple and straightforward to open a free discussion, e.g. “do you understand what this mechanism does?”, “does this give you a sense of control over RS?”, etc. As is common in qualitative studies, we put effort to let participants explain their positions in more detail, asked them to reflect on those of others, and to avoid group thinking by confronting them with opposed (hypothetical) positions when necessary. The focus groups were audio recorded, anonymously transcribed and inductively analyzed making use of the qualitative data analysis software `Atlas.Ti` [7].

Obviously, speaking about control mechanisms is different from actually using them. The fact that we could only showcase the control mechanism with a dummy interface (due to limited resources) is a drawback of our study, and may impact its external validity. However, as we will explain hereafter, the NRP did function realistically: each adjustment in the control mechanisms had direct effects on the recommendations. Moreover, as control mechanisms in RS are relatively underdeveloped, it is important to consider people and their opinions, even without actually using them, in order to better align their development. Future studies can test whether our findings replicate when participants use control mechanisms in real life RS.

3.2 News Recommender Prototype

We developed a NRP using the commercial prototyping service `proto.io.` due to its simplicity and ability to render a realistic image of such control mechanisms. Our NRP (see Figure 1) was designed to resemble the layout of a typical, modern news website like the of one of *The New York Times* [32]. The homepage is divided into two sections: a traditional editorial ordering of news articles on the left, and a "recommended for you" section on the right. The recommendation section consists of a list design and three distinct control mechanisms. The recommendations list is scrollable and contains nine items. All contents (articles) are filled with *Lorem ipsum* placeholder text because we want participants to focus on the control mechanisms and not be influenced by actual news contents. We will now discuss the design of the recommended items and the controls mechanisms.

Items Each article is represented by a title and three features $\{n_s, n_d, t\}$: n_s the amount of similar users (by reading history) that have read it, n_d the amount of dissimilar users that have read it, and t the main topic of the article $\in \{politics, sports, entertainment\}$, represented by three respective symbols (globe, two table tennis

Figure 2: The three pop-ups allowing for different control mechanisms: the input at the top left, the process at the bottom and the output at the top right.

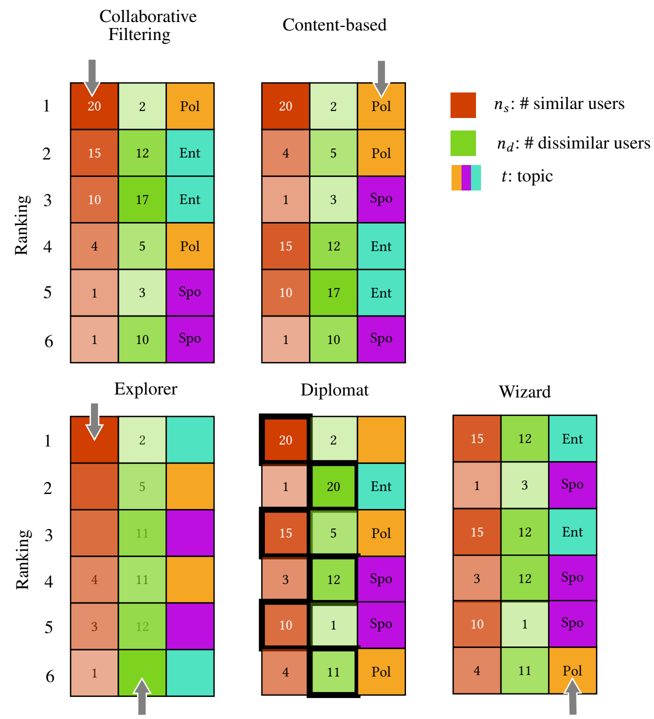

Figure 3: Examples of rankings of article (represented as triplets $\{n_s, n_d, t\}$) using five different recommender algorithms as provided by our prototype. Arrows and bold lines identify the main features for each algorithm. For this example, similar to the prototype, we assume that the user's reading history is focused on politics.

bats, cocktail glass). These features should be sufficient to understand how altering the control mechanisms influences the recommended articles.

Control mechanisms As explained before, our interface provides separate control mechanisms for three different phases in the recommendation process: data input, process and output (see Figure 2). Each has their own button on the screen, and when clicked on, a specific pop-up screen emerges which allows the user to adjust each respective settings. *All* adjustments have immediate effects, and result in a new list of recommended articles.

The "Welcome back user" button (see Figure 1) corresponds to the **input** control mechanism. When clicked on, this pop-up screen allows the user to *a)* examine their summarized reading history in the form of topic-distribution (in our prototype a reading history of mostly politics), and *b)* control their topic preferences by adjusting the importance of each topic (via sliders and on/off switches).

The "Adjust recommendation settings" button corresponds to the **process**, or the algorithmic computation phase. When clicked on, this pop-up screen allows the user to select a recommendation algorithm from a list of six options: the first three correspond to the popular collaborative-filtering, content-based and random algorithms [10]. The last three, correspond to three distinct *types* of recommendation algorithms, which are anthropomorphized as corresponding personae in order to provide an intuitive understanding of their functioning [15]. The *Explorer* offers "news from unexplored territories", is inspired by the notion of diversity, helps readers to expand their horizon, and generates a list of recommended items ordered by n_d (high to low) and reversed-ordered by n_s (low to high). The *Diplomat* offers 'news from the other ideological side", is inspired by the notion of intellectual diplomacy, helps people to understand their ideological counterparts, and generates a list of recommendations such that the $2i^{th}$ items are ordered by n_s (high to low), while the $(2i+1)^{th}$ items ordered by n_d (high to low), where $i \in \mathbb{W}$. The *Wizard* offers "surprising news", inspired by the notion of serendipity, generates a random order of items.

The "Filter recommended articles" button corresponds to the **output** phase. When clicked on, this pop-up screen allows the user to re-order the *already established* list of recommendations. Sliders allow the user to adjust the importance of each topic, or by similar/dissimilar users. Figure 3 presents examples of rankings for the aforementioned algorithms (excluding random) similar to the ones used in the NRP.

4 RESULTS

4.1 Recommender Systems: curse and blessing

The emergence of RS in the news domain is seen as an interesting and useful development by our participants. However, filter bubble and personal data abuse concerns, aggravated by the Facebook/Cambridge Analytica scandals, taint their ideas about these technologies. Although these concerns were not the topic of our focus groups, it was hard to keep these out of the discussions.

Interestingly, our own efforts to explore ways for people to exercise more control over recommender systems were sometimes met with suspicion and distrust. Talking about this project, one participant said that "even if these [control mechanisms] are implemented, it would only give you an illusion of transparency and control, like, 'hey, look at what you're reading and what influence you have!'. But behind it, there is a business model which is all about gathering as much personal data as possible, connecting them with advertisers and making more money"(R1, FG2). Most participants were less critical and argued that "it all matters how it is being used, it is a curse and blessing" (R2, FG2), and "if it is a good company, then they would want to help their customers, otherwise they will walk away and go to another" (R3, FG2). This observation shows how trust in RS relates not only to the technical system itself, but that it operates in a societal context: public debates and the image of the company are important factors as well.

A second general observation relates to the perceived usefulness of having RS in the news context at all. Filter bubble fears, or feelings of going to miss out on important or interesting news due to personalization, were very present: "that's a bit my fear yeah, that I will miss something. I am open for something completely outside of my ordinary reading orbit, but what could interest me nevertheless" (R1, FG4). Other people also questioned whether algorithmic curation does any good to us, actively thinking human beings, "I wonder whether this does not make us lazy, that you have someone else filter for you" (R3, FG1). More specifically, some respondents wondered whether they would actually use enhanced control mechanisms: "I wouldn't make such an effort to change all such settings, since a news app, or a website, is already structured and organized with headings, columns and sections that I don't mind scrolling down to see what I find interesting, without it being suggested to me" (R4, FG2).

The value of RS to select relevant items from an abundance of content was often brought up: we were frequently asked whether the NRP is for one specific news outlet only, or for more outlets combined (e.g. a news aggregator). When asked about their willingness to use control mechanisms in news recommenders they explicitly stated that "it all depends on the quantity of content that I can retrieve, if the quantity is very big, and if it can help me make better selections, then I would" (R1, FG4).

However, people's willingness to use control mechanisms also has to do with how they are designed and where they are featured on a website/app. As this respondent emphasized: "yes, I would like to use those [mechanisms], I find it handy, but how often do I want to use those? If it is not often presented in your face, would you then tinker with those? I don't think so. So the art is: how to confront users in a non-annoying way with the possibilities of what they can do by changing their recommender settings?" (R1, FG4). A fellow respondent confirmed: "if you can directly click on stuff, then I would use it, but if I have to open a menu every time, then it is one click too far" (R5, FG4). A solution is offered: "if your 'recommended for you' box is also the cockpit for your settings, then people would use it" (R1, FG4).

These quotes all show the well-known difficulties RS designers face: they have to design control mechanisms in such a way that they will be easy to use, have clear advantages for the user, and should be smartly brought under their attention. Otherwise, they will not be used, just as privacy statements are rarely consulted. In order to shed more light on these difficult issues, the following section goes deeper into the three distinct phases where users can be offered enhanced control mechanisms.

4.2 Input control mechanisms

Most *input* of RS is basically people's online reading history, but that overview is hardly shown to users. Sometimes they can indicate their preferences, but whether and how these are actually included as input for the recommender system is generally unclear. In our NRP we showed people's user profile (aggregated reading history across three broad categories) and provided sliders for people to indicate their preferences. At first sight, people found this an easy way to exercise control, the sliders are quite straightforward, most said, although some people would find it "more clear if there would be percentages connected" (R3, FG1); if "there would be indicators above it, like in a questionnaire, from completely off to completely on" (R2, FG3); or if we would choose instead for "a pie chart, so you can better see how the whole is divided" (R1, FG2).

However, when talking longer about this control mechanism, more questions arose. Respondents asked about how RS operators create the user profile: "do they track me as well when I am not registered" (R4, FG1), and "is that based on my reading history, or also on other data?" (R3, FG3). Another participant confirms: "are they using data from Facebook, Twitter and all that gang? I would like to know, it would surprise me after all. So it would be good if they are clear about this" (R2, FG3). Data collection concerns are very present, and people express a desire and need for clear explanations about how their personal data is collected and used.

Obtaining an intelligible overview of one's reading history is much appreciated by people. For example, to learn more about oneself: "because I don't really know what I click on, how the distributions between categories are, if there's a bias in what I am reading. I would like to see that" (R2, FG4). But also "to help yourself not to get lost in gossip and entertainment" (R3, FG1). Another participant argued, "I find it refreshing actually, to see how *they* see me. I mean, you get something back of what you did. You give in fact information to the news provider, so it's nice to get it back, it's your information after all"(R1, FG4). Ekstrand and Willemsen highlight the importance of *reciprocity* in user-technological system relations [11], which is clearly emphasized by our participants. Moreover, creating a useful dashboard for users with their reading history is an excellent pro-active example of how to meet the new European General Data Protection Regulation (GDPR) requirements regarding the right to receive information [12].

Such a dashboard is not just informative, participants emphasized how they can use this information (with aligning control mechanisms) to improve themselves, to realize longer term goals. As this participant said: "what I like about it is that you see your manifest behavior, but next to it, you have your ambitions, 'I should read more about art', well now you can express that, and change the settings, I really like that" (R1, FG4). Another participant agreed and added: "I would like to see how I change over time, like every three months, to see, 'oh hey I moved that way, my interests have shifted, or I am very stuck in this bubble, perhaps I should actively do something, such sliders are a great way to do so then" (R2, FG4). These comments show that designing RS for self-actualization [29] is not just a nice thought, but much appreciated by people. These findings corroborate with other efforts to connect explanations of user profiles with the epistemic goals of users [46]

Questions were also raised about the relationship between the user profile and the sliders indicating current preferences: "what is not clear to me, is what it means if you move them [sliders], would you say that you can influence your history with these sliders?" (R2, FG3). Or another: "so this is one's reading history, and "politics" is very high, but what happens when I put that "politics" slider all the way down, will it still remember that I read a lot of politics?" (R2, FG2). Indeed, a fellow focus group participant added, "I would expect that the sliders are how you want it from now on. I would find it very weird if your reading history is still taken along, that they are implicitly still connected" (R1, FG3). The relationship between the user profile and control mechanisms is essential, but that was in our NRP not that clear. Some of the participants therefore suggested to have people's user profile and preference integrated: "for me it would be most clear if the sliders start with how it is up there [reading history], and that you can adjust it from there, now I want more or less of this" (R3, FG2). The direct effects of adjusting preferences on the recommendations is in that way indeed much more clear. Otherwise, there is more need for explanations.

The problem of content categorization was an additional prominent issue. For this prototype we used three broad news categories (politics, sports and entertainment) for exemplary purposes only, but that aroused controversy and debate. Obviously, people found those three too limited, "these are really too crude, too general" (R1, FG3), and expressed their need for more "specific categories, the more specific it gets, the more useful it will be for me" (R2, FG3). But as another respondent argued, "the number of categories should be limited, otherwise you say, 'oh, too many sliders, never mind!'. It is the art to find the right balance" (R1, FG4). Indeed, it is a recurrent problem in User Experience (UX) design to find the right balance between more categories and ease of navigation.

Interestingly, people argued that they should have control over the categories they wish to deploy in their interface, "I would like this thing to be more personalized actually. I don't know much about arts and culture, so that would be a good container category for me, but with science and technology I would like to specify more. I think this is really personal" (R2, FG4). At the same time, the whole process of constructing one's own interest list before starting was not appreciated by everyone, "too much hassle" (R1, FG1). Suggestions were made instead to "exclude certain categories, to filter them away" (R4, FG2), or to select categories post-hoc: "after reading an article that you liked, for example, that you can indicate, this topic I find interesting, and I would like to get more of that" (R3, FG3). Categorization is never easy, and always runs into troubles, or so it seems.

To conclude, showing and adjusting one's preferences to get different recommendations is generally seen as an important way to control RS. It offers ways to make better use of the extraction powers of RS, and when coupled with a cogent and clear user profile, these control mechanism can be deployed for specific information needs and to realize longer term goals.

4.3 Process control mechanisms

When evaluating possibilities to control the recommendation process by choosing from different pre-configured types of algorithms participants reacted rather enthusiastically:"ohhhh, but this is really fun" (R3, FG2), "one-click only, so this is much easier" (R2, FG1), and "do I understand it well that the results directly change when you click on it? What a nice way to exercise influence over a big bunch of articles" (R1, FG3). The idea that one can deploy different recommender algorithms is seen as novel and exciting, yet easy to grasp and relate to. Here are some exemplary reactions: "haha, yeah, comfortably in your own bubble" (R4, FG2)(about collaborative filtering); "I find it really interesting that I won't end up with tunnel vision, but have options to find articles written from different points of view" (R2, FG1)(about recommender personae); "I would found it a wonderful surprise if that Wizard would suddenly come with an excellent review of an interesting political movie, simply because I am interested in politics" (R4, FG3). Participants quickly developed (high) expectations about what this control feature can do.

But while people found this control mechanism "easy... simply because you press on one button, and don't need to tweak multiple settings", others found them "less transparent" (R2, FG1) as well. The functioning of the different algorithms gave rise to many questions. About collaborative filtering, participants questioned what "similar readers are, are that people who search for the same contents? Or are those people with the same income, education, sex, neighbourhood? On the basis of what data are we similar? " (R3, FG3). Or about the personae: "I understand what it is says, but at the same time there's much that I can't see. How are things computed? I mean I have no idea what it sees as my 'horizon' or beyond?" (R1, FG1). And: "this is really nice, but what is that, the diplomat? Opposing perspectives, but how do they know what my political leaning is?" (R2, FG3). Or: "emotionally, I keep having the feeling that someone else decided for me what I am seeing. As if I am not in control. They decide for me what is 'broad' news" (R2, FG1). Participants argued that these constructs need to be very well yet cogently explained.

The contrast between the first three descriptive and the three anthropomorphised algorithms spurred divergent opinions. Some participants "really liked the intuitive, the human like [personae], they are more inviting than those technical ones" (R3, FG3), and said that "their names are very catchy, have something open minded, they don't give me the feeling of being put in a box, but as a broadly interested citizens instead, it has something cool" (R2, FG1). They found the personae "nicely imagined, very inviting, I would use those" (R4, FG3) and "much more interesting, they are really challenging me" (R1, FG4). Others said the complete opposite: "I really don't need that anthropormorphism, I understand what content based means, with the diplomat or wizard, I need more explanation" (R2, FG3).

These divergent opinions about antropomorphized algorithms translate into different evaluations of this control mechanism. Some find "especially the personae vague, because I understand the technology behind the other ones, but when it says 'surprises' or 'gradually expand your horizon', then I am not sure what happens" (R2, FG2). For others it "is the complete opposite, I will be honest, I am not good with computers, so the [personae] are much more clear:

'surprise me, so the Wizard', or (R5, FG4). These opinions do not merely rely on technical literacy skills, this participant describes himself as a "technician, I understand what it does, it is more clear, but I find these [personae] much more interesting, they challenge me: 'expand my horizon', yes Explorer, something completely different, then 'Wizard"'.

While the personae are generally valued for taking people out of their filter bubble or ordinary reading habits, the value of collaborative filtering is much more disputed. While some joke about it, "haha, people like me, yeah, I need to have that' (R3, FG2), other people found it "annoying to read what similar people are reading" (R3, FG4) or expressed their concern about this functionality: "don't you think this could be dangerous? If people can actively select news only from their own circles, then I think that it could really dangerous, it would only enhance 'us-them' thinking and polarization"(R4, FG2). But that position got challenged as well: "I compare it with how it is now, at least [now] you choose to enter a filter bubble, instead of unwittingly being pushed in" (R5, FG2).

Despite these differences, participants argued that they would not be using these control mechanisms all the time: "I won't be using it all days of the week, that's really something for when you go sit down to read the news" (R1, FG2). Indeed, another adds, "I would use this on a Saturday afternoon, not on weekdays" (R1, FG1). But UX design is important as well: "same here, would I use it? Yes. But I would use it more often if I would be smartly reminded" (R1, FG4). While not seen as becoming part of people's daily news routine, participants valued these control mechanism for doing something different: to explore beyond one's ordinary reading, to get out of one's filter bubble, to grow and to learn more about other topics and perspectives. Perhaps more a tool for self-actualization goals than for everyday use, choosing between different recommender algorithms is for our participants a powerful and fun way to control RS.

4.4 Output control mechanisms

The last control mechanism we discussed, filtering the recommended items, was the least popular. Confusion firstly arose as to "what exactly the difference was between this and the dashboard above [user profile or first control mechanism]?" (R3, FG1). Participants found this "quite confusing", it often took some time for them to realize that "with the first you set your preferences, and still can get mixed recommendations, here you can say much harder, I just want entertainment" (R1, FG2). Some suggested to "not use sliders [as in *input*], but to have a drop-down menu 'Order by' to put them in different order" (R2, FG3), which makes good sense.

When asked about having feelings of being in control over the recommendations given, most participants reacted negatively about the output control mechanism: "well, you don't really have control right? Per definition. Just over the ranking" (R4, FG2). Some emphasized how they "would prefer to have control over the algorithms" (R1, FG2), and found this control mechanism "superfluous, one would be enough" (R4, FG2). But others found "it really useful, despite not having control over the recommendations, it's more of an extra service" (R2, FG3). They emphasized how this "quickly helps you adjust the ordering, rendering visible what you are really interested in at that moment. I would use it to be able to make faster choices" (R1, FG4). Indeed, another participant confirms: "sometimes I just wanna read about politics, so than this would be a handy tool" (R3, FG3). But that would then only work, as mentioned before, "with enough items in the list, otherwise it has little effect" (R2, FG4). In conclusion, the option to filter the output is seen as a welcome tool, but not sufficient to give people (a sense of) control over the RS.

5 DISCUSSION

The focus groups highlighted a strong desire for having more control over RS. Recent public discussions about the rise and power of algorithms and Big Data, often sensationalized by media, have left their marks on ordinary people. Participants frequently expressed distrust about the 'real' objectives behind RS, and about how their data is being (ab)used. Obtaining more control over RS is not only appreciated by our participants who regard this an important strategy to mitigate aforementioned concerns, but is currently expected by policy makers and data protection regulations (like the GDPR) alike. RS scholars and designers would do well to acknowledge such concerns and discuss how these can be translated into more responsible RS designs.

A first solution to address concerns related to intrusive data collection by companies deploying RS is inspired by the normative concept of reciprocity [11]. Giving something of that pervasive information extraction back to users, is a logical and important step in the restructuring of the digital economy towards a more fair market and society [39]. Our participants agree, and highly desire obtaining the information that is collected of them. Dashboards with meaningful information are a great way to satisfy that desire, and to pro-actively comply with current data protection regulations regarding the right to receive information [12]. We can include current attempts by Spotify (a yearly overview of what users have (not) been listening to), Apple ("Time in front of the screen"), and Instagram ("Time on IG") in the same line of making users more aware of their digital activities.

Besides these privacy and data abuse concerns, participants in our study emphasized the alleged loss of human agency due to RS. By pre-selecting what people supposedly like (to read), RS would diminish their own critical thinking and creativity. Self-reflexive questions such as 'what do *I* want to read?', 'What are *my* preferences?' and 'how should *I* read the news' are, according to our participants, easily sidestepped by using RS. These ideas resonate with academic concerns that people tend to unreflexively obey RS with the consequence of them undermining human creativity [22, 29]. Thinking about activating people and supporting their autonomy and creativity [8], instead of simply suggesting items with the highest success rates, may therefore be crucial when designing (*and* communicating) responsible , if only to align it better with people's ideas of how to maintain and/or integrate human creativity with the potentials of digital technologies.

The several user control mechanism we discussed in this study are particularly appreciated when they can be deployed to develop and achieve longer term goals. Obtaining more control over RS is, according to our participants, thus not only a goal in itself, but also a means towards a certain end. Whatever that end is, can be left to the user, but media organizations (or other content providers) could also step in here and align such ends with their own professional missions or societal objectives [46]. The important point is that RS can help defining and realizing normative goals. Knijnenburg et al. have made a good start with thinking about such "Recommender Systems for Self-Actualization (RSSA)" and the different optimization strategies they should embody [29]. Interestingly, many of their suggestions to design RS differently have come up in our study as well, which confirms that these are not just theoretical solutions, but have real empirical grounding in peoples concerns, needs and wishes.

6 CONCLUSIONS

RS are pervasive in contemporary societies, but they are not uncontroversial due to filter bubble concerns, and a lack of transparency, diversity, and user control. In this paper, we align with the normative turn in computer science which scrutinizes the ethical and societal implications of RS. We focus and elaborate on the concept of user control because that mitigates all those concerns at once: it empowers users to deploy RS more to their needs and interests, increases trust and satisfaction, requires RS to be more transparent and explainable, and lessens algorithmically induced blind spots.

Taking the news as our domain, we conducted four focus groups, or moderated think-aloud sessions, with Dutch news readers (N = 21) to systematically study how people evaluate different control mechanisms (at the input, algorithm, and output phase) in a NRP. While these mechanisms are sometimes met with distrust about the actual control they offer, we found that an intelligible user profile (including reading history and flexible preferences settings), coupled with possibilities to choose the recommendation algorithms is extremely valued, especially when these control mechanisms can be operated in relation to achieving personal goals. Post-recommendation filtering is only seen as nice extra option, but not sufficient to make users feel that they have control over RS.

This paper contributes to a richer understanding of why and how to design for user control in RS. We do so by showing the diversity and complexity of users' concerns, motivations and aspirations. Quantitative studies can build forth on our work and measure how these views are distributed between different societal groups, and in different contexts. We hope to have shown that thinking about user control is important in the development of responsible RS.

ACKNOWLEDGMENTS

This study is part of the Fair News project, which is financed by the Netherlands Organisation for Scientific Research (NWO) as part of the recently created National Research Agenda in the Netherlands. Project number: 400.17.605. Cooperating partner of this project is De Persgroep Nederland BV, with whom we productively collaborated

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-aware recommender systems. In *Recommender systems handbook*. Springer, 217–253.
[2] 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*. ACM, 35–42.
[3] Danah Boyd and Kate Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15, 5 (2012), 662–679.
[4] Jenna Burrell. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data Society* 3, 1 (2016).
[5] Bruno Cardoso, Gayane Sedrakyan, Francisco Gutiérrez, Denis Parra, Peter Brusilovsky, and Katrien Verbert. 2019. IntersectionExplorer, a multi-perspective approach for exploring recommendations. *International Journal of Human-Computer Studies* 121 (2019), 73–92.
[6] Pablo Castells, Neil J Hurley, and Saul Vargas. 2015. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*. Springer, 881–918.
[7] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. Sage.
[8] John Christman. 2004. Relational autonomy, liberal individualism, and the social constitution of selves. *Philosophical studies* 117, 1 (2004), 143–164.
[9] Nicholas Diakopoulos. 2016. Accountability in algorithmic decision making. *Commun. ACM* 59, 2 (2016), 56–62.
[10] Michael D Ekstrand, Daniel Kluver, F Maxwell Harper, and Joseph A Konstan. 2015. Letting users choose recommender algorithms: An experimental study. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 11–18.
[11] Michael D Ekstrand and Martijn C Willemsen. 2016. Behaviorism is not enough: better recommendations through listening to users. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 221–224.
[12] Sarah Eskens, Natali Helberger, and Judith Moeller. 2017. Challenged by news personalisation: five perspectives on the right to receive information. *Journal of Media Law* 9, 2 (2017), 259–284.
[13] Daniel Fleder and Kartik Hosanagar. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science* 55, 5 (2009), 697–712.
[14] Mario Haim, Andreas Graefe, and Hans-Bernd Brosius. 2018. Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital journalism* 6, 3 (2018), 330–343.
[15] Jaron Harambam, Natali Helberger, and Joris van Hoboken. 2018. Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2133 (2018), 20180088.
[16] F Maxwell Harper, Funing Xu, Harmanpreet Kaur, Kyle Condiff, Shuo Chang, and Loren Terveen. 2015. Putting users in control of their recommendations. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 3–10.
[17] Chen He, Denis Parra, and Katrien Verbert. 2016. Interactive recommender systems: A survey of the sota and future research challenges and opportunities. *Expert Systems With Applications* 56 (2016), 9–27.
[18] Natali Helberger. 2019. On the democratic role of news recommenders. *Digital Journalism* (2019), 1–20.
[19] Natali Helberger, Kari Karpinnen, and Lucia D'Acunto. 2018. Exposure diversity as a design principle for recommender systems. *Information, Communication Society* 21, 2 (2018), 191–207.
[20] Anthony Jameson, Martijn C Willemsen, Alexander Felfernig, Marco de Gemmis, Pasquale Lops, Giovanni Semeraro, and Li Chen. 2015. Human decision making and recommender systems. In *Recommender Systems Handbook*. Springer, 611–648.
[21] Dietmar Jannach, Sidra Naveed, and Michael Jugovac. 2017. User control in recommender systems: Overview and interaction challenges. In *Proceedings of 17th International Conference on E-Commerce and Web Technologies 2016*. Springer, Cham, SH, 21–33.
[22] Lanier Jaron. 2010. You Are Not a Gadget: A Manifesto. *New York: Knopf* (2010).
[23] Yucheng Jin, Nava Tintarev, and Katrien Verbert. 2018. Effects of personal characteristics on music recommender systems with different levels of controllability. In *18 Proceedings of the 12th ACM Conference on Recommender Systems. ACM Press, New York, NA, US*. RecSys, 13–21.
[24] Michael Jugovac and Dietmar Jannach. 2017. Interacting with recommenders - overview and research directions. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 3 (2017), 1–10.
[25] Marius Kaminskas and Derek Bridge. 2017. Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 1 (2017), 2.
[26] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems–Survey and roads ahead. *Information Processing & Management* 54, 6 (2018), 1203–1227.
[27] Rob Kitchin. 2017. Thinking critically about and researching algorithms. *Information, Communication Society* 20, 1 (2017), 14–29.
[28] Bart Knijnenburg, Niels Reijmer, and Martijn Willemsen. 2011. Each to his own: How different users call for different interaction methods in recommender systems. In *11 Proceedings of the fifth ACM conference on Recommender systems. ACM Press, New York, NA, US*. RecSys, 141–148.
[29] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender systems for self-actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 11–14.
[30] Joseph Konstan and John Riedl. 2012. Recommender systems: From algorithms to user experience. *User modeling and user-adapted interaction* 22, 1-2 (2012), 101–123.

[31] Denis Kotkov, Shuaiqiang Wang, and Jari Veijalainen. 2016. A survey of serendipity in recommender systems. *Knowledge-Based Systems* 111 (2016), 180–192.

[32] Sara Leckner. 2012. Presentation factors affecting reading behaviour in readers of newspaper media: an eye-tracking perspective. *Visual Communication* 11, 2 (2012), 163–184.

[33] Jin Ha Lee and Rachel Price. [n.d.]. Understanding users of commercial music services through personas: Design implications. In *Proceedings of 16th International Society for Music Information Retrieval Conference*. ACM Press, New York, NA, US, 476–482.

[34] Sean McNee, John Riedl, and Joseph Konstan. 2006. Making recommendations better: An analytic model for Human-Recommender Interaction. In *Abstracts of the 2006 ACM Conference on Human Factors in Computing Systems*. ACM Press, New York, NY, US, 1003–1008.

[35] David L Morgan. 1996. *Focus groups as qualitative research*. Vol. 16. Sage publications.

[36] Cathy O'Neil. 2017. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

[37] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. Penguin, London, UK.

[38] Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67.

[39] Eric A Posner and E Glen Weyl. 2018. *Radical markets: Uprooting capitalism and democracy for a just society*. Princeton University Press.

[40] Pearl Pu, Li Chen, and Rong Hu. 2012. Evaluating recommender systems from the userâĂŹs perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 317–355.

[41] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. *Recommender systems handbook*. Springer, Cham, SH.

[42] James Schaffer, Tobias Hollerer, and John O'Donovan. 2015. Hypothetical recommendation: A study of interactive profile manipulation behavior for recommender systems. In *The Twenty-Eighth International Flairs Conference*.

[43] Douglas Schuler and Aki Namioka. 1993. *Participatory design: Principles and practices*. CRC Press.

[44] David Silverman. 2016. *Qualitative research*. Sage.

[45] Rashmi Sinha and Kirsten Swearingen. 2002. The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems*. ACM, 830–831.

[46] Emily Sullivan, Dimitrios Bountouridis, Jaron Harambam, Shabnam Najafian, Felicia Loecherbach, Mykola Makhortykh, Domokos Kelen, Daricia Wilkinson, David Graus, and Nava Tintarev. [n.d.]. Reading News with a Purpose: Explaining User Profiles for Self-Actualization. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. ACM, 241–245. http://doi.acm.org/10.1145/3314183.3323456

[47] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook*, Lior Rokach and Bracha Shapira (Eds.). Springer, Cham, SH, 353–382.

[48] Max van Kleek Ulrik Lyngs, Reuben Binns and Nigel Shadbolt. 2018. So, tell me what users want, what they really, really want!. In *CHI EA '18 (alt.chi) Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM Press, New York, NY, US, 1–10.

[49] Malte Ziewitz. 2016. Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values* 41, 1 (2016), 3–16.