Is this a click towards diversity? Explaining when and why news users make diverse choices

Felicia Loecherbach
f.loecherbach@vu.nl
Vrije Universiteit Amsterdam,
Department of Communication
Science
Amsterdam, NL

Kasper Welbers
k.welbers@vu.nl
Vrije Universiteit Amsterdam,
Department of Communication
Science
Amsterdam, NL

Judith Moeller
j.e.moller1@uva.nl
University of Amsterdam,
Department of Communication
Science
Amsterdam, NL

Damian Trilling
d.c.trilling@uva.nl
University of Amsterdam,
Department of Communication
Science
Amsterdam, NL

Wouter van Atteveldt
wouter@vanattelveldt.com
Vrije Universiteit Amsterdam,
Department of Communication
Science
Amsterdam, NL

ABSTRACT

Modelling the different factors that lead people to choose news articles is one of the key challenges for understanding the diversity of news diets – as a news diet is the result of a series of decisions for certain articles over others, a sequence of choices that was made by the individual consumer. This study sheds light on the interplay between content-related (past behavior, habits, preferences) and situational factors (positioning, saturation, control). The latter could offer possibilities to promote more unexpected news encounters that diverge from past news consumption. To test this, a Python-based web application for interactively testing different forms of news personalization over time was used. 247 respondents used the system over a two-week period, in total making almost 23,000 choices. Our results show that: (1) Selections are influenced by a strong positioning effect that follows a reading pattern (left-right, up-down). This effect is stable across devices, topics, and preferences. (2) How much control people are given influences the length and the amount of different sessions (personalization leads to fewer and shorter sessions). (3) With high control, the diversity of preferences influenced the diversity of selected news more, possibly widening gaps between diversity-seeking and -averse users. (4) How often a topic was chosen in the last hour negatively impacts whether it gets chosen again, showing saturation effects. (5) Clicks on sports and economic articles can be explained by preferences, but not past behavior; for political news the opposite is found. (6) There is no significant correlation between the actual diversity (presented or selected topics) and the topic diversity perceived by the users – in spite of clear differences in actual diversity between the groups. From this we can conclude the importance of situational factors in modelling news selection and their potential to narrow or widen the diversity corridor. In sum, our results contribute to a better understanding of the interaction of news recommender systems and humans and how this shapes which news articles get chosen.

CCS CONCEPTS

• **Information systems** → **Personalization**; • **Applied computing** → *Law, social and behavioral sciences*; • **Human-centered computing** → **User studies**; Web-based interaction.

KEYWORDS

diversity, news recommender systems

1 INTRODUCTION

News selection online is shaped by a high-choice environment. In theory, users are able to choose from a large amount of news articles that are being produced every second of every day. To deal with this information overflow, news selection is more and more shaped by different forms of news personalization. Thus, adapting the presentation of news articles on a page to the individual user based on their explicit wishes, past behavior, or the behavior of similar users. This has consequences for the way we are exposed to and consume news. Often concerns are voiced about the impact that news personalization has on the diversity of information that a person sees and reads (such as in discussions on 'filter bubbles'). The diversity of consumed news is seen as important building block in shaping citizens' knowledge and attitudes. Therefore, scholars call for diversity as a design principle in recommender systems [9].

However, in order to design those systems effectively, it is important to take a close look at factors influencing news selection choices of users and how they might contribute to a diverse news diet. News personalization takes place in interaction with users

who have varying levels of control over which news items are displayed. In the end, via explicit choices for personalization settings or by changing news selection patterns used as basis for a recommendation system, users also shape the way the system is personalized. Understanding why users make certain choices over others and how this happens in interaction with different forms of personalization can thus be seen as critical for the development of news recommender systems. We assume that there is an interplay between content-based, habitual factors that lead people to stay within their comfort zone and situational factors that can help in disrupting this process and lead people to select news stories that do not match their preferences or habits. By using an experimental design that allows the interaction of users with different kinds of news personalization over a longer period of time, we disentangle those different factors in predicting news choices and give insights for further development of news recommender systems.

2 PREVIOUS WORK

2.1 News consumption and diversity

The diversity of media systems and exposure is essential to a democratic formation of public opinion. While the focus in academic research was put mostly on supply diversity (breadth of content being spread by media outlets), during the last decade the attention has increasingly shifted towards exposure or consumption diversity [15, 22]. In a high-choice media environment, it becomes more and more relevant to study the full range of news consumed by users. For this purpose, researchers often investigate all news exposures in a given time frame simultaneously. From those, measures of diversity can be calculated such as Shannon's entropy. This can indeed give a good overview of how diverse the media diet of individuals is and in how far they are comparable.

What is lost in this process, however, is how this aggregate being studied came about. A more or less diverse news diet is the result of a series of decisions for certain news articles over others, a sequence of choices that was made by the individual consumer. They can be explored by looking at chains of choices in digital traces of news consumption [18, 32]. In this case, diversity becomes something that is less focused on the absolute level but is more brought back to an individual measure. One can for example imagine a user with a generally diverse news consumption – a news-junkie or omnivore [5]. At the opposite we have a user who only ever consumes sports news about a particular team they support. For the first user, most clicks they make are in line with expectations of their already broad baseline. For the second user, however, every click that is outside of their limited news diet substantially broadens their horizon, for example by clicking on an article on background information on a corruption scandal in sports with economic impact. Thus, the potential of a news selection to broaden the horizon of the individual is dependent on how far from the expected baseline the selected item can be placed. Every person has their own diversity corridor, a set of news that are within the acceptable range of news topics and content that they normally would want to consume. This ties in with the general idea of users being either diversity-seeking or challenge-averse [21]. Here an analogy to the latitudes of acceptance and rejection from the social judgment theory (SJT, [26]) can be made. In this theory, an individuals' own attitudes are

seen as anchoring points against which all other attitudes are being evaluated, some opinions are deemed acceptable while others are not – however, this range is not set and it can shift over time.

This corridor is not seen as static, it can widen and narrow over time. An improvement regarding diverse news consumption would be indicated by seeing users stepping out of their normal consumption patterns to select news that are unexpected. One possible goal of diversity-oriented recommender systems could thus be to increase the unexpectedness of news choices for each individual, taking into account the respective baselines. However, in order to design such systems that nudge users into more diverse consumption, a model is needed that explains how we arrive at this baseline and what different factors can enhance the chances of unexpectedness. In the following, we will thus first explain the theoretical underpinnings of a diversity corridor as baseline before proposing three factors explaining unexpected choices that can potentially widen that corridor.

2.2 Diversity corridor – it's all about the baseline

A plethora of different theories and frameworks have helped in better understanding media choices (for an overview see [8]) – among the most prominent are the Uses and Gratifications approach [13], Informational Utility Approach [2], and Selective Exposure/Cognitive dissonance [4]. Those theories have in common that they assume a (more or less) rational user who makes choices based on characteristics of media content they are confronted with. When being given the choice between media article A and B, elements such as the topic or stance of each article are evaluated and the "better" option is chosen. What the better option will be can for example be predicted by earlier experiences, preferences, and attitudes. News consumption is highly habitualized or even automated behaviour, which makes news choices more predictable [7]. The predicted choices fall within what we call the diversity corridor of a person. Selective exposure and uses and gratifications thus generally predict that news consumers will be drawn towards "more of the same", to choose the topics that they prefer over and over again – a "sports person" choosing predominantly sports news etc.

However, even with a perfect estimation of a person's interests, this approach of explaining media choices – relying on general content preferences/avoidances to predict news choices – has received considerable criticism over the years. The media landscape looks entirely different compared to a few decades ago: Information is available 24/7 for most users with new content flooding in every second, requiring choices out of vast catalogues of available news stories to choose from. In handling this information overload, relying on users making decisions based on content factors falls too short since "being a motivated reasoner takes effort" [29, p. 757]. In this case, habitual factors are of importance for reflecting a persons' interests. But apart from rational decisions and habits, the context of the news selection also plays a large role in a high-choice media environment.

This shifts the focus much more towards situational factors in media selection that can cause deviations from the baseline. Those factors have in common that they do not concern the content of

the news (the topic, the stance, the source etc.) but are nonetheless crucial in influencing news selection patterns in a high choice environment. In this paper we are focusing on three in particular: Positioning effects, saturation, and control of the news environment.

2.2.1 Positioning effects. Positioning effects are probably among the most underestimated factors for news selection in theoretical academic models on news choice. While in the research on advertisements, recommender systems and search engines it is widely accepted that one of the most important factors for why something is read or clicked on is the position [3, 27], it often plays a minor role in theoretical explanations of news choices. Still, much is known about why users choose earlier items first and often follow a pattern that is similar to the reading habits of their culture (in Western cultures left to right, top to bottom [1]). Those positioning effects can be placed in the area of decision biases, they affect human memory and have been studied and found for centuries (for an overview see [30]). Joachims and Radlinski [10] show that the rank of a search machine result link largely determines whether it is chosen, while Collins et al. [3] in researching digital library recommender systems showed that "articles recommended at higher positions received significantly higher click-through rates than expected, regardless of their actual relevance" (p. 343). Again, those mechanisms can also be explained by habitual factors in news usage – being used to clicking on the first item on the page as a routine to stay informed about breaking news for example. Thus, it can be assumed that items appearing higher up on the page will be clicked on more often, irrespective of their content. However, it might be that certain other factors weaken this effect: if a topic is of high interest to a person, one might still click on it even though it is lower on the page. Likewise, some topics might simply be seen as more important or more urgent than others, interfering with positioning effects. Lastly, it can be assumed that the kind of device (mobile phone vs laptop/desktop [PC]) plays a large role due to screen size. Therefore, we propose the following first research question:

RQ1: To what extent can positioning effects be found in news selection and how consistent are they across (1) preferences, (2) topics, and (3) devices?

2.2.2 Control of the news environment. Increasingly, both algorithmic solutions and customization options for personalizing news environments are being developed, enhanced, and tested in both academic and commercial settings [12]. Users can explicitly alter settings (e.g. by choosing preferred topics) or the system adapts to them via feedback loops – every choice sets a new stage for the next decisions. This means that the news to choose from become more tailored to the individual. This could impact the usage of a news website in two different directions: On the one hand the news recommendations will be closer to the diversity corridor of the user, making them more interesting and prompting them to stay on the website. Therefore, personalization systems are often seen as useful for news providers due to the "need to maximize the relevance of content to individual users" [31, p. 785] in advertisement-based business models. This can also be found more prominently in the entertainment industry – for example Netflix keeping users on the

platform by suggesting relevant new content to watch [6]. Longer sessions also mean the potential to encounter new things and news stories outside of one's diversity corridor. On the other hand, news are often consumed for reasons other than entertainment and rather fulfill a specific informational or surveillance purpose (as the theories of Informational Utility and Uses and Gratifications predict). When only wanting a quick update on the news of the day on specific topics, more tailored news help in finding what was searched for, potentially shortening the browsing session.

RQ2: To what extent does news personalisation (explicit and implicit) impact the amount and length of browsing sessions?

In the last few years, discussions about the danger of algorithmic personalization on reducing the diversity of news diets were prominent in academic and public discourse [35]. Although many scholars have shown these concerns to be overrated in current media systems, others call for a stronger involvement of the user in the news personalization process to avoid potential issues in the future [25]. The pessimistic stance of overpersonalization is at odds with the position that sees user control as favorable from both normative standpoints and the user perspective [20]. However, there seems to be a gap between those theoretical ideas and the practical testing in experimental settings: User control also means that factors such as selective exposure (e.g. primarily choosing content that agrees with one's preferences) become more pronounced. Therefore, those with a wider diversity corridor (meaning: interest in many different topics) also customize to see many different topics while those with a more narrow baseline also choose for a more limited set of topics. Potentially, this could lead to a widening of the differences between those already interested in diversity and those that are not.

RQ3: To what extent does the amount of control affect the relationship between the diversity of topic preferences (the baseline) and the diversity of read topics?

2.2.3 Saturation. News selection decisions by the same user can have a mental impact on future decisions – especially in the short term. After having read a few articles on a topic of a persons' choice, all information on this topic has been consumed and new content is needed. Even an avid football supporter can only read that many articles about their team in a row before 1) there is no more information left, their information need (as predicted by the Informational Utility approach) is satisfied and 2) getting the same content again gets boring and less stimulating (satisfaction of e.g. entertainment gratifications). In this case, the user needs to search for new, different content [12, 17]. This phenomenon of having to make adjustments based on situational factors is very similar to the theory of mood management [34]. As is known from mood adjustment strategies, "goals will vary with the context and individual motivation to meet perceived situational requirements" [14, p. 236]. In those moments, unexpected or serendipitous media choices might occur that can be explained by the past choices of a specific preceding time frame – a very homogeneous news choice often can only be interesting for so long before becoming too monotonous and boring, requiring adjustment from the user.

RQ4: Under which conditions does the selection of a topic in the recent past reduce its likelihood of being selected again, showing saturation effects?

2.3 Explaining diverse choices: A theoretical model

The main purpose of this paper is to develop and test a model for explaining the interplay of different factors leading to the choice of some news items over others. The specific focus is put on the interplay between content-related factors (i.e. explicit preferences and long-term behavior) and situational factors (position, saturation). Situational factors are seen as possibly promoting more unexpected news exposures, helping us to explain how diverse media diets develop and can be promoted. By this we contribute to the research on news consumption by shedding the light on how to model diverse media choices. It can also help in showing possible pathways for developing more diversity-promoting news recommendation systems.

RQ5: To what extent can content-wise and situational factors explain topic choice?

3 METHODS

To answer the research questions, a platform for testing an interactive news recommender system for a longer period of time is used [16]. It is based on an open-source Python application and presents the user with a web interface showing a selection of nine different news articles in tiles arranged in a 3x3 grid. For each article, the title and a short teaser are shown, additionally the topic is indicated with a colored tag. For more information on the interface and setup of the website, a detailed description can be found in [16]. The news are retrieved from several RSS feeds of different Dutch news providers (similar to strategies employed in [24]).

This ensures that the articles presented are recent as well as actual news (and by that newsworthy) — a crucial aspect since "in contrast to other domains like movie recommendation, the relevance of news items can change very rapidly" [12, p. 1204]. Thus, the website is intended to resemble a news website in a controlled design (i.e. without distracting advertisements and additional cues such as pictures and other tags). A set of nine items is used per recommendation round since the amount and placement of stories plays a crucial role in recommender systems [11], following an inverted U-shape where too few and too many choice options negatively impact the satisfaction with the system. On a PC and tablet, all elements can be displayed on one page without scrolling – which was not possible for mobile devices, where all articles are displayed below each other. In an initial questionnaire, participants were asked about their news usage, news topic preferences for a set of 8 different topics (e.g. politics, sports [23]) and the strength of their news preferences [33].

After the initial questionnaire, participants were redirected to the website and prompted to create an account. During the registration they were randomly assigned to experimental conditions changing the amount of control over the news environment. One group received random news stories (*random*: *n* = 57), for two groups different recommendation algorithms personalized the news environment: one similarity-based recommender (*Rec_A*: *n* = 81) and one topic-based recommender (*Rec_B*: *n* = 67). The last group was in charge of the personalization by having the option to state and adapt their topic preferences (*customization*: *n* = 42). For more detailed information on the customization and recommendation algorithms used see [16]. For all groups but the random group 6 out of the 9 stories were personalized, either according to their explicit wishes (customization) or based on their behavior (recommendation algorithm). Information on and timestamps of each set of news articles an individual could choose from (exposure) was saved, as well as the selection made (consumption) and the rating of this selection (5-star rating). After having gathered 80 interaction points (awarded for reading and rating news stories over the course of the whole study) and having logged in on at least 7 different days, participants could finish the study.

The data were collected between October and December of 2019 in the Netherlands. Dutch-speaking participants were recruited via a commercial online panel company. Participants were told that they were testing a first version of a new news website but were given no specific task but to use the website over the course of a week. The sample had an average age of 46.72 (*SD* = 13.83, range 18-86), with balanced gender representation (113 male, 134 female). 43 respondents had a secondary degree (VMBO, HAVO/VWO), 71 a middle-level applied education (MBO), 85 a higher professional education (HBO), and 48 a scientific education (WO). Randomization checks showed no significant differences between the experimental groups regarding sociodemographic variables.

All data and code used for the analyses reported in this article can be found at https://github.com/FeLoe/diversity-corridor/releases/tag/v.1.0.

4 RESULTS

4.1 Usage of the system

Table 1 shows an overview of the topics displayed and selected by the users. As can be seen, in total nine different topic categories (plus an "Other" category) could be presented to the users. Three topics (Immigration, Environment and Science) were presented to the users less than 3,000 times, so in less than 15% of all exposures. It turned out that those topics were simply not that present in the news sources we scraped to fill the page. In consequence, they also could not be selected much, each topic less than 500 times out of the 23,000 selections. Therefore, we decided to exclude those three topics from the analysis since respondents did not often even have the option of choosing that topic, posing problems with the subsequent analyses. The "other" category was also excluded from due to its variable content and since no preferences for this category had been asked in the preceding survey. Hence, all results that we present do not include these four categories.

In total, 248 participants finished the study. One user was excluded due to having an unusually large amount of selections (1232), more than double compared to the second most active user. On average, each user made 87.82 news selections (*SD* = 65.04), ranging from 23 to 456. The news selections are grouped in sessions, after at least 15 minutes of inactivity on the website a new session starts. The average user visited the website in total almost 20 times (*M* = 16.95, *SD* = 9.38), ranging from 6 to 74. An average session is 5.58 news selections long (*SD* = 3.74) – however, the length steadily

decreases from 7.33 stories, after about 40 sessions the average is down to reading 1-2 articles in one session. For some users the length of sessions sharply increased again after 40-50 sessions – most likely due to a lack of understanding of the incentivization leading to the attempt to gather points by clicking on many stories on one day (a behavior that was not rewarded). For all other users, the usage of the website went down over time.

Table 1: Frequency of displayed and selected topics

Topic	Displayed	Selected	Percentage Chosen
Sports	55,922	3,909	6.99
Crime	42,932	5,587	13.01
Domestic Politics	41,877	5,368	12.82
Entertainment	33,192	3,621	10.91
Foreign Politics	15,946	1,887	12.87
Economics	10,250	1,319	12.87
Science	2,870	477	16.62
Environment	2,087	311	14.90
Immigration	490	61	12.45
Other	29,172	3,542	12.14

4.2 Positioning effects

Turning to RQ1 (the prevalence of positioning effects) shows that overall news articles that were placed on the top left corner of the page – where the reading starts in Western cultures – were also selected most often. Figure 1 shows an overview of the chances of an article being positioned on a particular position – in general, but also split up by device, topic and preferences. Overall, articles on the first position were selected in 31% of the cases. After this, the chances of an article to get picked go down until the last position on the page (right lower corner), which only gets selected in 8% of the cases. But the question is how "stable" this pattern is. As could be expected, positioning plays the largest role for mobile devices, the articles in the top position were selected in more than 50% of the cases. On tablets and PC's the positioning effect is still clearly visible but less pronounced. On all devices stories placed lower than on the fourth spot (for PC and tablet: the first row and first article of the second row on the page) have a 1 in 10 chance or lower of being clicked on. When looking at the different topics, almost identical patterns can be found for all of them – with one notable exception: sports news. They were much less likely to be selected in general and also showed a less pronounced positioning effect. It seems that respondents in general were rather avoiding sports news – possibly caused by them being too frequent, too repetitive or not interesting for the user (e.g. about a sports or team they do not like). Lastly, it is interesting to see whether the positioning effect is different for topics that respondents stated they are really interested in compared to those they do not usually follow. Here, it can be seen that the pattern is similar for high interest (5 and above on a 7-point scale) and medium interest (2-4), but clearly diverges for the topics that a user was not at all interested in (1). Those topics got picked less often, even if they appeared high on the page. All of this shows that the positioning effect is strong and quite stable. It is even stronger on mobile devices and less pronounced for sports articles and topics far out of a person's personal preferences. For
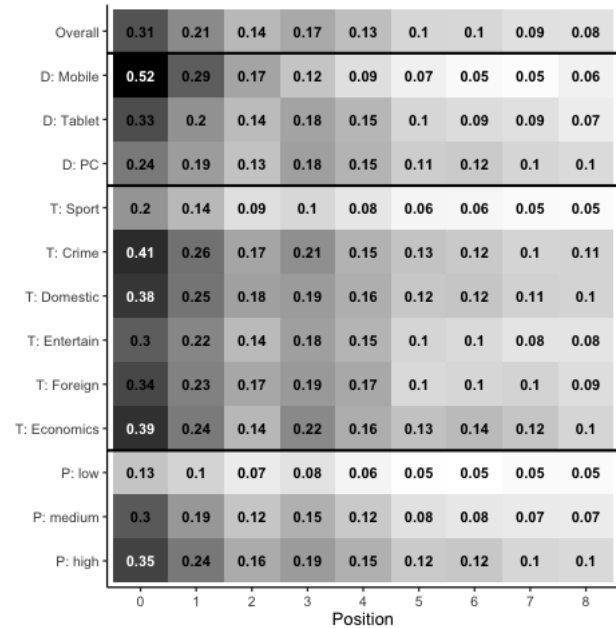

Figure 1: Proportion article was selected when on specific position (from upper left corner (0) to down right corner (8)). D = Device, T = Topic, P = Preference. Preferences were aggregated with low = 1, medium = 2-4, high = 5-7.

all other topics, simply being placed higher on the page greatly enhances the chances of being selected.

4.3 Control

The second research question refers to the impact of news personalization on the amount and length of the sessions. The four experimental groups had different amounts of control, starting with a random selection of news (*random*), followed by two recommender systems that offered implicit control via the stories a user clicked on (*Rec_A, Rec_B*) and one group fully in control via customization (*custom*). Table 2 shows the results of a Kruskal-Wallis one-way ANOVA testing for differences in number of selections (how many news stories were selected) and number of sessions (how often users came back after a break of at least 15 minutes). As can be seen, the random group selected more news stories compared to all other groups, followed by Rec_A. The two groups with clearer implicit and explicit personalization (Rec_B and custom) selected less news stories. A similar pattern can be observed for the amount of sessions: People who were in control of the information environment (custom) came back the least, followed by Rec_B and Rec_A. The random group had more sessions than the other groups. In terms of session length (how many news stories were selected in one go), the random group has the highest average ($M = 7.16$, $SD = 6.25$) while all other groups have average session lengths between 4 and 5 articles. All of this shows that users when being given random news selected more news items and returned more often to the website – the more explicit the personalization got the less they stayed on the website. This might indicate that people given random stories

had to spend more time on the website until they found content relevant for them that satisfied their information needs.

Table 2: Kruskal–Wallis one-way ANOVA with Dunn Test for pairwise comparisons (Holm-corrected p values)

Variable	Median	Comparison (p-value)		
		Rec A	Rec B	Custom
N Selections (df =3, H = 35.95, ε^2=.15, p<.001)				
Random	94	.019	<.001	<.001
Rec A	71	-	.019	.01
Rec B	58		-	.521
Custom	56			-
N Sessions (df =3, H = 33.19, ε^2=.13, p<.001)				
Random	17	.199	.018	<.001
Rec A	15	-	.2	<.001
Rec B	14		-	.005
Custom	11			-

The third research question refers to the relationship between the diversity of topic preferences and the diversity of read topics and whether this was influenced by the amount of control. When looking at the news preferences participants stated in the survey, the most popular topic was domestic news (M = 5.64, SD = 1.17), only eleven people rated it lower than 3 on the 7-point scale. The two least popular topics were sports (M = 3.85, SD = 2.12) and entertainment (M = 3.69, SD = 1.64). Especially sports polarized with more than a fifth of respondents giving it a 1 on the scale (n = 52), and a quarter rating it 6 or higher (n = 69). Looking back at the amounts of topics displayed and selected (see Table 1) this large group of respondents disliking sports topics can also be found in the low percentage of sports articles chosen when displayed compared to all other topics. A diversity baseline variable was calculated by summing the topics that were rated higher as 4 on the preference scale – it thus ranges from 0 to 8. The consumption diversity was defined as the topic entropy of selected articles by a person (using Shannon's Entropy). On average, people indicated interest in 4.47 topics (SD = 2.18). A randomization check in form of a Kruskal Wallis test showed no differences between the groups regarding this diversity baseline (H(3) = 3.87, p = .276). Multiple regression analysis was used to test if the experimental groups and the diversity baseline significantly predicted participants' consumption diversity. The model controlled for gender, age, education, political knowledge and political interest. The results of the regression indicated the predictors explained 18% of the variance ($R2$ = .18, F(12,235)=4.33, p <.001). It was found that the diversity baseline did not significantly predict the consumption diversity (β = .01, p =.29). However, it does show a significant interaction effect between the control group and the diversity baseline (β = .04, p =.02). As Figure 2 shows, there is indeed a positive relationship between baseline and consumption diversity in this group. In other words: When being in control, it matters whether the person has diverse preferences or not. For the other groups, no tendency can be seen.

4.4 The integrated model

Research questions four and five – referring to saturation effects and the combination of the different factors – were investigated
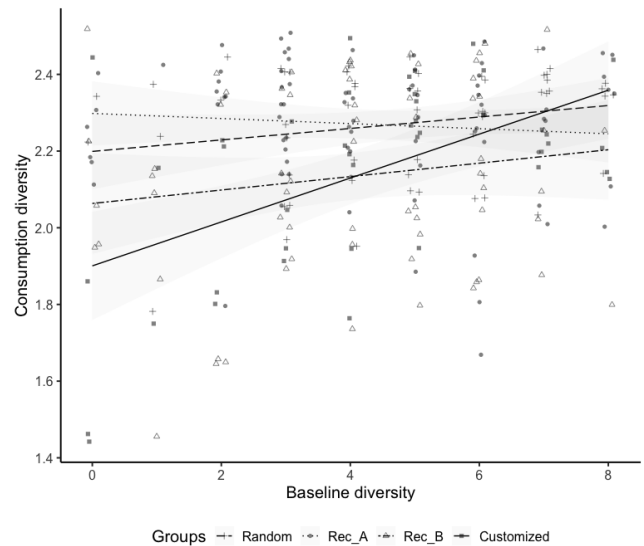

Figure 2: Relationship between baseline diversity and consumption diversity per experimental group (higher values = more diverse).

with one integrated model for predicting topic selection, which we present per topic as well as in a single model with topics pooled (Table 3). The cases in this model are the unique topics that were presented to a participant for every exposure to 9 articles, and the dependent variable is whether (1) or not (0) the topic was selected. In other words, the model predicts how likely a participant is to select a topic given that the topic was present in the current recommendation set. We used a multilevel logistic regression model to analyze how the odds of selecting a topic was affected by how the topic was presented in the app, the user's self indicated topic preferences, and the proportion of cases that the user recently selected the same topic. We included random intercepts for users to account for user level differences in the odds of selecting a topic. Unless specified, we report the coefficients of the pooled model.

For the presentation of the topic we included the average position of the topic in the exposure[1] and controlled for the number of times the topic occurred in the exposure. The effect of position (odds ratio = 0.86; p < .001) extends the previous observation that articles in the lower positions are much more likely to be selected with the observation that this affects the selection of topics, also when controlling for topic preference and topic proportion. The predicted probability of selecting a topic in the first position (0.358) is more than two times as high compared to the last (9th) position (0.144). The effect of *topic preference* (odds ratio = 1.11, p < .001), as a user level variable, indicates whether higher self-reported preference for a topic also predicts a higher likelihood of selecting it. Interestingly, this was overall the case, but not for the topics *Domestic* and *Foreign*. Without moving beyond tentative speculation, it is interesting to consider that this might result from these topics being

---

[1]Since a topic can occur in multiple articles within each exposure of 9 articles, article-level predictors need to be aggregated.

Table 3: Multilevel logistic regression analysis, predicting topic selection

	Was a topic selected when it was an option in a selection?						Pooled topics
	Domestic	Foreign	Entertainment	Law & Crime	Sports	Economy	
Fixed Effects							
(Intercept)	0.18***	0.11***	0.10***	0.14***	0.03***	0.08***	0.09***
Position	0.86***	0.87***	0.87***	0.86***	0.84***	0.88***	0.86***
Topic in selection	1.77***	1.94***	1.82***	1.76***	1.65***	2.17***	1.68***
Topic preference	1.02	1.01	1.07**	1.10***	1.34***	1.10**	1.11***
Lag topic selection:							
last 3 days	2.26***	7.15***	3.09***	2.16***	1.62	2.67*	6.33***
last hour	0.64***	0.48***	0.58***	0.71***	0.75**	0.41***	0.83***
Random Effects							
σ^2_{event}	3.29	3.29	3.29	3.29	3.29	3.29	3.29
τ_{user}	0.05	0.03	0.11	0.10	0.50	0.00	0.00
τ_{topic}							0.05
ICC	0.01	0.01	0.03	0.03	0.13	0.00	0.02
AIC	11400.14	4020.01	9073.68	11790.87	10360.18	2496.99	49620.92
N	10068	4381	9036	10225	11596	2670	47976
N_{users}	245	220	237	244	244	185	247
N_{topics}							6

Note:

*p<0.05; **p<0.01; ***p<0.001
Coefficients transformed to odds ratios

rather broad and unspecific, which could makes topic preference less informative.

To measure saturation effects we looked at the proportion of cases that the users selected the same topic, which we measured separately for a short term (1 hour) and longer term (3 days) time window.[2] To calculate these *lag topic selection* variables, the user had to be exposed to the topic at least once in both the 3 day and 1 hour window. We therefore dropped cases where this criterion is not met, and furthermore dropped cases with less than 5 exposures in the 3-day window to get better estimates of the longer term topic selection. The 3-day lag serves as a behavioral measure of topic preference, so the positive effect (odds ratio = 6.33, p < .001) makes sense. The negative effect of the 1-hour lag (odds ratio = 0.83, p < .001) indeed seems to indicate a saturation effect, as it means that topics selected within the past hour are less likely to be selected, controlled for the other variables. Finally, an interesting observation is that the effect of the 3-day window is only absent for sports news. This could be related to the aforementioned speculation that topic preference is more informative for certain topics. If we remove topic preference from the model (not reported), the effect of the 3-day lag (as a behavioral measure of topic preference) does become significant for the sports category (odds ratio = 2.47, p < .01).

4.5 Additional result: Perceived diversity

An additional question that emerges when studying the diversity of news exposure and consumption is whether users do perceive the differences in diversity. Exposure diversity is the entropy of topics *seen*, while consumption diversity is the entropy of topics *chosen* by a person (using Shannon's entropy). Additionally, users were asked at the end of the study to indicate their *perceived* diversity of the website (with items such as "The articles on the website covered many different topics."). The question is how those three measures interact: Does someone with a diverse exposure also make consume diversely and perceive the system as diverse?

Consumption and exposure diversity correlate positively on a medium level (r = 0.44, p < .001) – being exposed to diverse news does thus not necessarily go hand in hand with the consumption of diverse news. Where it becomes even more interesting is when looking at the correlation between perceived diversity and actual (consumption and exposure) diversity: For both consumption diversity (r = -0.01, p = .83) and exposure diversity (r = 0.02, p = .76) no correlation with perceived diversity could be found. Interestingly, there were significant differences regarding actual consumption diversity between the experimental groups with both the custom group and the topic recommender group leading to less diverse consumption[3] – however, no differences in perception resulted from this. Thus it can be concluded that seeing or selecting diverse news does not automatically result in also perceiving the system as more diverse.

5 DISCUSSION

In this paper, we investigated the influence of different content-related and situational factors on topic selection in a news recommender system, giving new insights regarding our understanding of how people interact with news recommenders. This was done within a novel experimental setting that allowed the selection of live-scraped news by respondents over a longer period of time, more closely modelling actual news selection than most other experimental studies. What can we take from the results for future studies on news recommender systems?

Firstly, we showed that positioning effects play a substantive role in the news selection process, hardly influenced by news preferences or topics and becoming even more pronounced for smaller devices. This stresses the need to control for positioning in experimental designs on news personalization. Also, it shows the potential of positioning for diversity-promoting recommender systems – even non-preferred articles will be more likely selected selected if they are only high enough on the page.

---

[2]The observations of the 1-hour window are not included in the 3-day window. The correlation between the variables is 0.269

[3]H(3) = 17.57, p < .001, pairwise comparisons show significant differences between the custom group and both the similarity recommender and random baseline as well as between the topic recommender and both the similarity recommender and random baseline

Secondly, a personalized news environment led to shorter and less sessions compared to random news. This stresses that personalization might not keep people longer on the website (as often is assumed) but rather fulfills the needs as predicted by for example the Informational Utility approach faster, preventing users from further searching for the information they were interested in. This also ties in with the saturation effects found: the more a topic was chosen in the last hour, the less likely it is going to be selected again. Thus one of the challenges for news recommender systems is to find a good balance between giving the user relevant information and keeping them long enough on the website to start looking at news outside of their diversity corridor.

Thirdly, it was shown that when users are given explicit customization options, their diversity corridor (diversity of interest in different news topics) plays a larger role than with implicit or no personalization. This indicates that explicit personalization might rather amplify existing differences between users, giving those with diverse interests diverse news while others stay within their more narrow lane. If the goal of a diversity-enhancing recommender system is to widen the diversity corridor of all users, additional strategies (such as playing with personalization effects through re-ranking stories) should be applied. It also again stresses the need for individualized diversity benchmarks in evaluation that take into account a person's diversity corridor.

Furthermore, we give insights into how explicitly stated preferences and implicit behavior can help in predicting topic choice. We showed that for broader, political topics users were less able to judge their own preferences beforehand. In contrast, for sports news behavioral information was less useful. Those insights can help with building good user profiles for news personalization – and stress that even for news topics a "one size fits all" approach is not desirable.

Lastly, we showed that there is a gap between the actual, measurable diversity that people are exposed to and consume and their perception of diversity. This stresses that all abstract measurements of diversity (such as entropy) might not even correlate with what ends up at the user. This explorative result should receive more attention in news diversity research in the future – disentangling what might cause this inaccurate perception. One possible reason might be that users only start perceiving a lack in diversity if it becomes very obvious (e.g. only sports articles are shown on the page) but not when it is more subtle and there are still several choices left. Getting more insights into the perception and processing of the user can help also in better understanding user behavior and possible incentives for increasing the diversity of one's news diet.

Apart from those results, this study also gave insights into a research design that enhances the ecological validity through using live-scraped news. This brings a lot of advantages when it comes to better modelling user behavior outside of lab settings – but it also leads to many challenges. What this design (in its current form) cannot provide is the possibility to model the different ways that people use for accessing the news (e.g. via social media, via different outlets and actors). This social context of selecting news certainly interferes with a lot of the factors that we looked at in this study. For the sake of a clean experimental design no further context about sources and social endorsements were given, but this also leads to disadvantages for understanding news use in the "real

world". This can better be attempted by studying actual browsing behavior (as done in [19, 28]) – which then in turn is lacking the possibility of varying and control for specific factors. We believe that our study thus provided a good start for finding a balance between experimental and observational designs for studying news decision processes.

Another challenge was the usage of common statistical models for the data: We had to exclude only sparsely appearing topics, and for the multilevel logistic regression model cases with too few exposures in the defined time windows had to be dropped (e.g. if people did not see any economics articles in the past hour, they also could not select them). Overall, quite some data could not be analyzed because of the imbalance of topics in the data set – it would be advisable to make sure that a selection out of all options (e.g. topics) is possible for the user at least most times to accurately model whether a person even had the chance to choose the topic.

In addition, it would be advisable to add an option to not make a choice in case none of the options suited the user, in our setting a choice was forced – however, if no fitting option was present this might not be the ideal solution. Lastly, we for now assumed that the choice for an article was based on a news topic category. But of course there are many more ways of categorizing news that are closer to the actual news content (e.g. entities present, emotionality of headline, whether it is breaking news content) – we could not control for this due to prioritizing having ecological validity with the newest stories being presented to the user. However, for a more "clean" experimental design, controlling for more of those factors would be advisable. Nonetheless, this study provided a blueprint for how different factors influencing news selection in news personalization environments could be analysed and modelled.

ACKNOWLEDGMENTS

REFERENCES

[1] Farough Abed. 1991. Cultural influences on visual scanning patterns. *Journal of Cross-Cultural Psychology* 22, 4 (1991), 525–534. https://doi.org/10.1177/0022022191224006

[2] Charles Atkin. 1973. Instrumental Utilities and Information Seeking. In *New Models for Mass Communication*, Peter Clarke (Ed.). SAGE Publications, Beverly Hills, 205–242.

[3] Andrew Collins, Dominika Tkaczyk, Akiko Aizawa, and Joeran Beel. 2018. Position Bias in Recommender Systems for Digital Libraries. In *Transforming Digital Worlds*, Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett (Eds.). Springer International Publishing, Cham, 335–344. https://doi.org/10.1007/978-3-319-78105-1_37

[4] Leon Festinger. 1957. *A theory of cognitive dissonance*. Stanford university press, California.

[5] Sabine Geers. 2020. News Consumption across Media Platforms and Content. *Public Opinion Quarterly* 84, S1 (2020), 332–354. https://doi.org/10.1093/poq/nfaa010

[6] Carlos A Gomez-Uribe and Neil Hunt. 2015. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)* 6, 4 (2015), 1–19. https://doi.org/10.1145/2843948

[7] Tim Groot Kormelink and Irene Costera Meijer. 2014. Tailor-Made News: Meeting the demands of news users on mobile and social media. *Journalism Studies* 15, 5 (2014), 632–641. https://doi.org/10.1080/1461670x.2014.894367

[8] Tilo Hartmann (Ed.). 2009. *Media choice: a theoretical and empirical overview.* Routledge, New York.

[9] Natali Helberger, Kari Karppinen, and Lucia D'Acunto. 2018. Exposure diversity as a design principle for recommender systems. *Information, Communication & Society* 21, 2 (2018), 191–207. https://doi.org/10.1080/1369118X.2016.1271900

[10] Thorsten Joachims and Filip Radlinski. 2007. Search Engines that Learn from Implicit Feedback. *Computer* 40, 8 (2007), 34–40. https://doi.org/10.1109/mc.2007.289

[11] Michael Jugovac and Dietmar Jannach. 2017. Interacting with recommenders—overview and research directions. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 3 (2017), 1–46. https://doi.org/10.1145/3001837

[12] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems – Survey and roads ahead. *Information Processing & Management* 54, 6 (2018), 1203–1227. https://doi.org/10.1016/j.ipm.2018.04.008

[13] Elihu Katz, Jay G. Blumler, and Michael Gurevitch. 1973. Uses and Gratifications Research. *Public Opinion Quarterly* 37, 4 (1973), 509. https://doi.org/10.1086/268109

[14] Silvia Knobloch. 2003. Mood Adjustment via Mass Communication. *Journal of Communication* 53, 2 (2003), 233–250. https://doi.org/10.1111/j.1460-2466.2003.tb02588.x

[15] Felicia Loecherbach, Judith Moeller, Damian Trilling, and Wouter van Atteveldt. 2020. The Unified Framework of Media Diversity: A Systematic Literature Review. *Digital Journalism* 8, 5 (2020), 605–642. https://doi.org/10.1080/21670811.2020.1764374

[16] Felicia Loecherbach and Damian Trilling. 2020. 3bij3 - Developing a framework for researching recommender systems and their effects. *Computational Communication Research* 2, 1 (2020), 53–79. https://doi.org/10.5117/ccr2020.1.003.loec

[17] Hao Ma, Xueqing Liu, and Zhihong Shen. 2016. User Fatigue in Online News Recommendation. In *Proceedings of the 25th International Conference on World Wide Web - WWW '16.* ACM Press, Montréal, Québec, Canada, 1363–1372. https://doi.org/10.1145/2872427.2874813

[18] Mykola Makhortykh, Claes de Vreese, Natali Helberger, Jaron Harambam, and Dimitrios Bountouridis. 2020. We are what we click: Understanding time and content-based habits of online news readers. *New Media & Society* (2020), 146144482093322. https://doi.org/10.1177/1461444820933221

[19] Judith Möller, Robbert Nicolai van de Velde, Lisa Merten, and Cornelius Puschmann. 2020. Explaining online news engagement based on browsing behavior: Creatures of habit? *Social Science Computer Review* 38, 5 (2020), 616–632. https://doi.org/10.1177/0894439319828012

[20] Cristina Monzer, Judith Moeller, Natali Helberger, and Sarah Eskens. 2020. User Perspectives on the News Personalisation Process: Agency, Trust and Utility as Building Blocks. *Digital Journalism* 8, 9 (2020), 1142–1162. https://doi.org/10.1080/21670811.2020.1773291

[21] Sean A. Munson and Paul Resnick. 2010. Presenting diverse political opinions. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10.* ACM Press, Atlanta, Georgia, 1457–1466. https://doi.org/10.1145/1753326.1753543

[22] Philip M. Napoli. 2011. Exposure Diversity Reconsidered. *Journal of Information Policy* 1 (2011), 246. https://doi.org/10.5325/jinfopoli.1.2011.0246

[23] Nic Newman, David A. L. Levy, and Rasmus Kleis Nielsen. 2016. Questionnaire Reuters Digital News Report 2016. http://media.digitalnewsreport.org/wp-content/uploads/2016/06/Reuters-Institute-2016-Questionnaire-UK270144-.pdf?x57165

[24] Owen Phelan, Kevin McCarthy, and Barry Smyth. 2009. Using twitter to recommend real-time topical news. In *Proceedings of the third ACM conference on Recommender systems - RecSys '09.* ACM Press, New York, 385–388. https://doi.org/10.1145/1639714.1639794

[25] Sander Andreas Schwartz and Martina Skrubbeltrang Mahnke. 2020. Facebook use as a communicative relation: exploring the relation between Facebook users and the algorithmic news feed. *Information, Communication & Society* (2020), 1–16. https://doi.org/10.1080/1369118x.2020.1718179

[26] Muzafer Sherif and Carl I Hovland. 1961. *Social judgment: Assimilation and contrast effects in communication and attitude change.* Yale Univer. Press, New Haven, CT.

[27] Edith G. Smit, Peter C. Neijens, and Robert Heath. 2013. The differential effects of position, ad and reader characteristics on readers' processing of newspaper ads. *International Journal of Advertising* 32, 1 (2013), 65–84. https://doi.org/10.2501/ija-32-1-065-084

[28] Sebastian Stier, Nora Kirkizh, Caterina Froio, and Ralph Schroeder. 2020. Populist Attitudes and Selective Exposure to Online News: A Cross-Country Analysis Combining Web Tracking and Surveys. *The International Journal of Press/Politics* 25, 3 (2020), 426–446. https://doi.org/10.1177/1940161220907018

[29] Charles S Taber and Milton Lodge. 2006. Motivated skepticism in the evaluation of political beliefs. *American journal of political science* 50, 3 (2006), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

[30] Erich Christian Teppan and Markus Zanker. 2015. Decision Biases in Recommender Systems. *Journal of Internet Commerce* 14, 2 (2015), 255–275. https://doi.org/10.1080/15332861.2015.1018703

[31] Neil Thurman and Steve Schifferes. 2012. The future of personalization at news websites: Lessons from a longitudinal study. *Journalism Studies* 13, 5-6 (2012), 775–790. https://doi.org/10.1080/1461670x.2012.664341

[32] Susan Vermeer, Damian Trilling, Sanne Kruikemeier, and Claes de Vreese. 2020. Online News User Journeys: The Role of Social Media, News Websites, and Topics. *Digital Journalism* 8, 9 (2020), 1–28. https://doi.org/10.1080/21670811.2020.1767509

[33] Martijn C. Willemsen, Mark P. Graus, and Bart P. Knijnenburg. 2016. Understanding the role of latent feature diversification on choice difficulty and satisfaction. *User Modeling and User-Adapted Interaction* 26, 4 (2016), 347–389. https://doi.org/10.1007/s11257-016-9178-6

[34] Dolf Zillmann. 1988. Mood management through communication choices. *American Behavioral Scientist* 31, 3 (1988), 327–340. https://doi.org/10.1177/000276488031003005

[35] Frederik Zuiderveen Borgesius, Damian Trilling, Judith Möller, Balázs Bodó, Claes H De Vreese, and Natali Helberger. 2016. Should we worry about filter bubbles? *Internet Policy Review. Journal on Internet Regulation* 5, 1 (2016), 1–16. https://doi.org/10.14763/2016.1.401