It ain't easy: using normatively motivated news diversification to facilitate policy support, tolerance, and political participation

Nicolas Mattis, Philipp K. Masur, Judith Moeller & Wouter van Atteveldt

View supplementary material

Published online: 11 Nov 2024.

Submit your article to this journal

Article views: 223

View related articles

View Crossmark data

ⓐ OPEN ACCESS Check for updates

It ain't easy: using normatively motivated news diversification to facilitate policy support, tolerance, and political participation

Nicolas Mattis[a], Philipp K. Masur[a], Judith Moeller[b] and Wouter van Atteveldt[a]

[a]Department of Communication Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; [b]University of Hamburg in Cooperation with the Leibniz Institute for Media Research — Hans-Bredow-Institut (HBI), Hamburg, Germany

ABSTRACT

Proponents of 'democratic news recommender design' argue that algorithmic news diversification may facilitate democratic participation. However, while various news diversification metrics have been proposed in recent years, few of them have been put to the test with real users. To assess the promises and pitfalls of algorithmic news diversification, we conduct a 2 (low vs. high levels of activating language) by 3 (low vs medium vs high levels of alternative voices) between subjects experiment with $N = 715$ respondents to test how normatively driven news diversification affects readers' (a) policy support, (b) outcome tolerance, (c) outgroup tolerance, and (d) political participation. Results show that in a one-off experiment, exposure diversity has at best very small effects on the dependent variables when demographic and attitudinal characteristics are controlled for. We also find that extreme forms of news diversification may impede user satisfaction.

1. Introduction

In light of worries about filter bubbles and echo-chambers, academics and practitioners alike have begun paying increasing attention to the notion of exposure diversity. Due to its tradition in democratic theory and free speech, exposure diversity can be seen as 'a concept that serves a democratic mission of realising freedom of expression and contributing to a functioning democratic debate' (Helberger & Wojcieszak, 2018) and has been posited as a core design principle of 'democratic news recommenders' (Helberger, 2019). The core idea behind democratic news recommenders is optimising towards exposure diversity to facilitate normatively desirable outcomes such as greater levels of tolerance and political participation. To explore whether this argument holds true, we carried out a preregistered[1] 2 by 3 between-subjects experiment in the Netherlands to test

---

how news that score either high or low on two recently proposed diversification metrics affect readers' policy support, tolerance and political participation.

While rooted in democratic theory and media research (Loecherbach et al., 2020), the notion of exposure diversity has increasingly attracted attention in more technical areas such as natural language processing (NLP) (Reuver, Fokkens et al., 2021), information retrieval (Draws et al., 2022) and recommender system research (Treuillier et al., 2022; Vrijenhoek et al., 2022). Indeed, recent reviews name diversity as a core 'beyond-accuracy' metric for (news) recommender design and evaluation (Karimi et al., 2018). As such, it comes as no surprise that diversity metrics are increasingly being developed and implemented in practical settings.

However, implementing news diversity metrics is challenging, as it requires translating various complex conceptualisations of diversity (see Loecherbach et al., 2020 into concrete technical metrics that practitioners can operationalise (Reuver, Fokkens et al., 2021; Reuver, Mattis et al., 2021). Moreover, we still know rather little about the extent to which the diversity dimensions that are measured affect citizens – especially when moving beyond a diversity of viewpoints.

As of now, most news diversification metrics seem to be (implicitly) informed by ideals of deliberative democracy, with diversity being predominantly conceptualised in terms of viewpoints. According to Sax (2022), this underrepresentation of other, potentially more critical models of democracy may rob us of opportunities to explore interesting alternatives for how news can be presented and how they may spark reactions among readers.

To address this gap in the literature, as well as concerns about potentially 'monolithic' news coverage and the negligence of less powerful groups (Blumler & Cushion, 2014), this paper thus approaches news diversity from the lens of critical models of democracy. Specifically, we operationalise two of Vrijenhoek et al. (2021)'s news diversification metrics, namely activation and alternative voices, that reflect variations in journalistic writing style and the extent to which minorities are being represented. By translating the mathematical ideas behind the two metrics into a stringent experimental design, we provide a careful operationalisation of both metrics in the context of a new transgender law that is currently being discussed in Dutch parliament.

By testing the effects of an internally valid operationalisation of the two metrics in a controlled experimental setting, we are able to explore causal effects on policy support, tolerance, participation and user satisfaction and engagement in isolation. Thereby, we hope to provide a starting point for realistically assessing the extent to which news diversification can facilitate particular outcomes.

2. Theoretical framework

Following Helberger and Wojcieszak (2018), we understand exposure diversity as a 'concept with a mission and considerable societal values' (p. 536). Its importance is rooted in democratic theory, which highlights the value of pluralism and argues that for the system to function, diverse voices must be expressed, disseminated and received by others (Helberger & Wojcieszak, 2018).

Not in small part due to the attention that Eli Pariser (2011)'s filter bubble hypothesis has attracted (whether justified or not), exposure diversity also attracts increasing

attention in more technical fields, such as computer science, NLP and literature on recommender systems (e.g., Draws et al., 2020; Michiels et al., 2022; Reuver, Fokkens et al., 2021). As these disciplines grapple with how diversity can be safeguarded in their models, several (news) diversification metrics have recently been proposed and applied (Draws et al., 2022; Lunardi et al., 2020; Vrijenhoek et al., 2021).

Among the more social-scientifically grounded metrics of them are the ones by Vrijenhoek et al. (2021). That is, because their five metrics, namely calibration, fragmentation, activation, representation, and alternative voices, can be mapped onto four different models of democracy as discussed by Helberger (2019). These are liberal models, which champion a 'marketplace of ideas' (Napoli, 1999), participatory models that focus on shared civic culture and committed citizenship, deliberative models that highlight rational exchange of ideas and consensus-finding and critical notions which argue that conflict is necessary and can be productive (Helberger, 2019; Sax, 2022).

Originally, Vrijenhoek et al. (2021)'s metrics were designed to assess the diversity of existing recommendations. Examples of such approaches include the work of Vrijenhoek et al. (2022), who analyse MIND and Polimeno et al. (2023), who apply NLP techniques to assess fragmentation in the HeadLine Grouping Dataset. In this paper, we argue that in addition to analysing existing recommendations, these metrics can also be employed for deliberate news diversification that may nudge readers to engage with more diverse news by making them more prominent and accessible (Mattis et al., 2024). For similar study designs, albeit with different underlying metrics and concepts, see Heitz et al. (2022) and Heitz et al. (2023).

2.1. Operationalising news diversification from a critical angle

In their work, Vrijenhoek et al. (2021) distinguish metrics for calibration (reflecting 'to what extent the issued recommendations reflect the user's preferences'; p. 4), fragmentation (capturing 'the amount of overlap between news story chains shown to different users'; p. 5), representation (expressing 'whether the issued recommendations provide a balance of different opinions and perspectives'; p. 6), as well as activation and alternative voices.

Among them, we chose to isolate the activation and alternative voices metrics, because they can be operationalised without prior data on user interactions and because they relate particularly to more critical normative angles, which so far have only received limited attention in work on news diversity (Sax, 2022).

The activation metric 'expresses whether the issued recommendations are aimed at inspiring the users to take action [with the underlying assumption that] an emotional article may activate [readers] to undertake action' (Vrijenhoek et al., 2021, p. 177). In contrast, the alternative voices metric 'measures the relative presence of people from a minority or marginalised group' (Vrijenhoek et al., 2021, p. 179). As such, this metric is particularly concerned with who is represented in the news.

Within the context of our study, focusing on these two metrics has two important advantages: First, variations in the two metrics exemplify different styles of news reporting that can be mapped onto and evaluated by different models of democracy. For example, low levels of activation and no explicit focus on marginalised voices reflects the normative ideals of deliberative models of democracy, while high levels of both

align closely with the type of content that critical models of democracy prioritise (Vrijenhoek et al., 2021). As such, examining the effects of these metrics can help tease out the different effects that particular types of democratic news recommenders may have.

Second, compared to the more common approach of diversifying topics (e.g., Loecherbach et al., 2021) or viewpoints (e.g., Draws et al., 2022; Heitz et al., 2022), our focus on more stylistic elements of how news is being reported extends the breadth of research into news diversification metrics and their effects. While prior research has already explored some of Vrijenhoek et al. (2021)'s metrics for the evaluation of news recommendations (e.g., see Polimeno et al., 2023; Vrijenhoek et al., 2022), this study is, to the best of our knowledge, the first that explores the effects of news diversification according to these metrics on readers.

2.2. Effects of exposure diversity

Our research builds on two strands of literature: (1) normative assumptions of democratic theory and supporting evidence (e.g., Heitz et al., 2023, 2022; Lu & Gall Myrick, 2016; Mutz, 2002) and (2) insights from news framing studies (e.g., Aarøe, 2011; Boyer et al., 2022; Gross, 2008).

Regarding the former, previous studies found that, at least under some conditions, higher exposure diversity can increase tolerance (Heitz et al., 2022; Mutz, 2002), and political participation (Lu & Gall Myrick, 2016). Mutz (2002) suggests perspective taking as a potential moderator such that people high in perspective-taking are more open towards embracing opposing viewpoints. While perspective taking can be conceptualised as a character trait, we argue that the style of news may also affect if and to what extent people are willing and able to view issues from others' perspectives. This applies especially to the alternative voices metric, as giving minorities the room to discuss their lived experiences may make it easier for readers to put themselves in their shoes.

Another relevant line of research is work on cross-cutting exposure, which denotes readers' exposure to opposing viewpoints. Prior cross-cutting exposure studies examined outcome variables such as out-party hostility (Amsalem et al., 2021) and political participation (Chen & Lin, 2021; Lu & Gall Myrick, 2016). Studies that operationalise cross cutting exposure as conversations with people who hold different political opinions tend to show depolarising effects and increased tolerance (Amsalem et al., 2021) as well as more flexible attitudes (Himmelroos & Christensen, 2020). However, studies that operationalise cross-cutting exposure in terms of news exposure yield less clear-cut results. A recent meta-analysis by Matthes et al. (2019) found no significant effects of cross-cutting exposure on political participation, while a follow-up study by Chen and Lin (2021) suggests a demobilising effect, albeit only in the case of low issue importance. They explain their findings with increased political ambivalence.

Taken together, prior work suggests that the exposure to diverse viewpoints can facilitate individuals understanding of others' arguments or prompt them to reconsider their own ones. However, the distinction between effects of discussion networks and news exposure also suggests relatively weak effects of mere exposure to particular news articles, where instances of selective-exposure and motivated reasoning are also common.

Given the operationalisation of our two diversification metrics under study, insights from framing research can also be informative. That is, because the combination of

activating language and alternative voices aligns closely with the notion of episodic or human interest news frames. According to Iyengar (1990), episodic frames illustrate broader issues through a focus on particular cases or by zooming in on a concrete event, while human interest frames bring a human face or an emotional angle to the presentation of an event, issue, or problem' (Semetko & Valkenburg, 2000, p. 95) and often feature individualised and more emotional stories (Boyer et al., 2022).

According to Gross (2008, p. 169), 'framing effects on policy opinion operate through both affective and cognitive channels'. Affective channels seem to be particularly important in the case of episodic frames. For example, work by Gross (2008) and Aarøe (2011) suggests that reading an episodically framed news article elicits stronger affective reactions, such as empathy, anger and pity, than thematically framed ones. According to Van Doorn (2014), empathy and threat perceptions also constitute important predictors of tolerance towards specific outgroups. Moreover, if strong affective reactions are present, they also make such frames more persuasive (Aarøe, 2011), suggesting that such articles could increase policy support. On a related note, Boyer et al. (2022) suggest that human-interest frames can mitigate motivated reasoning under some conditions, as individualisation impedes 'both the motivation for in-group bias and the ability to counter-argue identity-threatening messages' (p. 4).

Taken together, extant work thus suggests that emotional personalised stories can elicit strong emotional reactions which in turn may increase policy support, tolerance and political participation.

2.3. Hypotheses and research questions

In light of the previously discussed literature, we pose two hypotheses predicting positive effects of news diversification on policy support, outcome tolerance, outgroup tolerance, and political participation, as well as two research questions on potential mediators and effects on user satisfaction and engagement respectively:

> **H1**: *Higher levels of exposure to news that features alternative voices leads to higher levels of (a) policy support, (b) issue-specific tolerance, (c) outgroup tolerance and (d) political participation.*
>
> **H2**: *Higher levels of exposure to news that features activating language leads to higher levels of (a) policy support, (b) issue-specific tolerance, (c) outgroup tolerance and (d) political participation.*
>
> **RQ1**: *Can we find empirical evidence suggesting that the effects of news diversification are mediated by (a) empathy and (b) threat perceptions?*
>
> **RQ2**: *How do different diversification metrics relate to users' (a) engagement with the recommender (overall reading time and number of articles selected) and (b) satisfaction?*

In addition, we also wondered whether mere exposure diversity is sufficient, or whether respondents need to engage with diverse news more actively – for example by clicking on and reading the whole article (Mattis et al., 2024). While the limitations of clicks as a measure for true engagement are well established (see Groot Kormelink & Costera Meijer, 2018), in the absence of eye-tracking or qualitative methods it remains one of the more meaningful measures of deeper engagement.

Accounting for differences in the extent to which readers did engage with the diverse articles respectively is not only interesting from a methodological point of view, but also

from a selective exposure angle. Indeed, a long line of research has established that people tend to prefer content that aligns with their preferences and opinions (Stroud, 2017) and a recent analysis of MIND by Treuillier et al. (2022), as well as a field experiment on Facebook and Instagram (Guess et al., 2023) suggest that users' news selections tend to be less diverse than their initial, algorithmically curated supply. Thus, to disentangle the effects from news that people are exposed to from the effects of the news that they click on, we ask:

> *RQ3: Do the effects of news diversification according to (a) alternative voices and (b) activation vary depending on whether we measure users' exposure or consumption diversity?*

2.4. Case description

Our study focused on an ongoing political issue in the Netherlands, namely the discussion around a new transgender law that has been stuck in parliament for several years now. The objective of the law, which was first proposed in 2019, is to clear hurdles towards changing one's gender through several updated regulations, such as a lower minimum age and the scrapping of a so-called expert statement from a psychologist, in order to speed up a currently rather long and complex procedure. The bill has stirred quite some media attention throughout the last year, including critical commentaries in prominent news media outlets and a highly controversial poster campaign launched by a group of opponents in summer 2022 (Dollé, 2023; NA, 2023). Taken together, we argue that the salience of and controversy around this particular topic make for an interesting research context. Moreover, due to its focus on improving the lives of a marginalised group, the topic also lends itself well towards the objective to foreground the lived experiences of minorities in the news.

3. Methodology

We carried out a preregistered 2 (low vs. high levels of activating language) by 3 (low vs. medium vs. high levels of alternative voices) between-subjects experiment with $N = 715$ respondents from the Netherlands. The experiment involved three different steps. First, we measured relevant control variables in Qualtrics. Next we told respondents that they would be asked to vote in a referendum on the topic, but that beforehand, they first needed to interact with an external news feed for at least two minutes. For this step, we forwarded participants to an external experimental news website[2] where we exposed them to differently diversified news feeds. After spending at least two minutes browsing the news website and reading at least one article, participants could return to the Qualtrics survey where we asked them to vote in a fictitious referendum on a new transgender law that was being debated in the Dutch parliament when we conducted the experiment and measured relevant dependent variables. For screenshots of the news website see Figure 1.

3.1. Independent variables

Our main manipulation concerned the extent to which news feeds included articles that featured activating language and alternative voices. To maximise internal validity, we
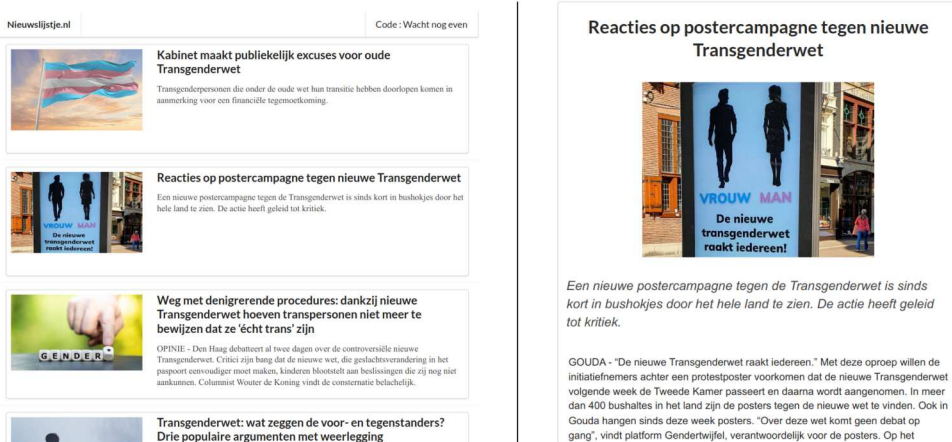
Figure 1. Screenshot of the experimental news website's home screen (left) and article view (right).

manipulated existing articles on the topic to create four different versions that matched our experimental manipulation. All articles were edited by a Dutch student assistant with a background in Journalism and pretested to ensure that the manipulations were apparent from the title and teaser alone.[3]

After the pretest, we identified eight articles which we assembled into six news feeds that varied in the extent to which they included (a) articles that featured transgender people as active agents (87% vs. 38% vs 0%) and (b) articles that featured activating language (75% vs. 25%). Upon opening the experimental news app, participants were automatically assigned to one of the six experimental conditions and presented with a randomised order of articles form the according article pool. An overview of all article pools, as well as the individual articles are available in the stimulus material on OSF.[4]

The exact percentages that we used for the different diversification conditions were the result of an attempt to create notably different levels of exposure diversity with a limited number of pretested stimuli.

3.2. Dependent variables

We measured four main dependent variables, namely respondents' policy support, outcome tolerance, outgroup tolerance, and political participation, as well as respondents' satisfaction, perceived diversity and symbolic and realistic threat perceptions.

Policy support was measured by asking people to vote either in favour (1) or against (0) the new transgender law ($M = 0.68$; $SD = .47$). For **outcome tolerance**, participants indicated to what extent they agreed that the opposite outcome of what they had voted for would make them angry on a scale from 1 (completely disagree) to 7 (completely agree); ($M = 4.12$; $SD = 1.59$).

To measure **outgroup tolerance** ($M = 4.58$; $SD = 1.25$; $\alpha = .83$), we adapted four items taken from Stroud and Muddiman (2013) that tapped into respondents' propensity to engage with proponents of the opposite election outcome (e.g. being interested in

hearing what they have to say or being willing to become friends) on a scale from 1 (completely disagree) to 7 (completely agree).

Political participation was measured by asking respondents how willing they were to engage in various political behaviours on a scale from 1 (very unwilling) to 7 (very willing) ($M = 3.32$; $SD = 1.22$; $\alpha = .90$). We used 9 items such as participating in a demonstration regarding the new transgender law or posting about it on social media.

For **user satisfaction**, we measured *future use intentions* ('I would like to use this website again') ($M = 3.83$; $SD = 1.52$), *perceived quality* ('The articles I could choose from were of good quality') ($M = 4.66$; $SD = 1.31$) and *overall satisfaction* ('I liked the selection of articles I could choose from') ($M = 4.26$; $SD = 1.37$) with one item each. Users indicated their (dis-)agreement on a 7-point Likert scale.

Perceived diversity was measured by asking respondents to what extent they agreed that the news feed featured a diverse range of articles, opinions, and journalistic styles respectively on a 7-point Likert scale ($M = 4.46$; $SD = 1.22$; $\alpha = .83$).

Following Li (2023), we measured symbolic and realistic threat perceptions separately. For each index, we asked respondents to indicate their agreement with a set of statements on a scale from 1 (completely disagree) to 7 (completely agree). **Symbolic threat perceptions** ($M = 2.44$, $SD = 1.65$, $\alpha = .97$) were measured with three items that tapped into perceived threats to Dutch culture, while **realistic threat perceptions** ($M = 2.11$, $SD = 1.28$, $\alpha = .95$) were measured with six items related to concerns over the political and economic influence of transgender people.

3.3. Control variables

For demographic control variables, we measured age, gender, and education. Given the political nature of the topic and outcome variables, we also included common control variables from political communication science, namely political interest, internal political efficacy and political ideology.

For **political interest** ($M = 4.79$; $SD = 1.58$), we asked to what extent respondents agreed with the statement 'I am interested in news over national, international, and European politics' on a 7-point Likert-scale. For **political ideology** ($M = 5.09$; $SD = 2.15$), we asked respondents to locate themselves on a 10-point left to right spectrum where 10 stands for right and 0 for left.

Political efficacy ($M = 3.93$; $SD = 1.18$, $\alpha = .80$) was measured with a 5-item index taken from Niemi et al. (1991), including items such as 'I consider myself well-qualified to participate in politics' or 'I feel that I have a pretty good understanding of the important political issues facing our country'.

Given the polarising nature of the issue and the potential importance of previous opinions, we also measured issue importance and transphobia. To measure **issue importance** ($M = 4.58$; $SD = 1.67$), we asked how important participants considered the issue of transgender rights to be on a 7-point Likert scale.

Transphobia was measured with a slightly adapted scale from Li (2023). It included eight items, such as 'A man who dresses as a woman is disgusting' or 'If I was to find out that my best friend wants to change their gender I would flip

out'. Respondents expressed their agreement on a 7-point Likert scale ($M = 1.93$; $SD = 1.15$; $(\alpha = .89)$.

4. Results

Since we had manipulated overall levels of exposure diversity within a news feed, we first explored whether this affected how many diverse articles readers clicked on. Results of linear regression analyses with the experimental factors as independent and the number of read articles featuring alternative voices ($R^2 = .48$, $F(10, 704) = 65.06$, $p < .001$) and activation ($R^2 = .23$, $F(10, 704) = 21.46$, $p < .001$) as dependent variables suggest that greater exposure diversity did indeed lead to greater consumption diversity. Specifically, we found that exposure to a newsfeed featuring 75% instead of 25% of activating news significantly positively predicted increased selection of such articles ($b = 1.26$, $t(704) = 14.08$, $p < .001$). Similarly, we also found that compared to the low alternative voices condition where no news featuring transgender people as active agents were present, both the medium ($b = .77$, $t(704) = 8.36$, $p < .001$) and the high ($b = 2.17$, $t(704) = 24.33$, $p < .001$) alternative voice condition increased the number of news featuring alternative voices that people read.

4.1. Hypothesis tests

To test the effects of our experimental manipulations, readers' consumption diversity and relevant control variables, we used hierarchical regression analyses. For each dependent variable, we tested three baseline models with (a) dummy variables for respondents' exposure diversity, (b) variables for respondents' consumption diversity, and (c) a combination of the two. We then compared these models against a full model that included respondents' age, gender, education, issue importance, transphobia, political ideology, and internal political efficacy as control variables.

This enabled us to explore how much variation in the outcome variable was explained by exposure and consumption diversity as compared to person-specific characteristics. In the following, we report the results of the full regression models per dependent variable. For the binary dependent variable policy support we report odds ratios (see Figure 2A). For all other dependent variables we report unstandardised regression coefficients (see Figure 2B). All regression models are available in our supplementary materials (Tables 1 through 4).

4.1.1. Effects on policy support

H1a and H2a posited that that higher levels of exposure to news that features alternative voices and activation lead to higher levels of policy support. We find support for H2a, but not H1a. Specifically, exposure to a news feed featuring 75% instead of 25% of activating articles significantly increases the likelihood of voting in favour of the new transgender law ($OR = 1.68$, 95% CI [1.05, 2.71], $p = .032$). In contrast, neither medium ($OR = 0.85$, 95% CI [0.50, 1.45], $p = .555$), nor high ($OR = 0.76$, 95% CI [0.37, 1.54], $p = .441$) levels of alternative voices significantly affect policy support when compared to the low alternative voices condition.

That said, a comparison with the baseline models shows that once control variables are taken into account, the models' explained variance increases from $R^2 = .01$ to $R^2 = .38$. Accordingly, we do find significant negative effects of gender ($OR = 0.52$, 95% CI [0.33,
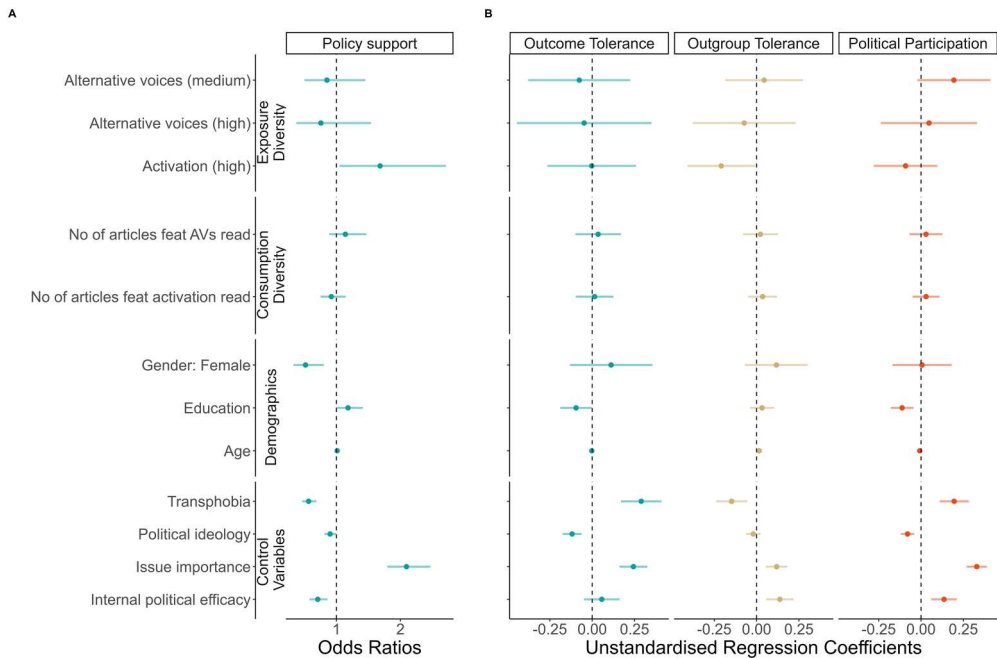
Figure 2. Coefficient plots for linear regression analyses. Results are grouped by type of independent variable: (1) exposure diversity/experimental conditions, (2) consumption diversity/readers clicks on diverse articles, (3) demographics, and (4) control variables.

0.80], $p = .003$), efficacy ($OR = 0.71$, 95% CI [0.58, 0.86], $p = .001$), transphobia ($OR = 0.56$, 95% CI [0.46, 0.69], $p < .001$) and political ideology ($OR = 0.90$, 95% CI [0.81, 1.00], $p = .045$), as well as a significant positive effect of issue importance ($OR = 2.09$, 95% CI [1.80, 2.47], $p < .001$).

4.1.2. Effects on outcome tolerance

H1b and H2b posited that that higher levels of exposure to news that features alternative voices and activation leads to higher levels of outcome tolerance. However, similar to our results for policy support, we again see that person-specific characteristics matter a lot more, the model with control variables explained notably more variance ($R^2 = .09$) than the base model ($R^2 = .01$).

Among our control variables, we find significant positive effects of respondents' transphobia ($b = .29$, $t(702) = 4.74$, $p < .001$) and issue importance ($b = .24$, $t(702) = 5.78$, $p < .001$), as well as significant negative effects of education ($b = −.10$, $t(702) = −2.01$, $p = .045$) and political ideology ($b = −.12$, $t(702) = −4.17$, $p < .001$). In contrast, neither the alternative voices nor the activation manipulations had a significant effect on outcome tolerance, with the same being true for respondents' consumption diversity (see Figure 2B). Thus, we find no support for H1b and H2b.

4.1.3. Effects on outgroup tolerance

H1c and H2c posited that that higher levels of exposure to news that features alternative voices and activation leads to higher levels of outgroup tolerance. However, we again find

no support for either hypothesis. In fact, regarding H2c, we even find the opposite, namely a significant negative effect of the high vs low activation condition on respondents self-reported tolerance towards people who support the opposite referendum outcome ($b = -.21$, $t(702) = -2.06$, $p = .040$).

Among our control variables we find significant positive effects of age ($b = .02$, $t(702) = 4.39$, $p < .001$), issue importance ($b = .12$, $t(702) = 3.71$, $p < .001$) and internal political efficacy ($b = .14$, $t(702) = 3.68$, $p < .001$) and a significant negative effect of transphobia ($b = -.15$, $t(702) = -3.17$, $p = .002$). Moreover, we once again see that the full model ($R^2 = .14$) with control variables explains a lot more variance than the base model ($R^2 = .02$).

4.1.4. Effects on political participation

H1d and H2d posited that that higher levels of exposure to news that features alternative voices and activation leads to higher levels of political participation. However, apart from a marginally significant positive effect of medium levels of alternative voices ($b = .20$, $t(702) = 1.76$, $p = .078$), we again find no support for either hypothesis.

Similar to the analyses discussed above, we again find that the model with control variables ($R^2 = .22$) explains notably more variance than the one without ($R^2 = .01$). Among our control variables, age ($b = -.01$, $t(702) = -2.32$, $p = .021$), education ($b = -.11$, $t(702) = -3.29$, $p = .001$), and political ideology ($b = -.08$, $t(702) = -3.90$, $p < .001$) emerge as negative predictors, while higher levels of issue importance ($b = .33$, $t(702) = 10.91$, $p < .001$), transphobia ($b = .20$, $t(702) = 4.48$, $p < .001$) and internal political efficacy ($b = .14$, $t(702) = 3.54$, $p < .001$) emerge as positive predictors of respondents' intention to become politically active.

4.2. Exploratory analyses

If present, the effects of news diversification are likely to be mediated in some form, for example because engaging with diverse news may affect how people perceive the issue at hand. Thus, RQ1 asked whether respondents' levels of empathy as well as symbolic and realistic threat perceptions may mediate potential effects of exposure diversity. Given the lack of significant effects of the alternative voices manipulations, we could only test this for the effects of activation on policy support and outgroup tolerance. However, we found no evidence for a mediation in either case (see supplementary materials; Tables 10 and 11).

A common concern regarding increased exposure diversity is that at some point, users may become alienated. Therefore, RQ2 asked how the experimental conditions affected readers' engagement and satisfaction. In terms of engagement, linear regression analyses indicate that high levels of activation ($b = -.31$, $t(704) = -2.41$, $p = .016$) and medium ($b = -.32$, $t(704) = -2.04$, $p = .042$), but not high ($b = -.24$, $t(704) = -1.57$, $p = .117$) levels of alternative voices led people to read fewer articles overall. However, when looking at respondents' overall reading time, no such differences emerged (see supplementary materials; Table 8).

Regarding user satisfaction, we had asked respondents to indicate their future use intentions, their perceived quality of the articles, and their overall satisfaction with news supply. Overall, few significant effects emerged. However, we did find a negative

effect of the high activation condition on the perceived quality ($b = -.19$, $t(704) = -2.11$, $p = .035$) and marginally significant negative effects of the high alternative voices condition on future use intentions ($b = -.23$, $t(704) = -1.75$, $p = .080$) and overall satisfaction ($b = -.20$, $t(704) = -1.71$, $p = .087$). Thus, it seems that increased exposure diversity may backfire either by making respondents read fewer articles or perceiving diverse articles to be of lower quality.

In an additional exploratory analysis, we also found a significant negative effect of the high alternative voices condition on readers' perceived diversity ($b = -.23$, $t(704) = -2.19$, $p = .029$). This may suggest that one can only amplify the voices of a specific minority (in this case transgender people) to a certain point before readers perceive the news feed as notably less diverse.

Lastly, responding to RQ3, we find that exposure diversity has at best small and consumption diversity seemingly no significant effects on (a) policy support, (b) outcome tolerance, (c) outgroup tolerance, and (d) political participation, when relevant control variables are taken into account.

5. Discussion

In this study, we tested an often voiced assumption of democratic theory, namely that greater exposure diversity benefits society at large by facilitating pro-democratic outcomes such as greater tolerance and political participation. To facilitate interdisciplinary conversation and integrate more technical literature on news diversification, we did so by operationalising two normatively motivated news diversification metrics as proposed by Vrijenhoek et al. (2021), namely the extent to which news feeds feature minorities and activating language. High levels of activation and alternative voices reflect the type of news content that critical models of democracy promote, whereas lower levels align more closely with the content that deliberative and participatory models of democracy would favour (Helberger, 2019; Vrijenhoek et al., 2021). As such, our results provide insights into how different types of democratic news recommenders may affect news readers.

5.1. Limited effects of exposure and consumption diversity

Our results indicate that mere exposure to diverse news, at least when operationalised in terms of alternative voices and activation, has at best small effects on respondents' policy support, issue tolerance, outgroup tolerance, and political participation. Indeed, when controlling for age, gender, education, issue importance, transphobia, internal political efficacy, and political ideology, we only find significant effects for the activation manipulation and only for policy support and outcome tolerance.

Specifically, we found that those who were exposed to a news feed with 75% rather than 25% articles that featured activating language were more likely to vote in favour of the proposed transgender law, but also less willing to tolerate an unwanted referendum outcome. Coupled with the finding that respondents in the high activation condition perceived the articles to be of lower quality, this casts doubt on whether facilitating activation can contribute to productive democratic discourse – although, as argued

further below, the effects of news diversification likely take more than a single exposure to materialise.

When it comes to the extent to which transgender people are represented in the news as active agents, we do not find any significant effects on our dependent variables. The same is true for respondents' consumption diversity as measured in their number of read articles that featured activating language and transgender people as active agents. Given what we know about selective exposure, the latter may suggest that respondents interacted with the news for different reasons, possibly ranging from looking to learn more to counterarguing with arguments one already disagrees with Stroud (2017).

Thus, our results highlight that, at least in a one-shot experiment, both the exposure to and consumption of diversified news hardly suffices to change people's levels of policy support, outcome and outgroup tolerance and political participation – especially when compared to person-specific characteristics such as demographics or issue-related attitudes.

A possible explanation is that the normatively desired media effects that we tested for only manifest after repeated exposure over longer periods of time. Indeed, societal beliefs tend to change gradually and for long term media effects to emerge, powerful messages and continuous contextual support are necessary (Shehata et al., 2021). In this reading, the (marginally) statistically significant effects that we do find could even be interpreted as evidence that effects may well occur and accrue over time.

5.2. Obstacles towards facilitating critical notions of exposure diversity

Our results indicate that too much exposure diversity can backfire. Specifically, we found that high levels of activating language lead to a lower perceived quality of the articles and that a high representation of minorities ultimately decreases readers' perceived diversity. In regard to the latter, we do however see that this effect only emerges in the more extreme condition. This may suggest that news diversification does not decrease user satisfaction and perceived diversity per se, but only has negative effects once it surpasses a particular threshold – namely an individual's 'latitude of diversity' (Reuver, Mattis et al., 2021).

On a more abstract level, our study also revealed interesting insights into the production and demand side of diverse news. On the demand side, our data suggests that neutral written articles tend to be more popular than articles that feature activating language (possibly due to lower perceived quality), or foreground the lived experiences of minorities. Coupled with the effects on engagement and perceived article quality, this may indicate that audiences only have a limited appetite for extreme versions of news recommenders that adhere to ideals of critical models of democracy.

To an extent, this may also have to do with the content that readers are used to. When creating the stimuli, we noticed that especially the supply of news that featured transgender people as active agents was rather limited. Although far from a complete empirical analyses, this anecdotal evidence suggests that, as Blumler and Cushion (2014) argue, the tendency of journalists to rely on elite sources limits the breadth of news coverage from the perspective of the less powerful. When viewed from the lens of critical democratic theory (Sax, 2022), this poses a considerable challenge for algorithmic news diversification, as such algorithms can do little if no one produces the kinds of news that they are meant to amplify.

Thus, overall we find that both the supply and the audience's appetite for news that adhere to critical notions of democracy is rather limited, thus calling into question the extent to which critical news recommenders could be technically realised and commercially viable.

5.3. Limitations and future research

Our results must be interpreted in light of several limitations. First, as we chose to maximise internal, rather than external validity, our study approach to news diversification falls short of the scope and mathematical nuance of existing news diversification metrics (Vrijenhoek et al., 2021, 2022). Thus, although there are ways to adapt and scale our approach for live news recommender systems – for example by creating binary classifiers that capture our experimental manipulations in real news, or by introducing additional nuance through the mathematical formulations proposed by Vrijenhoek et al. (2022) – this study predominantly provides insights into the effects of news diversification, rather than into their technical implementation. Accordingly, future research should attempt to further bridge the divide between research in communication science, information retrieval, and recommender systems. A particularly promising approach would be longitudinal field experiments (e.g., see Heitz et al., 2022, 2023).

Another limitation pertains to the issue under study. While the controversy surrounding a new Dutch transgender law ensured sufficient variation in terms of policy support and attitudes towards transgender people, it also meant that most people likely already exhibited somewhat strong and stable attitudes prior to the study, possibly suppressing the effects of exposure diversity. Future studies could use more artificial scenarios or less salient issues to explore whether weaker attitudes on the target issue on the whole would change the effects that we found.

Despite these limitations, by combining normative and technical work on news recommendation and testing the effects of different diversification methods within a communication science framework, this study highlights the value of interdisciplinary approaches to complex questions surrounding new technologies. Additionally, the results also provide valuable insights into the promises and pitfalls of normatively driven news diversification that can inform both, future research into the effects of as well as new technical methods for the operationalisation of news diversification.

6. Conclusion

Drawing on handcrafted and carefully pretested news articles on the topic of a new transgender law in the Netherlands, we created six different news feeds that varied in the extent to which they included articles that featured alternative voices and activating language. Thereby, we tested how diversification according to these metrics affects readers policy support, outcome tolerance, outgroup tolerance, and political participation. In a controlled forced exposure experiment, we find that only activating language has significant effects on some of the outcome variables, namely policy support and outgroup tolerance. We also find that high levels of alternative voices negatively affect the number of articles that people read. Overall, our results suggest that normatively motivated news diversification has at best limited short-term effects, thus highlighting the complexity in

strategically employing normative news diversification metrics to facilitate democratically desirable effects.

Notes

1. https://osf.io/zevc3.
2. see https://github.com/nickma101/NewsWebsite.
3. see our preregistration for details: https://osf.io/zevc3.
4. https://osf.io/zqw6s/.

Acknowledgments

Disclosure statement

Funding

Notes on contributors

Nicolas Mattis is a postdoctoral researcher at the Amsterdam School of Communication Research, University of Amsterdam.

Philipp K. Masur is an Assistant Professor of Persuasive Communication at the Vrije Universiteit Amsterdam. His research focuses on different aspects of digital communication including social influence and persuasion processes on social media, privacy and self-disclosure in networked publics, and media literacy.

Judith Moeller is Professor for 'Empirical Communication Research, especially Media Use and Social Media Effects' at the University of Hamburg in cooperation with the Leibniz Institute for Media Research | Hans-Bredow-Institut (HBI).

Wouter van Atteveldt is Professor of Computational Communication Science & Political Communication at the Vrije Universiteit Amsterdam.

Data availability statement

An anonymised version of the data used for the analyses that are reported in this paper can be obtained under the following link: https://osf.io/zqw6s/

References

Aarøe, L. (2011, April). Investigating frame strength: The case of episodic and thematic frames. *Political Communication*, *28*(2), 207–226. https://doi.org/10.1080/10584609.2011.568041

Amsalem, E., Merkley, E., & Loewen, P. J. (2021, August). Does talking to the other side reduce inter-party hostility? Evidence from three studies. *Political Communication*, 1–18. https://doi.org/10.1080/10584609.2021.1955056

Blumler, J. G., & Cushion, S. (2014, April). Normative perspectives on journalism studies: Stock-taking and future directions. *Journalism*, *15*(3), 259–272. https://doi.org/10.1177/1464884913498689

Boyer, M. M., Lecheler, S., & Aaldering, L. (2022, June). Don't throw the frame out with the bathwater: How episodic news frames can prevent identity-motivated reasoning. *Journalism & Mass Communication Quarterly*, 1–22. https://doi.org/10.1177/10776990221097057

Chen, H. T., & Lin, J. S. (2021, April). Cross-cutting and like-minded discussion on social media: The moderating role of issue importance in the (de)mobilizing effect of political discussion on political participation. *Journal of Broadcasting & Electronic Media*, *65*(1), 135–156. https://doi.org/10.1080/08838151.2021.1897822

Dollé. (2023, September). The True T of It All.

Draws, T., Inel, O., Tintarev, N., Baden, C., & Timmermans, B. (2022, March). Comprehensive viewpoint representations for a deeper understanding of user interactions with debated topics. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* (pp. 135–145). ACM. https://doi.org/10.1145/3498366.3505812

Draws, T., Liu, J., & Tintarev, N. (2020, November). Helping users discover perspectives: Enhancing opinion mining with joint topic models. In *2020 International Conference on Data Mining Workshops (ICDMW)* (pp. 23–30). IEEE. https://doi.org/10.1109/ICDMW51313.2020.00013

Groot Kormelink, T., & Costera Meijer, I. (2018, May). What clicks actually mean: Exploring digital news user practices. *Journalism*, *19*(5), 668–683. https://doi.org/10.1177/1464884916688290

Gross, K. (2008). Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. *Political Psychology*, *29*(2), 169–192. https://doi.org/10.1111/j.1467-9221.2008.00622.x

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., & Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science (New York, N.Y.)*, *381*(6656), 398–404. https://doi.org/10.1126/science.abp9364

Heitz, L., Lischka, J. A., Abdullah, R., Laugwitz, L., Meyer, H., & Bernstein, A. (2023, September). Deliberative diversity for news recommendations: Operationalization and experimental user study. In *Proceedings of the 17th ACM Conference on Recommender Systems* (pp. 813–819). ACM. https://doi.org/10.1145/3604915.3608834

Heitz, L., Lischka, J. A., Birrer, A., Paudel, B., Tolmeijer, S., Laugwitz, L., & Bernstein, A. (2022, February). Benefits of diverse news recommendations for democracy: A user study. *Digital Journalism*, *10*(10), 1710–1730. https://doi.org/10.1080/21670811.2021.2021804

Helberger, N. (2019, September). On the democratic role of news recommenders. *Digital Journalism*, *7*(8), 993–1012. https://doi.org/10.1080/21670811.2019.1623700

Helberger, N., & Wojcieszak, M. (2018, September). Exposure diversity. In P. M. Napoli (Ed.), *Mediated communication* (pp. 535–560). : De Gruyter Mouton. https://doi.org/10.1515/9783110481129-029

Himmelroos, S., & Christensen, H. S. (2020, April). The potential of deliberative reasoning: Patterns of attitude change and consistency in cross-cutting and like-minded deliberation. *Acta Politica*, *55*(2), 135–155. https://doi.org/10.1057/s41269-018-0103-3

Iyengar, S. (1990, March). Framing responsibility for political issues: The case of poverty. *Political Behavior*, *12*(1), 19–40. https://doi.org/10.1007/BF00992330

Karimi, M., Jannach, D., & Jugovac, M. (2018, November). News recommender systems – Survey and roads ahead. *Information Processing & Management*, *54*(6), 1203–1227. https://doi.org/10.1016/j.ipm.2018.04.008

Li, M. (2023, March). Exemplifying power matters: The impact of power exemplification of transgender people in the news on issue attribution, dehumanization, and aggression tendencies. *Journalism Practice*, *17*(3), 525–553. https://doi.org/10.1080/17512786.2021.1930104

Loecherbach, F., van Atteveldt, W., Moeller, J., & Trilling, D. (2020). The unified framework of media diversity: A systematic literature review. *Digital Journalism*, *8*(5), 605–642. https://doi.org/10.1080/1369118X.2016.1271900

Loecherbach, F., Welbers, K., Moeller, J., Trilling, D., & Van Atteveldt, W. (2021, June). Is this a click towards diversity? Explaining when and why news users make diverse choices. In *13th ACM Web Science Conference 2021* (pp. 282–290). ACM. https://doi.org/10.1145/3447535.3462506

Lu, Y., & Gall Myrick, J. (2016, July). Cross-cutting exposure on Facebook and political participation: Unraveling the effects of emotional responses and online incivility. *Journal of Media Psychology*, *28*(3), 100–110. https://doi.org/10.1027/1864-1105/a000203

Lunardi, G. M., Machado, G. M., Maran, V., & de Oliveira, J. P. M. (2020, December). A metric for filter bubble measurement in recommender algorithms considering the news domain. *Applied Soft Computing*, *97*(A), 106771. https://doi.org/10.1016/j.asoc.2020.106771

Matthes, J., Knoll, J., Valenzuela, S., Hopmann, D. N., & C. Von Sikorski (2019, October). A meta-analysis of the effects of cross-cutting exposure on political participation. *Political Communication*, *36*(4), 523–542. https://doi.org/10.1080/10584609.2019.1619638

Mattis, N., Masur, P., Möller, J., & van Atteveldt, W. (2024, June). Nudging towards news diversity: A theoretical framework for facilitating diverse news consumption through recommender design. *New Media & Society*, *26*(7), 3681–3706. https://doi.org/10.1177/14614448221104413

Michiels, L., Leysen, J., Smets, A., & Goethals, B. (2022, July). What are filter bubbles really? A review of the conceptual and empirical work. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 274–279). ACM. https://doi.org/10.1145/3511047.3538028

Mutz, D. C. (2002, March). Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, *96*(1), 111–126. https://doi.org/10.1017/S0003055402004264

NA. (2023, October). Wijziging van Boek 1 van het Burgerlijk Wetboek [Text]. https://www.tweedekamer.nl/kamerstukken/wetsvoorstellen/detail.

Napoli, P. (1999). Deconstructing the diversity principle. *Journal of Communication*, *49*(4), 7–34. https://doi.org/10.1111/j.1460-2466.1999.tb02815.xopen_in_new ISSN0021-9916

Niemi, R. G., Craig, S. C., & Mattei, F. (1991, December). Measuring internal political efficacy in the 1988 national election study. *American Political Science Review*, *85*(4), 1407–1413. https://doi.org/10.2307/1963953

Pariser, E (2011). *The filter bubble: What the internet is hiding from you*. Viking/Penguin Press.

Polimeno, A., Reuver, M., Vrijenhoek, S., & Fokkens, A (2023). Improving and evaluating the detection of fragmentation in news recommendations with the clustering of news story chains. In *The First Workshop on Normative Design and Evaluation of Recommender Systems*.

Reuver, M., Fokkens, A., & Verberne, S (2021). No NLP task should be an island: Multi-disciplinarity for diversity in news recommender systems. In *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation* (pp. 45–55).

Reuver, M., Mattis, N., Sax, M., Verberne, S., Tintarev, N., Helberger, N., & van Atteveldt, W (2021). Are we human, or are we users? The role of natural language processing in human-centric news recommenders that nudge users to diverse content. In *Proceedings of the 1st Workshop on NLP for Positive Impact* (pp. 47–59). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.nlp4posimpact-1.6

Sax, M. (2022, September). Algorithmic news diversity and democratic theory: Adding agonism to the mix. *Digital Journalism*, *10*(10), 1650–1670. https://doi.org/10.1080/21670811.2022.2114919

Semetko, H., & Valkenburg, P. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, *50*(2), 93–109. https://doi.org/10.1111/j.1460-2466.2000.tb02843.x

Shehata, A., Andersson, D., Glogger, I., Hopmann, D. N., Andersen, K., Kruikemeier, S., & Johansson, J. (2021, January). Conceptualizing long-term media effects on societal beliefs. *Annals of the International Communication Association*, *45*(1), 75–93. https://doi.org/10.1080/23808985.2021.1921610

Stroud, N. J (2017). Selective exposure theories. In K. Kenski, & K. H. Jamieson (Eds.), *The Oxford Handbook of Political Communication*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199793471.013.009

Stroud, N. J., & Muddiman, A. (2013). Selective exposure, tolerance, and satirical news. *International Journal of Public Opinion Research*, *25*(3), 271–290. https://doi.org/10.1093/ijpor/edt013

Treuillier, C., Castagnos, S., Dufraisse, E., & Brun, A. (2022, July). Being diverse is not enough: Rethinking diversity evaluation to meet challenges of news recommender systems. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 222–233). ACM. https://doi.org/10.1145/3511047.3538030

Van Doorn, M. (2014, October). The nature of tolerance and the social circumstances in which it emerges. *Current Sociology*, *62*(6), 905–927. https://doi.org/10.1177/0011392114537281

Vrijenhoek, S., Bénédict, G., Gutierrez Granada, M., Odijk, D., & De Rijke, M.. (2022, September). RADio – Rank-aware divergence metrics to measure normative diversity in news recommendations. In *Proceedings of the 16th ACM Conference on Recommender Systems* (pp. 208–219). Association for Computing Machinery. https://doi.org/10.1145/3523227.3546780

Vrijenhoek, S., Kaya, M., Metoui, N., Möller, J., Odijk, D., & Helberger, N. (2021, March). Recommenders with a mission: Assessing diversity in news recommendations. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (pp. 905–927). ACM. https://doi.org/10.1145/3406522.3446019