Newsalyze: Effective Communication of Person-Targeting Biases in News Articles

Felix Hamborg*‡, Kim Heinser‡, Anastasia Zhukova§, Karsten Donnay*† and Bela Gipp*§

*Heidelberg Academy of Sciences and Humanities, Germany, ‡University of Konstanz, Germany,
†University of Zurich, Switzerland, §University of Wuppertal, Germany
felix.hamborg@uni-konstanz.de

Abstract—Media bias and its extreme form, fake news, can decisively affect public opinion. Especially when reporting on policy issues, slanted news coverage may strongly influence societal decisions, e.g., in democratic elections. Our paper makes three contributions to address this issue. First, we present a system for bias identification, which combines state-of-the-art methods from natural language understanding. Second, we devise bias-sensitive visualizations to communicate bias in news articles to non-expert news consumers. Third, our main contribution is a large-scale user study that measures bias-awareness in a setting that approximates daily news consumption, e.g., we present respondents with a news overview and individual articles. We not only measure the visualizations' effect on respondents' bias-awareness, but we can also pinpoint the effects on individual components of the visualizations by employing a conjoint design. Our bias-sensitive overviews strongly and significantly increase bias-awareness in respondents. Our study further suggests that our content-driven identification method detects groups of similarly slanted news articles due to substantial biases present in individual news articles. In contrast, the reviewed prior work rather only facilitates the visibility of biases, e.g., by distinguishing left- and right-wing outlets.

Index Terms—news bias, conjoint experiment, Google News, AllSides, frames, framing, news aggregator, user interface, HCI

I. INTRODUCTION

News articles serve as an essential source of information on current events. While a rich diversity of opinions is, of course, desirable also in the news, systematically biased information can be problematic as a basis for collective decision-making, e.g., in democratic elections, if not recognized as such. Empowering newsreaders in recognizing biases in coverage, especially on policy issues, is thus crucial.

We take this as a motivation to define our research question: *How can we effectively communicate instances of bias in a set of news articles reporting on the same political event to non-expert news consumers?* The main contributions of this paper are the first *large-scale user study (contribution C1)* that employs a conjoint design to investigate how visualizations and also individual components therein can help news consumers to become aware of biases. Our study employs a setting that resembles daily news consumption. Before the study, we introduce our *bias identification system (C2)*, which combines state-of-the-art methods to identify media bias. Moreover, we introduce layouts and components to build *modular visualizations to communicate biases (C3)*.

We publish the survey materials, including questionnaires, articles, and visualizations: https://zenodo.org/record/4704891

II. RELATED WORK

A. Definitions

Defining media bias is a challenging task [1] due to overlapping or even contrary bias theories [2]. We rely on a frequent (though not consistent) concept among bias definitions developed by social science researchers, where *media bias* is a relative concept, i.e., bias can only be evaluated comparatively to other information, e.g., news articles (cf. [2]–[4]).

We define *bias-awareness* generally as an effect of bias communication. In practical terms, we define bias-awareness in this paper as an individual's motivation and ability to relate and contrast perspectives present in news coverage to another [1] and also to the individual's views [5].

B. Approaches

We limit our problem statement to the analysis and communication of media bias (also called bias diagnosis, measurement, and mitigation [1]). We exclude other means to address media bias, such as bias prevention during news production [1]. Bias-sensitive visualizations may support news consumers in making more informed choices [6]. However, we find that the prior work suffers from the following shortcomings.

High cost and lack of recency: Content analyses and frame analyses are among the most effective bias analysis tools. Decade-long research in the social sciences has proven them effective and reliable, e.g., to capture also subtle yet powerful biases (cf. [2], [7]). However, because researchers need to conduct these analyses mostly manually, the analyses do not scale with the vast amount of news [2]. In turn, such studies are always conducted for (few) topics in the past and do not deliver insights for the current day [8], [9]; this would, however, be of primary interest to people reading the news.

Superficial results: Many automated approaches suffer from superficial results, especially when compared to the results of cumbersome analyses as conducted in the social sciences [2]. Reasons include that the approaches treat media bias as a rather vaguely or broadly defined concept, e.g., "differences of [news] coverage" [10], "diverse opinions" [11], or "topic diversity" [12]. Further, especially early approaches [1], [12] suffer from poor performance since word-, dictionary-, or rule-based methods as commonly employed in traditional machine learning fail to capture the "meaning between the lines" [2]. To improve performance, some approaches employ crowdsourcing [13]–[15], e.g., to gain bias ratings. Crowdsourced data can

be an effective means to gather labeled data. However, such data can be problematic if not carefully reviewed for biases [16], e.g., if users are not a representative sample or already biased through earlier exposure to systematically biased news coverage. Recent methods that employ deep learning or word embeddings can yield more substantial results. However, the creation of large-scale datasets required for their training is very costly [17], [18], and semi-automated approaches require careful, manual revision of the automatically identified bias categories [19].

Inconsistency: The design of some approaches only facilitates the visibility of biases that might be in the data rather than determining whether and which biases are indeed present. For example, when not analyzing the articles' content but using the outlets' political orientation [14] or only the headline [20].

In sum, many studies confirm the effectiveness of communicating biases to news consumers. However, prior work suffers from various shortcomings, such as requiring manual analyses, yielding superficial results, or only facilitating the visibility of media bias that *might* be in the data. To address these issues, we propose an approach that automates parts of frame analysis, the established procedure in the social sciences to analyze media bias.

III. System Description

Given a set of news articles reporting on the same political event, our system seeks to find and visualize groups of articles that similarly frame the event using a three-phase workflow (cf. [21], [22]): article gathering, bias analysis, and bias communication. This section summarizes our previous research concerning article gathering [23] and bias analysis [2]. Afterward, we briefly describe the grouping method used in our bias analysis. Section IV then introduces our novel visualizations for bias communication. For article gathering, we integrate a crawler and extractor for news articles [23]. Users provide the system with a set of URLs linking to news articles reporting on the same event. The news crawler then extracts the required information from the articles' web pages, i.e., title, lead paragraph, and main text. Alternatively, users can directly provide news articles to the system, e.g., through JSON files.

Bias analysis performs three tasks as depicted in Figure 1. First, we perform state-of-the-art NLP *preprocessing*, including part-of-speech (POS) tagging, dependency parsing, full parsing, named entity recognition (NER), and coreference resolution [24], [25]. We use CoreNLP with neural models where available and the defaults for the English language [26]. In the following, we describe the subsequent tasks, i.e., target concept analysis and frame analysis.

Target concept analysis finds and resolves persons mentioned across the topic's articles, including broadly defined and highly event-specific coreferences as they frequently occur in person-targeting bias forms. Especially in the presence of bias by word choice and labeling, person mentions may be coreferential only in news coverage on a specific event, but otherwise not or even opposing, such as "freedom fighters" and "terrorists" [2]. To resolve such mentions, we use the sieve-based method for context-driven cross-document coreference (CDCR) proposed by Hamborg et al. [2].

In sum, our method performs two tasks: candidate extraction and candidate merging. In candidate extraction, we create a base set of coreferential chains from two sources. First, we take the chains from CoreNLP's coreference resolution on the individual news articles (see *preprocessing*). We extend this base set by adding any noun phrase (NP) as singleton coreference chains. The candidate merging task uses six sieves, where each analyzes specific characteristics of two candidates to determine whether they refer to the same semantic concept. For example, the first sieve matches two chains' representative phrases, which represent the chains' core meaning [26]. The second sieve determines the semantic similarity of all mentions of two chains [28]. The CDCR method achieves a $F1_m = 81.7$ compared to 75.3 achieved by the best baseline, which extracts each noun phrase (NP) and each mention of coreference chains as single candidates chains and clusters the candidate chains in the word2vec space using affinity propagation. Hamborg et al. provide more information on the approach and the baseline [21]. The output of target concept analysis is the set of persons involved in the news coverage of the event, and for each person, all the person's mention across all news articles.

Frame identification determines how news articles portray the persons involved in the event and then finds groups of articles that similarly portray these persons. This task centers around (political) framing [7], where a frame represents a specific perspective on an event. Identifying frames would approximate content analyses, the standard tool used in the social sciences to analyze media bias [2]. However, doing so would require infeasible effort since researchers in the social sciences typically create frames for a specific research question [2], [7]. Our system, however, is meant to analyze media bias on any coverage reporting on policy issues. Thus, we seek to determine a fundamental bias effect resulting from framing: polarity of individual persons, which we identify for each person mention (on the sentence level) and aggregate to article-level. To achieve state-of-the-art performance in target-dependent sentiment classification (TSC) on news articles, we use a RoBERTa-based neural model trained on more than 11k manually labeled sentences sampled from news articles ($F1_m = 83.1$) [18].

The last step of frame analysis is to determine groups of articles that similarly frame the event, i.e., the persons involved in the event. We propose two methods for grouping. (1) *Grouping-MFA*, a simple, polarity-based method, first determines the single person that occurs most frequently across all articles, called most frequent actor (*MFA*). Then, the method assigns each article to one of three groups, depending on whether the article mentions the MFA mostly positively, ambivalently, or negatively. (2) *Grouping-ALL* considers the polarity of all persons instead of only the MFA. Specifically, the method uses k-means with $k = 3$ on a set of vectors where each vector a represents a single news article: $a = \begin{pmatrix} s_0 & \cdots & s_{|P|-1} \end{pmatrix}$ where P is the set of all persons, a person's
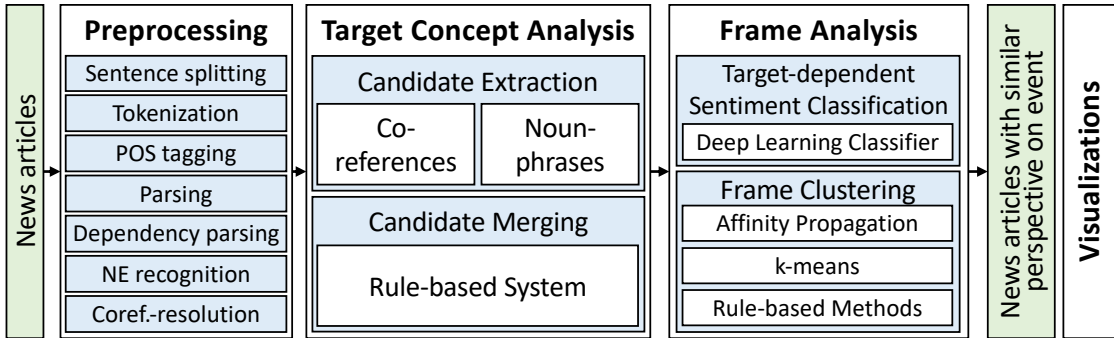
Fig. 1. Shown is the three-tasks analysis workflow as it preprocesses news articles reporting on the same event, extracts and resolves phrases referring to persons involved in the event, and groups articles reporting similarly on these persons. Adapted from: [27]

sentiment polarity s_i in a is

$$s_i = \sum_{m \in M} w(m)s(m)/m_{\max,a} \qquad (1)$$

where m is each mention of all the person's mention in a, $w(m)$ is a weight depending on the position of the mention (mentions in the beginning of an article are considered more important [29]), $s(m)$ yields the polarity score of m (1 for positive, -1 for negative, 0 else). To consider the individual persons' frequency in an article for clustering, we normalize by $m_{\max,a}$, which is the number of mentions of the most frequent person in a.

In addition to grouping, we calculate each article's relevance concerning the event and concerning the article's group using simple word-embedding scoring.

IV. VISUALIZATIONS FOR THE CONJOINT EXPERIMENT

Our visualizations resemble typical online news consumption, i.e., an *overview* enables users to first get a synopsis of news events and articles (Section IV-A) and an *article view* shows an individual news article (Section IV-B). To measure the effectiveness not only of our visualizations but also their constituents (see our conjoint experiment design [30] described in Section V-A), we design them so that visual features can be altered or exchanged. To more precisely measure the change in bias-awareness concerning only the textual content, we apply changes compared to typical news consumption. For example, the visualizations show the texts of articles (and information about biases in the texts) but no other content, e.g., no photos or outlet names. Further, in our study, the overview shows only a single topic instead of multiple. The visualizations show brief explanations for all visual features.

A. Overview

The overview aims to enable users to get a synopsis of a news event quickly. We devise a modular, bias-sensitive visualization layout, which we use to implement and test specific visualizations. The comparative layout aims to support users in quickly understanding the frames present in coverage of the event. The bias-sensitive layout is vertically divided

into three parts (Figure 2 shows A and B). The event's *main article* (part A) shows the event's most representative article. The *comparative groups* part (B) shows up to three perspectives present in event coverage by showcasing each perspective's most representative article. To determine the bias-groups, the system uses one of the grouping methods described in Section III, i.e., Grouping-MFA or Grouping-ALL. Finally, a list shows the headlines of *further articles* reporting on the event (bottom, not shown in Figure 2).

In each overview, two types of visual clues conveying bias information can be enabled and altered depending on the conjoint profile (cf. Section V-A). First, zero or more headline tags are shown next to each article's headline. They indicate the political orientation of the article's outlet (*PolSides tags*, see "2" in Figure 2), the article's overall polarity regarding the MFA due to Grouping-MFA (*MFAP tags*, see "3"), and the article's group according to its polarity regarding all persons due to Grouping-ALL (*ALLP tags*), respectively. Second, labels and explanations in the visualization are either *generic* or *specific*. The specific variants explain how the grouping was specifically performed (see "1") and provide specific group labels (see "4"). In contrast, all generic variants use the same universal explanation, e.g., only mentioning that our system automatically determined the three perspectives, and use generic coloring and labels, e.g., "Perspective 1."

B. Article View

The article view visualizes a single news article. It thus represents the typically second step in news consumption, i.e., after getting an overview, users subsequently may read individual articles of interest. The article view shows a given article's headline, lead paragraph, and main text. Optionally, the following visual clues to communicate bias information: (1) in-text polarity highlights, (2) polarity context bar, (3) article's bias-groups, and (4) headline tags. These clues are enabled, disabled, or altered depending on the conjoint profile.

In-text polarity highlights aim to enable users to quickly comprehend how the individual sentences of a news article portray the mentioned persons. To achieve this, we visually mark mentions of individual persons within the news article's

Fig. 2. Excerpt of the MFAP overview showing three primary frames present in news coverage on a debt-ceiling event.
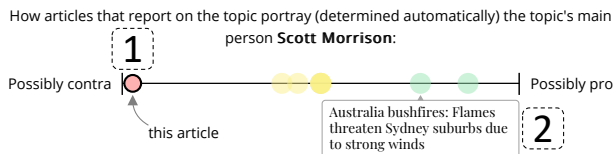

Fig. 3. Polarity context bar showing the current and other articles' polarity regarding the MFA and a tooltip showing the headline of a hovered article.

text. We test the effectiveness of the following modes: *single-color* (visually marking a person mention using a neutral color, i.e., gray, if the respective sentence mentions the person positively or negatively), *two-color* (using green and red colors for positive and negative mentions, respectively), *three-color* (same as two-color and additionally showing neutral polarity as gray), and *disabled* (no highlights are shown). For example, in the sentence "The Mueller report was tough on Trump," the person mention "Trump" has a negative polarity and would be highlighted red in the two- and three-colors modes.

The *polarity context bar* aims to enable users to quickly contrast how the current article and other articles portray the event's MFA. The 1D scatter plot depicted in Figure 3 represents each article as a circle, where the article visualized in the article view is highlighted with a bold circle (see "1" in Figure 3). The polarity context bar places each circle depending on its article's overall polarity regarding the MFA. Users can interactively, i.e., by hovering their cursor of circles, view individual articles' headlines (see "2").

The *article's bias-groups* part is a set of basic indicators that show the article's bias-group, analogously to the headline tags, i.e., the outlet's political orientation (PolSides), how the article reports on the MFA (MFAP), or all persons (ALLP).

In contrast to the headline tags, which are shown besides all headlines, each bias-group indicator is a more prominent component that prominently shows only the bias-group of the article currently shown in the article view. Depending on the conjoint profile (identical to headline tags), individual indicators are shown or disabled. The *headline tags* have identical purpose and function as when shown in the overview (see Section IV-A).

V. STUDY DESIGN

A. Methodology

We used a conjoint design [31], which is especially suitable for estimating the effect of individual components. Traditional survey experiments are limited to only identifying the "catch-all effect" due to confounding of the treatment components [30]. In contrast, conjoint experiments identify "component-specific causal effects by randomly manipulating multiple attributes of alternatives simultaneously" [30]. In a conjoint design, respondents are asked to rate so-called "profiles," which consist of multiple "attributes." In our study, such attributes are, for example, the overview, which topic it shows (or which article the article view shows), and if or which tags or in-text highlights are shown.

Conjoint experiments rest on three core assumptions: (1) stability and no carry-over effects, (2) no profile-order effects, (3) randomization of the profiles [30]. In our evaluation, (2) holds by design for all tasks except the forced-choice (see (6) in Section V-E). We ensure (3) by randomly choosing the attributes independently of another and for each respondent. To confirm the randomization is successful, we employed a Shapiro-Wilk test [32] and compared relative frequencies of the demographic dimensions. We found that the data were

nearly balanced in each of the three experiments concerning essential demographic dimensions, such as age, sex, education, and political orientation. Thus, also the respondents' demographic data adhere to the conjoint randomization assumption. The political orientation was slightly skewed to the left as is expected with MTurk respondents (cf. [33]), but still approximates the US distribution [34].

To ensure (1), i.e., the absence of carry-over effects from one task set to another, we applied the diagnostics proposed by Hainmueller et al. [30]. We refer to a task set as all tasks shown to a respondent for a single topic, e.g., in our main study, we show respondents for each topic one overview and subsequently two article views. We then calculate if there are meaningful differences across the task sets by building a subgroup for each task set. We found weak carry-over effects when comparing the individual attributes' effects (using our main overview question across the task sets) and when testing the effect of the task set's order (for all overview questions combined (Est. $= 3.28\%$, $p = 0.018$), i.e., respondents are on average more bias-aware in the second task set). Further, in our main study, the other attributes' effects differ when sub-grouping for the task set. However, this is not necessarily problematic. A learning effect is expected and desirable in bias communication. Since we randomized the attributes within each task set, we can include the task sets in the analysis and thus control for the effects, regardless of the task set.

Since in our experiments the three assumptions hold, our design allows for an estimation of the relative influence of each component on the bias-awareness, which is called average marginal component effects (AMCE) [30]. AMCE "represents the marginal effect of attribute l [such as a visualization component] averaged over the joint distribution of the remaining attributes" [30].

In our questionnaires, we employ discrete choice (DC) as well as rating questions (cf. Section V-E) to measure bias-awareness on a behavioral as well as attitudinal level [35]. DC questions are widely used within the conjoint design and found to have high external validity in mimicking real-world behavior [36]. Additionally, DC questions elicit behavior, i.e., which news article respondents prefer to read or rely on for decision-making [37]. In contrast, rating questions capture attitudes and personal viewpoints better [38].

B. Data

We selected four news topics with varying degrees of expected biases among the news articles reporting on the topic. To approximate the degree of bias, we used the topics' expected polarization. Specifically, we selected three topics expected to be highly polarizing for US news readers: gun control (Orlando shooting in 2016), debt ceiling (discussions in July 2019), and abortion rights (Tennessee abortion ban in June 2020). To better approximate regular news consumption, where consumers typically are exposed to news coverage on single events, we selected a single event for each of these topics (shown in parentheses). We added a fourth event, which

we expected to be only mildly polarizing: Australian bush fires, i.e., a foreign event without direct US involvement.

For each event, we selected ten articles from left-, center, and right-wing online US outlets (political orientation as self-identified by the outlets or from [14]). We added the abortion topic during our pre-studies due to a strong, negative influence of the debt-ceiling topic on bias-awareness. We could trace this back qualitatively to respondents' critique of the topics being too "boring" and "complicated," which also manifested in lower reading times on average. To ensure high quality, we manually retrieved the articles' content. For the second pre-study and our main study, we carefully shortened all articles so that they were of similar length (300–400 words) to address the high reading times and noise in the responses, a key finding of the first pre-study (see Section VI-A). We consistently applied the same shortening procedure to preserve the perspectives of the original articles. For example, by maintaining the relative frequency of person mentions and by discarding only redundant sentences that do not contribute to the overall tone. In all experiments, we removed any non-textual content, such as images, to isolate the effects in the change of bias-awareness due to the text content, our text-centric bias analysis, and visualization.

C. Setup and Quality

We conducted our experiments as a series of online studies on Amazon Mechanical Turk (MTurk). To participate in our studies, crowd workers had to be located in the US. To ensure high quality, we further required that participants possess MTurk's "Masters" qualification, i.e., have a history of high-quality work. While we compensated respondents always, we discarded data of any respondent who failed to meet all quality criteria, including a minimum study duration, and correctly answering questions checking attention and seriousness [39]. Depending on the study's duration, participants received compensation that approximated an hourly wage of $10.

D. Baselines

We compare our system and variants with baselines that resemble news aggregators popular among news consumers and an established bias-sensitive news aggregator. Screenshots of their visualizations can be found in the online repository (Section I). *Plain* is an overview variant that resembles popular news aggregators. Using a bias-agnostic design similar to Google News, this baseline shows article headlines and excerpts in a list sorted by the articles' relevance to the event (see relevance calculation described in Section III). *PolSides* is a strong overview variant that represents a bias-sensitive news aggregator [14]. PolSides yields bias-groups (see "B" in Figure 2) by grouping articles depending on their outlets' political orientation (left, center, and right, as self-identified by them or taken from [14]). Conceptually, PolSides employs the left-right dichotomy, a simple yet often effective means to partition the media into distinctive slants. However, this dichotomy is determined only on the outlet level. It thus may incorrectly classify the biases indeed present in a specific

event (see Section II-B), e.g., articles shown to be of different slants having indeed similar perspectives (and vice versa). We investigate this issue in our study (see Section VI). Because of its similar problem statement, we deem NewsCube [1] a highly relevant approach. However, we omit it from our study since we suspect the system would achieve poor performance since it was devised using rather quantitative text features. Further, its technique devised for the Korean language cannot be transferred directly to English.

To understand how visualizations, including their layout and explanations, affect bias-awareness compared to what they visualize, i.e., the bias groups resulting from our analysis, we introduce two baseline concepts. First, we include, for most overviews, including the baselines previously mentioned, generic variants (see Section IV-A). Second, we test an overview with generic explanations that randomly assign individual news articles to either of the three bias groups.

For the article view, we test in our conjoint design only the individual visual features described in Section IV-B, since to our knowledge, there are no easy-to-use visualizations for bias communication in single news articles (see Section II). We test only PolSides tags and MFAP tags and exclude Cluster headline tags in our experiments.

E. Workflow and Questions

Our study consists of seven steps. We refer to a *task set* as a sequence of steps associated with one topic, i.e., task set 1 refers to the first topic shown to a respondent, including the overview, the two article views, and respective questionnaires (steps 2–6). The *(1) pre-study questionnaire* asks for demographic and background data [15], such as age, political orientation, news consumption, and attitudes toward the topics we used [40].

Afterward, we show an *(2) overview* as described in Section IV-A and Section V-D including instructions shown prior to the overview. The *(3) post-overview questionnaire* then operationalizes the bias-awareness in respondents (see Section II-A) by asking about their perception of the diversity and disagreement in viewpoints, if the visualization encouraged them to contrast individual headlines, and how many perspectives of the public discourse were shown, e.g., *Do you think the coverage shown in the previous visualization represents all main viewpoints in the public discourse (independent of whether you agree with them or not) [from Not at all to very much]? Overall, how did you perceive the articles shown in the previous visualization [very different – very similar; very opposing – very agreeing]?*. To match our definition of bias-awareness, we use as our main question (cf. [13]): *When viewing the topic visualization, did you have the desire to compare and contrast articles' [Not at all – very much]?*

Afterward, we show an *(4) article view* as described in Section IV-B. A *(5) post-article questionnaire* operationalizes bias-awareness in respondents on an article-level [15], i.e. *How did you perceive the presented news article? [very unfair – very fair; very partial – very impartial; very unacceptable – very acceptable; very untrustworthy – very trustworthy; very*

unpersuasive – very persuasive; very biased – very unbiased]. We also ask whether the article contains political bias and biases against persons mentioned in the article. We repeated steps 2–5 two times since we showed two task sets. After each overview, we showed two articles, i.e., we repeated steps 4 and 5 two times. To measure the effect of seeing an overview before an article, we also introduce a variant where we skip the overview. In such cases, the overview steps (2, 3) are skipped entirely. Afterward, a *(6) discrete choice question* asks respondents to choose between two articles, i.e., which one they consider to be more biased. In a *(7) post-study questionnaire*, respondents give feedback on the study, i.e., what they liked and disliked.

In the two pre-studies, where we tested the study design and usability of the visualizations (see Section VI-A), we repeated the same procedure with only one article after each overview and excluding step (5).

VI. Evaluation

A. Pre-Studies

Before our main study, we conducted two pre-studies (E1 and E2) [41]. E1 consisted of 260 respondents recruited on MTurk (we discarded 3% discarded from 268 due to the quality criteria described in Section V-C). E2 consisted of 98 respondents (we discarded 11% from 110). The pre-studies aimed at testing the study design and usability of the visualizations. We also used the pre-study to find a set of well-performing overviews, including representative baselines. The latter was necessary to satisfy the conjoint assumption *randomization of profiles*, which requires that all profiles have the same set of attributes (each with a randomly selected value for each profile, see Section V-A). Across our overviews, the number of attributes differs (cf. Section V-D), e.g., Plain has only two attributes (one for each headline tag), our bias-sensitive overview layout (see Section IV-A) has an additional grouping attribute, and "no overview" naturally has no attributes (see Section V-E).

In E1, we tested only variants using our bias-sensitive overview layout, where we randomly varied all attributes, i.e., grouping and the two headline tags. We identified (primarily insignificant) trends that indicated well-performing variants. In E2, we then tested the same design as planned for the main study (see Section V-E), including the article view and the other baselines (see below).

We also used the pre-studies to improve our design and visualizations. Reasons for partially mixed results in both pre-studies were various usability issues interfering with the effectiveness. For example, in E1, respondents reported they wanted to know how the grouping was performed and by whom. Before conducting E2, we addressed these shortcomings, e.g., by adding explanations (specific and generic) about how our system derives the classifications. After addressing these issues, we found positive, significant effects of our bias-sensitive overviews in the second pre-study, confirming our research design concerning the overview. E2 revealed that headline tags are most effective in improving bias-awareness in

the Plain baseline. In contrast, for the bias-sensitive overviews, the bias-awareness remained unchanged or decreased. We suspected that users might feel overwhelmed if many visual clues are present (cognitive load).

Using the pre-study findings, we defined the following overview variants for the main study. *(1) No overview. (2) Plain* as described in Section V-D, *(3) PolSides* as described in Section V-D with PolSides headline tags enabled to closely resemble the bias-sensitive news aggregator AllSides.com [14]. *(4) MFA* using the bias-sensitive layout (Section IV-A), Grouping-MFA (Section III), and polarity headline tags enabled, which was the best performing variant of MFA in our pre-studies. *(5) PolSides-generic* being identical to (3) but using generic explanations. *(6) MFA-generic* being identical to (4) but using generic explanations. *(7) Random-generic* using the bias-sensitive layout and random grouping. *(8) ALL-generic* using the bias-sensitive layout, ALL-generic (Section III), and cluster headline tags enabled. Note that we did not test a variant of grouping-ALL with specific explanations.

B. Results

In our main study, we recruited 174 respondents on MTurk from which we discarded 8% using our quality measures. In sum, the $n = 160$ respondents (age: $[23, 77], m = 45.5$, 72f/88m/0d, 100% native speakers, liberal (1)–conservative (10): $m = 4.83, sd = 2.98$) provided answers to 283 post-overview questionnaires (excluding "no overview"), 320 discrete choices on article views, and 640 post-article view questionnaires. Our sample size $n < 245$ as suggested by Cochran's Formula [42], [43], but they assume one observation per respondent, whereas we have two observations per respondent. The average study duration was 15min ($sd = 6.22$).

Our user study shows that the bias-sensitive overviews strongly and significantly increase respondents' bias-awareness compared to the Plain baseline. The estimate (*Est.*) in Table I shows the percentage increase in bias-awareness compared to the attributes' baselines, which is CoreNLP, Plain, and bushfire for the attributes CDCR, Overview, and Topic, respectively. PolSides achieves overall the highest effect when shown with specific explanations (Est. = 21.34). If shown with generic explanations PolSides has no significant effect (Est. = 8.46, $p = 0.17$). In contrast, our methods for determining bias-groups strongly and significantly increase bias-awareness for specific explanations (MFA: Est. = 17.80) as well as generic explanations (MFA: 13.35, ALL: 17.54).

We qualitatively investigated the visualized information of the bias-groups, their articles' content, and respondents' comments. Overall, we found that all grouping methods that achieved significant effects yielded meaningful frames for most topics, especially those polarizing among the political spectrum from left to right. For example, in the gun control topic, the bias-groups achieved by any grouping resemble the frames "gun control" and "gun rights" with subtle differences between the groups. Grouping-MFA yields two "gun control" frames (one being argumentative, the other using factual language) and one "gun right" frame (focusing on cruelty

TABLE I
EFFECTS ON BIAS-AWARENESS AFTER OVERVIEW EXPOSURE

Attr.	Level	Est.	SE	z	p
CDCR	TCA	1.05	2.61	0.40	0.69
Over-view	Random-gen.	6.73	5.85	1.15	0.28
	PolSides	**21.34**	4.84	4.40	***
	MFA	**17.80**	4.90	3.63	***
	PolSides-gen.	8.46	6.19	1.36	0.17
	MFA-gen.	13.35	5.03	2.64	**
	ALL-gen.	**17.54**	5.61	3.12	**
Topic	abortion law	0.78	2.96	0.26	0.79
	gun control	3.16	2.94	1.07	0.25

and the shooter). In PolSides and grouping-ALL, the "gun right" frame focuses on the victims and their right to defend themselves.

However, why is there a loss of effectiveness of PolSides when using generic explanations? Fully elucidating this question would require a larger sample size concerning respondents and topics but we qualitatively and quantitatively identified three potential, partially related causes. *(1) Popularity and intuition:* The left-right dichotomy employed by PolSides is a well-known concept and easily understood by news consumers (none reported they did not understand the concept and 20% of respondents exposed to PolSides praised that grouping was easy to understand, e.g., "I liked that it was laid out with left, center, right. It was intuitive."). In contrast, our grouping techniques are novel and technical, as are their explanations. For example, 30% of respondents exposed to MFA found the descriptions (slightly) confusing and too "technical." For MFA and ALL, showing generic explanations improved respondents' comprehension, likely because the generic explanations are more conceptual and high-level (10% for each of MFA-generic and ALL-generic). For PolSides, however, the effect is reversed, potentially indicating a large proportion of the bias-awareness effect is simply due to the well-known dichotomy rather than the visualized bias-groups.

(2) Learning effect: We think that understanding how framing works, i.e. how and through what means news articles can have different perspectives on an event, helps to comprehend the idea of our grouping methods work or at least what they aim to achieve. We hypothesize that while all respondents were aware of the study's focus on media bias (and framing) after the post-overview questionnaire at the end of task set 1, some respondents might not have been before.[1] This, in turn, would have strongly facilitated comprehending how our grouping methods work or at least what they aim to achieve and thus have a similar effect as described previously for cause (1). While in task set 1, only PolSides employing the well-known left-right dichotomy increased bias-awareness significantly (Est. = 21.55, $p = .0007$), in the subsequent task set 2, our approaches yielded the strongest, most significant

---

[1]Albeit we also informed respondents about the study's focus on perspectives in the news, at the latest in the post-overview questionnaire they must understand the study's focus to be able to proceed.

TABLE II
EFFECTS AFTER OVERVIEW EXPOSURE IN TASK SET 2

Attr.	Level	Est.	SE	z	p
CDCR	TCA	2.78	3.91	0.71	0.48
Over-view	Random-gen.	8.40	9.32	0.90	0.37
	PolSides	23.54	7.45	3.16	**
	MFA	**28.12**	7.03	4.00	***
	PolSides-gen.	12.41	8.48	1.46	0.14
	MFA-gen.	21.78	7.45	2.92	**
	ALL-gen.	**26.46**	8.30	3.19	**
Topic	abortion law	-2.13	5.11	-0.42	0.68
	gun control	0.85	4.80	0.18	0.86

effects (see Table II). Specifically, MFA achieved the strongest effect (Est. = 28.12) among the overviews with specific explanations, and the best performing overview with generic explanations (ALL-generic: 26.46) performed still better than PolSides with specific explanations (23.54).

(3) Substantiality of bias-groups: We qualitatively analyzed the bias-groups yielded by individual grouping methods We found that all grouping methods, including PolSides, determine substantial bias-groups determined in overall for all topics. For some topics, however, the bias-groups determined by MFA and ALL seem to be more substantial compared to bias-groups from PolSides. This is intuitive since PolSides determines bias-groups through the political orientation of the articles' outlets and thus is content-agnostic. In contrast, MFA and ALL analyze in-text features. For example, the debt-ceiling topic employed in our pre-studies highlights this substantiality issue. Our methods yield coherent bias-groups framing the deal negatively, e.g., as political hypocrisy (frame 1), or positively by focusing on (positive) effects for the economy (frame 2) and military (frame 3). In contrast, the bias-groups by PolSides—despite the topic's assumed left-right polarization—resemble rather superficial frames, all framing the deal positively (two groups framing the issue highly similarly, the other focusing on the overall implications of the deal). We suspect that the issue of non-substantial frames is amplified further in news coverage on topics where the left and right wings lack opposing positions.

Overall, the article view does not significantly increase bias-awareness. There are no significant effects for neither component in the article view. Only in-text highlights significantly increase bias-awareness if 3–7 of such highlights are shown and 2-color mode is used (Est. = 2.58, $p = 0.046$). In our qualitative analysis, we identified two potential main causes. *(1) Bias is context-dependent* at least to some degree and thus also depends on relating and contrasting news items [3]. While the bias-sensitive overviews facilitate contrasting news articles, the article view shows only a single news article. Further, the overall tone of an article, e.g., as visualized in the overview, might be more important for bias-perception than individual features (cf. [44]). *(2) User experience (UX) issues:* When qualitatively reviewing respondents' feedback, we identified various UX issues, especially when too many visual clues were shown, e.g., 3-colors mode or many in-text highlights may have caused mental overload [45], and 30% of the respondents that had more than seven highlights reported that they felt "overwhelmed." We also identified issues of other components, e.g., the polarity context bar did not increase bias-awareness albeit conceptually facilitating the comparison of news articles. Often, many articles were placed as circles on the same spot (due to the grouping-MFA). Further, it prominently visualizes only the articles' MFA polarity, but users would need to see actual content rather than such a derived characteristic.

We identified two potential causes concerning why highlights were only effective when their frequency was in a specific range. An article with less than three highlights might not contain enough sentiment to be perceived as biased. On the other hand, showing more than seven highlights could lead to a "mental overload" could be reached [45]. For example, 30 % of respondents reported feeling "overwhelmed" by the highlights and thus could not derive a consistent conclusion on whether the article contained bias.

Showing an overview before the individual news articles had inconclusive effects. We found a significant, mild effect caused by only the MFA overview (Est. = 4.64, $p = 0.003$); other overviews had no significant effects.

C. Limitations and Future Work

In our view, the main limitations of our experiments and results are their *representativeness and generalizability*, mainly due to three partially related factors. *(1) Study design,* e.g., respondents had to view given events and articles rather than deciding what to read. An interactive design might more closely resemble real-world news consumption and address further issues we faced in our experiments, such as study fatigue. While our study's duration is well below where one would expect study fatigue [46], users on MTurk typically work on many online tasks. Rather than querying respondents for the subjective concept of bias-awareness, a long-term usage study could directly measure the effects of our approach on news consumption. For example, if respondents will read more articles portraying events from different perspectives [1].

(2) Respondent sample: While our sample approximately resembles the US distribution concerning a dimension important in this study, i.e., political affiliation (cf. [47], [48]), the sample contains selection biases, e.g., since we recruited respondents only on one platform and from only the US. Thus, we cannot conclude findings for other news consumers of countries with systematically different political or media landscapes. For example, while the two-party system may lead to more polarizing news coverage in the US, countries with multi-party systems typically often have more diversified media landscapes [49]. Further, we seek to increase the sample size since we found inconclusive or insignificant effects for respondents' demographics, their attitudes toward the topics, and in task sets or other sub-groups.

Our *(3) event and article sample* yields similar limitations as described for (2) due to its small size and systematic creation. We suggest increasing the number of events and

articles per event and use a random sample. Lastly, our study did not relate bias-awareness to the articles' content but only to our approach, the respondents, and a topic's expected degree of polarization. To measure the effect of content and biases therein, a future study could relate bias-awareness to a ground-truth created using manual frame analysis.

Besides the previously mentioned limitations and future improvements concerning our study, we plan to address the following issues concerning our approach and the usability of our visualizations. We think that our article view's inconclusive effects are partly due to non-optimal UX, e.g., the view may visualize too few or too many in-text highlights (5% article views did not contain any in-text highlights) or ineffective visual clues, such as the polarity context bar. Another reason for the inconclusive effectiveness might be because bias-awareness is a comparative concept, and the article view shows only a single article. We think that showing representative summaries of each bias group will increase the overview's effectiveness compared to showing the headlines of representative articles. Headlines may contain journalistic hooks and, if at all, are representative for their article but not the group. While our approach yields high effectiveness overall, we also found that our analysis is sub-optimal in topics that are not person-oriented, e.g., in the bushfire topic, where news coverage focused on the consequences for society, economy, and nature. We seek to extend our analysis to other semantic concepts and investigate topic-independent frames and derivations (cf. [17], [19], [21]).

VII. Conclusion

We present the first system to automatically identify and then communicate person-targeting forms of bias in news articles reporting on policy issues. Earlier, researchers could reliably identify these biases only using content analyses, which—despite their effectiveness in capturing also subtle biases—could only be conducted for few topics in the past due to their high cost, manual effort, and required expertise. In a large-scale user study, we employ a conjoint design to measure the effectiveness of visualizations and individual components. We find that our overviews strongly and significantly increase bias-awareness in respondents. In particular, and in contrast to prior work, our bias-identification method seems to reveal biases that emerge from the content of news coverage and individual articles. In practical terms, the study results suggest that the biases found and communicated by our method are indeed *present* in the news articles, whereas the reviewed prior work rather *facilitates* the detection of biases, e.g., by distinguishing between left- and right-wing outlets.

Acknowledgment

References

[1] S. Park, S. Kang, S. Chung, and J. Song, "NewsCube: Delivering multiple aspects of news to mitigate media bias," in *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. New York, New York, USA: ACM Press, 2009, p. 443. [Online]. Available: http://dl.acm.org/citation.cfm?doid=1518701.1518772

[2] F. Hamborg, K. Donnay, and B. Gipp, "Automated identification of media bias in news articles: an interdisciplinary literature review," *International Journal on Digital Libraries*, vol. 20, no. 4, pp. 391–415, 2019. [Online]. Available: https://doi.org/10.1007/s00799-018-0261-y

[3] E. Elejalde, L. Ferres, and E. Herder, "On the nature of real and perceived bias in the mainstream media," *PLOS ONE*, vol. 13, no. 3, p. e0193765, 3 2018. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0193765

[4] E. Pitoura, P. Tsaparas, G. Flouris, I. Fundulaki, P. Papadakos, S. Abiteboul, and G. Weikum, "On Measuring Bias in Online Information," *ACM SIGMOD Record*, vol. 46, no. 4, pp. 16–21, 2 2018. [Online]. Available: https://dl.acm.org/doi/10.1145/3186549.3186553

[5] R. Giner-Sorolla and S. Chaiken, "The Causes of Hostile Media Judgments," *Journal of Experimental Social Psychology*, vol. 30, no. 2, pp. 165–180, 3 1994. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0022103184710080

[6] E. P. S. Baumer, F. Polletta, N. Pierski, and G. K. Gay, "A Simple Intervention to Reduce Framing Effects in Perceptions of Global Climate Change," *Environmental Communication*, vol. 11, no. 3, pp. 289–310, 5 2017. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/17524032.2015.1084015

[7] R. M. Entman, "Framing: Toward Clarification of a Fractured Paradigm," *Journal of Communication*, vol. 43, no. 4, pp. 51–58, 12 1993. [Online]. Available: https://academic.oup.com/joc/article/43/4/51-58/4160153

[8] J. McCarthy, L. Titarenko, C. McPhail, P. Rafail, and B. Augustyn, "Assessing stability in the patterns of selection bias in newspaper coverage of protest during the transition from communism in Belarus," *Mobilization: An International Quarterly*, vol. 13, no. 2, pp. 127–146, 2008.

[9] P. E. Oliver and G. M. Maney, "Political Processes and Local Newspaper Coverage of Protest Events: From Selection Bias to Triadic Interactions," *American Journal of Sociology*, vol. 106, no. 2, pp. 463–505, 9 2000. [Online]. Available: http://www.journals.uchicago.edu/doi/10.1086/316964

[10] S. Park, M. Ko, J. Kim, Y. Liu, and J. Song, "The politics of comments," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11*, ACM. New York, New York, USA: ACM Press, 2011, p. 113. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1958824.1958842

[11] S. A. Munson and P. Resnick, "Presenting diverse political opinions," in *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, ACM. New York, New York, USA: ACM Press, 2010, p. 1457. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1753326.1753543

[12] S. A. Munson, D. X. Zhou, and P. Resnick, "Sidelines: An Algorithm for Increasing Diversity in News and Opinion Aggregators." in *ICWSM*, 2009.

[13] S. Park, M. Ko, J. Kim, H. Choi, and J. Song, "NewsCube 2.0: An Exploratory Design of a Social News Website for Media Bias Mitigation," in *Workshop on Social Recommender Systems*, 2011.

[14] AllSides.com, "AllSides - balanced news," 2021. [Online]. Available: https://www.allsides.com/unbiased-balanced-news

[15] T. Spinde, F. Hamborg, K. Donnay, A. Becerra, and B. Gipp, "Enabling News Consumers to View and Understand Biased News Coverage: A Study on the Perception and Visualization of Media Bias," in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. New York, NY, USA: ACM, 8 2020, pp. 389–392. [Online]. Available: https://dl.acm.org/doi/10.1145/3383583.3398619

[16] C. Hube, B. Fetahu, and U. Gadiraju, "Understanding and Mitigating Worker Biases in the Crowdsourced Collection of Subjective Judgments," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 5 2019, pp. 1–12. [Online]. Available: https://dl.acm.org/doi/10.1145/3290605.3300637

[17] D. Card, A. E. Boydstun, J. H. Gross, P. Resnik, and 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)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2015, pp. 438–444. [Online]. Available: http://aclweb.org/anthology/P15-2072

[18] F. Hamborg and K. Donnay, "Newsmtsc: (multi-)target-dependent sentiment classification in news articles," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, Apr. 2021, pp. 1663–1675.

[19] H. Kwak, J. An, E. Jing, and Y.-Y. Ahn, "FrameAxis: Characterizing Framing Bias and Intensity with Word Embedding," pp. 1–24, 2 2020. [Online]. Available: http://arxiv.org/abs/2002.08608

[20] H.-K. Kong, Z. Liu, and K. Karahalios, "Frames and Slants in Titles of Visualizations on Controversial Topics," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 4 2018, pp. 1–12. [Online]. Available: https://dl.acm.org/doi/10.1145/3173574.3174012

[21] F. Hamborg, A. Zhukova, and B. Gipp, "Automated Identification of Media Bias by Word Choice and Labeling in News Articles," in *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. Wuhan, CN: IEEE, 6 2019, pp. 196–205. [Online]. Available: https://ieeexplore.ieee.org/document/8791197/

[22] F. Hamborg, A. Zhukova, K. Donnay, and B. Gipp, "Newsalyze: Enabling News Consumers to Understand Media Bias," in *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. New York, NY, USA: ACM, 8 2020, pp. 455–456. [Online]. Available: https://dl.acm.org/doi/10.1145/3383583.3398561

[23] F. Hamborg, N. Meuschke, C. Breitinger, and B. Gipp, "news-please: A Generic News Crawler and Extractor," in *Proceedings of the 15th International Symposium of Information Science*. Verlag Werner Hülsbusch, 2017, pp. 218–223.

[24] K. Clark and C. D. Manning, "Deep Reinforcement Learning for Mention-Ranking Coreference Models," *Empirical Methods on Natural Language Processing*, 9 2016. [Online]. Available: http://arxiv.org/abs/1609.08667

[25] ——, "Improving Coreference Resolution by Learning Entity-Level Distributed Representations," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2016, pp. 643–653. [Online]. Available: http://aclweb.org/anthology/P16-1061

[26] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky, "The Stanford CoreNLP Natural Language Processing Toolkit," in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2014, pp. 55–60. [Online]. Available: http://aclweb.org/anthology/P14-5010

[27] F. Hamborg, "Media Bias, the Social Sciences, and NLP: Automating Frame Analyses to Identify Bias by Word Choice and Labeling," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 79–87. [Online]. Available: https://www.aclweb.org/anthology/2020.acl-srw.12

[28] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," in *Proceedings of the 1st International Conference on Learning Representations (ICRL 2013)*, 2013, pp. 1–12. [Online]. Available: http://arxiv.org/abs/1301.3781

[29] D. Christian, P. Froke, S. Jacobsen, and D. Minthorn, *The Associated Press stylebook and briefing on media law*. The Associated Press, 2014.

[30] J. Hainmueller, D. J. Hopkins, and T. Yamamoto, "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments," *Political Analysis*, vol. 22, no. 1, pp. 1–30, 1 2014. [Online]. Available: https://www.cambridge.org/core/product/identifier/S1047198700013589/type/journal_article

[31] R. D. Luce and J. W. Tukey, "Simultaneous conjoint measurement: A new type of fundamental measurement," *Journal of Mathematical Psychology*, vol. 1, no. 1, pp. 1–27, 1964.

[32] N. M. Razali and B. W. Yap, "Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov , Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 21–33, 2011.

[33] C. Huff and D. Tingley, ""Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents," *Research & Politics*, vol. 2, no. 3, pp. 1–12,

8 2015. [Online]. Available: http://journals.sagepub.com/doi/10.1177/2053168015604648

[34] J. Jones, "U.S. Party Preferences Have Swung Sharply Toward Democrats," 2020. [Online]. Available: https://news.gallup.com/poll/315734/party-preferences-swung-sharply-toward-democrats.aspx

[35] S. Ronen and O. Shenkar, "Attitudinal and Behavioral Dimensions," in *Navigating Global Business*. Cambridge: Cambridge University Press, 2017, pp. 189–244. [Online]. Available: http://ebooks.cambridge.org/ref/id/CBO9781316107034A037

[36] J. Hainmueller, D. Hangartner, and T. Yamamoto, "Validating vignette and conjoint survey experiments against real-world behavior," *Proceedings of the National Academy of Sciences*, vol. 112, no. 8, pp. 2395–2400, 2 2015. [Online]. Available: http://www.pnas.org/lookup/doi/10.1073/pnas.1416587112

[37] K. A. Phillips, F. R. Johnson, and T. Maddala, "Measuring What People Value: A Comparison of "Attitude" and "Preference" Surveys," *Health Services Research*, vol. 37, no. 6, pp. 1659–1679, 12 2002. [Online]. Available: http://doi.wiley.com/10.1111/1475-6773.01116

[38] B. F. Wijnen, I. M. van der Putten, S. Groothuis, R. J. de Kinderen, C. Y. Noben, A. T. Paulus, B. L. Ramaekers, G. C. Vogel, and M. Hiligsmann, "Discrete-choice experiments versus rating scale exercises to evaluate the importance of attributes," *Expert Review of Pharmacoeconomics & Outcomes Research*, vol. 15, no. 4, pp. 721–728, 7 2015. [Online]. Available: http://www.tandfonline.com/doi/full/10.1586/14737167.2015.1033406

[39] F. Aust, B. Diedenhofen, S. Ullrich, and J. Musch, "Seriousness checks are useful to improve data validity in online research," *Behavior Research Methods*, vol. 45, no. 2, pp. 527–535, 6 2013. [Online]. Available: http://link.springer.com/10.3758/s13428-012-0265-2

[40] M. Gao, Z. Xiao, K. Karahalios, and W.-T. Fu, "To Label or Not to Label: The Effect of Stance and Credibility Labels on Readers' Selection and Perception of News Articles," *Proceedings of the ACM on Human-Computer Interaction*, vol. 2, no. CSCW, pp. 1–16, 11 2018. [Online]. Available: https://dl.acm.org/doi/10.1145/3274324

[41] F. Hamborg, C. Breitinger, and B. Gipp, "How to effectively identify and communicate person-targeting media bias in daily news consumption?" in *Proceedings of the 15th ACM Conference on Recommender Systems, 9th International Workshop on News Recommendation and Analytics (INRA 2021)*, Amsterdam, Netherlands, 2021, pp. 1–8.

[42] W. G. Cochran, "The Combination of Estimates from Different Experiments," *Biometrics*, vol. 10, no. 1, pp. 101–129, 3 1954. [Online]. Available: https://www.jstor.org/stable/3001666?origin=crossref

[43] R. F. Woolson, J. A. Bean, and P. B. Rojas, "Sample Size for Case-Control Studies Using Cochran's Statistic," *Biometrics*, vol. 42, no. 4, p. 927, 12 1986. [Online]. Available: https://www.jstor.org/stable/2530706?origin=crossref

[44] L. Feick, K. Donnay, and K. T. McCabe, "The Subconscious Effect of Subtle Media Bias on Perceptions of Terrorism," *American Politics Research*, vol. 49, no. 3, pp. 313–318, 5 2021. [Online]. Available: http://journals.sagepub.com/doi/10.1177/1532673X20972105

[45] B. Rogowitz and L. Treinish, "Data visualization: the end of the rainbow," *IEEE Spectrum*, vol. 35, no. 12, pp. 52–59, 12 1998. [Online]. Available: https://ieeexplore.ieee.org/document/736450/

[46] R. Schatz, S. Egger, and K. Masuch, "The Impact of Test Duration on User Fatigue and Reliability of Subjective Quality Ratings," *Journal of the Audio Engineering Society*, vol. 60, pp. 63–73, 2012. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=16167

[47] Pew Research Center, "In Changing U.S. Electorate, Race and Education Remain Stark Dividing Lines," Tech. Rep., 2020. [Online]. Available: https://www.pewresearch.org/politics/2020/06/02/in-changing-u-s-electorate-race-and-education-remain-stark-dividing-lines/

[48] R. M. Perloff, "A Three-Decade Retrospective on the Hostile Media Effect," *Mass Communication and Society*, vol. 18, no. 6, pp. 701–729, 11 2015. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/15205436.2015.1051234

[49] J. Yang, H. Rojas, M. Wojcieszak, T. Aalberg, S. Coen, J. Curran, K. Hayashi, S. Iyengar, P. K. Jones, G. Mazzoleni, S. Papathanassopoulos, J. W. Rhee, D. Rowe, S. Soroka, and R. Tiffen, "Why Are "Others" So Polarized? Perceived Political Polarization and Media Use in 10 Countries," *Journal of Computer-Mediated Communication*, vol. 21, no. 5, pp. 349–367, 9 2016. [Online]. Available: https://academic.oup.com/jcmc/article/21/5/349-367/4161799

Please use the following BibTeX code to cite this paper:

```
@InProceedings{Hamborg2021a,
  title = {Newsalyze: Effective Communication of Person-Targeting Biases in News Articles},
  booktitle = {Proceedings of the {ACM}/{IEEE} {Joint} {Conference} on {Digital} {Libraries}
      ({JCDL})},
  author = {Hamborg, Felix and Heinser, Kim and Zhukova, Anastasia and Donnay, Karsten and
      Gipp, Bela},
  year = {2021},
  month = {Sep.},
  topic = {newsanalysis},
}
```