The Framing Loop: Do Users Repeatedly Read Similar Framed News Online?

Markus Reiter-Haas[1,*], Elisabeth Lex[1]

[1]Graz University of Technology, Institute of Interactive Systems and Data Science, 8010 Graz, Sandgasse 36/III

Abstract

It is well established in psychology that framing of content affects the behavior of people. This effect is, however, only sparsely explored in information-seeking and retrieval behavior. In the present work, we consider the diversity of consumed content and repetition patterns regarding their framing. We conduct a framing analysis in the MIcrosoft News Dataset (MIND) comprising textual content and user interaction behaviors. By extracting the frames of the item sequences, we uncover a tendency of users to consume similar framed news repeatedly when sticking to the same type of content. Consequently, framing biases are important to consider in information systems. We hope that our work inspires future research on corresponding debiasing methods.

Keywords

Framing Theory, User Behavior, Empirical Study, Content Bias, Repeat Consumption, Viewpoint Diversity

1. Introduction

The effects of framing on peoples' choices have been well established in psychology and can be traced back to the notable work of Tversky and Kahneman [1]. While the grounding of framing effect as a cognitive bias is solid, research on its effects on information seeking and retrieval behavior has only recently emerged [2]. Besides this sparsely explored area resides a vast body of research on both framing theory (see [3] for an overview) and biases in online information systems ([4] provides an overview of biases in Web data) to draw from. Regarding framing theory, a wide variety of computational methods are available to extract the framing of content [5]. Whereas, for analyzing biased behavior patterns, several approaches have been studied for information systems, such as to understand repeat consumption [6] and assessing viewpoint diversity [7] regarding web searches. Hence, the conflation of the two research strands to expand the framing research in information systems seems reasonable.

In the present work, we investigate the content consumption regarding the framing in the MIcrosoft News Dataset (MIND) [8], which is well researched and sparked an influx for news recommendation research [9]. As depicted in Figure 1, each user has a sequential history of consumed items, as well as impressions and interactions for a specific timestamp. Additionally, each news item is assigned a specific category, which can be used to represent the sequence

Joint Proceedings of the ACM IUI Workshops 2024, March 18-21, 2024, Greenville, South Carolina, USA

*Corresponding author.

reiter-haas@tugraz.at (M. Reiter-Haas); elisabeth.lex@tugraz.at (E. Lex)

https://iseratho.github.io/ (M. Reiter-Haas); https://elisabethlex.info/ (E. Lex)

0000-0001-9852-8206 (M. Reiter-Haas); 0000-0001-5293-2967 (E. Lex)

regarding consumed categories. Similarly, we can extract the framing based on the content and assign it to the items, which results in sequences of consumed frames. We consider such sequences to uncover biased behavioral patterns regarding the framing. For frame extraction, we use the FrameFinder library [10], which extracts three types of frames, i.e., media frames, moral frames, and semantic frames.

We find that frame consumption depends on the consumed categories, the types of frames and the information system itself. In particular, users repeatedly consume the same frames when sticking to the same category, which could be counteracted by the information system. Overall, the consumption behavior is more balanced concerning moral frames compared to semantic frames, whereas media frames depend on the categories the most.

In sum, our main contributions are:

1. We connect two separate strands of research in computer science (i.e., computational framing analysis and biases in information systems) that are both *rooted in psychology*.
2. We introduce an approach to analyze biased behavior patterns based on *sequences of consumed frames*.
3. We provide *empirical evidence of behavioral biases* due to framing on a well-established recommendation dataset.

To the best of our knowledge, we are the first to directly investigate this *link between the framing of content and the consumption behavior* of users. For reproducibility reasons, we additionally open source the code (also containing the supplementary materials referenced in the paper), as well as the framing dataset used for our study[1].

2. Related Work

Framing Theory: Framing has long been considered as a fractured paradigm in literature [12]. According to [3], there are three types of framing relating to language, cognition, and communication, respectively. While our study touches all three types, its focus lies on *communicative frames* present in media. Herein, framing as a form of bias in media has identified [13] and been thoroughly studied. For example, Morstatter et al. [14] train a classifier to detect the framing bias in news articles and relate it to opinion bias. This already indicates the relation to *cognitive frames*, which is an explicit requirement of communicative frames [3]. Finally, *semantic frames* were established by Fillmore and Baker [15] and depend on the language structure, but also on cognitive frames.

Recently, a vast amount of research uses computational methods for framing detection on wide range of frames, e.g., war [16], terrorists [17], morality [18], or blame [19] frames. The range of **computational framing analysis** approaches mainly span topic modeling and neural networks models (see Ali and Hassan [5] for a comprehensive survey). Neural networks are especially suitable in supervised settings, such as at the SemEval Challenge of 2023 [20], where every best-performing team used Transformer models [21, 22, 23]. Besides, open-source libraries like OpenFraming [24] and FrameFinder [10] support the extraction of frames. Our approach

[1]Code: https://github.com/Iseratho/frameloop Dataset: https://zenodo.org/records/10509498 [11]

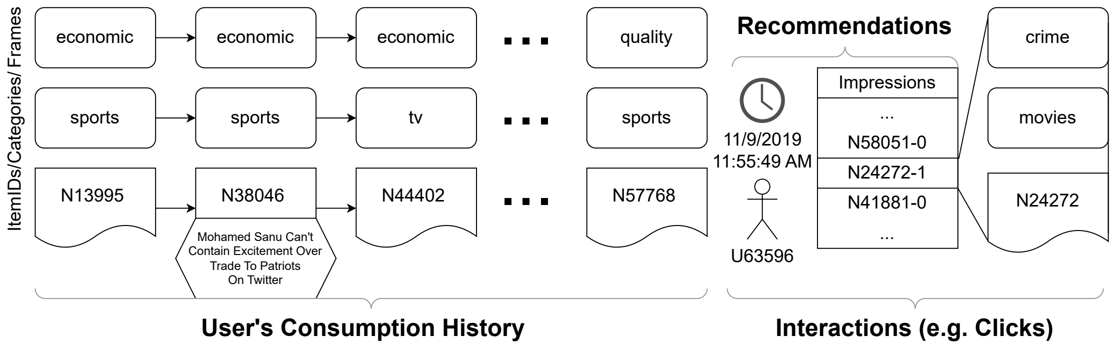


Figure 1: Real example from the dataset that shows how users interact with the system. Each user has a history of consumed items. Each item in the system contains textual content (depicted for the second item in the sequence), from which the frames can be extracted. At the top, the calculated media frame labels are represented. Here, the user shows a low viewpoint diversity regarding both categories and labels but even higher repeat consumption behavior regarding framing compared to categories. On the right, the user impression log is represented together with the clicked item (typically one). The impressions or clicked items can be seen as a continuation of the user history.

uses the latter to extract the framings present in news articles, as it also employs Transformer models [25] to extract frame representation in an unsupervised manner.

Biased Behavior Patterns: Cognitive psychology plays a vital role in information systems, which also provides the inspiration for various recommendation approaches [26]. As an example, a cognitive model of human memory (ACT-R) can predict music genre preferences [27]. Moreover, it has been shown that the cognitive-inspired ACT-R model also effectively predicts music relistening behavior [28], while also increasing the diversity of genres [29]. The relistening behavior is a type of **repeat consumption**, defined as "the act of consuming an enjoyable stimulus that one has already consumed in full in the past" in psychology [30]. Such biased repetition patterns have been found in a variety of domains and platforms, such as on Wikipedia, Google Maps, and YouTube [6]. Regarding diversity, assessing the viewpoints presented to users is another important bias in information systems to consider [7]. Herein, algorithmic diversification plays a key role in opinion forming domains, e.g., the news domain [31]. Moreover, the presence of distinct frames as a proxy for **viewpoint diversity** in news discourse is vital for high-quality debates [32].

In the present work, we investigate biased behavior patterns in news consumption sequences due to framing concerning both repeat consumption and viewpoint diversity with frame labels.

3. Problem Formulation and Notation

In an information system, a set of users U interacts with a set of items I. Each user $u \in U$ has a consumption history H_u, which consists of a sequence of n_h consumed items $i \in I$ by the user. For simplicity, we consider all user histories from the same specified time, thus omitting an additional time index ($i^H_{u,t,1} = i^H_{u,1}$). To access the information, a user might be presented with a list of n_r potential items $i \in I$ given by the function $\mathcal{R}(u,t)$. The function takes as input the

Notation	Description	# in *MIND-small*	$\overline{AVG}$
U	set of users, represented by their user IDs: $u \in U$	$\lvert U\rvert = 50,000$	$\lvert R\rvert/\lvert U\rvert = 3.14$
I	set of items, represented by their item IDs: $i \in I$	$\lvert I\rvert = 51,282$	$\overline{w} = 10.75$
H	set of user histories, $H = \bigcup_{u \in U} H_u$	$\lvert H\rvert = 49,108$	$\overline{H} = 18.52$
R	set of impression logs from the function $\mathcal{R}(u,t)$	$\lvert R\rvert = 156,965$	$\overline{R} = 37.23$
C	set of click logs from the function $\mathcal{C}(u,t)$	$\lvert C\rvert = 236,344$	$\overline{C} = 1.51$
L	set of label spaces; $l \in \bigcup_{L_j \in L_1, L_2, \ldots} L_j$; # category labels: $\lvert L_{Cat}\rvert = 17$		
$f_j(\cdot)$	mapping function for label space L_j from a set of mapping functions $f_j(\cdot) \in F$		
n_h, n_r, n_c	lengths of specific item set (i.e., logs of history, impression, and click, respectively)		
i^H_x, i^R_x, i^C_x	item lookup from logs (i.e., H, R, and C) with x providing required indices		

Table 1

Description of symbols used throughout the paper and their according statistics in *MIND-small*. $\overline{w}$ denotes the average number of words in the title.

user u and a specific time t. After evaluating the potential items in $\mathcal{R}(u,t)$, a user then interacts (i.e., consumes) one (or more) items of the list of potential items $i_c \in \mathcal{R}(u,t)$. This interaction can be formalized by the function $\mathcal{C}(u,t)$. The number of interacted items is denoted by n_c, which is $n_c = 1$ in most cases (i.e., where we can omit the positional index: $\mathcal{C}(u,t) = \{i^C_{u,t}\}$). The three described equations are thus given by (a summary of the main symbols is in Table 1):

$$\begin{aligned} H_u &= [i^H_{u,1}, i^H_{u,2}, \ldots, i^H_{u,n_h}] \\ \mathcal{R}(u,t) &= [i^R_{u,t,1}, i^R_{u,t,2}, \ldots, i^R_{u,t,n_r}] \\ \mathcal{C}(u,t) &= \{i^C_{u,t,1}, i^C_{u,t,2}, \ldots, i^C_{u,t,n_c}\} \end{aligned}$$

Each item i contains some content and can additionally be assigned some metadata, such as labels. For instance, we can assign a category label l to each item i based on its content $f_j(i) = l$, where $f_j(\cdot)$ is the mapping function from the content to the label space from a list of potential categories $l \in L_j$. Note that the system can have multiple label spaces $L = L_1, L_2, \ldots$, each with their corresponding mapping function. Consequently, we can transform the previous equations to the label space L_j for analysis, as shown in Figure 1:

$$\begin{aligned} H_u^{L_j} &= [f_j(i^H_{u,1}), f_j(i^H_{u,2}), \ldots, f_j(i^H_{u,n_h})] &&= [l_{i^H_{u,1},j}, l_{i^H_{u,2},j}, \ldots, l_{i^H_{u,n_h},j}] \\ \mathcal{R}^{L_j}(u,t) &= [f_j(i^R_{u,t,1}), f_j(i^R_{u,t,2}), \ldots, f_j(i^R_{u,t,n_r})] &&= [l_{i^R_{u,t,1},j}, l_{i^R_{u,t,2},j}, \ldots, l_{i^R_{u,t,n_h},j}] \\ \mathcal{C}^{L_j}(u,t) &= \{f_j(i^C_{u,t,1}), f_j(i^C_{u,t,2}), \ldots, f_j(i^C_{u,t,n_c})\} &&= \{l_{i^C_{u,t,1},j}, l_{i^C_{u,t,2},j}, \ldots, l_{i^C_{u,t,n_h},j}\} \end{aligned}$$

4. Data and Methods

We employ a two-step approach to identify biased behavior patterns regarding framing in the MIND dataset [8]. Specifically, we first construct sequences of labels (see Figure 1), which we then use to calculate four metrics on the sequence of categorical data for the behavior analysis. To ensure a fair comparison, we implement several simplifications on the data representation and evaluation setting (described below).

4.1. MIND Dataset

The MIND dataset [8] is a *large-scale dataset for news recommendation research* released in 2020, which follows the structure outlined in Figure 1. We use the smaller version *MIND-small*, which is a subset consisting of $50,000$ randomly sampled users and their associated data. The most important statistics of the dataset are provided in Table 1. The dataset has a high sparsity of $\frac{|H|\times\overline{H}+|R|\times\overline{R}}{|I|\times|U|} = 2.63 \times 10^{-3}$. In the dataset, each item (i.e., news article) consists of a single category that was manually assigned. Note that while a timestamp is available for the impression log, neither the individual interactions nor the sequential items in the history have been assigned any temporal data besides the order.

4.2. Constructing Label Sequences

We use a two-step procedure to construct the label sequences. First, we use metadata assigned to the items to construct sequences of categories. Second, we extract framing representations from the textual data (specifically the titles, as the short text is partially incomplete). Here, we employ the FrameFinder library [10], which allows the extraction of three distinct types, i.e., (i) media frames, (ii) moral frames, and (iii) semantic frames. Each representation uses a Transformer [25] model from Hugging Face [33] as a basis, where we use the default setting for all three types (details below). As these representations are not directly comparable, we simplify them by only considering the most pronounced feature per item and using that as a label.

Categories: For each item in the sequence (e.g., user history), we look up the category as there is always exactly one and assign it. Thus, the sequence is transformed into a sequence of labels. In *MIND-small*, there are 17 distinct labels, which are: 'lifestyle', 'health', 'news', 'sports', 'weather', 'entertainment', 'autos', 'travel', 'foodanddrink', 'tv', 'finance', 'movies', 'video', 'music', 'kids', 'middleeast', and 'northamerica'.

Media Frames: For the media frames: we use the *facebook/bart-large-mnli* model for zero-shot learning [34, 35, 36] with label definitions from the media frame corpus [37]. This model transforms the textual data to label probability scores, where we take the label with the maximum score. It is thus similar to the categories, but the labels are computed automatically rather than assigned manually. The set of 15 labels comprises: 'morality', 'economic', 'quality', 'capacity', 'crime', 'security','health', 'political', 'public', 'other', 'cultural', 'fairness', 'policy', 'legality,', and 'external'.

Moral Frames: We use the *sentence-transformers/all-mpnet-base-v2* encoder model [38, 39] to extract the moral frames with the definitions derived from the moral foundation theory [40, 41]. Here, the textual data is transformed into alignment scores, which can be positive or negative, as each dimension is formed by an antagonistic label pair. Therefore, we take the maximum absolute value with a corresponding label (i.e., positive or negative, depending on the original sign). This forms a set of 10 labels: 'authority', 'cheating', 'subversion', 'degredation', 'harm', 'fairness', 'care', 'betrayal', 'loyalty', and 'sanctity'.

Semantic Frames: The model *Iseratho/model_parse_xfm_bart_base-v0_1_0*, which is a copy on Hugging Face of an AMRLib[2] model. The model is based on BART using abstract meaning

[2] https://amrlib.readthedocs.io/en/latest/

Name	Example sequence	DRR	$RRdist$	$Uniq$	$Gini$
Specific	$[a, b, a, b, b, c]$	0.2	0.5	0.4 if $\|L_j\| \geq 6$	0.61
All same	$[a, a, \ldots, a]$	1.0	1.0	0.0	0.0
Alternating	$[a, b, a, b, \ldots, a, b]$	0.0	0.5	$1/(\|L_j\| - 1)$	0.5
All different	$[a, b, c, \ldots, j]$	0.0	0.0	1.0	$lim_{n\to\infty}$ = 1
Encased	$[a, b, b, \ldots, b, a]$	$lim_{n\to\infty}$ = 1	0.5	$1/(\|L_j\| - 1)$	$lim_{n\to\infty}$ = 0
Random	$[rng(L_j), rng(L_j), \ldots, rng(L_j)]$	$lim_{n\to\infty}\mathbb{E}[S] = 1/\|L_j\|$	$lim_{n\to\infty}\mathbb{E}[S] = 1/\|L_j\|$	$lim_{n\to\infty}\mathbb{E}[S] = 1$	$lim_{n\to\infty}\mathbb{E}[S] = 1-(1/\|L_j\|)$

Table 2
Metrics on example sequences DRR and $RRdist$ tend to behave mostly opposite to $Uniq$ and $Gini$.

representations [42, 34] that transforms texts to semantic graphs comprising semantic frames[3]. From the semantic graphs, we extract the most pronounced frames. Due to the large size of the label space, we only consider frames that appear at least 200 times at the root (i.e., the most pronounced position). The resulting set contains 23 frames: 'say-01', 'possible-01', 'report-01', 'cause-01', 'die-01', 'find-01', 'have-degree-91', 'watch-01', 'contrast-01', 'get-01', 'arrest-01', 'be-located-at-91', 'charge-05', 'open-01', 'show-01', 'kill-01', 'have-03', 'reveal-01', 'recommend-01', 'announce-01', 'want-01', 'close-01', and 'win-01'. We then use the first frame of the set in the serialized form of the graph. If none of the frames are present, we insert a special 'other' frame (similar to how the media frames have an 'other' label) instead.

4.3. Behavior Sequence Analysis

For the behavior analysis, we use two metrics each (one coarse- and one fine-grained) as a proxy to measure repeat consumption behavior and viewpoint diversity, respectively. All metrics are normalized to fall in the range of $[0, 1]$. For repeat consumption behavior, a high value means that the same items are repeatedly consumed and thus indicate a less balanced consumption pattern. For viewpoint diversity, a high value means more diversity in consumed items and thus indicates a balanced consumption diet. For repeat consumption metrics, the sequence order is relevant while the label distribution is secondary, whereas for viewpoint diversity metrics, the sequential orderings are irrelevant.

The metrics are defined to work on arbitrary sequences S containing categorical data. In the most basic case, we evaluate the sequence of a user's history of a particular label space, i.e., $S = H_u^{L_j}$. For simplicity, we omit the details of the indices besides the positional index (i.e., $[l_{i^H_{u,1},j}, l_{i^H_{u,2},j}, \ldots, l_{i^H_{u,n_h},j}]$ becomes $[l_1, l_2, \ldots, l_n]$). Besides, we use $\mathbb{1}_{condition}$ as the indicator function, which returns 1 if the $condition$ is true and 0 otherwise.

Direct Repetition Ratio (DRR) measures the ratio of sequential item pairs having the same labels.

$$DRR(S) = 1/(n-1)\sum_{i=1}^{n-1} \mathbb{1}_{l_i = l_{i+1}} \tag{1}$$

[3]The representation also contains additional data beyond the scope of this work. The list of frames is available at: https://propbank.github.io/v3.4.0/frames/

When considering the example sequences in Table 2, we observe the in the *specific* example sequence one out of five sequential pairs is a direct repetition (i.e., $1/5$). Note that higher order patterns (e.g., *alternating* sequences) do not impact the value. Therefore, singular outliers (e.g., in the *encased* sequence) will only marginally affect the value. The convergence behavior of *random* sequences depends on the size of the label space.

Reciprocal Repeat Distance (RRdist) measures the average distance between neighboring repetitions (i.e., same labels while every label between them is different) and is normalized by the reciprocal value. Therefore, it can be seen as a sort of probability score that labels are repeated.

$$RRdist(S) = \frac{\sum_{i=1}^{n-1} \sum_{j=2}^{n} \mathbb{1}_{i<j, l_i=l_j \wedge l_i \neq l_k, \forall k, i<k<j}}{\sum_{i=1}^{n-1} \sum_{j=2}^{n} (j-i) \mathbb{1}_{i<j, l_i=l_j \wedge l_i \neq l_k, \forall k, i<k<j}} \tag{2}$$

Concerning the *specific* example in Table 2, a has a distances of one and two, while b has a distance of three, which results in an average distance of two (i.e., reciprocal value of 0.5). Note that metric capture higher order patterns, such as both the *alternating* and *encased* sequence having a distance of 0.5. In the former case, the distance is always two, while in the latter case, $n-3$ times a distance of one and one time a distance of $n-1$ resulting of $n-2$ repetition events (i.e., $\frac{n-2}{(n-3)*1+1*(n-1)}$). Similar to DDR, the limit of a *random* sequence approaches the reciprocal value of the label space.

Uniqueness Index (Uniq) determines how much of unique labels are present compared to the theoretical maximum. The maximum depends on the sequence length and label space and is bounded by whichever is smaller. Therefore, if $|L_j| < n$, then the maximum is reached when all labels are present, whereas if $n < |L_j|$, the maximum is reached when *all* labels are *different*.

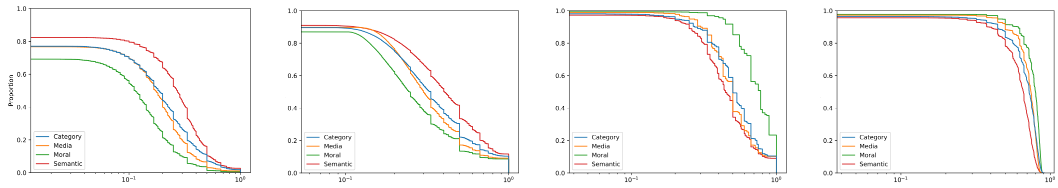
$$Uniq(S) = \frac{|\{S\}| - 1}{min(|S|, |L_j|) - 1} \tag{3}$$

In Table 2, the *specific* sequence is $(3-1)/(6-1)$ as three of potentially six labels are present. If *all* items are the *same*, then the minimum of zero is reached (which is why one is deducted from both the enumerator and denominator). The value tends towards one for long *random* sequences. Therefore, the metric is a form of coverage on the sequence level rather than system level.

Distribution Imbalance (Gini) uses the Gini index, which considers the probabilities of label occurrence. Therefore, uniform distribution lead to higher values than skewed distributions.

$$Gini(S) = 1 - \sum_{l \in L_j} (p_l)^2, \quad p_l = \frac{1}{|S|} \sum_{i=1}^{n} \mathbb{1}_{l_i=l} \tag{4}$$

The *specific* example of Table 2 is thus the result of $1 - ((1/6)^2 + (2/6)^2 + (3/6)^2)$. Gini is 0 with *all same* sequence, has 0.5 with two labels equally distributed (e.g., *alternating* sequence), and tends towards 1 as long sequence of *all different* labels. Similar to DDR, singular outliers do



(a) DRR ECDF Plot (b) RRdist ECDF (c) Uniq ECDF (d) Gini ECDF

Figure 2: Complementary empirical cumulative distribution function (ECDF) plots on a log scale (x-axis) on S_H. Category = blue, Media Frames = orange, Moral Frames = green, Semantic Frames = red.

not significantly affect the outcome on long sequences (e.g., consider *encased*). For long *random* sequences, the value depends on the size of the label space.

5. Experiments

We want to answer the following research questions by analyzing their corresponding label sequences (denoted by $\rightarrow$):

RQ1: **How is the repeat consumption behavior and viewpoint diversity of frames compared to categories?**
$\rightarrow S_H = H$: the set of user histories; also used for comparison in RQ2 and RQ3.

RQ2: **What is the interplay between frames and categories?**
Whether more of the same frames are consumed in per-category sub-sequences?
$\rightarrow S_{H/L_{Cat}}$: the subsets from the user histories per category

RQ3: **What are the effects of framing with regard to (a) retrieved, i.e., with impressions,**
($\rightarrow S_{H\oplus R}$: the user history enhanced with a single impression)
and (b) consumed, i.e., with clicked, content
($\rightarrow S_{H\oplus C}$: the user history enhanced with a single click)?

RQ1: Comparison of Framing Behavior

Concerning the user history S_H, we observe that categories and media frames are closely related (Table 3 and Figure 2), which can be the result of the set of media frames being defined in terms of topics (for which they were already criticized [5]). On the other hand, moral and semantic frames deviate notably and have the opposite tendency towards each other. Users show a low repeat consumption behavior (according to DRR and $RRdist$) in terms of moral frames and high viewpoint diversity (according to $Uniq$ and $Gini$). The effect is most pronounced for the uniqueness index, which becomes visually apparent in Figure 2c. Concerning the overall distribution of values (Figure 2), repeat behavior metrics are lower for all label types compared to viewpoint diversity. In fact, around 20% of sequences do not have any direct repetitions (left starting point in Figure 2a), and around 10% of sequences have all different labels, which results in an $RRdist = 0$ indicated in Figure 2b. Herein, the results first increase much quicker for DRR, whereas for $RRdist$, there is a noticeable jump at the end to the value of 1. For $Gini$ the values appear clustered around a high value close to 1 without actually reaching it (Figure 2d).

	Categories (L_{Cat})	Media Frames (L_{Media})	Moral Frames (L_{Moral})	Semantic Frames (L_{Sem})
User History (RQ1, $\lvert S_H \rvert = 49,108$, $\overline{S_H} = 18.85$)				
DRR	0.2194 ± 0.21	0.1908 ± 0.17	$\underline{0.1336} \pm 0.14$	0.2861 ± 0.22
$RRdist$	0.3880 ± 0.28	0.3641 ± 0.26	0.3164 ± 0.26	0.4532 ± 0.29
$Uniq$	0.5453 ± 0.23	0.5288 ± 0.21	0.7459 ± 0.20	$\mathbf{0.4883} \pm 0.23$
$Gini$	0.6715 ± 0.19	0.6956 ± 0.16	0.7464 ± 0.16	0.6115 ± 0.18
Per Category (RQ2, $\lvert S_{H/L_{Cat}} \rvert = 687,054$, $\overline{S_{H/L_{Cat}}} = 6.84$)				
DRR	-	$\underline{0.2990} \pm 0.35 \uparrow$	$0.1591 \pm 0.26 \uparrow$	$\mathbf{0.3217} \pm 0.35 \uparrow$
$RRdist$	-	$\underline{0.4703} \pm 0.40 \uparrow$	$0.3354 \pm 0.37 \uparrow$	$\mathbf{0.4979} \pm 0.40 \uparrow$
$Uniq$	-	$0.5523 \pm 0.35 \uparrow$	$0.7385 \pm 0.28 \downarrow$	$0.5454 \pm 0.35 \uparrow$
$Gini$	-	$\underline{0.5058} \pm 0.24 \downarrow$	$0.6051 \pm 0.19 \downarrow$	$\mathbf{0.4886} \pm 0.23 \downarrow$
With Impressions (RQ3a, $\lvert S_{H \oplus R} \rvert = 5,723,002$, $\overline{S_{H \oplus R}} = 37.26$)				
DRR	0.2182 ± 0.16	$0.1959 \pm 0.13 \uparrow\uparrow$	$\mathbf{0.1331} \pm 0.11$	$0.2978 \pm 0.17 \uparrow\uparrow$
$RRdist$	$0.3126 \pm 0.23 \downarrow$	$0.3007 \pm 0.20 \downarrow$	$\underline{0.2533} \pm 0.20 \downarrow$	$0.3813 \pm 0.24 \downarrow$
$Uniq$	$0.5563 \pm 0.19 \uparrow$	$0.5075 \pm 0.17 \downarrow\downarrow$	$\underline{0.8077} \pm 0.18 \uparrow$	$\underline{0.4902} \pm 0.19$
$Gini$	$0.7291 \pm 0.13 \uparrow\uparrow$	$0.7473 \pm 0.10 \uparrow$	$\underline{0.8061} \pm 0.09 \uparrow$	$0.6515 \pm 0.14 \uparrow$
With Clicks (RQ3b, $\lvert S_{H \oplus C} \rvert = 231,530$, $\overline{S_{H \oplus C}} = 41.09$)				
DRR	$0.2227 \pm 0.17 \uparrow$	$0.1946 \pm 0.13 \uparrow$	$\underline{0.1336} \pm 0.10$	$0.2914 \pm 0.16 \uparrow$
$RRdist$	$0.3093 \pm 0.22 \downarrow\downarrow$	$0.2942 \pm 0.20 \downarrow\downarrow$	$\mathbf{0.2490} \pm 0.20 \downarrow\downarrow$	$0.3674 \pm 0.25 \downarrow\downarrow$
$Uniq$	$0.5580 \pm 0.19 \uparrow\uparrow$	$0.5128 \pm 0.17 \downarrow$	$\mathbf{0.8113} \pm 0.18 \uparrow\uparrow$	$0.5080 \pm 0.20 \uparrow\uparrow$
$Gini$	$0.7253 \pm 0.14 \uparrow$	$0.7497 \pm 0.10 \uparrow\uparrow$	$\mathbf{0.8066} \pm 0.09 \uparrow\uparrow$	$0.6597 \pm 0.13 \uparrow\uparrow$

Table 3
Mean metrics of sequences with standard deviation ($\pm$). $\uparrow$ indicates a statistically significant increase ($p < 0.0005$ according to a t-test) in metric compared to user history sequences S_H, while $\downarrow$ indicates the opposite direction. In case that both impressions and clicks have the same effect on the direction, we denote the stronger effect with a double arrow (i.e., $\uparrow\uparrow$ or $\downarrow\downarrow$). The overall highest and lowest values per metric are highlighted in **bold**, while the second highest/lowest are <u>underlined</u>. For each set of sequences, we denote the amount and average length.

In sum, the repeat consumption and viewpoint diversity are frame-specific. Moral frames appear to be consumed in a more balanced way compared to semantic frames. Furthermore, categories and media frames seem to be closely related in terms of consumption behavior. Therefore, we investigate this relation in RQ2.

RQ2: Relation between Categories and Framing

All three types of frames are correlated with the categories on all four metrics (plots are provided in the code repository). The consideration of the subsequences per category ($S_{H/L_{Cat}}$) leads to statistically significant changes in all metrics and frames (Table 3). Specifically, the repeat consumption always increases (both DRR and $RRdist$), while $Gini$ always decreases. In fact, this results in the highest (bold in semantic frames) and second highest (underlined in media frames) values overall in terms of repeat consumption and similarly the lowest and second lowest for $Gini$. Therefore, the consumption behavior appears less balanced when considering individual categories. In other words, a balanced consumption behavior regarding framing appears to be partially the result of a more diverse set of categories consumed. Interestingly, the $Uniq$, while still affected, does not show such a tendency. Moreover, it even increases for media and semantic frames, thus indicating a still broad range of frames in these shorter sequences.

Overall, we can conclude that categories play a vital role in the consumption behavior of frames, as the same frames are consumed even more repeatedly. As information systems are also prone to narrow the content shown to users [43], e.g., by repeatedly recommending similar items in terms of categories, we investigate these effects more closely in RQ3.

RQ3: Framing Effects in Information Systems

To start, we investigate whether shown and click items are a mere repetition of the last item's label in the user history (i.e., whether DRR increases in Table 3). Apparently, the last category is not used to determine the shown items, while the users themselves, more often than not, stick to the same category. Here, user intent might play a role (see [44] for an example of intent modeling in sequential recommendation), which is beyond the scope of the current study. Nevertheless, the system seems to repeat the media and semantic frames, which also affects the user click behavior. The effect is more pronounced in $S_{H \oplus R}$ compared to $S_{H \oplus C}$, which might indicate that the system is the source of the bias rather than the users themselves. Interestingly, moral frames do not seem that affected (no statistically significant change of $p < 0.0005$) and stay low (being the lowest values of DRR overall). In comparison, $RRDist$ decreases for both sets of sequences, while viewpoint diversity tends to increase. This effect is most pronounced regarding the moral frames, especially on the click behavior. In general, the click behavior is more affected regarding $RRdist$, $Uniq$, and $Gini$. One outlier here is the uniqueness of media frames, which decreases and is more pronounced in the impressions rather than click behavior.

The results suggest that, although information systems tend to promote sticking to the same type of content, the effects on consumption behavior might be a net positive, as users could be supported in balancing their media consumption. Please note that the current study cannot deduce long-term effects and therefore urges for future work.

6. Conclusion

In the present, study we relate the framing of content to consumption behavior in information systems. Herein, we investigate the repeat consumption behavior and viewpoint diversity for three types of frames (i.e., media, moral, and semantic frames). Our findings suggest the relation to behavior is different per frame type, with media frames closely following categories. The repetition of frames also increases when investigating the categories separately, whereas the diversity tends to increase due to the effects of information systems.

Our study has broad implications for the design of information systems, as it suggests considering user behavior within particular types of content rather than diversifying through recommending a broad spectrum of types.

Limitations. Our study has two main limitations. First, the scope of the study is narrow, as we consider only a single dataset in the news domain, which was designed for recommendation research, with three specific models. Second, we performed a simplified analysis for better comparison, which omitted fine-grained details in content (e.g., the graph structure of semantic frames), metrics (e.g., the influence due to number of labels), and behavior (e.g., user intent).

Future Work. We hope our work sparks interest in considering framing as a form of bias in information systems. Most of all, we call for the development of debiasing methods concerning user behavior due to framing. Specifically, we see personalized user interfaces that support a balanced consumption diet, e.g., through transparency, as a promising research direction for future work.

References

[1] A. Tversky, D. Kahneman, The framing of decisions and the psychology of choice, science 211 (1981) 453–458.

[2] L. Azzopardi, Cognitive biases in search: a review and reflection of cognitive biases in information retrieval, in: Proceedings of the 2021 conference on human information interaction and retrieval, 2021, pp. 27–37.

[3] K. Sullivan, Three levels of framing, Wiley Interdisciplinary Reviews: Cognitive Science (2023) e1651.

[4] R. Baeza-Yates, Bias on the web, Communications of the ACM 61 (2018) 54–61.

[5] M. Ali, N. Hassan, A survey of computational framing analysis approaches, in: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, 2022, pp. 9335–9348.

[6] A. Anderson, R. Kumar, A. Tomkins, S. Vassilvitskii, The dynamics of repeat consumption, in: Proceedings of the 23rd international conference on World wide web, 2014, pp. 419–430.

[7] T. Draws, N. Tintarev, U. Gadiraju, Assessing viewpoint diversity in search results using ranking fairness metrics, ACM SIGKDD Explorations Newsletter 23 (2021) 50–58.

[8] F. Wu, Y. Qiao, J.-H. Chen, C. Wu, T. Qi, J. Lian, D. Liu, X. Xie, J. Gao, W. Wu, et al., Mind: A large-scale dataset for news recommendation, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 3597–3606.

[9] S. Raza, C. Ding, News recommender system: a review of recent progress, challenges, and opportunities, Artificial Intelligence Review (2022) 1–52.

[10] M. Reiter-Haas, B. Klösch, M. Hadler, E. Lex, Framefinder: Explorative multi-perspective framing extraction from news headlines, arXiv preprint arXiv:2312.08995 (2023).

[11] M. Reiter-Haas, Mind-small frames, 2024. URL: https://doi.org/10.5281/zenodo.10509498. doi:10.5281/zenodo.10509498.

[12] R. M. Entman, Framing: Toward clarification of a fractured paradigm, Journal of communication 43 (1993) 51–58.

[13] R. M. Entman, Framing bias: Media in the distribution of power, Journal of communication 57 (2007) 163–173.

[14] F. Morstatter, L. Wu, U. Yavanoglu, S. R. Corman, H. Liu, Identifying framing bias in online news, ACM Transactions on Social Computing 1 (2018) 1–18.

[15] C. J. Fillmore, C. F. Baker, Frame semantics for text understanding, in: Proceedings of WordNet and Other Lexical Resources Workshop, NAACL, volume 6, 2001.

[16] P. Wicke, M. M. Bolognesi, Framing covid-19: How we conceptualize and discuss the pandemic on twitter, PloS one 15 (2020) e0240010.

[17] D. Demszky, N. Garg, R. Voigt, J. Zou, M. Gentzkow, J. Shapiro, D. Jurafsky, Analyzing

polarization in social media: Method and application to tweets on 21 mass shootings, arXiv preprint arXiv:1904.01596 (2019).
[18] M. Reiter-Haas, S. Kopeinik, E. Lex, Studying moral-based differences in the framing of political tweets, arXiv preprint arXiv:2103.11853 (2021).
[19] C. Shurafa, K. Darwish, W. Zaghouani, Political framing: Us covid19 blame game, in: International Conference on Social Informatics, Springer, 2020, pp. 333–351.
[20] J. Piskorski, N. Stefanovitch, G. Da San Martino, P. Nakov, Semeval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup, in: Proceedings of the the 17th International Workshop on Semantic Evaluation (SemEval-2023), 2023, pp. 2343–2361.
[21] Q. Liao, M. Lai, P. Nakov, Marseclipse at semeval-2023 task 3: Multi-lingual and multi-label framing detection with contrastive learning, arXiv preprint arXiv:2304.14339 (2023).
[22] M. Reiter-Haas, A. Ertl, K. Innerebner, E. Lex, mcpt at semeval-2023 task 3: Multilingual label-aware contrastive pre-training of transformers for few-and zero-shot framing detection, arXiv preprint arXiv:2303.09901 (2023).
[23] B. Wu, O. Razuvayevskaya, F. Heppell, J. A. Leite, C. Scarton, K. Bontcheva, X. Song, Sheffieldveraai at semeval-2023 task 3: Mono and multilingual approaches for news genre, topic and persuasion technique classification, in: Proceedings of the The 17th International Workshop on Semantic Evaluation (SemEval-2023), 2023, pp. 1995–2008.
[24] V. Bhatia, V. P. Akavoor, S. Paik, L. Guo, M. Jalal, A. Smith, D. A. Tofu, E. E. Halim, Y. Sun, M. Betke, et al., Openframing: open-sourced tool for computational framing analysis of multilingual data, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 2021, pp. 242–250.
[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).
[26] E. Lex, D. Kowald, P. Seitlinger, T. N. T. Tran, A. Felfernig, M. Schedl, et al., Psychology-informed recommender systems, Foundations and Trends® in Information Retrieval 15 (2021) 134–242.
[27] D. Kowald, M. Reiter-Haas, S. Kopeinik, M. Schedl, E. Lex, Transparent music preference modeling and recommendation with a model of human memory theory (2024).
[28] M. Reiter-Haas, E. Parada-Cabaleiro, M. Schedl, E. Motamedi, M. Tkalcic, E. Lex, Predicting music relistening behavior using the act-r framework, in: Proceedings of the 15th ACM Conference on Recommender Systems, 2021, pp. 702–707.
[29] M. Moscati, C. Wallmann, M. Reiter-Haas, D. Kowald, E. Lex, M. Schedl, Integrating the act-r framework with collaborative filtering for explainable sequential music recommendation, in: Proceedings of the 17th ACM Conference on Recommender Systems, 2023, pp. 840–847.
[30] E. O'Brien, A mind stretched: The psychology of repeat consumption, Consumer Psychology Review 4 (2021) 42–58.
[31] N. Tintarev, E. Sullivan, D. Guldin, S. Qiu, D. Odjik, Same, same, but different: algorithmic diversification of viewpoints in news, in: Adjunct publication of the 26th conference on user modeling, adaptation and personalization, 2018, pp. 7–13.
[32] C. Baden, N. Springer, Conceptualizing viewpoint diversity in news discourse, Journalism 18 (2017) 176–194.

[33] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, et al., Transformers: State-of-the-art natural language processing, in: Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, 2020, pp. 38–45.
[34] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, L. Zettlemoyer, Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension, arXiv preprint arXiv:1910.13461 (2019).
[35] A. Williams, N. Nangia, S. Bowman, A broad-coverage challenge corpus for sentence understanding through inference, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018, pp. 1112–1122.
[36] W. Yin, J. Hay, D. Roth, Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach, in: 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Association for Computational Linguistics, 2019, pp. 3914–3923.
[37] D. Card, A. Boydstun, J. H. Gross, P. Resnik, N. A. Smith, The media frames corpus: Annotations of frames across issues, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015, pp. 438–444.
[38] K. Song, X. Tan, T. Qin, J. Lu, T.-Y. Liu, Mpnet: Masked and permuted pre-training for language understanding, Advances in Neural Information Processing Systems 33 (2020) 16857–16867.
[39] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, 2019, pp. 3982—-3992.
[40] J. Frimer, J. Haidt, J. Graham, M. Dehghani, R. Boghrati, Moral foundations dictionaries for linguistic analyses, 2.0, Unpublished Manuscript. Retrieved from: www. jeremyfrimer. com/uploads/2/1/2/7/21278832/summary. pdf (2017).
[41] J. Graham, J. Haidt, B. A. Nosek, Liberals and conservatives rely on different sets of moral foundations., Journal of personality and social psychology 96 (2009) 1029.
[42] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, N. Schneider, Abstract meaning representation for sembanking, in: Proceedings of the 7th linguistic annotation workshop and interoperability with discourse, 2013, pp. 178–186.
[43] T. T. Nguyen, P.-M. Hui, F. M. Harper, L. Terveen, J. A. Konstan, Exploring the filter bubble: the effect of using recommender systems on content diversity, in: Proceedings of the 23rd international conference on World wide web, 2014, pp. 677–686.
[44] Y. Chen, Z. Liu, J. Li, J. McAuley, C. Xiong, Intent contrastive learning for sequential recommendation, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 2172–2182.