Topical Preference Trumps Other Features in News Recommendation: A Conjoint Analysis on a Representative Sample from Norway

ERIK KNUDSEN, MediaFutures, University of Bergen, Norway

ALAIN D. STARKE, Amsterdam School of Communication Research, University of Amsterdam, Netherlands and MediaFutures, University of Bergen, Norway

CHRISTOPH TRATTNER, MediaFutures, University of Bergen, Norway

A variety of news articles features can be used to tailor news content. However, only a few studies have actually compared the relative importance of different features in predicting news reading behavior in the context of news recommender systems. This study reports the results of a conjoint experiment, where we examined the relative importance of seven features in predicting a user's intention to read, including: *topic headline* (Abortion vs Meat Eating), *reading time*, *recency*, *geographic distance*, *topical preference match*, *demographic similarity*, and *general popularity* in a news recommender system. To ensure an externally valid result, the study was distributed among a representative Norwegian sample (N = 1664), where users had to choose their preferred news article profile from four different pairs. We found that a topical preference match was by far the strongest predictor for choosing a news article, while recency and demographic similarity had no impact.

CCS Concepts: • **Information systems** → **Recommender systems**.

Additional Key Words and Phrases: news, recommender systems, choice-based conjoint experiment

ACM Reference Format:

Erik Knudsen, Alain D. Starke, and Christoph Trattner. 2023. Topical Preference Trumps Other Features in News Recommendation: A Conjoint Analysis on a Representative Sample from Norway. In *INRA '23: 11th International Workshop on News Recommendation and Analytics in Conjunction with ACM RecSys 2023 at the 17th ACM Conference on Recommender Systems (RecSys '23), September 18–22, 2023, Singapore.* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

News recommender systems face various domain-specific challenges [20, 28]. A user's interest can strongly depend on contextual factors [20], such as time of the day, the user's current location, or the technology used to read the article [6, 13, 21]. Moreover, while most recommender systems use historical data to present content that a user likes or needs [19], this is challenging in news. There is a fast churn of items, news articles may be updated, and many users do not log in at all [7, 8, 26].

Most news recommender systems face a 'permanent cold-start problem' [20]. Therefore, besides showing the most recent items [1], many news recommender applications focus on content-based recommendation. Central to such approaches are different news article features [20]. These can describe the article's content, authorship, and contextual factors. The list of possible features is long, for it can not only include news article content (e.g., title or headline,

Manuscript submitted to ACM

keywords), but also a user's location or time of the day [12]. Such features can be used in various methods similarity comparisons, such as TF-IDF and cosine similarity [30, 40].

Only a few of these features may be deemed useful by users. Studies have examined news recommender scenario through so-called similar-item recommendations [15, 20]. This is a common method of recommendation on news websites, using the rationale to show 'more like this' of a reference news article a user likes [34]. Recommendation approaches typically involve similarity functions, falling in line with studies on semantic similarity [38].

It is unclear how important each news article feature is, relatively speaking. Most studies involve offline evaluation [20], focusing on training a model using a subset of features with the goal of improving accuracy, instead performing a holistic evaluation of multiple features. One study [34], in the context of semantic similarity, has shown that title-based and body-text similarity functions seem to represent user similarity judgments better than recency-based or image-based features do. This suggests that users are more likely to rate recommendations based on news article text as accurate in personalization scenario. However, this study did not include all factors, such as the proximity of a reader to the news event [3, 32], whether the news article is among the most read articles [14], or whether it is clear why the news article is recommended to a user [10]. Moreover, such studies have not directly compared the relative importance of all news article features.

It is not entirely clear which recommendation mechanisms users would like a news recommender system to take into account. For example, while the offline performance of different recommendation approaches has been examined [20], it is less clear whether users would prefer recency-based recommendations (i.e., content-based) over crowdsourced-based recommendations (i.e., collaborative filtering). Moreover, it is also unclear how their importance relates to other factors, related to personalization based on demographics.

While numerous studies have looked into algorithmic optimization of news retrieval and recommendation [5, 20], less is known about how users evaluate presented recommendations. Also the study of Starke et al. [34] only involves a user's similarity judgment as a dependent variable, and does not include an evaluative measure that is directly related to a recommendation scenario.

1.1 Research Question and Contribution

In this study, we examine the relative importance of a variety of factors in a news recommendation scenario. Included are seven features, which have been identified in earlier studies on news recommendation [20]. First, a news article's *topic* and whether it aligns with a user's interests, because topic modeling and tags are at the core of many news recommender approaches [11], *transparency* about whether it aligns with a user's previous rating (cf. [4, 37]), and whether the news article is *recommended* or not. Moreover, we compare the importance of whether the news article is *recent*, whether it is among the most *popular* news articles, whether it is a short or long article (i.e., *reading time*), and whether the news article discusses events that are spatially close to the user (i.e., *proximity*). We present the following research question:

- **RQ:** In the context of news personalization, which news article features are the strongest predictors of a user's intention to read?

This research question is examined using a method that is novel for the recommendation domain. Whereas most studies will focus on offline evaluation or some form of online evaluation, possibly in the context of an evaluation framework [22, 31], this study presents a controlled experiment. We employ conjoint analysis [16], a type of experiment that is increasingly used in social science research, which aims for understanding multidimensional decision-making, through for instance a best-to-worst scaling of different factors by asking users to indicate a preferred alternative from multiple set of options. Since each of these items contains certain values for different attributes, the importance of these attributes can be determined [16]. This method is one of the most notable advancements in social science experimental research the last decade [23]. In this study, we present pairs of news articles between which users need to choose in terms of their preferences.

Another stand-out aspect of this study is its sample. Our news article vignettes have been evaluated by a representative sample of the Norwegian population, being part of a survey administered by the Norwegian Citizen Panel (NCP). Whereas many computer science studies currently rely on crowdsourcing participants from websites such as Amazon MTurk or Prolific due to convenience and an improvement in quality compared to university student samples [2], their demographic characteristics are far from representative for the larger population [35]. This is particularly important in the news recommender domain, where not only predictive accuracy should be considered, but also democratic and normative values [17, 39].

2 METHODS

In a choice-based conjoint experiment with a probability-based representative sample of Norwegian Internet users (N = 1664), we tested the effects of news headline features and news recommender system features on users' intention to click on and read a full news article.

2.1 Participants

Our data collection was part of the 24$^{\text{th}}$ wave of the Norwegian Citizen Panel (NCP) in June 2022. The NCP is a highly-respected time-sharing online survey panel which collects high-quality suvey data—representative of the Norwegian adult population—three times a year using probability-based sampling. While costly (i.e., our study cost ≈ 17.000 €), probability sampling is considered "the gold standard" of survey research [41], as the entire adult population of Norway have an equal and known probability of being invited. NCP's time-sharing strategy also ensures that the high cost of collecting such high-quality data is distributed over a large number of studies. The entire panel of the NCP's respondents are gathered through postal recruitment of individuals over 18 years, with regular new recruitment due to well-known issues of panel attrition over time [27]. These individuals were randomly selected for recruitment from Norway's National Registry: a list of all individuals who either are or have been a resident in Norway, maintained by the official Tax Administration. NCP's data are available free of cost for scholars via the Norwegian Social Science Data Archive. For more details about response rates or other methodological matters, please refer to the NCP methodology reports [33].

10.160 of NCP's panel respondents participated in the 24$^{\text{th}}$ wave of the NCP. Demographic information was collected for all panel respondents. A random sub-sample of 1664 participants in the NCP were randomly assigned to, and completed, our experiment. In our sub-sample, 49 % were female, 66 % had a higher education, and the median year of birth was between 1960-1989.

2.2 Procedure, Research Design, Materials

Our research question was addressed using a conjoint experiment. In such experiments, participants are typically asked to choose between Y number of alternatives in terms of a specific dependent variable, such as favorability or appropriateness [23]. Each alternative, referred to as a Profile, contained randomly assigned levels of different features.

Table 1. Overview of features.

Feature Type	Feature	Levels
Article	Topic in Headline	Meat prices
		Extension of Abortion Rights
Article	Reading time	2 minutes
		15 minutes
Article	Recency of Publication	2 minutes ago
		5 days ago
RS	Geographic Distance	Close
		Distant
RS	Topical Preference Match	Yes
		No
RS	Demographic Similarity	Read by people like you
		Not read by people like you
RS	Popularity-based	Among Most Read
		Among Least Read

Note: The Feature type 'Article' refers to article features while 'RS' refers to Recommender System features.

By predicting for each choice set what alternative was chosen in terms of its feature values, the relative importance of each feature can be determined.

The news profile task is depicted in Figure 1. For each choice task, participants were asked to choose between the two profiles, selecting the news story they would prefer to click on and read. Each respondent evaluated two profiles of news item recommendations through four consecutive choice tasks, resulting in a total of 13.311 observations.

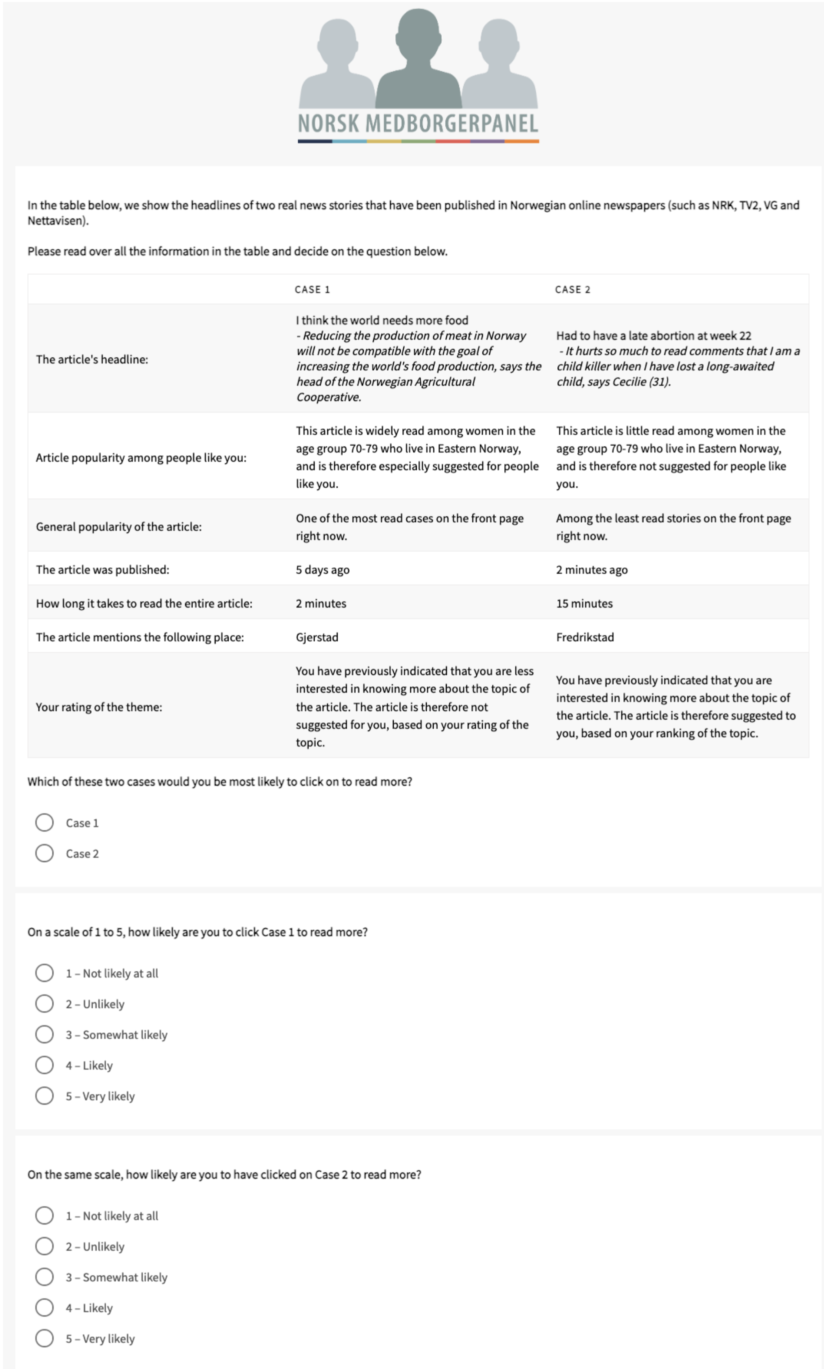
In this study, participants were presented pairs of news article profiles (i.e., descriptions). Each profile was composed of seven news article features with two levels. As can be observed from Figure 1, these features had distinct levels that were randomly assigned to each profile. As such, more than 2^7 combinations were possible, as it is formally a 2x2x2x2x2x2x2 research design.

The seven specific features, what they represented and the involved levels are outlines in Table 1. For instance, each profile featured a randomly selected headline (out of four headlines that were based on actual Norwegian news stories from the newspapers "Vårt Land", "TV2", "NRK", "Dagsavisen", and "ABC nyheter") on the topic of either abortion and meat prices (in response to mitigating climate change)[1]. In addition, each profile featured information whether it was recommended, or not, due to, for instance, the demographic similarity, compared to the respondents, of other users who had read the story. The features were selected based on relevance in earlier work on news recommender systems (cf. [11, 20]); more detail can be found in Section 2.3. As we opt for a statistically efficient conjoint design [9], and because each feature only varied between two levels, each profile pair always displayed different feature levels. To illustrate, if Profile 1 displayed a headline on abortion, Profile 2 always displayed a headline on meat prices, and vice versa.

2.3 Measures

Independent variables. All experimental treatments (i.e., our independent variables) are listed in Table 1. We provide additional detail on three of these attributes. First, the Topical Preference factor was generated through, earlier in the

[1]Note that although this strictly speaking involved more than two levels for the news headline feature, the two relevant levels for the research design were the differences in topic, not the specific 'example' headline. Variation was implemented due to the repeated measures per participant, while designated two levels per feature represented a statistically efficient research design.



NORSK MEDBORGERPANEL

In the table below, we show the headlines of two real news stories that have been published in Norwegian online newspapers (such as NRK, TV2, VG and Nettavisen).

Please read over all the information in the table and decide on the question below.

	CASE 1	CASE 2
The article's headline:	I think the world needs more food *- Reducing the production of meat in Norway will not be compatible with the goal of increasing the world's food production, says the head of the Norwegian Agricultural Cooperative.*	Had to have a late abortion at week 22 *- It hurts so much to read comments that I am a child killer when I have lost a long-awaited child, says Cecilie (31).*
Article popularity among people like you:	This article is widely read among women in the age group 70-79 who live in Eastern Norway, and is therefore especially suggested for people like you.	This article is little read among women in the age group 70-79 who live in Eastern Norway, and is therefore not suggested for people like you.
General popularity of the article:	One of the most read cases on the front page right now.	Among the least read stories on the front page right now.
The article was published:	5 days ago	2 minutes ago
How long it takes to read the entire article:	2 minutes	15 minutes
The article mentions the following place:	Gjerstad	Fredrikstad
Your rating of the theme:	You have previously indicated that you are less interested in knowing more about the topic of the article. The article is therefore not suggested for you, based on your rating of the topic.	You have previously indicated that you are interested in knowing more about the topic of the article. The article is therefore suggested to you, based on your ranking of the topic.

Which of these two cases would you be most likely to click on to read more?

- Case 1
- Case 2

On a scale of 1 to 5, how likely are you to click Case 1 to read more?

- 1 – Not likely at all
- 2 – Unlikely
- 3 – Somewhat likely
- 4 – Likely
- 5 – Very likely

On the same scale, how likely are you to have clicked on Case 2 to read more?

- 1 – Not likely at all
- 2 – Unlikely
- 3 – Somewhat likely
- 4 – Likely
- 5 – Very likely

Fig. 1. Screenshot of the experiment from the respondents' point of view.

survey, asking each respondent to rank a list of five topics through the question "Below are a number of topics that Norwegian newspapers write about. Please rank the topics according to which one you would like to know more about. You should rank the topic that you would most like to read about at the top and the one that you are least interested in reading at the bottom." Based on each respondent's ranking of these five topics, a script matched each respondent's ranking of topics with the topics presented in the experiment. If the respondent had ranked one topic higher than another, the experiment would present respondents with the following wording: "You have previously indicated that you are interested in knowing more about the subject of the article. The article has therefore been suggested to you, based on your ranking of the topic." Vice versa, respondent were presented with the following wording: "You have previously indicated that you are less interested in knowing more about the subject of the article. The article has therefore not been suggested to you, based on your ranking of the topic."

Second, the Geographical Distance factor mentioned a geographical place that was either close or distant to the respondent, based on panel information of each respondent regarding where they live. This information was exclusively displayed to the respondents and only recorded as either "close" or "distant" to guarantee the utmost privacy of personal information for each individual participant.

Third, the Demographic Similarity feature was created based on panel information of each participant. A recommended article featured the following wording "This article is widely read among [respondent's gender] in the [respondent's age group] age group living in [respondent's Region], and is therefore suggested for people like you." Vice versa, an article that was not recommended featured the following wording: "This article is not widely read among [respondent's gender] in the [respondent's age group] age group living in [respondent's Region], and is therefore suggested for people like you." The information on gender, age and region was exclusively displayed to the respondents and only recorded as either "recommended" or "not recommended" to guarantee the utmost privacy of personal information for each individual participant.

Dependent variable. Our outcome of interest in this paper is participants' binary choice between two news stories. The participants' task was to select which news item they would click on and read based on the presented choices through the question "Which of these two stories are you most likely to click on to read more?", with the binary choice between "News story 1" and "News story 2" as the dependent variable. The choice made in the study was considered a stated intention to read, relative to the other news article presented.

2.4 Analysis

The conjoint design enabled us to analyze the influence of individual news article features. To compare treatment effects, we estimated the average marginal component effects (AMCE) of each treatment [16]. This represented the marginal effect of one feature averaged across the joint distribution of the other factors. While such designs produce a high number of possible treatment combinations, Hainmueller and colleagues [16] demonstrate that not all specific combinations are required to estimate the AMCEs of each component.

We estimated the AMCEs through regressing each feature on news story selection. To this end, a logistic regression with within-respondent clustering was used to ensure robust standard errors and get unbiased estimates of the variance [16].

3 RESULTS

The results of our AMCE-based logistic regression analysis is displayed in Figure 2. The dots indicated point estimates, bars illustrated 95%-confidence intervals, and dots without bars represented the reference categories. The AMCEs could be interpreted as percentage points [23].

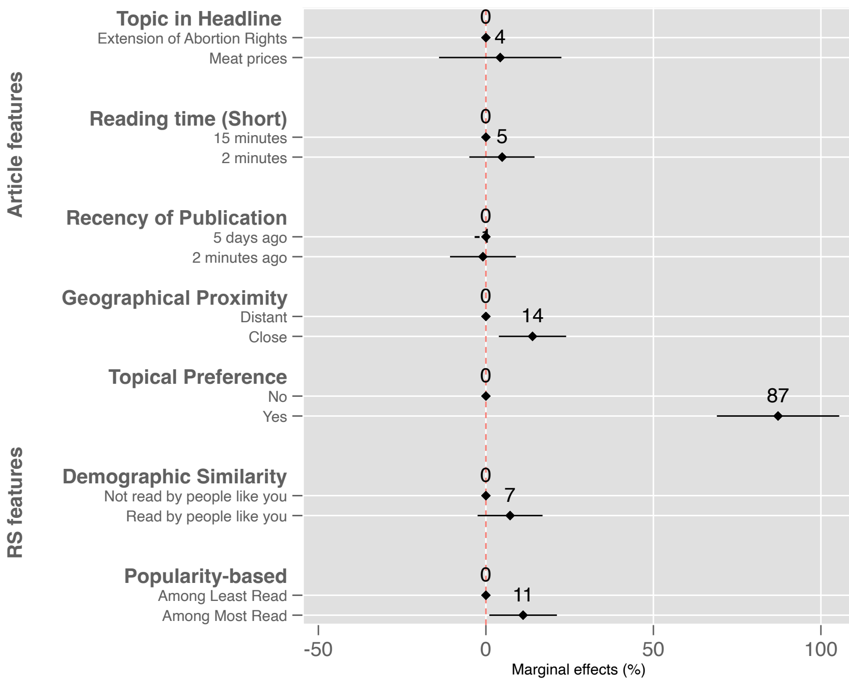


Fig. 2. Effects (AMCE) of news recommender features and article features on probability to click on and read a news story.

The coefficients and significance levels are also presented in Table 2. We observed no statistically significant treatment effects between any of the levels of the *news article features* on preferring any news profile. This involved the specific headline, reading time, and recency of publication. This suggested that although these features were used in earlier studies, they were less important in determining user preferences for news articles than features that emphasize recommender aspects.

Hence, in contrast, the *news recommender system features*, showed substantial effects. Working from top to bottom in Figure 2 and Table 2, we observed that respondents are 14 percentage points more likely to click on a news story if a news recommender highlights a location that is geographically close to the user, compared to an alternative that is more distant. This suggested that news article preferences are in part determined by local relevance.

The most substantial effect was found through topical preference recommendation. Respondents were 87 percentage points more likely to chose a news story if it was recommended them due a highlighted match in their topical interests, compared to a news story with a topic that was lower ranked ($p < 0.001$). This effect was much stronger than those induced by any of the other features, which suggested that transparent topical preference matching was the most important feature.

The two remaining features showed mixed results. Highlighting demographic similarity by indicating 'people like you' read this, did not lead to significantly more preference choices than 'not read by people like you' (a non-significant difference of 7 percentage points). In contrast, highlighting the popularity of a news article, compared to unpopularity, led to a small, significant increase in user preferences by 11 percentage points ($p < 0.05$).

Table 2. Marginal effects (AMCE) of news recommender and news article features on a user's intention to read, i.e., the probability to click on and read a news story. Results stem from a logistic regression model, clustered at the user level.

Feature	β **(S.E.)**
Topic in Headline (Meat prices, vs Abortion Rights)	0.04 (0.09)
Short Reading Time (2 minutes, vs 15 minutes)	0.05 (0.05)
Recency of Publication (2 minutes vs 5 days ago)	-0.01 (0.05)
Geographical Proximity (Close, vs distant)	0.14^{**} (0.05)
Topical Preference (Yes vs no)	0.87^{***} (0.09)
Demographic Similarity (Similar vs Dissimilar)	0.07 (0.05)
Popularity-Based (Most vs Least Popular)	0.11^{*} (0.05)
Observations	13380
Pseudo R2 (Cragg & Uhler)	0.067
AIC	17876.75
BIC	17929.27

Note: $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

3.1 Conclusion

Overall, one feature had the largest impact. Comparing the treatment effects across all features in our conjoint experiment, we observed that Topical Preference was the most important one (i.e., 87 percentage points) for predicting news story selection, followed by Geographical Distance and Popularity-based recommendations.

4 DISCUSSION

We have presented results of a user study with a novel approach in the context of news recommender systems, employing a novel experimental design adopted from the social sciences. Through a conjoint analysis, we have compared the impact of different news article features and recommender system features on a user's intention to read a news story. By doing so, we have highlighted the relative importance of different news article features and recommender system features that have been used in the past in news recommender studies [20], either in offline learning tasks (cf. [12, 36]) or online evaluation studies (cf. [34, 40]).

Our study has involved a scenario in which both news article features and recommendation methods are presented transparently to the user. In this context, we find that Topical Preference is particularly important for a user's news story selection. This indicates that news recommender systems that focus on surveying users' topic preferences, and recommend stories based on the answers from such surveys, will likely have a higher chance or success-rate in terms of predicting clicks or reads. While prior work has argued that a topic match between the user and news content should be more effective than a mismatch [11], our results contribute to this literature by showing that it is far more effective than other forms of similarity, revealing large differences with other features. Topical preference seems to trump similarity based on demographic similarity, general popularity, and recency, all of which are also often used in news recommendation [20].

Regarding the comparative evaluation of features, our findings are complementary to other studies that examined multiple features simultaneously. For example, whereas Starke et al. [34] show that title-based and body-text based similarity resonate with a user's similarity judgment, we show which underlying mechanisms of recommendation are important in building user preferences. Whereas other work in semantic similarity is focused on validating the 'correctness' of algorithmic functions [38], we have examined which features are preferred by users.

These findings are important because they provide direction for designers of news recommender systems. They illustrate that some features are indeed more important than others when it comes to predicting what a user may read. This might narrow what factors and features should be considered in offline evaluation approaches, which are still the sole method in many papers [20], to generate an algorithmic approach that actually resonates with users in online evaluation [34, 38].

Another strong point of this study is its research population and sample. Very few studies in recommender research have tapped into representative samples for their evaluation. One reason may be that such evaluations are costly for session-based studies, where the behavioral implications remain unclear [19], and therefore the sample is of less importance. Another reason might be a systemic bias when it comes to how important such demographic characteristics are. It could be, for example, that some studies with news recommenders have found favorable results for a specific algorithmic approach because of a relatively young sample. By using a representative sample from Norway, which is a typical 'Western-type country', the chance of such a confounding bias is strongly reduced.

The extent to which our effect sizes hold up in more traditional Recommender studies is less clear. For this study, some features have been subject to two 'ends of the scale'. For example, popularity was compared by showing either 'least popular' or 'most popular', which is unlikely to happen in a recommender field experiment with news. There, popularity is one of many of possible approaches [11, 20], which would be juxtaposed to other 'best guesses', not a different, 'intentionally poor' approach. Thus, we caution that our results should not be used to make absolute statements on the probability that a recommender system feature influence news use. At the same time, the goal of this study has been to assess the relative importance of different features. Taking the 'whole spectrum' of a feature, such as by using least and most popular, does provide a clear assessment of a feature's importance. While this importance *could* be reduced with a different baseline, all features except the headline had such extreme values, deeming it a fair comparison.

4.1 Limitations & Future Work

The main limitation of our design is the use of only two news topics, which are both controversial. Abortion and meat eating have led to polarized discussions in society [18, 29], which might have amplified the importance of a match in the topic at hand. Hence, such a strong effect may not be found for more nuanced topics, as our desing is limited to two (controversial) topics and might not generalize to other news topics. In addition, the design does not mimic real news use decision and thus it is hard to determine the extent to which our findings would be reproduced outside the experimentally controlled environment in which we conduct our study.

Another limitation may explain the lack of a recency effect, which has been observed in previous studies [20, 24]. Our headlines have not included breaking-news elements, being more "timeless" than many breaking-news stories. This can, perhaps, have implications for recency, for its importance may be amplified if a news story has just broken.

Moreover, the fact that this study has been rather hypothetical in the sense that only the headline, and not the entire news article, were read, might have exacerbated the lack of a recency effect.

Considering all the limitations mentioned, we encourage future studies to employ a more naturalistic reproduction of this study. We argue that the conjoint method could still be used, since it provides a fair feature-based comparison, but that the user interaction with the recommender system should be part of a news website context. This way, the user's profile may be more durable, and not formed this a single preference elicitation session. Moreover, more recent news articles could be used, making the dependent variable also more relevant, in the sense that the presented news articles would be novel to a user. This way, a user's intention could also be related to actual reading behavior, which has been investigated in some news recommender stories [11, 20]. Finally, assessing the effectiveness of a conjoint method for preference elicitation would be a valuable line of research. While only a few non-news domains have used this approach in recommender research [25], this would be new to the news recommender system domain, and a promising line of research.

ACKNOWLEDGMENTS

This work was supported by the NewsRec Project (project number: 324835) and by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the centers for Research-based Innovation scheme, project number 309339.

REFERENCES

[1] Jae-wook Ahn, Peter Brusilovsky, Jonathan Grady, Daqing He, and Sue Yeon Syn. 2007. Open user profiles for adaptive news systems: help or harm?. In *Proceedings of the 16th international conference on World Wide Web*. 11–20.

[2] Tara S Behrend, David J Sharek, Adam W Meade, and Eric N Wiebe. 2011. The viability of crowdsourcing for survey research. *Behavior research methods* 43 (2011), 800–813.

[3] Dan Berkowitz and Douglas W Beach. 1993. News sources and news context: The effect of routine news, conflict and proximity. *Journalism Quarterly* 70, 1 (1993), 4–12.

[4] Roi Blanco, Diego Ceccarelli, Claudio Lucchese, Raffaele Perego, and Fabrizio Silvestri. 2012. You should read this! let me explain you why: explaining news recommendations to users. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. 1995–1999.

[5] Toine Bogers and Antal Van Den Bosch. 2007. Comparing and evaluating information retrieval algorithms for news recommendation. In *RecSys'07: Proceedings of the 2007 ACM Conference on Recommender Systems*. https://doi.org/10.1145/1297231.1297256

[6] Pedro G Campos, Ignacio Fernández-Tobías, Iván Cantador, and Fernando Díez. 2013. Context-aware movie recommendations: an empirical comparison of pre-filtering, post-filtering and contextual modeling approaches. In *E-Commerce and Web Technologies: 14th International Conference, EC-Web 2013, Prague, Czech Republic, August 27-28, 2013. Proceedings 14*. Springer, 137–149.

[7] Wei Chu, Seung-Taek Park, Todd Beaupre, Nitin Motgi, Amit Phadke, Seinjuti Chakraborty, and Joe Zachariah. 2009. A case study of behavior-driven conjoint analysis on Yahoo! Front Page Today module. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1097–1104.

[8] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*. 271–280.

[9] Brandon De la Cuesta, Naoki Egami, and Kosuke Imai. 2022. Improving the external validity of conjoint analysis: The essential role of profile distribution. *Political Analysis* 30, 1 (2022), 19–45.

[10] Nicholas Diakopoulos and Michael Koliska. 2017. Algorithmic transparency in the news media. *Digital journalism* 5, 7 (2017), 809–828.

[11] Chong Feng, Muzammil Khan, Arif Ur Rahman, and Arshad Ahmad. 2020. News recommendation systems-accomplishments, challenges & future directions. *IEEE Access* 8 (2020), 16702–16725.

[12] Len Feremans, Robin Verachtert, and Bart Goethals. 2022. A neighbourhood-based location-and time-aware recommender system. (2022).

[13] Blaž Fortuna, Carolina Fortuna, and Dunja Mladenić. 2010. Real-time news recommender system. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 583–586.

[14] Florent Garcin, Boi Faltings, Olivier Donatsch, Ayar Alazzawi, Christophe Bruttin, and Amr Huber. 2014. Offline and online evaluation of news recommender systems at swissinfo. ch. In *Proceedings of the 8th ACM Conference on Recommender systems*. 169–176.

[15] Florent Garcin, Kai Zhou, Boi Faltings, and Vincent Schickel. 2012. Personalized news recommendation based on collaborative filtering. In *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Vol. 1. IEEE, 437–441.

[16] Jens Hainmueller, Daniel J Hopkins, and Teppei Yamamoto. 2014. Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political analysis* 22, 1 (2014), 1–30.

[17] Natali Helberger, Max van Drunen, Judith Moeller, Sanne Vrijenhoek, and Sarah Eskens. 2022. Towards a Normative Perspective on Journalistic AI: Embracing the Messy Reality of Normative Ideals. , 1605–1626 pages.

[18] Mark Romeo Hoffarth and Gordon Hodson. 2016. Green on the outside, red on the inside: Perceived environmentalist threat as a factor explaining political polarization of climate change. *Journal of Environmental Psychology* 45 (2016), 40–49.

[19] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. 2010. *Recommender systems: an introduction*. Cambridge University Press, Cambridge, UK.

[20] Mozhgan Karimi, Dietmar Jannach, and Michael Jugovac. 2018. News recommender systems–Survey and roads ahead. *Information Processing & Management* 54, 6 (2018), 1203–1227.

[21] Benjamin Kille, Frank Hopfgartner, Torben Brodt, and Tobias Heintz. 2013. The plista dataset. In *Proceedings of the 2013 international news recommender systems workshop and challenge*. 16–23.

[22] Bart P Knijnenburg and Martijn C Willemsen. 2015. Evaluating recommender systems with user experiments. *Recommender systems handbook* (2015), 309–352.

[23] Erik Knudsen and Mikael Poul Johannesson. 2019. Beyond the limits of survey experiments: How conjoint designs advance causal inference in political communication research. *Political Communication* 36, 2 (2019), 259–271.

[24] Lei Li, Ding-Ding Wang, Shun-Zhi Zhu, and Tao Li. 2011. Personalized news recommendation: a review and an experimental investigation. *Journal of computer science and technology* 26, 5 (2011), 754–766.

[25] Benedikt Loepp, Tim Hussein, and Jüergen Ziegler. 2014. Choice-based preference elicitation for collaborative filtering recommender systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3085–3094.

[26] Tapio Luostarinen and Oskar Kohonen. 2013. Using topic models in content-based news recommender systems. In *Proceedings of the 19th Nordic conference of computational linguistics (NODALIDA 2013)*. 239–251.

[27] Peter Lynn. 2018. Tackling panel attrition. *The Palgrave handbook of survey research* (2018), 143–153.

[28] Eliza Mitova, Sina Blassnig, Edina Strikovic, Aleksandra Urman, Aniko Hannak, Claes H de Vreese, and Frank Esser. 2023. News recommender systems: A programmatic research review. *Annals of the International Communication Association* 47, 1 (2023), 84–113.

[29] Ted Mouw and Michael E Sobel. 2001. Culture wars and opinion polarization: the case of abortion. *Amer. J. Sociology* 106, 4 (2001), 913–943.

[30] Raymond K Pon, Alfonso F Cardenas, David Buttler, and Terence Critchlow. 2007. Tracking multiple topics for finding interesting articles. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. 560–569.

[31] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. 157–164.

[32] Pamela J Shoemaker, Jong H Lee, Gang Han, and Akiba A Cohen. 2007. Proximity and scope as news values. *Media studies: Key issues and debates* (2007), 231–248.

[33] Øivind Skjervheim and Asle Høgestøl. 2022. *Norwegian Citzen Panel Methodology Report wave 24*. Technical Report. Ideas 2 Evidence, Bergen.

[34] Alain D Starke, Sebastian Øverhaug, and Christoph Trattner. 2021. Predicting Feature-based Similarity in the News Domain Using Human Judgments. In *15th ACM Conference on Recommender Systems, RecSys 2021*.

[35] Neil Stewart, Jesse Chandler, and Gabriele Paolacci. 2017. Crowdsourcing samples in cognitive science. *Trends in cognitive sciences* 21, 10 (2017), 736–748.

[36] Mozhgan Tavakolifard, Jon Atle Gulla, Kevin C Almeroth, Jon Espen Ingvaldesn, Gaute Nygreen, and Erik Berg. 2013. Tailored news in the palm of your hand: a multi-perspective transparent approach to news recommendation. In *Proceedings of the 22nd international conference on world wide web*. 305–308.

[37] Maartje Ter Hoeve, Mathieu Heruer, Daan Odijk, Anne Schuth, and Maarten de Rijke. 2017. Do news consumers want explanations for personalized news rankings. In *FATREC Workshop on Responsible Recommendation Proceedings*.

[38] Nava Tintarev and Judith Masthoff. 2006. Similarity for news recommender systems. In *In Proceedings of the AH'06 Workshop on Recommender Systems and Intelligent User Interfaces*. Citeseer.

[39] Sanne Vrijenhoek, Gabriel Bénédict, Mateo Gutierrez Granada, Daan Odijk, and Maarten De Rijke. 2022. RADio–Rank-Aware Divergence Metrics to Measure Normative Diversity in News Recommendations. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 208–219.

[40] Kam Fung Yeung and Yanyan Yang. 2010. A proactive personalized mobile news recommendation system. In *Proceedings - 3rd International Conference on Developments in eSystems Engineering, DeSE 2010*.

[41] Elizabeth S Zack, John Kennedy, and J Scott Long. 2019. Can nonprobability samples be used for social science research? A cautionary tale. In *Survey Research Methods*, Vol. 13. 215–227.