Nudging Towards Exposure Diversity: Examining the Effects of News Recommender Design on Audiences' News Exposure Behaviours and Perceptions

Abstract

Scholars are increasingly studying how news recommenders can provide audiences with diverse news offers. However, despite their increased interest, little research has been conducted on how news audiences engage with news recommenders. This article assesses how a news recommender's design affects people's news exposure behaviours and perceptions. To do so, we conducted an online experiment (n = 337) in a virtual news environment with four different news recommenders, including three news recommenders that aimed to stimulate diverse news exposure. The results show that all diversity-based news recommenders steered users towards more diverse exposure behaviour, with a personalised diversity-based news recommender being the most effective. We also found that people using a diversity-based news recommender did not think they read more diversely, pointing towards a so-called 'diversity paradox'. We discuss what this paradox means and how it relates to the secretive nature of news algorithms. We also make a call to policymakers, news organisations and scholars to give transparency and diversity-based news recommenders a more pronounced place in the news environment and in future news recommendation research.

Keywords: diversity paradox; diversity-based news recommender; exposure diversity; news personalisation; personalised diversity; public-service algorithms

Introduction

As news organisations have shifted from a mass communication model towards a personalised news and recommendation model (Bodó, 2019; Thurman & Schifferes, 2012), scholars are increasingly paying attention to how news recommenders work and how their selection mechanisms should be designed. Traditionally, news recommenders use selection mechanisms, such as content-based and collaborative news filtering, that strive for similarity by looking at people's previous news exposure behaviour and that of similar user profiles (Karimi et al., 2018). In doing so, they aim to provide users with the most relevant suggestions possible (Castells et al., 2015). However, in more recent literature, there is a growing idea among scholars to embed democratic values, such as news diversity, into news recommenders to promote better-informed citizens and an improved public sphere (see, e.g., Helberger, 2019; Helberger et al., 2018). Here, news articles are not recommended based on similarity but on what might make an individual read more diversely. This idea also corresponds with the concept of a 'public-service algorithm' in which public-service media explore the use of algorithmic power to preserve societal values in a digital environment (Jones & Jones, 2019).

Despite high aspirations towards the functions and outcomes of public-service news recommenders (Joris et al., 2021), little research has been done on how people engage with news recommenders. As Loecherbach and Trilling (2019) explained, most studies lack a realistic setting (with real-time news or existing news recommenders) or do not focus on how users interact with or evaluate these news recommenders. As a result, it is unclear how news recommenders and their selection mechanisms affect people's news exposure behaviour. Especially in case of public service news recommenders that target diverse news exposure, this question is unexplored and increasingly pertinent to assess their their value.

Therefore, this article provides insights into audiences' news exposure behaviours and perceptions when using different news recommenders. To do so, we conducted a web-based experiment in which we differentiated between four news selection mechanisms: content-based similarity, random diversity, open diversity and reflective diversity. The first is one of the most dominant mechanisms used by news organisations and focuses on giving people news topics in which they are interested. The other mechanisms, in contrast, aim to make a user's news exposure more diverse in terms of topics. The random diversity news selection mechanism randomly recommends news articles, while the open diversity and reflective diversity mechanisms also make use of personalisation features that underlie content-based similarity news filtering to achieve their goals. This is what Joris et al., (2021) calls 'personalised diversity'.

Using news diversity as a selection mechanism, our article also challenges people's selective exposure patterns. Studies have shown that news consumers generally prefer attitude-consistent topic information over counter-attitudinal information (Knobloch-Westerwick, 2014; Stroud, 2017), indicating that news exposure is strongly biased by people's interests and pre-existing beliefs. As a diversity-based news recommender aspires to overcome these biases, our study tested to what degree an algorithm may 'push' audiences towards more exposure diversity.

Literature Review

Our literature review begins with an overview of how news recommenders work and how their outcomes have traditionally been evaluated. We then discuss the concept of news exposure, which has been increasingly considered in the context of news recommendation. Finally, we focus on the idea underlying diversity-based news recommenders and formulate hypotheses for our study.

News Recommenders and Their Designs

News recommenders are increasingly prevalent in the current news environment to automate the content distribution process and filter the growing abundance of online news. Due to their practical relevance, several types of news recommenders have been explored and developed, leading to a wide variety of how these systems are designed. In general, news recommender systems consist of two design attributes that determine how news articles are selected, sorted and ranked (Joris et al., 2021). A first design attribute is concerned with the data source being used (Karimi et al., 2018). These data sources may range from individuals' previous consumption behaviours (i.e., content-based news selection) to the previous consumption behaviour of others, such as friends or people with similar reading behaviour (i.e., collaborative news selection). Most commonly, researchers use a combination of both approaches (i.e., hybrid news selection) (Karimi et. al., 2018).

A second important design attribute is concerned with a recommendation goal or the evaluation metric used to assess an outcome (Joris et al., 2021). Most commonly, researchers use accuracy measures (Jannach et al., 2012; Karimi et al., 2018). These aim to provide individuals with the most relevant news items and, therefore, look at the similarity between users' interests and the news content (e.g., Maximal Marginal Relevance, see e.g., Karimi et al., 2018). Recently, the importance of other evaluation metrics, such as novelty and diversity, has also been stressed (De Pessemier et al., 2015). Novelty is understood as the difference between users' present and past experiences, whereas diversity relates to internal differences within a set of recommended items (Castells et al., 2015).

Despite this evolution towards more user-centric evaluations, most evaluations are still conducted offline in an experimental setting that makes use of existing datasets and a protocol that models user behaviour (see, e.g., Garcin et al., 2014). As a result, the external validity of these studies is relatively low (Karimi et al., 2018; Shani & Gunawardana, 2011). Moreover, these studies often focus on diversity *within* a diverse set of news recommendations (see, e.g., Haim et al., 2018; Möller et al., 2018), while recommending a diverse set of news articles does not guarantee that people effectively *consume* a diverse range of news articles. As a result, little is known about the effect of news recommender systems on the diversity to which people are exposed in terms of news content (i.e., exposure diversity).

Exposure Diversity

The concept of exposure diversity has received increasing scholarly attention in the last couple of years. In particular, as the news landscape has shifted from traditional mass media towards personalised news and platforms (Thurman & Schifferes, 2012), an increasing number of scholars are concerned that personalised content and services could limit the diversity of news content people are exposed to (i.e., filter bubble, Zuiderveen Borgesius et al., 2016) and could stimulate people to more easily ignore stories they deem irrelevant or counter-attitudinal (i.e., echo chambers) (Beam, 2014;

Dylko et al., 2017). Although many of these concerns have been nuanced (see, e.g., Bruns, 2019), exposure diversity remains an increasingly popular topic in the field of communication science.

Exposure diversity can be easily defined as all the content that an individual selects, as opposed to all the content that is available (McQuail, 1992). It is thus concerned with the extent to which selected news content is diverse in terms of content dimensions, such as demographics, topics, ideas and viewpoints (Joris et al., 2020). From a normative viewpoint, exposure diversity can be understood as a necessary condition for human progress and a well-functioning democracy (Helberger, 2011). It is not an end in itself but a means for users to be engaged in society and inform themselves about the various viewpoints and perspectives (Helberger, 2012).

From a conceptual viewpoint, the concept of exposure diversity must be seen within the general concept of 'news diversity', as it focuses on one particular side of diversity—the consumption side—and what is being selected. As such, its conceptual meaning should also be seen in relation to the various conceptual and normative assumptions related to the general concept of news diversity. As Joris et al., (2020) argued, this means that several normative and conceptual dimensions should be considered to develop a well-founded conceptual understanding. First, a researcher should decide what the 'optimal outcome' of diversity is. In general, a researcher may choose between two normative benchmarks: 'open diversity' and 'reflective diversity'. Open diversity claims that diversity is an equal representation of all possible categories, whereas reflective diversity argues that diversity should reflect proportions in society or in journalistic offerings. Second, a researcher should select what is being studied in terms of content, also called 'diversity dimensions'. Here, scholars can generally choose between various content dimensions, such as news topics, the actors involved or the political parties mentioned in the text. However, these content dimensions are not limited to quantitative dimensions. They could also include more qualitative aspects of news content, such as the discourses and arguments used to describe news event. The latter aspects are more difficult to detect, although they are at least as important from a democratic viewpoint (Joris et al., 2020)

Diversity-Based News Recommenders

Within communication science, scholars are increasingly reconsidering how recommender systems should be developed and how they can be used for purposes that support news organisations in their normative role of informing the public (Fields et al., 2018; Jones & Jones, 2019; Van den Bulck & Moe, 2018). A prominent idea in this discussion is the idea of so-called diversity-based news recommenders, also called diversity-enhancing news recommenders or diversity-oriented news recommenders (Bodó, 2019; Loecherbach et al., 2021). In contrast to other news recommender systems, the ultimate goal of these systems is not to provide individuals with more of the same news but to achieve diversity *within* individuals' news exposure behaviour. Hence, diversity-based news recommenders primarily focus on the concept of 'exposure diversity' and what might make an individual read more diversely. This also means that a diversity-based news recommender does not simply provide individuals with a diverse set of news recommendations (as this does not guarantee diverse news exposure) but actively searches for ways to make individuals' news exposure more diverse. As Joris et al. (2021) explained, these ways could be like similarity-based news recommenders, although their goals are significantly different. Indeed, diversity-based news recommender could also make use of users' interests but only as a means to reach their goals (see the concept of 'personalised diversity', Joris et al. (2021)).

Due to the recency of this idea, the literature lacks clear insights into the effects of diversity-based news recommenders on people's news exposure behaviours and perceptions. Most audience-centred studies have focused on traditional news recommenders and their impact on exposure behaviour (see, e.g., Yang, 2016). Only a few studies have focused on news recommenders and editorial values. For example, Lu et al. (2020) tested a news recommender that integrated two editorial concepts: dynamism and serendipity. They found that the editorial news recommender effectively steered users to more diverse exposure behaviour, with increased item coverage from the provider's perspective (i.e., more engagement). Based on these insights, we hypothesise the following:

- **H1:** Diversity-based news recommenders stimulate measured exposure diversity.
- **H2**: Diversity-based news recommenders stimulate audiences to select more articles.

Lu et al. (2020) did not measure how these types of news recommenders affected audiences' perceptions towards their own exposure diversity behaviour or to the news recommender in general. Joris et al. (2021) argued that perceptions are increasingly important, as they may determine how and when audiences accept or reject recommendations. Previous research on digital news consumption has shown that there might be a discrepancy between the logged behaviour of individuals and self-reported behaviour (Prior, 2009; Parry et al., 2021). Therefore, we argue that it is also relevant to take audiences' perceived exposure diversity and appreciation of a news recommender into account. Audiences' perceived exposure diversity can be understood as a self-reported measure in which participants reflect upon their exposure diversity behaviour (Parry et al., 2021). In the current study, news recommender appreciation is defined as the participants' attitude towards a news recommender's selection choices (Joris et al., 2021).

As research has shown that logged behaviour moderately correlates with self-reported behaviour (Prior, 2009; Parry et al., 2021), we hypothesised that diversity-based news selection has less impact on perceived exposure diversity, in contrast to its impact on measured exposure diversity. However, based on Lu et al. (2020), we hypothesised that it would still have a positive impact on perceived exposure diversity:

- **H3:** Diversity-based news recommenders stimulate perceived exposure diversity.

Research has also found evidence that diversity-based news selection is rarely preferred by news audiences, in contrast to similarity-based news recommenders (Joris et al., 2021). Therefore, we hypothesised that diversity-based news recommenders would negatively affect audiences' attitudes:

- **H4**: Diversity-based news recommenders lower audiences' appreciation of the news recommender.

Methodology

As the literature review shows, several questions and hypotheses can be drawn about how audiences' behaviours and perceptions might be affected by the selection mechanisms underlying news recommenders. To test these hypotheses, we conducted a web-based experiment in which participants were asked to read online news articles. Below, we describe our sample and stimulus material, explain and justify the data collection procedure and discuss how we measured our (in)dependent variables. To limit bias and distinguish between predictions and postdictions (Nosek et al., 2018), we registered our research plan in advance via the Open Science Framework.[1]

Sample

Based on a convenience sampling strategy, 16 students from Ghent University were asked to invite people to participate in this study. To determine the ideal sample size (N = 341), we conducted a power analysis in *G*Plus* with five parameters: a statistical power of 0.80 (Cohen, 2013), a statistical significance (α) of 0.05, four conditions, three covariates and a moderate effect size (f) of 0.18. For the latter parameter, we looked at the results of previous research on selective exposure to information (Hart et al., 2009) and news recommender systems (Yang, 2016), in which effect sizes tended to be within the range of small to moderate.

The inclusion criteria for participating in this study were as follows: (1) being older than 15 years and (2) having sufficient knowledge of the Flemish language to read news articles. The response rate was 45.96% (n = 415) before data cleaning and 37.32% after data cleaning (n = 337). On average, the participants in our sample were 35 years old (range = 18–84, SD = 16.74). 61.40% were female and 38.60% were male (see Appendix 2 for a full overview).

Stimulus Material

To collect the data for this study, we built an online news website that aggregated news articles from different news outlets. To design this website, we looked at the characteristics of popular news aggregator platforms that use implicit forms of news personalisation to recommend news articles (e.g., Google News and Apple News). Based on these platforms, we formulated three design features for our stimulus material: (1) a scrollable website, (2) no external redirects and (3) the use of images and text snippets (see Figure 1). Clicking on a respective article enabled the participant to read the entire article. To overcome selection bias, the author's name and source information were removed.

The news articles in our stimulus material were delivered by four Flemish news organisations: VRT, DPG Media, Mediahuis and Apache. The first three are the leading news companies in Flanders in terms of audience engagement (Picone and Donders, 2020), and the fourth is a digital-only news organisation offering investigative journalism and in-depth stories. Combining the news articles from these four news organisations enabled us to provide an externally valid sample of news articles in the Flemish news market. To manage the news retrieval process, we used an online database that retrieved all news articles, including paid content, via RSS feeds or data dumbs. The system refreshed its content daily and only considered the 5,000 most recent articles. News articles with fewer than 200 characters were automatically removed.

[1] To read our pregistration, see: https://osf.io/ytjem/?view_only=4a346c86488c4b82b1b75b240f42f20c.

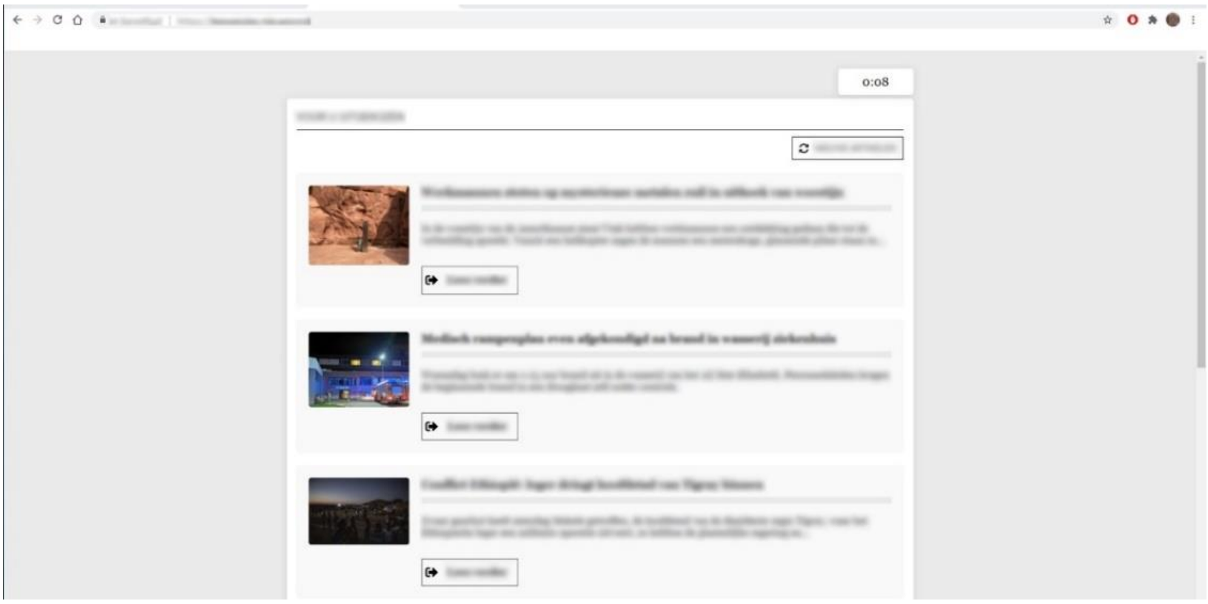


Figure 1. The homepage of our online news website

To label all news articles with topics, a multilabel classification system was built (De Clercq, De Bruyne & Hoste, 2020). The technical details of this system can be found in Appendix 1.

To examine the effect of news recommender design on news exposure behaviour, we developed four recommenders, each using its own mechanism to select and present news articles. The technical aspects of these recommenders are described in Appendix 1. On a conceptual level, these selection mechanisms can be described as follows:

1. Content-based similarity news selection: This news recommender design focuses on users' topical interests and aims to provide news articles that align with these interests. For instance, if a user is exclusively interested in sports, this news recommender design will primarily provide sports news. To reliably measure users' topical interests, we asked respondents to give a score between 1 (= no interest) and 10 (= very interested) on how frequently they read particular news categories. Here, we made use of an IPTC topic list that consists of 17 news categories such as politics, sports or celebrity news. IPTC can be interpreted as a topic taxonomy that helps media organisations and researchers to categorise text and articles (International Press Telecommunications Council, n.d.). This question was part of the pre-survey in which we asked questions about their news exposure and cognitive personality traits (see below: Data Collection Process).

2. *Random diversity-based news selection:* This news recommender design aims to make news exposure more diverse by selecting news articles randomly. It does not consider a participant's topical interests, but simply make random choices.

3. *Open diversity-based news selection*: This news recommender design aims to make news exposure more diverse but uses a step-by-step approach to nudge people towards a diverse news diet (i.e., personalised news diversity, Joris et al. (2020)). This means that the recommender first aims to find touchpoints in a participant's topical interests and then guides him or her into a more diverse news diet. For instance, for people who dislike politics news or foreign affairs, this news recommender design will not simply start their recommendations with these topics. Instead, it will start with providing news articles about education or economics, which may lie in the closer area of someone's interests, but significantly differs from political news. Over time, the news recommender will slowly

guide users into a more diverse diet, by also providing news articles that lie more closely to topics that the participant dislikes (see Joris et al., 2020). Furthermore, it is important to note that in this news recommender design, diversity is perceived as an equal representation of all possible categories. As such, it strives for equality in terms of topics (i.e., open diversity, see Literature Study for more information) (McQuail & Van Cuilenburg, 1983).

4. *Reflective diversity-based news selection:* This news recommender design is similar as the previous news recommender, as it also uses a step-by-step approach to guide people towards a more diverse exposure repertoire. However, In contrast to the previous design, it perceives news diversity as a mirror of society (i.e., reflective diversity, see Literature Study for more information) (McQuail & Van Cuilenburg, 1983). This means that it does not nudge users towards an equal distribution of news topics (i.e., open diversity), but towards the topic distribution of the journalistic offer. To determine the topic distribution of the journalistic offer, we took a construct week sample of news articles from three news organisations in 2018. In Appendix 2, the results of this assessment are presented with the benchmark distribution of open diversity-based news selection to explicate the difference between these two mechanisms.

Data Collection Process

The participants first had to fill in a pre-survey in which general information was asked about their news exposure (e.g., topical news interests) and cognitive personality traits (e.g., news information overload). Next, the participants were invited to complete a reading session of 12 minutes using our news website, during which we logged their exposure behaviour. To limit sequence effects, invitations were sent one day after the participants completed the pre-survey. Only desktop users could participate. All of the participants were randomly assigned to one of the four conditions and were not informed about this assignment or condition (i.e., blinded experiment). Once completed, the respondents were asked to fill in a post-survey in which they could evaluate the news recommender and its output. Data collection took place between 16–30 November 2020.

Measures

We used a 5-point Likert scale for all dependent measures, where 1 is ‘strongly disagree’ and 5 is ‘strongly agree’ (unless noted). Appendix 3 summarizes all of the independent variables’ descriptives. In Appendix 4, we describe the reliability and validity of the latent constructs. We retained all items with an item-to-total correlation above 0.3 and a Cronbach’s alpha exceeding 0.7 (DeVellis, 2016).

Dependent Measures

Measured exposure diversity: We used a log-based and standardised measure that calculated the similarity between the individual’s topic exposure behaviour and the individual’s ideal topic exposure. The result is a score between 0 and 1, in which 1 stands for a high similarity between the individual’s topic exposure behaviour and the individual’s ideal topic exposure. To determine the ideal of an individual’s topic exposure, we looked at two normative concepts concerned with the optimal outcome of diversity: reflective diversity and open diversity (McQuail & Van Cuilenburg, 1983). Open diversity claims that diversity is an equal representation of all possible categories, whereas reflective diversity argues that diversity should reflect the proportions in society or journalistic offerings. In Appendix 2, a concrete example of both concepts is presented. As both concepts have other normative implications (see, e.g., Joris et al., 2020), we calculated two measures: (1) *measured open diversity* and (2) *measured reflective diversity*. The technical details of these measures can be found in Appendix 1.

Perceived exposure diversity: We used a single-item measure to assess the extent to which people thought they read diverse topics during the experiment.

Appreciation news recommender: We used four items that focused on the affective attitudes towards the selection mechanism of a news recommender. These items were based on the general attitude scale of the revisited UTAUT model (Venkatesh et al., 2003).

Selected articles: We used a log-based measure that calculated the number of articles clicked on by the participants (i.e., the sum index).

Results

We traditionally start our result section with a manipulation check in which we assess the participants' awareness of the conditions to which they were exposed. If successful, it allows us to draw more accurate conclusions related to the relationship between our independent and dependent variables. Next, we move to the results of our analysis of covariance (ANCOVA) in which we evaluate our hypotheses and whether the means of our dependent variables are equal across our independent variables. If so, we will also present a more in-depth analysis that describes which means differ (significantly).

Manipulation Check

To determine the effectiveness of the manipulation, we posed four questions that were related to the participants' cognizance regarding the conditions to which they were exposed. Two questions were concerned with the extent to which a news offer was perceived as personalised (i.e., perceived personalisation), and two questions tested the degree to which the participants perceived the system's output as diverse (i.e., perceived diversity). Based on a Wilcoxon Rank-Sum Test and a Mann-Whitness Test, our analysis confirmed that the personalised conditions (M = 4.32, SE = 1.68; M = 4.26, SE = 1.52; M = 4.71, SE = 1.62) were perceived as more personalised than the random diversity-based condition (M = 3.52, SE = 1.52).

Table 1. Estimated marginal means (M), standard errors (*SE*) and pairwise mean difference (MD) for all manipulation check questions.

	Perceived Personalisation				Perceived Diversity			
	A system that considered my interests (i.e., sports)		A system that did not consider my interests and randomly recommended news articles		A system that stimulated me to read a diverse range of news topics		A system that stimulated me to read a selective range of news topics	
Condition	M (*SE*)	MD	M (*SE*)	MD	M (*SE*)	MD	M (*SE*)	MD
Content similarity (n = 103)	4.32 (0.17)	0.80***	3.30 (0.18)	-0.66*	4.33 (0.16)	n/a	3.80 (0.18)	n/a
Random diversity (n = 75)	3.52 (0.18)	n/a	3.96 (0.19)	n/a	4.77 (0.18)	0.44*	3.23 (0.19)	-0.57*
Reflective diversity (n=86)	4.26 (0.16)	0.74***	3.55 (0.19)	-0.41	4.67 (0.18)	0.34	3.58 (0.18)	-0.22
Open diversity (n = 73)	4.71 (0.19)	1.19***	2.96 (0.20)	-1.00***	4.51 (0.18)	0.18	3.63 (0.20)	-0.17

Notes: a. The reference category for pairwise MDs is 'random diversity' for perceived personalisation and 'content-based similarity' for perceived diversity; b. MDs differ significantly from zero when the p-value is < 0.05 (*). To counter the problem of familywise error rates, the following corrections were calculated for significance: Holm correction: p-value < 0.0043 (**); Bonferroni correction: p-value < 0.0042 (***).

Our analysis showed that the random diversity-based condition was perceived as more diversity-stimulating (M = 4.77, SE = 1.56) than the content-based similarity condition (M = 4.33, SE = 1.57). For the other two diversity conditions (i.e., open diversity and reflective diversity), no significant differences were found (M = 4.67, SE = 1.62; M = 4.51, SE = 1.56). The latter result means that the participants were unaware of the diversity manipulation in the two treatment conditions. Although this may raise questions regarding causality, we believe that the gradual and secretive nature of personalised diversity recommenders explains why we found no difference.

ANCOVA

To test our hypotheses and understand the interaction patterns between our independent variables, we conducted a series of covariance analyses. In these analyses, we computed for each dependent variable the main effect of our independent variable, 'recommender design', on five dependent variables (see Table 2). Our findings show there were significant differences in at least one of the condition means of three dependent variables: 'measured exposure diversity (open)', 'measured exposure diversity (reflective)' and 'appreciation news recommenders' (F = 44.00, *p* = 0.00; F = 47.89, *p* = 0.00; F = 7.17, *p* = 0.00). No significant differences in the condition means of the dependent variables 'perceived exposure diversity' and 'selected articles' were found (F = 1.43, *p* = 0.23; F = 2.02 *p* = 0.11).

Table 2. Between-subject effects for all dependent variables

Dependent Variable	Independent Variable	df	F	p	η2
Measured exposure diversity (open)	Recommender design	3	44.00	0.00*	0.29
Measured exposure diversity (reflective)	Recommender design	3	47.89	0.00*	0.30
Perceived exposure diversity	Recommender design	3	1.43	0.23	0.01
Appreciation news recommender	Recommender design	3	7.17	0.00*	0.06
Number of selected news articles	Recommender design	3	2.02	0.11	0.02

Note: n = 337 for all independent variables; MD differs significantly from zero when the p-value is < 0.05 (*).

To understand which means and condition groups differed significantly, we calculated the estimated marginal means of all dependent variables across all categories of the independent variable 'recommender design' (see Table 3). Then, we compared each treatment group with the control group (i.e., pairwise MDs). The findings show that participants read more diversely in the diversity-based conditions (M = 0.59, *SE* = 0.02, *p* < 0.00; M = 0.71, *SE* = 0.01, *p* < 0.00; M = 0.69, *SE* = 0.01, *p* < 0.00) than in the content-based similarity condition (M = 0.51, *SE* = 0.02). However, when evaluating their own exposure behaviour (i.e., perceived exposure diversity), the participants in the diversity-based conditions did not think they read more diversely than in the content-based similarity condition (M =

3.55, *SE* = 0.11, *p* > 0.05; M = 3.70, *SE* = 0.09, *p* > 0.05; M = 3.59, *SE* = 0.12, *p* > 0.05). Based on these results, we accepted H1 and rejected H3. In H1, we stated that diversity-based news recommenders stimulate measured exposure diversity. In H3, we hypothesized that diversity-based news recommenders stimulate perceived exposure diversity.

For the other dependent variables and the estimated marginal means (see Table 3), we found only significant MDs for the treatment group in which news articles were provided on a random basis. In particular, when evaluating the news recommender on a scale from 1 to 5, the participants in the random diversity condition significantly rated the news recommender lower (M = 3.17, *SE* = 0.09, *p* > 0.05) than the participants in the content-based similarity condition (M = 3.49, *SE* = 0.06). The participants in other conditions, such as the reflective diversity condition or the open diversity condition, evaluated the news recommender on par with participants in the content-based similarity condition (M = 3.58, *SE* = 0.09, *p* > 0.05; M = 3.73, *SE* = 0.10, *p* > 0.05). Looking to the dependent variable, 'selected articles', no significant differences were found between the treatment conditions (M = 11.95, *SE* = 0.40, *p* < 0.05; M = 12.50, *SE* = 0.40, *p* > 0.05; M = 13.33, *SE* = 0.55, *p* > 0.05) and the content-based similarity condition (M = 13.35, *SE* = 0.48, *p* > 0.05). Based on these results, we rejected H2 and only partially accepted H4 because significant differences were found only for the random diversity condition. In H2, we argued that diversity-based news recommenders stimulate audiences to select more articles. In H4, we hypothesized that diversity-based news recommenders lower audiences' appreciation of the news recommender.

Table 3. Estimated marginal means (M), standard errors (SE) and pairwise mean differences (MD) of all the dependent variables across all categories of the independent variable 'recommender design'.

	Dependent Variables									
	Measured exposure diversity (open)		Measured exposure diversity (reflective)		Perceived exposure diversity		Appreciation news recommender		Selected articles	
Independent Variable: Design	M (*SE*)	MD	M (*SE*)	MD	M (*SE*)	MD	M (*SE*)	MD	M (*SE*)	MD
Content similarity (n = 103)	0.51 (0.02)	n/a	0.59 (0.02)	n/a	3.42 (0.10)	n/a	3.49 (0.06)	n/a	13.35 (0.48)	n/a
Random diversity (n = 75)	0.59 (0.02)	0.08***	0.82 (0.01)	0.23***	3.55 (0.11)	0.13	3.17 (0.09)	-0.32*	11.95 (0.40)	-1.40
Reflective diversity (n=86)	0.71 (0.01)	0.20***	0.77 (0.01)	0.18***	3.70 (0.09)	0.28	3.58 (0.09)	0.09	12.50 (0.43)	-0.85
Open diversity (n = 73)	0.69 (0.01)	0.18***	0.68 (0.02)	0.09***	3.59 (0.12)	0.17	3.73 (0.10)	0.23	13.33 (0.55)	-0.02

Notes: a. The reference category for all pairwise MDs is 'content-based similarity'; b. MDs are significantly different from zero when the p-value is <0.05 (*). To counter the problem of familywise error rates, the following corrections were calculated for significance: Holm correction: p-value < 0.0034 (**); Bonferroni correction: p-value < 0.0033 (***).

Discussion

This research investigated how news recommender design impacts individuals' news exposure behaviours and perceptions. The need to examine this in more detail was rising, as news organisations are increasingly shifting towards personalised news services in which they make use of algorithmic news recommenders to select and distribute news content (Bodó, 2019). Academic research in this area also lacks clear insights into how people engage with news recommenders. Most research and evaluations were conducted offline in an experimental setting that makes use of existing datasets and a protocol that models user behaviour. In contrast, this study developed four news recommenders and allowed the participants to engage one of them to empirically assess how it changed their exposure behaviours and perceptions. To assess the values and expectations of so-called 'diversity-based news recommenders', three of these news recommenders were designed to make users' exposure behaviour more diverse. Based on the results of this study, some conclusions and recommendations can be made.

First, we found evidence that the type of recommender design can influence people's news exposure behaviour and shape how individuals inform themselves online. In particular, our experiment shows that the design of the news recommender largely affects the diversity of the content that people consume (η^2=0.29; η^2=0.30). Although this effect was expected (see H1), the magnitude of the effect was not. This clarifies that news recommenders and their designs are important factors in the study of news audiences and their exposure repertoires, particularly when news recommenders are used as the main news source to follow the news.

This result also indicates that recommendation algorithms may have different societal effects and, thus, could not all be seen as the same. This is extremely important from a policy perspective because it points attention to the inner workings of algorithms and the design features being used. Subsequently, we recommend that policymakers focus more on the design features of news-recommendation algorithms. More specifically, we suggest looking at how public values, such as diversity, could be embedded in news recommenders that are used by platforms and news organisations. For instance, policymakers could enforce platforms and news organisations to include diversity in their algorithmic designs and provide audiences with a minimum amount of diversity in a news offer. Policymakers have several instruments at their disposal, such as financial rewards or subsidy programmes, which can be used for such purposes. For instance, in subsidy programmes part of European stimulus packages for media, diversity could be included as a design requirement for receiving subsidies. However, in practice, we note that few requirements are being formulated. We believe this is a missed opportunity to give direction to how the news landscape should evolve and how news organisations should design their recommendation technologies.

Second, we found that although people read more diversely when using a diversity-based news recommender, they did not think they read more diversely. In particular, our study found no differences between the different conditions in terms of perceived exposure diversity. This result indicates a so-called 'paradox' between measured and perceived exposure diversity, which we also call the 'diversity paradox'. There are two explanations for this diversity paradox. On the one hand, previous news and media consumption research has shown that participants have difficulty reliably recalling their past (consumption) behaviour (see, e.g., de Vreese & Neijens, 2016). Similarly, people might experience difficulty in accurately assessing the diversity of the content they selected. However, in contrast to media consumption research in general (Prior, 2009; Parry et al., 2021), people do not overestimate their exposure diversity in self-reporting. Instead, our results show that people generally estimate their exposure diversity to be lower than their actual exposure diversity. On the other hand, the 'diversity paradox' can be explained by looking at the invisible nature of news algorithms. As

Bucher (2018) explained, algorithms are black boxes whose inner function and decision-making are impenetrable and secretive. People receive no information about how their information is used to influence the decisions they and the system make. As a consequence, people might be unaware of how news recommenders work and how they influence their news exposure. The manipulation checks in our experiment supported this rationale, as they showed that people were unaware of the diversity treatment in the two experimental conditions.

Building further on the latter explanation, we offer several recommendations. On an academic level, more research should be conducted on algorithmic (non-)transparency and how it impacts exposure to news content. Previous research on algorithmic transparency has found that algorithmic transparency can help people increase their algorithmic awareness and reflect on their behaviour (Eslami et al., 2015; Munson et al., 2013; Nagulendra & Vassileva, 2014). However, to the best of our knowledge, little research has been done on how algorithmic transparency can steer people towards more diverse news exposure. It is, for instance, unclear how algorithmic transparency can be used as a digital nudge in a news recommendation context. What should it look like and how could it be used in a real-life context? It is also not clear how audiences respond to algorithmic transparency. Could it be used as a means to stimulate users to read more diverse texts and, if so, under which conditions?

On a policy level, it is clear that transparency should become a more prominent feature in news recommender systems. Although current policy regulations already express the importance of transparency in recommender systems, we argue that, based on these research results, policymakers should take additional steps to force platforms and organisations to remediate this issue. For instance, they could formulate concrete requirements that current and future technologies must meet, such as an obligation to explain how news articles are selected and which data are used. This may help audiences know how recommender systems work and how they are steered.

Third, our research has shown that news recommenders can stimulate diverse news exposure while maintaining user satisfaction and the number of selected articles. In particular, our study found no attitude or interaction differences between the various recommender designs, except for the random diversity-based news recommender (for which significantly lower scores on user satisfaction were found). As a result, our experiment debunks the conventional idea that news recommenders should focus on similarity and only provide news articles that align with users' interests. In particular, this study shows that news recommenders can provide a diverse range of news articles and sufficiently appeal to and satisfy people with news content that lies within their individual interest zones. This result should inspire news media to experiment with different news recommender designs and use exposure diversity as a quality criterion to assess the value of news recommender designs. Therefore, we strongly recommend that news organisations take up their (corporate social) responsibility and use diversity as an additional/key value in their recommenders' designs. From a commercial viewpoint, it can even be used as an additional selling proposition as part of their communication strategy as a 'reliable news partner'.

Limitations

Our research design was strongly focused on internal validity or how accurately we could draw conclusions about the effect of recommender design. As a consequence, our research design and experimental conditions were highly specific and artificial in contrast to real-world conditions. For instance, we forced participants to read news content for 12 minutes. However, this restriction does not align with a natural environment in which news content is, most commonly, consumed in a short, easy-going and incidental fashion (i.e., snacking: Costera Meijer & Groot Kormelink, 2015). By using a time restriction, our study focused on short-term effects, while news recommenders may also have significant long-term effects on, for instance, attitudes or the number of selected articles. Further research should be conducted to map these long-term effects. To make this possible, we declare that we are open to research collaborations in which other researchers can make use of our stimulus material for their own research purposes. We hope to inspire future scholars who want to replicate our study design and hypotheses.

In conclusion, we also want to point attention to our dependent variable, 'selected articles'. In this study, we used the number of clicks to measure how people engaged with news recommenders' designs. However, users can briefly skim and click through news article headlines without really engaging with them, while others may click on fewer stories but engage deeply with news articles. Although we initially planned to capture these cognitive differences in reading behaviour, we could not do so due to COVID-19 (i.e., no physical contact allowed during data collection). We controlled for this in our data analysis, but we recommend that scholars note this limitation and explore the use of other techniques (e.g., eye tracking) to capture 'interaction' in future research. In a similar vein, we also suggest that scholars develop and apply a multi-item construct for the dependent variable 'perceived exposure diversity'. This will help create a more valid measurement instrument for this construct.

Appendices

Appendix 1. News recommenders' technical description.

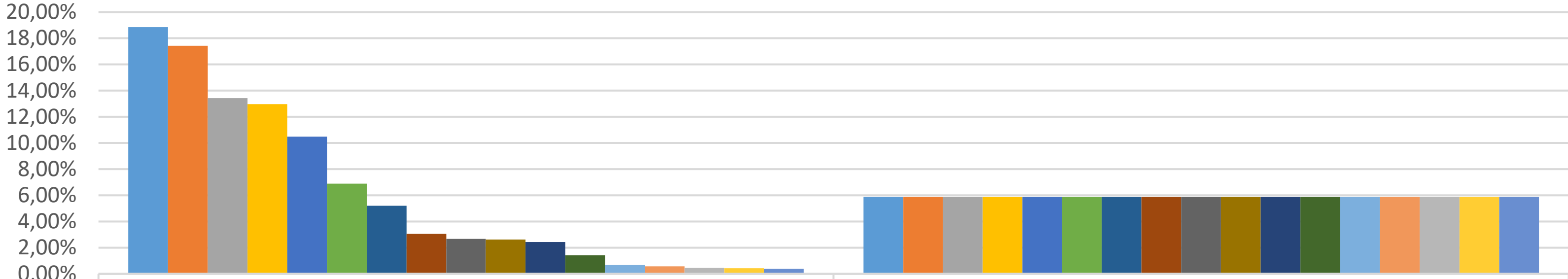
News selection mechanisms	Technical description
Content-based news selection	For each article, we extracted all mentioned concepts(i.e., topics categories, named entities, and nouns) and averaged their corresponding word2vec embeddings(i.e., real-valued vector representations that encode the meaning of words such that the words that are closer in the vector space are expected to be similar in meaning). During averaging, we used TF-IDF(term frequency-inverse document frequency) as a weighting scheme to minimize the weighting of frequent terms while making infrequent terms have a higher impact. Similarly, we constructed user profiles by summing up the article embeddings found in the user's reading history. Recommending articles was done via Maximum Marginal Relevance (MMR, see, e.g., Carbonell & Goldstein, 1998). MMR is an iterative reranking algorithm that maximizes both the relevance and novelty of the top-n items. A higher MMR score means that an item is both more relevant to the query(i.e., similar to user profile) and contains minimal similarity to the previously selected items. Finally, we used cosine similarity to measure the distance between articles and the similarity between user-profiles and articles.
Random news selection	In this news recommender design, we made use of computer-generated random numbers to select random subsets of news articles.
Open diversity news selection	We extended the user profiles with "related nudging concepts" to improve the range of recommended categories while maintaining sufficient personalization. These related nudging concepts can be seen as small deliberate steps outside the user's existing sphere of interests to steer him/her towards specific categories. Like the content-based selection, articles and user profiles are defined as TF-IDF weighted average over word2vec embeddings and sums of article embeddings, respectively. Instead of MMR, a similar taxonomic approach to AI-select from Agrawal et al. (2009) was applied to reranking and selection recommendations. In open diversity news selection, the aim is to nudge the user towards an open topic distribution.
Reflective diversity news selection	In reflective diversity news selection, we use the same method of "open diversity news selection" but nudge the user towards a reflective topic distribution instead of an open distribution.

Appendix 2. Descriptive statistics independent variables.

Variabele	Question text	Mean/mode	Standard deviation	Variance	Range	N
'Diversity-based news selection preference' (DIV)	having stories selected for me based on various opinions	4.26	.747	.558	1-5	337
	Having stories selected for me based on various perspectives	4.25	.702	.493	1-5	337
	having stories selected for me based on various topics	4.18	.683	.466	1-5	337
	having stories selected for me based on various sentiments on a certain topic	4.38	.693	.480	1-5	337
'News information overload' (NIO)	I often felt overwhelmed about the large amount of daily news	2.87	1.078	1.161	1-5	337
	I give up following the news due to the large amount of news	2.36	1.021	1.042	1-5	337
	I often felt that there was more news than I could process	3.29	1.086	1.178	1-5	337
	I often doubt whether I do not miss out the most important news of the day due to the large amount of news	2.80	1.065	1.134	1-5	337
	I often do not know where to start due to the large amount of news	2.70	1.036	1.072	1-5	337
	I often felt stressed about the large speed of news coverage	2.53	1.015	1.030	1-5	337
'Need for cognition' (NFC)	I would prefer complex to simple problems	3.24	.840	.705	1-5	337
	I like to have the responsibility of handling a situation that requires a lot of thinking	3.43	.839	.705	1-5	337
	Thinking is not my idea of fun (R)	3.29	.851	.723	1-5	337
	I would rather do something that requires little thought than something that is sure to challenge my thinking abilities (R)	2.36	.797	.635	1-5	337
	I really enjoy a task that involved coming up with new solutions to problems	3.72	.771	.595	1-5	337
	I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought	3.51	.791	.626	1-5	337
Age	What is your age?	34.99	16.74	280.32	n/a	337
Sex	What sex is on your identity card?	2	.487	0.24	n/a	337
Nationality	Is uw geboorteland België?	1	.161	0.03	n/a	337
Language	Is your mother language Dutch?	1	.193	0.04	n/a	337
Education	What is your highest level of education?	5	.861	0.74	n/a	337
Employment	What is your current occupation status?	1	1.153	1.33	n/a	337

Note: 1. All questions and variables were in the questionnaire clarified with an introduction text and several examples. To keep this table clear, we removed these texts. The complete questionnaire is available from the corresponding author, upon reasonable request; 2. Items marked with an asterisk were characterized with little variation or skewed means. These items were removed from further analysis.

Appendix 3. Benchmark distribution reflective diversity-based news selection and open diversity-based news selection.



	Reflective diversity	Open diversity
Economy	18,86%	5,88%
Art	17,44%	5,88%
Sport	13,43%	5,88%
Crime	12,97%	5,88%
Politics	10,49%	5,88%
Human Interest	6,89%	5,88%
Disasters	5,20%	5,88%
Education	3,05%	5,88%
Health	2,67%	5,88%
Society	2,62%	5,88%
Lifestyule	2,43%	5,88%
Environment	1,42%	5,88%
Science	0,68%	5,88%
Religion	0,57%	5,88%
Weather	0,46%	5,88%
Conflict	0,44%	5,88%
Labour	0,38%	5,88%

Appendix 4. Construct reliability and validity.

Latent variable	Observed variable	Cronbach's alpha	Factor loadings	Item variance	Squared multiple correlation
'Preference diversity-based news selection'	having stories selected for me based on various opinions	0.841	0.762	0.580	0.572
	Having stories selected for me based on various perspectives		0.816	0.666	
	having stories selected for me based on various topics		0.718	0.515	
	having stories selected for me based on various sentiments on a certain topic		0.726	0.527	
'News information overload'	I often felt overwhelmed about the large amount of daily news	0.815	0.786	0.618	0.431
	I give up following the news due to the large amount of news		0.576	0.332	
	I often felt that there was more news than I could process		0.645	0.416	
	I often doubt whether I do not miss out the most important news of the day due to the large amount of news		0.559	0.312	
	I often do not know where to start due to the large amount of news		0.732	0.536	
	I often felt stressed about the large speed of news coverage		0.610	0.372	
'Need for cognition'	I would prefer complex to simple problems	0.826	0.677	0.459	0.446
	I like to have the responsibility of handling a situation that requires a lot of thinking		0.644	0.415	
	Thinking is not my idea of fun (R)		0.602	0.363	
	I would rather do something that requires little thought than something that is sure to challenge my thinking abilities (R)		0.636	0.404	
	I really enjoy a task that involved coming up with new solutions to problems		0.705	0.498	
	I would prfer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought		0.732	0.536	

Note: Thresholds used to consider removing items (DeVellis, 2016): Cronbach's alpha: > .7; Factor loadings (CFA): > .5; Item variance (R-square): > .4; Squared multiple correlation: > .4

List of tables, figures and appendices

Bibliography

Agrawal, R., Gollapudi, S., Halverson, A., & Ieong, S. (2009). Diversifying search results. *Proceedings of the second ACM international conference on web search and data mining*, 5-14.

Beam, M. A. (2014). Automating the news: How personalized news recommender system design choices impact news reception. *Communication Research*, *41*(8), 1019-1041.

Bodó, B. (2019). Selling News to Audiences–A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media. *Digital Journalism, 7*(8), 1054-1075.

Bruns, A. (2019). Filter bubble. *Internet Policy Review*, *8*(4).

Bucher, T. (2018). *If... then: Algorithmic power and politics*. Oxford University Press.

Carbonell, J., and J. Goldstein. (1998, August). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval (pp. 335–336)

Castells, P., Hurley, N. J., & Vargas, S. (2015). Novelty and diversity in recommender systems. In *Recommender systems handbook* (pp.881-918). Springer.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Costera Meijer, I., & Groot Kormelink, T. (2015). Checking, sharing, clicking and linking: Changing patterns of news use between 2004 and 2014. *Digital Journalism*, *3*(5), 664-679.

De Clercq, O., L. De Bruyne, and V. Hoste. 2020. News topic classification as a first step towards diverse news recommendation. COMPUTATIONAL LINGUISTICS IN THE NETHERLANDS JOURNAL, 10, 37–55.

De Pessemier, T., Leroux, S., Vanhecke, K., & Martens, L. (2015). Combining collaborative filtering and search engine into hybrid news recommendations. *Proceedings of the third International Workshop on News Recommendation and Analytics, in conjunction with the 9th ACM Conference on Recommender Systems*, 13-18.

de Vreese, C. H., & Neijens, P. (2016). Measuring Media Exposure in a Changing Communications Environment. *Communication Methods and Measures*, *10*(2-3), 69-80.

DeVellis, R. F. (2016). *Scale development: Theory and applications* (Vol. 26). Sage publications.

Dylko, I., Dolgov, I., Hoffman, W., Eckhart, N., Molina, M., & Aaziz, O. (2017). The dark side of technology: An experimental investigation of the influence of customizability technology on online political selective exposure. *Computers in Human Behavior*, *73*, 181-190.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). " I always assumed that I wasn't really that close to [her]" Reasoning about Invisible Algorithms in News Feeds. Proceedings of the 33rd annual ACM conference on human factors in computing systems, 153-162.

Fields, B., Jones, R., & Cowlishaw, T. (2018). The case for public service recommender algorithms. Academia.
https://www.academia.edu/43929218/The_case_for_public_service_recommender_algorithms

Garcin, F., Faltings, B., Donatsch, O., Alazzawi, A., Bruttin, C., & Huber, A. (2014). Offline and online evaluation of news recommender systems at swissinfo. ch. *Proceedings of the 8th ACM Conference on Recommender systems,* 169-174.

Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism*, *6*(3), 330-343.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological bulletin*, *135*(4), 555-588.

Helberger, N. (2011). Diversity by design. Journal of Information Policy, 1, 441-469.

Helberger, N. (2012). Exposure diversity as a policy goal. Journal of Media Law, 4(1), 65-92.

Helberger, N. (2019). On the democratic role of news recommenders. *Digital Journalism*, 1-20.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191-207.

International Press Telecommunications Council. (n.d.). Media Topics. Retrieved January 21, 2021, from https://iptc.org/standards/media-topics/

Jannach, D., Zanker, M., Ge, M., & Gröning, M. (2012). Recommender Systems in Computer Science and Information Systems – A Landscape of Research. In C. Huemer & P. Lops, *E-Commerce and Web Technologies* Berlin, Heidelberg.

Jones, B., & Jones, R. (2019). Public Service Chatbots: Automating Conversation with BBC News. *Digital Journalism*, *7*(8), 1032-1053.

Karimi, M., Jannach, D., & Jugovac, M. (2018). News recommender systems – Survey and roads ahead. *Information Processing & Management*, *54*(6), 1203-1227.

Knobloch-Westerwick, S. (2014). *Choice and preference in media use: Advances in selective exposure theory and research*. Routledge.

Loecherbach, F., & Trilling, D. (2020). 3bij3–Developing a framework for researching recommender systems and their effects. *Computational Communication Research*, *2*(1), 53-79.

Loecherbach, F., Welbers, K., Moeller, J., Trilling, D., & Van Atteveldt, W. (2021). Is this a click towards diversity? Explaining when and why news users make diverse choices. *Proceedings of the 13th ACM Web Science Conference 2021.*

Lu, F., Dumitrache, A., & Graus, D. (2020). Beyond Optimizing for Clicks: Incorporating Editorial Values. *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*, 1-9.

McQuail, D., & Van Cuilenburg, J. J. (1983). Diversity as a media policy goal: A strategy for evaluative research and a Netherlands case study. *International Communication Gazette*, *31*(3), 145-162.

Möller, J., Trilling, D., Helberger, N., & van Es, B. (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.

Munson, S., Lee, S., & Resnick, P. (2013). Encouraging reading of diverse political viewpoints with a browser widget. *Proceedings of The International AAAI Conference on Web and Social Media*, 419-428.

Nagulendra, S., & Vassileva, J. (2014). Understanding and controlling the filter bubble through interactive visualization: a user study. *Proceedings of the 25th ACM conference on Hypertext and social media*, 107-115.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, *115*(11), 2600-2606.

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour.*

Prior, M. (2009). The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public opinion quarterly*, *73*(1), 130-143.

Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender systems handbook*. Springer.

Stroud, N.J. (2017). Selective exposure theories. In *The Oxford handbook of political communication*. Oxford.

Thurman, N., & Schifferes, S. (2012). The future of personalization at news websites: lessons from a longitudinal study. *Journalism Studies*, *13*(5-6), 775-790.

Van den Bulck, H., & Moe, H. (2018). Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas. Media, Culture & Society, 40(6), 875-892. https://doi.org/10.1177/0163443717734407

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425-478.

Yang, J. (2016). Effects of popularity-based news recommendations (“most-viewed”) on users' exposure to online news. *Media Psychology*, *19*(2), 243-271.

Zuiderveen Borgesius, F., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1), 1 - 16.