

Benefits of Diverse News Recommendations for Democracy: A User Study

Lucien Heitz^{a,b} , Juliane A. Lischka^c , Alena Birrer^d , Bibek Paudel^{e#} ,
Suzanne Tolmeijer^f , Laura Laugwitz^c  and Abraham Bernstein^f 

^aDepartment of Informatics, University of Zurich, Zürich, Switzerland; ^bDigital Society Initiative, University of Zurich, Zürich, Switzerland; ^cJournalism and Communication Studies, Universität Hamburg, Hamburg, Germany; ^dDepartment of Communication and Media Research, University of Zurich, Zürich, Switzerland; ^eDepartment of Biomedical Data Science, Stanford University, Stanford, CA, USA; ^fDepartment of Informatics, University of Zurich, Zürich, Switzerland

ABSTRACT

News recommender systems provide a technological architecture that helps shaping public discourse. Following a normative approach to news recommender system design, we test utility and external effects of a diversity-aware news recommender algorithm. In an experimental study using a custom-built news app, we show that diversity-optimized recommendations (1) perform similar to methods optimizing for user preferences regarding user utility, (2) that diverse news recommendations are related to a higher tolerance for opposing views, especially for politically conservative users, and (3) that diverse news recommender systems may nudge users towards preferring news with differing or even opposing views. We conclude that diverse news recommendations can have a depolarizing capacity for democratic societies.

KEYWORDS

Algorithmic curation; ethics; journalism; political polarization; public sphere; recommender systems; user preferences

Introduction

Social media, search engines, and news aggregators have increasingly become relevant intermediaries for news consumption (Newman et al. 2020). These intermediaries provide critical infrastructures to democratic societies, for which pluralistic debates are crucial processes (Müller 2021). Recommender systems are one element of intermediary infrastructure that select personalized news for users and determine news consumption (Hennig-Thurau et al. 2012). Using machine learning, they curate content to personalize recommendations on the basis of specific optimization criteria (Heuer 2020). Optimized for user preferences, recommender systems increase exposure to news that users want to know, which may come at the cost of what users should

CONTACT Lucien Heitz  heitz@ifi.uzh.ch

#Work partially performed whilst at the Department of Informatics, University of Zurich, Zürich, Switzerland.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

know from a normative point of view (Diakopoulos 2019; Nechushtai and Lewis 2019). While little evidence of filter bubbles in online news environments is found (Haim, Graefe, and Brosius 2018; Mummolo 2016), users' selective exposure and avoidance of opposing views could reinforce particular politicized stances (de Benedictis-Kessner et al. 2019; Hellmueller, Lischka, and Humprecht 2020; Peterson, Goel, and Iyengar 2021). Such user behavior may lead to political polarization, which preference-based recommender systems could further enhance.

Socially responsible designs of news recommender systems have increasingly become a focus in research (Harambam et al. 2019; Helberger 2019; Helberger et al. 2018; Milano et al. 2020; Nechushtai and Lewis 2019; Thurman et al. 2019). Exposure diversity has been discussed as one critical optimization criterion for responsible news recommender systems that can enhance cross-cutting exposure (Bernstein et al. 2021; Helberger et al. 2018). Previous research suggests that diverse news recommendations are appreciated by users (Christoffel et al. 2015; Paudel et al. 2017). Diversity-optimized recommendations may involve greater exposure to counter-attitudinal news and diverse news recommendations may broaden the horizon of users and increase tolerance towards opposing views (Möller et al. 2018; Mutz 2020), providing positive societal externalities. Such effects can be viewed as nudging since the design of "choice architecture" induces changes in people's attitudes and behavior that are normatively desirable (Thaler and Sunstein 2009). However, little is known about the normative capacity of diverse news recommendations compared to accurate preference-based recommendation approaches for users and society. We ask the following research questions (RQs):

RQ 1.: How do diverse news recommendations affect news consumption?

RQ 2.: How do diverse news recommendations affect individual utility of news consumption?

RQ 3.: How do diverse news recommendations affect societal externalities of news consumption?

Building on the notion of socially responsible designs of news recommenders (Bernstein et al. 2021; Helberger 2019; Helberger et al. 2018; Thurman et al. 2019), this study aims to understand the potential of diversity as a normative design criterion of news recommender systems for democratic public spheres. It thus contributes to the discourse architecture and algorithmic responsibility literature, providing an example to enhance the algorithmic social contract (Rahwan 2018). It is important to note, however, that a theoretical discussion of how to implement diversity in recommender systems to best support society and democratic institutions is beyond the scope of this article. Instead, our goal is to highlight potential implications of diversity-optimized recommendations for news readers and the public sphere, when contrasted with accuracy-based algorithms. To do so, we provide a customizable algorithm that can serve as a toolbox to define and compare different notions of diversity. What the precise optimization goal looks like, however, that is the subject of future work.

We conduct a single-factor between-groups experimental design, featuring chronological, accuracy- or diversity-optimized recommendations on a smartphone app with $n = 151$ news users in Switzerland. As a direct democracy integrating regional plurality in political decisions, Switzerland represents a context in which diversity is highly relevant (Linder and Mueller 2021).

Literature Review

Recommender Systems as Socio-Technical Regimes

From a socio-technical perspective, multiple actors have agency over machine-learning curation systems (Heuer 2020). In technology systems, “networks of agents interact [...] in a specific technology area under a particular institutional infrastructure to generate, diffuse and utilize technology” (Carlsson and Stankiewicz 1991, p. 111).

One central feature of technology systems is the ability to develop and exploit business opportunities (Carlsson and Stankiewicz 1991) following technical standards and corporate goals (Geels 2004). In a corporate-technological regime, agents adhere to technical standards as well as business goals. Regarding recommender systems, optimizing for user preference represents a manifestation of corporate-technological norms. Lacking the implementation of socio-cultural norms, the design of preference-optimizing algorithms has been criticized as incorporating capitalist ideology (Fuchs 2014; Mager 2012). However, while recommender systems have comprised the accurate modeling of user preferences (Karimi et al. 2018; Kunaver and Pozrl 2017), the major drawback of accuracy-optimized recommendations is that they can lead to a monotonous list of recommendations, eventually resulting in user fatigue (Ma, Liu, and Shen 2016). There are efforts trying to combat this monotony by accounting for serendipity in recommender systems and recommend items that are relevant but unexpected from a user’s perspective (Ziarani and Ravanmehr 2021). In this regard, recommender systems can promote long-tail items on top of a recommendation list to increase diversity (Paudel et al. 2017). From a socio-cultural perspective, news recommender systems determine the content available in public spheres, and norms of public spheres can be ascribed from prototypical models of democracy (Ferree et al. 2002). Instead of serendipity, majority as well as minority perspectives should be constantly represented to ensure viewpoint diversity (Helberger 2019).

Corporate-technological and socio-cultural norms compete in socio-technical regimes, requiring meta-coordination within and between the two regimes (Geels 2004). The debate around changes in optimization criteria of recommender systems indicates the destabilization of the previously dominant corporate-technological regime. By moving away from accuracy-based optimization towards criteria that may also enhance public spheres, a novel socio-technical regime is established. That is, news recommender systems “should not just maximize for clicks and short-term revenue, but, mindful of the democratic function of the media, also optimize for values that align with the overall mission of a news outlet” (Bernstein et al. 2021, p. 3). In this debate, diversity may represent a design criterion complying with socio-cultural norms and providing positive externalities.

Externalities of Diversity in News Recommender Systems

Externalities represent positive or negative impacts of the production or usage of a product or service for society that is neither compensated by the producer nor the user (Pigou 1932). News production and use is understood to provide positive externalities for society due to holding the powerful accountable, providing a public

sphere, and informing citizenry (McQuail 2005). Whether journalism can provide positive externalities also depends on its societal acceptance and perceived performance. From the users' perspective, media performance includes diverse and impartial coverage in addition to a feeling that users' own views are represented (Steppat, Castro Herrero, and Esser 2020). Users accordingly select news that provide instrumental utility for them, such as to understand society, to learn about multiple perspectives, or to reinforce their attitudes (Atkin 1973). Attitude-reinforcing news usage increases the perception of positive media performance (Steppat, Castro Herrero, and Esser 2020) but could hold negative externalities. For instance, low levels of exposure to non-like-minded content are related to lower voter turnout (Castro Herrero and Hopmann 2018). Selective exposure to such cross-cutting content may be asymmetrical across political positions. Highly committed conservatives are found to more often use pro-attitudinal news (Hmielowski, Hutchens, and Beam 2020; Mothes and Ohme 2019) but also to follow counter-attitudinal sources more often than liberals (Eady et al. 2019). Recommender systems can enhance a partisan selective exposure by "insulating users from exposure to different viewpoints, creating self-reinforcing biases and 'filter bubbles' that are damaging to the normal functioning of public debate, group deliberation, and democratic institutions more generally" (Milano et al. 2020, p. 964). Providing positive externalities, news recommenders could be "powerful tools to help users find their way in the plethora of available news, shape public opinion, and serve as a foundation for public cohesion" (Bernstein et al. 2021). The design of recommender systems is therefore identified as one ethically crucial area (Milano et al. 2020).

Diversity is regarded as one design principle enhancing positive externalities of news recommender systems (Helberger et al. 2018; Vrijenhoek et al. 2020). Diversity in news recommender systems includes how news topics, article complexity, or perspectives are selected according to user preferences, how fragmented recommendations are across different users, how balanced the set of recommendations is in terms of opinion plurality, and how much alternative voices are part of the recommendations (Vrijenhoek et al. 2020). Regarding the potential of lowering exposure to different viewpoints, opinion plurality represents a critical design principle (Bernstein et al. 2021). Previous research argues that exposure to dissonant political views increases tolerance, i.e., the ability to see and follow the arguments of the counter party (Mutz 2020). Yet, exposure to counter-attitudinal news can also backfire, i.e., reinforce existing political positions, and thus potentially increase political polarization (Bail et al. 2018). While users generally prefer pro- over counter-attitudinal views, automated recommendations can increase openness towards counter-attitudinal messages (Wojcieszak et al. 2021), especially when the topic holds utility for a user (Mummolo 2016). A recommender system diversifying exposure to cross-cutting news could increase or reduce openness towards opposing views, given the news provides instrumental utility. Hence, this study focuses on a recommendation method to vary exposure to individual opinion diversity.

Method

We conduct a single-factor between-groups experimental design in which the treatment is a recommender algorithm of a custom-built news app. Participants used the

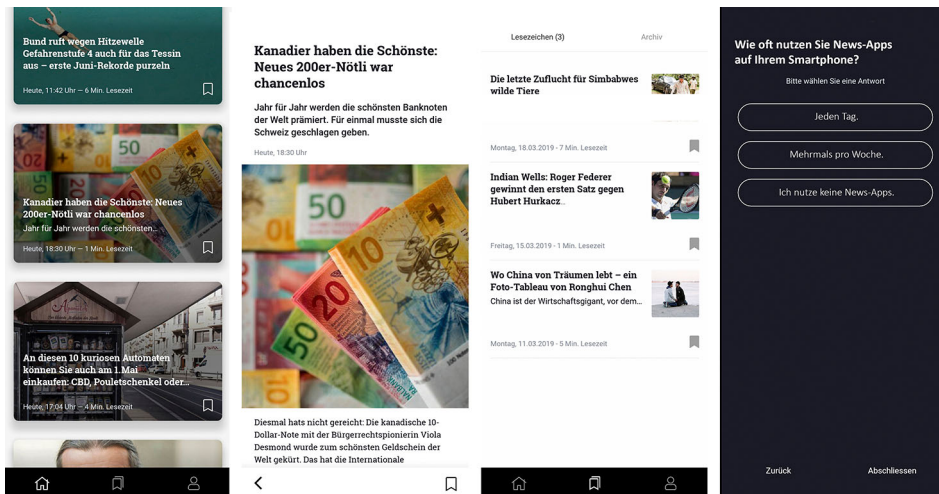


Figure 1. The app's home screen (left), article view (left to the center), bookmark list (right to the center), and the built-in survey tool (right). The look and functionality of the app is identical across all platforms (Google Android and Apple iOS).

news app for five weeks, being part of the narrow, diverse, or control group. In the beginning of the experiment, 20% of the participants were assigned to the narrow group, 20% to the diverse group and the remaining 60% to the control group. The narrow group received accuracy-optimized news ($n=35$), the diverse group received diversity-optimized articles ($n=28$), and the control group ($n=88$) chronologically ordered news recommendations, with the most recent items at the top of the reading list. (The numbers state final sample sizes.) The majority of people were in the control group; it has the additional purpose of providing the recommender system with the interaction data required for labeling news articles.

The study was approved by the ethics commission of the University of Zurich (approval number 19.8.10).

News App

We built a news app that aggregates real-time news from six major German-language Swiss news outlets representing a broad political spectrum, and private as well as public-service outlets (*Blick*, *Neue Zürcher Zeitung*, *Tages-Anzeiger*, SRF Online, *Weltwoche*, and *WOZ Die Wochenzeitung*). Such a multiple-outlet news app featuring paid and subscription content does not exist for the Swiss market and the users were the first to experience receiving paid content from multiple outlets in one mobile app.

Figure 1 provides an overview of the news reader app. On the home screen (far left), users are given their list of recommended news items. The second screenshot on the left shows the detailed reading view of an article once the user has made their selection on the home screen. Here, they are able to access the full article, to rate the article with “like” or “dislike,” and to bookmark the article. The second screenshot from the right shows the bookmark list. Here, users are presented with an overview of all

bookmarked articles. Finally, the rightmost screenshot shows how in-app surveys are presented to the users throughout the experiment.

The app (1) collects news from a variety of sources representing the overall political spectrum within a media system, (2) hides a news story's origin to avoid bias in news consumption, (3) devises a fully controllable recommendation list, (4) observes the readers' interaction with the news articles including their explicit statements (e.g., likes, choosing to read or ignore an article, etc.) and implicit behavior (e.g., scrolling through an article or skimming headlines), and (5) prompts user surveys.

News Recommender System

Our approach to diversity-optimized news recommendations is based on the idea of defining diversity as a discretized distribution of news articles over a political spectrum. We created a hybrid, human-centered scoring pipeline for news articles that is based on the political position of users. After aggregating all reading metrics of users, each news article gets assigned a political label based on the average political score of its readership. Assessing political alignment of an article based on averaging user labels has already been shown to work reliably in the online domain (Bakshy, Messing, and Adamic 2015).

Our proposed approach, hence, does not rate the *actual* political stance of an article, but rather *labels* articles based on the political stance of people who 'like' to read it, which is assessed on the basis of both implicit and explicit user feedback. This way of labeling articles has the advantage that it does not rely on any secondary process (with its own bias and error rates) for predicting the political stance of an article based on content analysis (e.g., labeling articles by experts).

User Scoring

Measuring a person's political stance is a much discussed topic in political science (Michael 2020). We established the political score of participants with a standardized survey. The calculation of the political score is based on the method of *Smartvote*¹ using the questions of *Parteienkompass*.²

To summarize the method succinctly, we assessed two political dimensions—left/right and liberal/conservative—for each user based on more than 20 questions. For every question and dimension, a maximum of 100 points are assigned. These are multiplied with a modifier based on the answer given: yes ($\times 1.0$), rather yes ($\times 0.75$), rather no ($\times 0.25$), and no ($\times 0.0$). Next, the resulting modified number of points is weighed based on question-specific weights (either $+1.0$ or -1.0). Finally, points are added for each dimension and normalized between -1.0 and $+1.0$.

Article Scoring

Having calculated political scores for users, the algorithm utilized these scores as a basis to assign a political label to news articles. The political label assigned to an article is a weighted average political user score of its readership. During the experiment, the article score $score_a$ of article a was calculated as follows:

$$score_a = average_{u:(t_{read,a,u} > 10 \text{ sec})} [score_u * factor_like_{a,u} * factor_list_{a,u}] \quad (1)$$

which is the average weighted score for all users u , who read an article a for more than 10 seconds, where

$$factor_like_{a,u} \begin{cases} 3, & \text{if } liked_{a,u} \\ 0, & \text{if } disliked_{a,u} \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

is the weight assigned when user u likes or dislikes article a and

$$factor_list_{a,u} \begin{cases} 2, & \text{if } \exists rl_{a,u} \wedge \neg \exists removed_{a,u} \vee archived_{a,u} \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

is the weight assigned when a user u has either added the article to the reading list ($\exists rl_{a,u}$) without having removed it from that list ($\neg \exists removed_{a,u}$) or has archived it ($archived_{a,u}$).

To increase the reliability of this approach, any news article recommended to the narrow or diverse group had to be read by at least five different users in the control group. Note that this setup confronts us with a cold-start problem; when the article is not read yet, it lacks a score. We address this issue by first presenting articles to our extra-large control group to derive the political article label.

The feedback system for leaving a like or dislike was complemented by a short survey where participants had to provide additional information on their rating where they were given a selection of reasons (e.g., disliking writing style or political stance). For the purpose of calculating recommendations, we only considered the rated articles that were liked or disliked because of their political content.

Recommending

The recommender algorithm requires a political score $score_u$ for each user u as well as an article distribution $art_distribution$, which contains a collection of tuples $[score, n_a]$ that model the discrete target distribution, i.e., the number of articles n_a required for each political orientation $score$. The algorithm takes these inputs and computes a desired (score-)distribution of recommended articles $desired_distribution$ as a one-dimensional array. The details of the algorithm are shown in [Appendix A](#).

Treatment

Diverse Recommendations

The normative recommender algorithm presented in the previous section was used to provide diverse as well as narrow recommendations to users. Both diverse and narrow recommendations were defined as discretized article distributions across a two-dimensional political spectrum (left-/right-leaning and liberal/conservative). User scores were rounded and grouped together to increase the performance of the recommender system.

The system focused on providing each user in the diversity group with a selection of news articles that covers the entire political space. It is a non-uniform distribution that includes slightly more items from the political middle. The reason behind doing so is to create a set of articles everyone reads, in order for a discussion to emerge. A

detailed overview of the diversity distributions used during the experiment is available in [Appendix B](#).

When creating the diversity distribution used in the experiment, we followed the type classification presented by Helberger (2019). Our target distribution for diversity follows the characteristics of participatory recommenders, i.e., creating a multi-platform system that includes different styles of articles, focusing on political news, as well as content that speaks to a broader audience. As required, the system is inclusive of all viewpoints, for each existing political orientation, at least one article is included in the recommendations. It also features a proportional representation of main political viewpoints; we accounted for proportionality by slightly shifting the diversity distribution towards the political center, i.e., the place where, given the participating outlets, we saw most of the article being located.

Narrow Recommendations

The distribution strategy for narrow or accuracy-optimized recommendations focused on articles that are generally in line with the participant's political views. Given a user in the narrow group, they receive news articles that have the same political score they have. An overview of the narrow distributions is in [Appendix B](#).

Chronological Recommendations

The control group received chronological recommendations. These did not require any special algorithm, as the users simply received articles ordered by their publishing/update date, sorted from newest to oldest.

Participants

Participants were recruited by a Swiss market research company using an ISO 26362:2009 certified online access panel. Participants received an incentive of 42 CHF (about 43 USD) for app usage and participating in the post-hoc survey.

Besides living in the German-speaking area of Switzerland and owning a mobile phone, recruiting criteria were frequent news use (more than one time per week) and an even distribution across age groups, gender, and political position (regarding party preference and on a left/right scale). While $n = 265$ users were initially recruited, only the medium and heavy users who used the app at least three hours or had a high rating activity during five weeks were invited to the post-hoc questionnaire ($n = 151$). While $n = 148$ participants adhere to the three-hour usage criterion, additional $n = 3$ were included due to their high rating frequency, who used the app for 0.6-0.8 h in total.

Participants of the final sample ($n = 151$) are on average 43 years old ($SD = 16.45$, $Min: 18$, $Max: 77$). Almost half is female (48%). About one third indicated apprenticeship or a lower educational level (36%), four out of ten a higher vocational training (41%), and just over one fifth a Bachelor's or higher university degree (23%).

Measures

App Usage Metrics

During the five-week app usage period, we track users' reading behavior, i.e., app usage time, article reading time, and interaction (likes and dislikes of articles as well as their bookmarked favorite articles). App usage time includes scrolling down the home screen seeing the article previews, while article reading time indicates engagement with an article, that may be positively or negatively evaluated by likes and dislikes, respectively. We measure general (time spent in app), active (articles read more than 10 seconds), and explicit engagement (numbers of liked and disliked articles).

In-App Survey

Before the app usage starts, this survey assesses the political stance on two dimensions (left-right and liberal-conservative) on issues such as minimum wage, increasing the retirement age, or investment into public transportation. We used items from *Parteienkompass*, which translates official positions of Swiss political parties into a survey.

Post-Hoc Survey

App Usability. We measure usability on a 5-point Likert scale with seven items such as "easy to learn," "complicate," and "activating" (Holzinger 2008). Items are measured on a 5-point scale from 0 = not at all to 4 = very much. For analysis, the scale was converted to values from 1 to 5.

News Diversity. We assess the perceived diversity of news with four items and the perceived news agenda with an open question about the top-5 news issues participants encountered the most. Perceived diversity is measured by the following items, "the app covered a broad spectrum of news," "was impartial regarding political opinions," "covered multiple political opinions," and "fit to my own political opinion." These measurements serve as manipulation check. Items are measured on a 5-point scale from 0 = not at all to 4 = very much. For analysis, the scale was converted to values from 1 to 5.

Instrumental Utility. Instrumental utility is measured with seven items as perceived informational utility of news in the app. Information utility is given when a recipient acquires guidance (e.g., to understand complex issues), performance (e.g., to solve practical problems), reinforcement (e.g., to reconfirm attitudes), and surveillance (e.g., to perceive threats and opportunities) from news (Atkin 1973). Items include "news in the app helps me to understand our society," "helps me to make wise decisions," "provides me with a daily account of what is happening in the world" (Li 2014). Items are measured on a 5-point Likert scale from 0 = not at all to 4 = very much. For analysis, the scale is converted to values from 1 to 5. The items are then summarized to a mean index (Cronbach's $\alpha = .889$).

Societal Externalities. Externalities include (a) political factors (such as knowledgeability and participation), (b) tolerance of opposing views, (c) social performance of

journalism, and (d) news diversity preferences. Regarding political factors, we measure political knowledge, interest, and efficacy with one item each (Kruikemeier and Shehata 2017; Pingree 2011), which is combined to a mean index of political knowledgeability (Cronbach's $\alpha = .912$). Political participation includes voting, visiting a political website, reading a political blog, talking about politics with friends, and signing a political petition (Hameleers et al. 2018; Kruikemeier and Shehata 2017), which is summarized to a mean index (Cronbach's $\alpha = .774$). Tolerance refers to the scope of one's acceptance with regard to different views in society (Li 2014). We use the political-opinion statement, "I can relate to opinions that differ from my own political views." Social performance of journalism is measured with one item, "journalism helps society to solve its problems" (Peifer 2018). News diversity preferences is measured with two items, i.e., news that inform about counter-attitudinal news and inform about majority opinions (Helberger 2019; Vrijenhoek et al. 2020).

Demographics and External News Use. We ask for gender, age, highest level of education, and external news usage per news outlet during the app usage. External news usage is aggregated to a sum index over all news outlets that were used during the app usage period.

Results

App Usage and News Consumption (RQ1)

App Usage Metrics

On average, participants had the app opened for 6.68 h ($SD = 8.21$, $Min = .61$, $Max = 68.04$), opened 265 articles ($SD = 362.80$, $Min = 38$, $Max = 3,104$), liked 317 articles ($SD = 725.66$, $Min = 0$, $Max = 5,937$), and disliked 50 articles ($SD = 78.62$, $Min = 0$, $Max = 480$) during the span of five weeks. There are no significant usage differences between the control, narrow, and diverse conditions according to a one-way between subject ANOVA ($F(2, 148) = .198$ to $.838$, n.s.). However, due to the large range of usage time, we include it as a covariate to our estimations.

App Usability

Participants regarded the app to be very easy to learn ($M = 4.62$, $SD = .77$), uncomplicated ($M = 4.39$, $SD = .98$), and useful ($M = 4.12$, $SD = 1.03$), rather valuable ($M = 3.78$, $SD = .96$), activating ($M = 3.60$, $SD = 1.07$), and clean ($M = 3.53$, $SD = 1.03$), and rather the way they wished a news app to be ($M = 3.58$, $SD = 1.09$) on a five-point Likert scale from 1 "not at all" to 5 "very much." A one-way between-subjects ANOVA reveals no significant differences between the control, narrow, and diverse conditions at the $p < .05$ level ($F(2, 148) = .002$ to 2.734 , n.s.).

Article Score Distribution

Figure 2 shows an overview of the mean political score of all articles. Each dot represents the average political score of the readership of one article. Over the course of the experiment, 7,287 unique news items were collected from the websites of our media partners (an average of 214 items per day), of which 6,282 different articles were read by users. In total, users accessed articles 55,344 times.

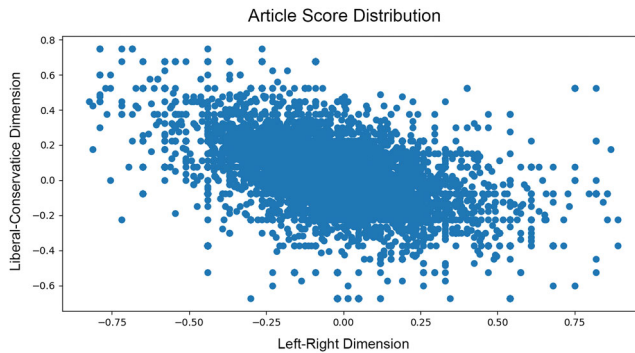


Figure 2. Distribution of news articles across the two-dimensional political spectrum. The x-axis shows the left-right dimension (with left being -1 and right $+1$). The y-axis shows the conservative-liberal dimension (with conservative being -1 and liberal $+1$).

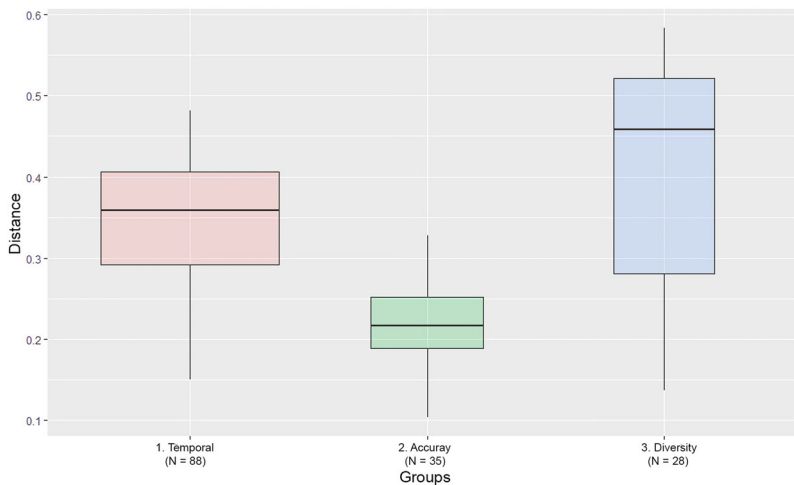


Figure 3. Standard box plot with quartiles showing the average distances between the political score of a user and the average score of articles that they read.

A post-hoc manual content analysis of $n = 100$ articles (25 each for the left-conservative, left-liberal, right-liberal, and right-conservative space) reveals that topics and stance are largely prototypical for each camp according to Parteienkompass. For instance, articles in the left-liberal space criticize former President of the United States Trump or favored social security while right-conservative articles were critical about immigration or supported Trump.

User Article Distances

Figure 3 shows the distance between the political scores of users and the political score assigned to the average newspaper article that they read. The average distance for people in the temporal group is 0.35, for people in the accuracy group it is 0.21 and for people in the diversity group 0.43. A pairwise t -test with Bonferroni's correction reveals significant differences between all group pairings (1/2, $p < .001$; 2/3, $p < .001$; 3/1, $p = .041$).

Manipulation Checks

To test whether the participants perceive the diversity of news differently across groups, we compare their evaluations in the post-hoc survey. Participants found that the news in the app covered a broad spectrum of news ($M = 4.37$, $SD = .89$), was rather impartial regarding political opinions ($M = 3.87$, $SD = .86$), rather covered multiple political opinions ($M = 3.80$, $SD = .92$), and moderately fit to one's own political opinion ($M = 3.26$, $SD = .87$). A one-way between-subjects ANOVA reveals no significant differences between the control, narrow, and diverse conditions ($F(2, 148) = .398$ to 1.09 , n.s.). Also, a one-way ANOVA including usage time and the number of articles read reveals no differences between groups ($F(8, 262) = 1.15$, n.s.). That is, narrow and diversity conditions were not perceived as such, independent from low or high app usage.

Instrumental Utility (RQ2)

On average, participants indicate their instrumental utility with news on the app to be moderately high ($M = 3.71$, $SD = .74$, $n = 150$).

We estimate fixed-effect ANOVAs to compare the experimental settings. We test the diversity and narrow groups compared to the control group. Control variables include the participants' conservative-liberal political position, app news usage time, external news consumption, gender, age, education, and interaction effects between the experimental groups and political position. Results show that the model fails to explain the variance of instrumental utility (see first column in [Table 1](#), $F = 1.862$, $p = .055$). Only the coefficient for app usage indicates a correspondence to a higher utility rating. The effect size suggests that app usage explains about 5% of variance in utility ($\eta_p^2 = .048$). Perceived utility of reading news in the app is similar when received chronologically, narrow, and diversely ordered news.

Societal Externalities (RQ3)

Political Aspects

Regarding the political categories, on average, all participants indicate their political knowledgeability to be moderate ($M = 3.37$, $SD = 1.00$, $n = 150$) and their average participation lies between "now and then" (scale point 4) and "often" (scale point 5) ($M = 4.25$, $SD = .85$, $n = 150$).

According to the fixed-effects results, 19% and 23% of variance of political knowledgeability and participation can be explained by the model in [Table 1](#), respectively. Political knowledgeability and participation do not differ across groups nor according to political position (second and third columns in [Table 1](#)). Instead, news consumption beyond the app ($\eta_p^2 = .036$) and gender ($\eta_p^2 = .123$) are predictors for political knowledgeability. That is, male as opposed to female participants and using additional news sources beyond the app are related to stronger political knowledgeability. Similarly, political participation is related to external news usage, gender, and education.

Social Performance of Journalism

On average, participants moderately agree that journalism has the capacity to solve society's problems ($M = 3.20$, $SD = .93$, $n = 148$).

Table 1. Instrumental utility and societal externalities of diverse news recommendations. Notes: Fixed effects ANOVA with controls and interaction terms. This table shows results tested against the control group. Additionally, differences between diversity and narrow group were tested and are found for tolerance for opposing views (not reported in this table). Political position (Pol. pos.) ranges from Conservative = 0 to Liberal = 1.

	Instrumental utility			Political interest			Political participation		
	β	ρ	η_p^2	β	ρ	η_p^2	β	ρ	η_p^2
Constant	3.668	0.000	0.468	1.227	0.004	0.057	2.242	0.000	0.233
Diversity Group	0.284	0.495	0.003	0.573	0.279	0.008	-0.136	0.754	0.001
Narrow Group	0.391	0.254	0.009	0.662	0.128	0.017	-0.144	0.686	0.001
Political position	0.037	0.900	0.000	0.270	0.476	0.004	-0.034	0.912	0.000
App usage (hrs)	0.020	0.009	0.048	-0.004	0.670	0.001	0.013	0.113	0.018
Ext. news usage	0.014	0.047	0.028	0.020	0.024	0.036	0.023	0.002	0.069
Gender (f = 1; m = 2)	-0.035	0.783	0.001	0.702	0.000	0.123	0.370	0.005	0.055
Age (y)	-0.005	0.241	0.010	-0.001	0.898	0.000	0.000	0.976	0.000
Education	-0.047	0.068	0.024	0.020	0.530	0.003	0.077	0.005	0.057
Diversity x Pol. pos.	-0.424	0.537	0.003	-0.855	0.329	0.007	-0.016	0.982	0.000
Narrow x Pol. pos.	-0.845	0.123	0.017	-0.558	0.421	0.005	0.185	0.745	0.001
F (p)	1.862 (.055)			4.467 (.000)			5.483 (.000)		
Adj. R ²	0.055			0.189			0.231		
df	139, 10			139, 10			139, 10		

	Social performance of justice			Tolerance for opposing views			News preferences for opposing views			News preferences for majority views		
	β	ρ	η_p^2	β	ρ	η_p^2	β	ρ	η_p^2	β	ρ	η_p^2
Constant	1.465	0.000	0.087	2.869	0.000	0.401	2.937	0.002	0.068	3.770	0.000	0.090
Diversity Group	1.098	0.032	0.033	1.443	0.000	0.097	0.903	0.054	0.026	0.748	0.159	0.014
Narrow Group	0.883	0.035	0.032	0.577	0.061	0.025	0.487	0.204	0.012	0.955	0.025	0.036
Political position	0.731	0.046	0.029	0.850	0.002	0.068	0.373	0.265	0.009	0.554	0.134	0.016
App usage (hrs)	0.012	0.195	0.012	0.013	0.062	0.025	0.022	0.009	0.049	0.017	0.060	0.025
Ext. news usage	0.009	0.263	0.009	0.018	0.005	0.055	0.026	0.000	0.088	0.013	0.095	0.020
Gender (f = 1; m = 2)	0.140	0.361	0.006	-0.059	0.603	0.002	-0.065	0.647	0.002	-0.144	0.359	0.006
Age (y)	-0.003	0.574	0.002	-0.011	0.005	0.056	-0.553	0.506	0.003	-0.708	0.441	0.004
Education	0.100	0.002	0.070	0.068	0.004	0.059	0.008	0.768	0.001	-0.043	0.173	0.013
Diversity x Pol. pos.	-1.628	0.053	0.027	-2.567	0.000	0.111	-1.200	0.122	0.017	-0.598	0.492	0.003
Narrow x Pol. pos.	-1.052	0.115	0.018	-0.813	0.098	0.02	-0.687	0.263	0.009	-1.289	0.058	0.026
F (p)	2.978 (.002)			4.971 (.000)			2.755 (.004)			1.635 (.103)		
Adj. R ²	0.119			0.210			0.105			0.041		
df	137, 10			139, 10			139, 10			139, 10		

The independent variables together explain 12% of variance of journalism's societal function. Participants in the narrow and diversity group ascribe a greater capacity to solve society's problems to journalism compared to the control group (part. $\eta_p^2 = .032$ and $.033$, respectively, see column four in Table 1). Also, the more liberal participants and the higher educated, the more participants agree to journalism's capacity. Additionally, the interaction terms reveal a negative sign for the diversity and narrow groups. That is, in both groups, conservative users agree more strongly to journalism's capacity than in the control group. Hence, both recommender situations are related to a greater appreciation of journalism than receiving news chronologically, which is especially true for conservative users.

Tolerance of Opposing Views

On average, participants agree to be able to tolerate politically opposing views ($M = 4.03$, $SD = .72$, $n = 150$).

The model explains 21% of variance in tolerance for opposing views (see column five in Table 1), which represents a comparatively high level of explained variance in our study. Tolerance regarding opposing views is clearly higher in the diversity group than in the control group ($\eta_p^2 = .097$) and tends to be higher in the narrow group ($\eta_p^2 = .025$). The difference between diversity and narrow group is significant regarding tolerance for opposing views according to a fixed-effects ANOVA testing the diversity against the narrow group ($F(10, 139) = 4.971, p < .001, \beta = .866, p = .043, \eta_p^2 = .029$). Moreover, younger age and higher education are related to higher tolerance for opposing views, explaining about 6% of variance in tolerance of opposing views each.

Furthermore, the political position explains variance in tolerance. A liberal as opposed to conservative political position is related to higher tolerance for opposing views ($\eta_p^2 = .068$). According to the interaction effect, an increase of tolerance is lower for users with a liberal political position as opposed to conservative political orientation in the diversity group ($\eta_p^2 = .11$). That is, only the diversity news recommendation condition is clearly related to tolerating opposing views (accounts for 10% of variance), and especially so for politically conservative users (11% of variance).

News Diversity Preferences

Overall, participants prefer to receive news with opposing views ($M = 3.72, SD = .852, n = 150$) over news with majority views ($M = 3.56, SD = .908, n = 150$) (one-sample t -test, $t(149) = 2.30, p = .023$).

Coefficients for the model explaining preferences for opposing views can be interpreted as different from zero ($F = 2.755, p = .004$, column six in Table 1). Yet, the independent variables fail to explain a sufficient amount of variance for the preference of majority views ($F = 1.635, p = .103$, column seven in Table 1). Still, the comparison might be insightful for future studies. Comparing news preferences across groups, users in the narrow group prefer majority opinions more strongly ($\eta_p^2 = .036$) while users in the diverse group tend to prefer being informed about opinions that are opposite to their own more strongly ($\eta_p^2 = .026$). Opposing views are also preferred by heavy news users, using the news app and external news more often. Majority views in news are not related to further independent variables in our model. These results suggest that recommender systems may nudge users to preferences according to a prevalent recommender design. While heavy news users across all groups prefer opposing views, the diversity condition may nudge all its users to appreciate cross-cutting content.

Discussion

This study set out to introduce the notion and implementation of a normative recommendation system and evaluate the utility of such a system for diverse news from a user perspective. Accuracy-optimized recommender algorithms are a popular choice for creating personalized news feeds for platforms, giving users more of what they already see. They are safe guesses with the purpose of increasing user engagement. Our results now challenge this practice.

First, similar usage rates across experimental conditions indicate that news consumption largely follows automated recommendations. A post-hoc content analysis

suggests that participants in the control group consumed pro-attitudinal news in line with previous findings (Wojcieszak et al. 2021) and the diverse condition thus included cross-cutting news. However, participants did not perceive diversity differences across conditions.

Regarding instrumental utility, the diversity-optimized algorithm is able to match the accuracy-optimized recommendation strategy. Utility similarity across groups may indicate the absence of a backfire effect as found in Bail et al. (2018). It also suggests that utility is a strong determinant for news selection as found in Mummolo (2016).

Concerning societal externalities, results reveal that diverse recommendations are related to ascribing journalism the capacity to solve problems in society, especially for conservative as opposed to liberal leaning participants. This indicates that a more diverse news supply may restore a social consensus among conservatives concerning the performance of journalism. Such a supply may signal impartiality to users, which is a media performance dimension (Steppat, Castro Herrero, and Esser 2020). Diverse news exposure thus could contribute to social cohesion as suggested in Bernstein et al. (2021). Results also suggest that recommendations enhance tolerance for opposing views in line with Mutz (2020), especially for politically conservative users. Showing stronger effects for conservatives may be a result of a stronger selective exposure of partisan news of conservatives as shown in previous research (Hmielowski, Hutchens, and Beam 2020). Our conservative participants may have appreciated cross-cutting content recommended in the diversity condition, manifesting in their attitudes towards journalism and opposing views. Hence, suggesting that diverse recommendations asymmetrically increase openness towards opposing views and appreciation of journalism as a societal institution, our study extends findings of Wojcieszak et al. (2021). However, despite an increased openness to counter-attitudinal views, diversity-optimized recommendations do not enhance political knowledgeability or political participation of users. These findings may be due to instrumental utility similarity. Moreover, while in Wojcieszak et al. (2021), participants were aware of an automated recommendation situation, our participants were not. Hiding the experimental conditions may have prevented a backfire effect (Bail et al. 2018) that could manifest in lower instrumental utility or increased political activism. While previous research suggests that moderate levels of cross-cutting content increase voter turnout (Castro Herrero and Hopmann 2018), our results show no relation between diverse recommendations and political knowledgeability or participation.

Results concerning changing news preferences suggest that recommender systems may nudge users to preferences according to a prevalent recommender design as assumed previously (Thaler and Sunstein 2009), despite people not being aware of an automated news recommendation situation. While these effects are small and indicate a tendency, they should not be overlooked. Our results suggest that accuracy-optimized recommendations may reinforce users' own news preference while experiencing diverse recommendations may nudge users towards preferring opinion diversity. The latter can prevent partisan filter bubbles as suspected (Milano et al. 2020).

Overall, results suggest that diversity as a design criterion can comply with the corporate-technological and the socio-cultural regimes (Geels 2004), securing user utility and providing positive externalities. Specifically, diverse news recommendations may entail a de-polarizing capacity for democratic societies.

Limitations and Future Research

We want to reiterate that the interpretation of diversity used in the experiment is but one possibility of how such a distribution across the political spectrum could look like. The algorithm presented here offers a sandbox to define an interpretation of diversity that can be custom-tailored to a specific political landscape. Being such a tool, the algorithm is unable to answer the question of what distribution best benefits society and democratic institutions. We see this challenge as part of a normative and ethical discussion outside the scope of this article.

The current rating and recommendation pipeline ignores the content of news articles. It solely focuses on reading metrics and ratings of the participants. Hence, the political labels of articles might say more about their readership than about the article contents. In order to account for this shortcoming, entity and sentiment recognition would need to be implemented.

Regarding the interpretability of effects of diverse news recommendations on news consumption, the manipulation check advises caution. While news usage adheres to the recommended articles in each group, the news consumed in each treatment situation was not perceived as differently according to the manipulation test. Future research could vary levels of cross-cutting news to identify a possible diversity perception threshold.

Present results rely on small group sizes. Future research needs to validate present results with greater samples and within different political and media system contexts. This could go hand in hand with extending the duration of the experiment to half a year or even longer periods. Since present results suggest that diversity-optimized news recommenders might provide normative utility specifically for users on the conservative political spectrum, it can be worth for future research to consider political positions to a greater extent, for instance by over-sampling users from politically extreme positions. Also, as results indicate a de-polarizing capacity of diversity recommenders, comparative studies including polarized political and media systems such as the U.S. will be valuable.

Conclusions

A healthy public sphere is often considered one that includes a variety of political positions while balancing majority and minority opinions within society (Helberger 2019; McQuail 2005). A normative, diversity-optimized recommender system can increase the variety of opinions that become part of public spheres by including distant items to a recommendation set. Our study suggests that society benefits from diverse news recommendations due to a de-polarizing potential reflecting in user attitudes and news preferences.

In this article, we present a model of how to operationalize the concept of diversity for news recommendations as a step towards normative recommender systems design. We developed an algorithm that is based on the idea that diversity can be defined as a distribution of news articles in an n -dimensional space of political orientation. Our approach focuses on exposure diversity, i.e., diversity through the use of news. The algorithm is highly customizable and can account for different diversity models.

We hope to bring about a positive change in the way news recommenders account not only for the preferences of the users, but also for the normative needs of society as a whole. We would like to continue the public, multi-lateral debate on accountability, responsibility, and transparency of recommender systems, discussing present as well as future threats and opportunities, further enhancing the algorithmic social contract (Rahwan 2018), and the field of recommender system ethics (Milano et al. 2020).

Notes

1. For the complete description of how political scores are computed, please see the official *Smartvote* documentation (German only): https://sv19.cdn.prismic.io/sv19%2Ffcff287e-aa5e-4f5c-ae46-f3d53d6e0b30_methodology_smartmap_de.pdf.
2. See the documentation of *Parteienkompass* for a complete catalogue of the questions (German only): https://d251wzmx05xjsf.cloudfront.net/assets/parteienkompass_fragenzeileung.pdf.
3. There were no users with scores of -1.0 and $+1.0$. The maximum rounded scores were -0.8 and $+0.8$.

Acknowledgements

We would like to express our gratitude to the participating news outlets *Blick*, *Neue Zürcher Zeitung*, *Tages-Anzeiger*, SRF Online, *Weltwoche*, and *WOZ Die Wochenzeitung*, as well as to both the *UZH Digital Society Initiative* and the *Hasler Foundation* for their support.

Disclosure Statement

No potential competing interest was reported by the author(s).

Funding

This work was supported by the UZH Digital Society Initiative under a Grant of the DSI Excellence Program; and the Hasler Foundation under a Grant on Political News Diversity.

ORCID

Lucien Heitz  <http://orcid.org/0000-0001-7987-8446>
 Juliane A. Lischka  <http://orcid.org/0000-0002-4845-0946>
 Alena Birrer  <http://orcid.org/0000-0002-7141-7028>
 Bibek Paudel  <http://orcid.org/0000-0002-1239-7207>
 Suzanne Tolmeijer  <http://orcid.org/0000-0002-8584-5656>
 Laura Laugwitz  <http://orcid.org/0000-0001-8527-2504>
 Abraham Bernstein  <http://orcid.org/0000-0002-0128-4602>

References

- Atkin, Charles. 1973. "Instrumental Utilities and Information Seeking." In *New Models for Mass Communication Research*, edited by Peter Clarke, 205–242. Beverly Hills, CA, USA: Sage.
- Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. "Exposure to Opposing Views on Social Media Can Increase Political Polarization." *Proceedings of the National Academy of Sciences of the United States of America* 115 (37): 9216–9221.

- Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. 2015. "Political science. Exposure to ideologically diverse news and opinion on Facebook." *Science (New York, N.Y.)* 348 (6239): 1130–1132.
- Bernstein, Abraham, Claes de Vreese, Natali Helberger, Wolfgang Schulz, Katharina Zweig, Christian Baden, Michael A. Beam, et al. 2021. "Diversity in News Recommendation." *Dagstuhl Manifestos* 9 (1): 43–61.
- Carlsson, B., and R. Stankiewicz. 1991. "On the Nature, Function and Composition of Technological Systems." *Journal of Evolutionary Economics* 1 (2): 93–118.
- Castro Herrero, Laia, and David Nicolas Hopmann. 2018. "The Virtue of Moderation: A Cross-National Analysis of Exposure to Cross-Cutting Information and Turnout." *International Journal of Public Opinion Research* 30 (4): 518–539.
- Christoffel, Fabian, Bibek Paudel, Chris Newell, and Abraham Bernstein. 2015. "Blockbusters and Wallflowers." In *RecSys'15*, edited by Hannes Werthner, Markus Zanker, Jennifer Golbeck, and Giovanni Semeraro, 163–170. New York, NY: RecSys and Association for Computing Machinery and ACM Conference on Recommender Systems and ACM Recommender Systems Conference, ACM Association for Computing Machinery.
- de Benedictis-Kessner, Justin, Matthew A. Baum, Adam J. Berinsky, and Teppei Yamamoto. 2019. "Persuading the Enemy: Estimating the Persuasive Effects of Partisan Media with the Preference-Incorporating Choice and Assignment Design." *American Political Science Review* 113 (4): 902–916.
- Diakopoulos, Nicholas. 2019. "Towards a Design Orientation on Algorithms and Automation in News Production." *Digital Journalism* 7 (8): 1180–1184.
- Eady, Gregory, Jonathan Nagler, Andy Guess, Jan Zilinsky, and Joshua A. Tucker. 2019. "How Many People Live in Political Bubbles on Social Media? Evidence from Linked Survey and Twitter Data." *SAGE Open* 9 (1): 215824401983270.
- Ferree, Myra M., William A. Gamson, Jürgen Gerhards, and Dieter Rucht. 2002. "Four Models of the Public Sphere in Modern Democracies." *Theory and Society* 31 (3): 289–324.
- Fuchs, Christian. 2014. *Social Media*. Beverly Hills, CA, USA: Sage.
- Geels, Frank W. 2004. "From Sectoral Systems of Innovation to Socio-Technical Systems." *Research Policy* 33 (6–7): 897–920.
- Haim, Mario, Andreas Graefe, and Hans-Bernd Brosius. 2018. Burst of the Filter Bubble? *Digital Journalism* 6 (3): 330–343.
- Hameleers, Michael, Bos Linda, Fawzi Nayla, Reinemann Carsten, Andreadis Ioannis, Corbu Nicoleta, Schemer Christian, et al. 2018. "Start Spreading the News." *The International Journal of Press/Politics* 23 (4): 517–538.
- Harambam, Jaron, 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." *Proceedings of the 13th ACM Conference on Recommender Systems* 2019 (2019): 69–77.
- Helberger, Natali. 2019. "On the Democratic Role of News Recommenders." *Digital Journalism* 7 (8): 993–1012.
- Helberger, Natali, Karl Karppinen, and Lucia D'Acunato. 2018. "Exposure Diversity as a Design Principle for Recommender Systems." *Information, Communication and Society* 21 (2): 191–207.
- Hellmueller, Lea, Juliane A. Lischka, and Edda Humprecht. 2020. "Shaping (Non)-Discursive Social Media Spaces: Cross-national Typologies of News Organizations' Heavy Commenters." *New Media & Society*. Advanced online publication.
- Hennig-Thurau, Thorsten, André Marchand, and Paul Marx. 2012. "Can Automated Group Recommender Systems Help Consumers Make Better Choices?" *Journal of Marketing* 76 (5): 89–109.
- Heuer, Hendrik. 2020. *Users & Machine Learning-Based Curation Systems*. Bremen, Germany: Universität Bremen.
- Hmielowski, Jay D., Myiah J. Hutchens, and Michael A. Beam. 2020. "Asymmetry of Partisan Media Effects?: Examining the Reinforcing Process of Conservative and Liberal Media with Political Beliefs." *Political Communication* 37 (6): 852–868.
- Holzinger, Andreas, ed. 2008. *HCI and Usability for Education and Work*. Berlin, Heidelberg: Springer.

- Karimi, Mozghan, Jannach Dietmar, and Jugovac Michael. 2018. "News Recommender Systems – Survey and Roads Ahead." *Information Processing & Management* 2018 (2018): 1203–1227.
- Kruikemeier, Sanne, and Adam Shehata. 2017. "News Media Use and Political Engagement among Adolescents: An Analysis of Virtuous Circles Using Panel Data." *Political Communication* 34 (2): 221–242.
- Kunaver, Matevž, and Tomaz Pozrl. 2017. "Diversity in Recommender Systems, a Survey." *Knowledge-Based Systems* 123 (2017): 154–162.
- Li, Xigen. 2014. "Perceived Channel Efficiency and Motivation and Orientation of Information Seeking as Predictors of Media Dependency." *Telematics and Informatics* 31 (4): 628–639.
- Linder, Wolf, and Sean Mueller, eds. 2021. *Swiss Democracy*. 4rd ed. Basingstoke, UK: Palgrave Macmillan.
- Ma, Hao, Xueqing Liu, and Zhihong Shen. 2016. "User Fatigue in Online News Recommendation." Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, Geneva, Switzerland, 1363–1372.
- Mager, Astrid. 2012. "Algorithmic Ideology." *Information, Communication and Society* 15 (5): 769–787.
- McQuail, Denis. 2005. *McQuail's Mass Communication Theory*. 5th ed. London, UK: Sage.
- Michael, Laver. 2020. "Measuring Policy Positions in Political Space." *Digital Journalism* 7 (8): 1180–1184.
- Milano, Silvia, Taddeo Mariarosaria, and Floridi Luciano. 2020. "Recommender Systems and Their Ethical Challenges." *AI & SOCIETY* 35 (4): 957–967.
- Möller, Judith, Damian Trilling, Natali Helberger, and Bram van Es. 2018. "Do Not Blame It on the Algorithm: An Empirical Assessment of Multiple Recommender Systems and Their Impact on Content Diversity." *Information, Communication & Society* 21 (7): 959–977.
- Mothes, Cornelia, and Jakob Ohme. 2019. "Partisan Selective Exposure in Times of Political and Technological Upheaval: A Social Media Field Experiment." *Media and Communication* 7 (3): 42–53.
- Müller, Jan-Werner. 2021. "Democracy's Critical Infrastructure: Rethinking Intermediary Powers." *Philosophy & Social Criticism* 2021: 0191453720987874.
- Mummolo, Jonathan. 2016. "News from the Other Side: How Topic Relevance Limits the Prevalence of Partisan Selective Exposure." *The Journal of Politics* 78 (3): 763–773.
- Mutz. 2020. "The Structural Transformation of the Public Sphere." *Digital Journalism* 7 (8): 1180–1184.
- Nechushtai, Efrat, and Seth C. Lewis. 2019. "What Kind of News Gatekeepers Do We Want Machines to Be?" *Computers in Human Behavior* 90 (2019): 298–307.
- Newman, Nic, Fletcher Richard, Kalogeropoulos Antonis, Schulz Anne, Sigme Andi, and Rasmus Kleis Nielsen. 2020. *Reuters Institute Digital News Report 2020*. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2020-06/DNR_2020_FINAL.pdf.
- Paudel, Bibek, Fabian Christoffel, Chris Newell, and Abraham Bernstein. 2017. "Updatable, Accurate, Diverse, and Scalable Recommendations for Interactive Applications." *ACM Transactions on Interactive Intelligent Systems* 7 (1): 1–34.
- Peifer, Jason T. 2018. "Perceived News Media Importance." *Communication Methods and Measures* 12 (1): 55–79.
- Peterson, Erik, Sharad Goel, and Shanto Iyengar. 2021. "Partisan Selective Exposure in Online News Consumption: Evidence from the 2016 Presidential Campaign." *Political Science Research and Methods* 9 (2): 242–258.
- Pigou, Arthur C. 1932. *The Economics of Welfare*. London, UK: Macmillan.
- Pingree, Raymond J. 2011. "Effects of Unresolved Factual Disputes in the News on Epistemic Political Efficacy." *Journal of Communication* 61 (1): 22–47.
- Rahwan, Iyad. 2018. "Society-in-the-Loop: Programming the Algorithmic Social Contract." *Ethics and Information Technology* 20 (1): 5–14.
- Steppat, Desiree, Laia Castro Herrero, and Frank Esser. 2020. "News Media Performance Evaluated by National Audiences: How Media Environments and User Preferences Matter." *Media and Communication* 8 (3): 321–334.
- Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge*. New York, NY, USA: Penguin Books.

- Thurman, Neil, Judith Moeller, Natali Helberger, and Damian Trilling. 2019. "My Friends, Editors, Algorithms, and I: Examining Audience Attitudes to News Selection." *Digital Journalism* 7 (4): 447–469.
- Vrijenhoek, Sanne, Mesut Kaya, Nadia Metoui, Judith Möller, Daan Odijk, and Natali Helberger. 2020. "Recommenders with a Mission: Assessing Diversity in Newsrecommendations." In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval: 173–183*.
- Wojcieszak, Magdalena, Arti Thakur, João Fernando Ferreira Gonçalves, Andreu Casas, Ericka Menchen-Trevino, and Miriam Boon. 2021. "Can AI Enhance People's Support for Online Moderation and Their Openness to Dissimilar Political Views?". *Journal of Computer-Mediated Communication* 26 (2021): 223–243.
- Ziarani, Reza Jafari, and Reza Ravanmehr. 2021. "Serendipity in Recommender Systems: A Systematic Literature Review." *Journal of Computer Science and Technology* 36 (2): 375–396.

Appendix A: Recommendation Algorithm

Algorithm 1 describes how to calculate *desired_distribution* based on *score_u*, *art_distribution*. It first looks for the largest value *max* for *n_a* among all tuples in *art_distribution* (Line 1) and initializes a new recommendation list *desired_distribution* (Line 2). In case there are articles left to recommend in *art_distribution*, all tuples where *n_a*=*max* are grouped together (Line 4) in *tuples_{filter}*. Next, the tuple with the score furthest away from *score_u* is selected (Line 5). The article score *score_a* of the currently selected tuple is then added to the recommendation list (Line 6). Afterwards, the tuple value *n_a* is decremented, for we now have one article of this score in *desired_distribution* (7). In case there are no more tuples left where *n_a*=*max*, we reduce *max* and move to the next set of tuples to recommend (Line 9). This process is repeated until no value *n_a* is larger than 1.

Algorithm 1: Procedure for translating distributions into reading lists

```

Input: scoreu, art_distribution // art_distribution defined as list of tuples
scorea, na
Output: desired_distribution
1 max=maxScore(art_distribution) // find largest value of na among all tuples
2 desired_distribution=[] // initialize recommendation list
3 while maxScore(art_distribution)>1 do
4 | tuplefilter=filterTuples(art_distribution, max) // get tuples of
art_distribution where na=max
5 | tuplemax=maxDistance(tuplesfilter, scoreu) // get tuple where |scorea -
userScore| is maximal
6 | appendScore(tuplemax, desired_distribution) // add scorea of tuple-
to desired_distribution
7 | reduceCount(tuplemax) // decrement na of tuplemax by 1
8 | if maxScore(art_distribution) ≠ max then
9 | | max=max - 1
10 return desired_distribution

```

Appendix B: Overview of Normative Diversity Distributions

The following section shows the specific values used for recommendation purposes. For each political dimension, a total of 9 groups or buckets are available: [-0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8].³ We operate with two-dimensional user scores *score_u*=*[art_distributionA, art_distributionB]*. To arrive at the two-dimensional distribution *art_distribution*, the algorithm iterates through all entry combinations in *art_distribution* and *art_distribution*, picking the smaller of the two values. One tuple of the two-dimensional *art_distribution* consists of *[score_{ij}, n_d]*

Table 2. Diversity-optimized recommendations used the following distributions for $art_distribution_A$ and $art_distribution_B$:

$userscore [-0.8]$:	[2	3	4	4	4	2	1	1	1]
$userscore [-0.6]$:	[1	3	4	4	4	3	1	1	1]
$userscore [-0.4]$:	[1	2	4	4	4	3	2	1	1]
$userscore [-0.2]$:	[1	2	3	4	4	4	2	1	1]
$userscore [+0.0]$:	[1	1	3	4	4	4	2	1	1]
$userscore [+0.2]$:	[1	1	2	4	4	4	3	2	1]
$userscore [+0.4]$:	[1	1	2	3	4	4	4	2	1]
$userscore [+0.6]$:	[1	1	1	3	4	4	4	3	1]
$userscore [+0.8]$:	[1	1	1	2	4	4	4	3	2]

Table 3. Accuracy-optimized recommendations used the following distributions for $art_distribution_A$ and $art_distribution_B$:

$userscore [-0.8]$:	[4	1	0	0	0	0	0	0	0]
$userscore [-0.6]$:	[1	4	1	0	0	0	0	0	0]
$userscore [-0.4]$:	[0	1	4	1	0	0	0	0	0]
$userscore [-0.2]$:	[0	0	1	4	1	0	0	0	0]
$userscore [+0.0]$:	[0	0	0	1	4	1	0	0	0]
$userscore [+0.2]$:	[0	0	0	0	1	4	1	0	0]
$userscore [+0.4]$:	[0	0	0	0	0	1	4	1	0]
$userscore [+0.6]$:	[0	0	0	0	0	0	1	4	1]
$userscore [+0.8]$:	[0	0	0	0	0	0	0	1	4]

(selected from either Table 2 or Table 3), where

$$score_{ij} = [art_distribution_{A,i}, art_distribution_{B,j}], i, j \in \{1 - 11\}, n_a = \min(dimension_{A_i}, dimension_{B_j}). \quad (4)$$