Social Media Spillover: Attitude-Inconsistent Tweets Reduce Memory for Subsequent Information

Reese Butterfuss (rbutter2@asu.edu)
Tracy Arner (tarnerr@asu.edu)
Danielle S. McNamara (dsmcnama@asu.edu)
Department of Psychology, 1000 S. Forest Mall
Tempe, AZ 85281

Laura K. Allen (Laura.Allen@unh.edu)
Department of Psychology, Murkland Hall
Durham, NH 03824

Abstract

Social media users are generally exposed to information that is predominantly consistent with their attitudes and beliefs (i.e., filter bubbles), which can increase polarization and decrease understanding of complex and controversial topics. One potential approach to mitigating the negative consequences of filter bubbles is intentional exposure to information that is inconsistent with attitudes. However, it is unclear how exposure to attitude-inconsistent information in social media contexts influences memory for controversial information. To fill this gap, this study examines the effects of presenting participants (n = 96) with Twitter content on a controversial topic (i.e., labor unions) that was either pro-union, anti-union, or neutral. Participants then read a media article including both pro-union and anti-union information. Participants who saw Twitter content that was inconsistent with their prior attitudes regarding labor unions recalled less of the article content compared to those who saw Twitter content that was consistent with their prior attitudes. The findings suggest that Twitter users’ memory for information related to controversial topics may not benefit from exposure to messages outside their filter bubble.

Keywords: Social media; filter bubble; comprehension; memory

Introduction

Social media environments afford users with unprecedented access to information about a variety of topics, ranging from the mundane to the highly controversial. Because of this often-overwhelming breadth of content, information ecosystems are becoming increasingly personalized such that users are exposed to only a subset of the available information (Zuiderveen Borgesius et al., 2016). Although this personalization may lead to increased engagement by users, it may also have negative consequences, particularly with respect to learning new information regarding complex topics covered on social media.

Prior work on discourse comprehension suggests that as individuals read information about a topic, new information reactivates related prior knowledge (Myers & O’Brien, 1998). The newly encoded information and reactivated prior knowledge are then integrated into an interconnected network, which comprises the individual’s memory representation of the topic (Kintsch, 1988). More coherent memory representations serve as the basis for effective subsequent learning about the topic.

Social media users are often sequestered into filter bubbles in which their stream of content consists of information that is consistent with their attitudes (Pariser, 2011; Zuiderveen Borgesius et al., 2016). As such, many individuals engage with only one side of complex and controversial issues (Kendeou et al., 2020). This intellectual isolation reduces individuals’ understanding of complex issues and may hinder subsequent learning, which in turn increases polarization (Resnick et al., 2013).

Existing research has offered potential approaches to mitigating the polarizing effect of filter bubbles on users’ understanding of controversial issues. One recommendation is for users to diversify their information exposure by deliberately seeking content from outside their filter bubble (Kendeou et al., 2020; Resnick et al., 2013). For example, a Twitter user who believes labor unions are unnecessary or detrimental could intentionally seek information (e.g., opinions, arguments, or links to other information sources) from the pages of Twitter users who believe that labor unions are beneficial (i.e., individuals outside of the Twitter user’s filter bubble). Whether exposure to information from outside one’s filter bubble is beneficial or detrimental may depend heavily on users’ pre-existing attitudes.

Attitudes are likely to be reactivated via exposure to social media content and may remain accessible even after exposure (Voss et al., 1993) and “spill over” to influence processing and memory of topic-relevant information that users subsequently encounter (Osatuyi, 2013). However, we lack an understanding of the potential positive or negative consequences of exposure to attitude-inconsistent information on users’ memory and learning of subsequent information.

In terms of positive consequences, exposure to attitude-inconsistent information about controversial topics may promote construction of a richer, more comprehensive understanding. However, this may only be the case for individuals who do not endorse particularly strong attitudes or beliefs about the topic. Such individuals may be less prone to biased processing and therefore less likely to endorse one
side of the issue and reject the other (Maier et al., 2018). In other words, individuals who are neutral about the topic may encode and integrate both sides of a controversial topic and therefore construct a more coherent mental representation that enables understanding and recall of both sides.

In terms of negative consequences, exposing individuals who already hold polarized attitudes and beliefs to attitude-inconsistent information about a controversial issue may “backfire” and elicit rejection of subsequent attitude-inconsistent information, which may further polarize their original stance on the topic (Nyhan & Reifler, 2010). When individuals reactivate strong attitudes, they tend to show biased processing such that they pay more attention to attitude-consistent information (Voss et al., 1993) and therefore better understand and more readily recall that information (i.e., text-belief consistency effect; Maier & Richter, 2013).

In sum, social media users who encounter information that conflicts with their attitudes may show biased processing of related information they subsequently access such that they attend more to attitude-consistent content. Consequently, their memory representation of the controversial information may consist of more attitude-consistent information than attitude-inconsistent information.

Current Study

The current study addressed the following question: to what extent does the consistency between social media content and users’ attitudes influence their memory for subsequent information? To this end, we simulated polarized social media filter bubble about a controversial topic (i.e., labor unions). Specifically, participants who varied in attitudes toward unions viewed a thread of tweets that were either anti-labor union, pro-labor union, or neutral. Afterward, they read a media article that presented pro-union and anti-union information and completed two recall measures.

We hypothesized that participants would recall less article content when they were primed with an attitude-inconsistent Twitter thread. Specifically, when participants who reported negative attitudes towards unions were first primed with pro-union Twitter messages, we expected that they would show worse memory for article content than when they were primed with either neutral or anti-union Twitter messages. Likewise, when participants who reported positive attitudes toward unions were first primed with anti-union Twitter messages, we expected that they would show worse memory than when primed with either pro-union or neutral Twitter messages. For participants who did not hold particularly strong attitudes towards unions, we anticipated relatively little consequence of priming in any condition on their memory for article content. To the best of our knowledge, no prior work has investigated how polarized social media content interacts with users’ attitudes about the topic to influence memory for subsequent content. The present study serves as a first step to address this gap in the literature.

Method

Participants

Data were collected from $n = 183$ US adults via Amazon MTurk. After excluding participants whose reading times indicated poor attention to the stimuli and/or media article (exclusion criteria detailed below), the final sample consisted of $n = 96$ adults (47 female, 48 male; 1 non-binary; $M_{age} = 39$, $SD = 11$ years). Fifty percent of the sample reported earning a 4-year college degree, 28% reported earning a graduate degree, 18% reported attending some college, and 3% reported earning a high-school diploma. The sample was 80% White, 7% Black, 3% Hispanic, 2% Asian, and 7% other.

Materials

Twitter Threads Participants were presented with a Twitter thread composed of six Tweets about labor unions involving three hypothetical Twitter users. Participants saw the Tweet thread in one of three polarity conditions: pro-union (170 words), anti-union (164 words), or neutral (150 words). Across the three conditions, the Tweets were from the same Twitter users and were comparable in length and difficulty. The final Tweet in each Twitter thread included a link to an article about unions that participants subsequently read. The Tweet thread was presented to participants as a single screenshot of the Twitter page embedded into Qualtrics survey software. Overall reading time on the Tweet thread was recorded. Data from $n = 47$ participants who did not spend at least 10 seconds reading the Tweet thread were excluded from analyses, as it is unlikely that they attended to the information (Brysbaert, 2019).

Media Article The article about labor unions was adapted from original content available from The Perspective, a news source that aims to provide balanced arguments about controversial issues. The original article presented a brief introduction to unions (90 words), followed by three anti-union arguments and three pro-union arguments. The language of the original article was modified to equate the three pro-union and three anti-union arguments for length, referential cohesion, and grade level. See Table I for values derived from Coh-Metrix Common Core Text Ease and Readability Assessor (T.E.R.A.; Jackson, Allen, & McNamara, 2016). We also added a neutral conclusion that was neither pro- nor anti-union (150 words). The presentation order of the three pro-union arguments and three anti-union arguments was counterbalanced, such that half saw the anti-union arguments first and half saw the pro-union arguments first. The article consisted of 757 words in total. Overall reading time on the article was recorded. Data from an additional $n = 40$ participants were omitted from analyses because their reading rates exceeded superior reading speed (i.e., 320 words per minute, Brysbaert, 2019), indicating that they did not attend to the article.
Measures

Free Recall Task Participants’ memory of the pro- and anti-union arguments in the media article was assessed with a free recall prompt to type everything they could remember from the article. Participants were told they could put things in their own words. Free-recall responses were analyzed for source overlap with the article participants read via the Tool for the Automatic Analysis of Cohesion (TAACO 2.0; Crossley, Kyle, & Dascalu, 2019). In particular, TAACO assesses text overlap in terms of semantic similarity and keywords between the participants’ free-recall responses and the article participants read. In the current study, the source overlap index reported by word2vec (Mikolov et al., 2013) was included in analyses. Overall, source overlap provides an indicator of the extent to which information from the article was represented in participants’ free-recall responses.

Sentence Verification Technique Participants’ memory of the pro- and anti-union arguments in the media article was also assessed with a 20-item sentence verification technique test (SVT; Royer, 2001). The Sentence Verification Technique evaluates the quality of the reader’s mental representation of the article rather than the exact recall of the text. The Sentence Verification Technique is supported by the construction integration theory suggesting that readers integrate text elements with their prior knowledge to construct a mental representation of the content (Kintsch, 1988; Royer, 2001). The SVT task consisted of four item types: five original sentences, five paraphrase sentences, five “meaning-change” sentences, and five distractor sentences. The original sentences and paraphrase sentences required a “yes” response to indicate that the sentence was either directly from the article or had the same meaning as a sentence in the article. The meaning-change and distractor items required a “no” response to indicate that the sentence had a different meaning from any sentence in the article. Participants’ sum scores were number of correct responses out of 20. Internal consistency of scores was acceptable (Cronbach’s $\alpha = .60$). For an exploratory analysis, we also calculated scores for items corresponding to the pro-union article content (7 items) and for the items corresponding to the anti-union article content (8 items). Note that the five distractor items were not included in the exploratory analysis.

General Union Attitudes Scale Attitudes toward labor unions were assessed with the General Union Attitudes Scale (McShane, 1986). The General Union Attitudes Scale is a unidimensional measure of attitudes toward unions consisting of eight Likert-type items (1 = strongly disagree to 7 = strongly agree). Participants’ sum scores were included in analyses. Internal consistency of scores was high (Cronbach’s $\alpha = .85$).

Procedure

Participants were randomly assigned to view a pro-union, anti-union, or neutral Tweet thread via Qualtrics survey software. Following the Twitter thread, participants read the news article with the order of the pro-union and anti-union arguments presented in a counterbalanced fashion across participants. Following the article, participants responded to the free-recall prompt. Next, participants completed the SVT, followed by the General Union Attitudes Scale. Finally, participants completed a demographics questionnaire.

Results

We first examined correlations among key variables (see Table 2 for descriptive statistics). Attitudes toward unions (sum scores on the General Union Attitudes scale) were significantly correlated with both source overlap ($r = .33, p < .001$) and SVT sum scores ($r = .31, p = .002$). Source overlap was not significantly correlated with SVT score ($r = .17, p = .09$). See Table 2 for descriptive statistics.

<table>
<thead>
<tr>
<th>Table 2: Descriptive statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Recall length</td>
</tr>
<tr>
<td>Source overlap</td>
</tr>
<tr>
<td>SVT scores</td>
</tr>
<tr>
<td>General Union Attitudes</td>
</tr>
</tbody>
</table>

Recall Length A mixed ANOVA with Twitter condition (pro-union vs. anti-union vs. neutral) as a between-subjects factor, attitudes towards unions as a continuous independent variable, and recall length (word count calculated by SiNL; Crossley et al., 2014) as the dependent variable was conducted to examine the extent to which exposure to polarized Twitter threads influenced the amount of article content participants recalled. Results revealed a main effect of attitudes, $F(1, 90) = 12.84, p < .001, \eta^2_p = .16$, such that participants with more positive attitudes tended to produce longer recall responses. This main effect was qualified with a Twitter polarity X attitude interaction, $F(2, 90) = 9.40, p < .001, \eta^2_p = .17$ (see Figure 1).

Figure 1: Recall as a function of Twitter polarity and attitudes
To understand how participants’ attitudes influenced the effect of Twitter polarity on the amount of article content they recalled, we estimated the mean recall length in each polarity condition at 1 SD above the mean attitude score (45.4), 1 SD below the mean attitude score (29.5), as well as at the mean attitude (37.4). For participants with negative (anti-union) attitudes, recall length in the anti-union Twitter condition (M = 70.8, SE = 7.8) was greater than in the pro-union Twitter condition (M = 42.3, SE = 8.3), p = .014, but not different from the neutral condition (M = 66.2, SE = 9.7), p = .72; additionally, recall length was marginally greater in the neutral condition than in the pro-union Twitter condition (p = .064). For participants with moderate attitudes, recall length in the pro-union Twitter condition (M = 73.1, SE = 5.7), neutral Twitter condition (M = 76.7, SE = 6.6), and anti-union Twitter condition (M = 67.6, SE = 5.9) were not significantly different (p > .30). For participants with positive attitudes, recall length in the pro-union Twitter condition (M = 103.9, SE = 7.6) was greater than in the anti-union Twitter condition (M = 64.5, SE = 8.8), p = .001, but not different from the neutral condition (M = 87.2, SE = 9.4), p = .17; recall length in the neutral condition was marginally greater than in the anti-union Twitter condition (p = .082).

**Source Overlap** A mixed ANOVA was used to examine the extent to which exposure to polarized Twitter threads influenced participants’ recall of article content. Twitter condition (positive vs. negative vs. neutral) was a between-subjects factor, attitudes towards unions was a continuous independent variable, and source overlap was the dependent variable. Results revealed a main effect of attitudes, F(2, 90) = 13.22, p < .001, η² = .13, such that participants with more positive (pro-union) attitudes showed greater overlap between their free-recall response and article content than those with negative (anti-union) attitudes (see Figure 2).

![Figure 2](image.png)  
**Figure 2**: Relation between attitudes and source overlap of recall responses

Results also revealed a marginal main effect of Twitter polarity, F(2, 90) = 2.54, p = .085, η² = .05. Pairwise comparisons indicated that similarity between free-recall responses and article content was marginally greater for participants who saw the pro-union Twitter thread (M = .85, SE = .02) compared to those who saw the neutral Twitter thread (M = .80, SE = .02), p = .088, neither of which differed from participants who saw the anti-union Twitter thread (M = .83, SE = .02), p > .35.

**Content Recognition** A mixed ANOVA was used to examine the extent to which exposure to polarized Twitter threads influenced participants’ recognition of article content. Twitter condition (positive vs. negative vs. neutral) was a between-subjects factor, attitudes towards unions was a continuous predictor, and SVT score was the dependent variable. Results revealed a main effect of attitudes, F(1, 90) = 8.64, p = .004, η² = .09, such that participants with more positive (pro-union) attitudes performed better on the SVT than those with more negative (anti-union) attitudes, regardless of the Twitter thread they saw. This main effect was qualified by a Twitter polarity X attitude interaction, F(2, 90) = 5.92, p = .04, η² = .12 (see Figure 3). To understand how participants’ attitudes influenced the effect of Twitter polarity on participants’ SVT scores, we estimated the mean SVT score in each Twitter polarity condition at 1 SD above the mean attitude score (45.4), 1 SD below the mean attitude score (29.5), as well as at the mean attitude (37.4). For participants with positive attitudes, SVT scores were higher in the anti-union condition (M = 12.37, SE = .65) than in the pro-union Twitter condition (M = 9.92, SE = .70), p = .012 but were not different from the neutral condition (M = 12.00, SE = .82), p = .72. SVT scores of participants in the neutral Twitter condition did not differ from those in the anti-union Twitter condition (p = .44).

![Figure 3](image.png)  
**Figure 3**: Memory for article content as a function of Twitter polarity and attitudes

Participants with positive attitudes (pro-union) in the pro-union Twitter condition had significantly higher SVT scores (M = 14.29, SE = .64) than participants in the anti-union Twitter condition (M = 12.20, SE = .74), p = .035, but were not different from the neutral condition (M = 13.04, SE = .79), p = .22. SVT scores of participants in the pro-union Twitter condition were marginally lower than those of participants in the neutral condition (p = .057). Finally, the
SVT scores of participants with moderate attitudes, who saw the anti-union Twitter thread (M = 12.28, SE = .49), the pro-union Twitter thread (M = 12.11, SE = .48), or the neutral Twitter thread (M = 12.52, SE = .56) did not differ (ps > .50). Additionally, as an exploratory analysis, we examined whether participants had better recognition of content from the pro-union side or anti-union side of the article depending on their attitudes and Twitter polarity. To do so, we isolated participants’ performance on the SVT items corresponding to the pro-union content and SVT items corresponding to the anti-union content. We conducted a mixed ANOVA with SVT side (pro-union items vs. anti-union items) as a within-subjects factor, Twitter polarity (pro-union vs. anti-union vs. neutral) as a between-subjects factor, attitudes as a continuous independent variable, and SVT scores on the pro-union and anti-union items as the dependent variable. The results revealed a main effect of attitudes, F(1, 90) = 7.30, p = .008, η²p = .08, such that participants with more positive attitudes tended to perform better overall than participants with more negative attitudes. This main effect was qualified by a Twitter polarity x attitudes interaction, F(2, 90) = 6.00, p = .004, η²p = .12, such that participants with more positive attitudes performed worse overall when they were primed with pro-union Tweets, whereas participants with more positive attitudes performed better when they were primed with anti-union Tweets. Finally, there was a SVT side x attitudes interaction, F(1, 90) = 4.05, p = .047, η²p = .04, such that participants with more positive attitudes performed better on the SVT items corresponding to the pro-union article content that did attitudes with more negative attitudes about unions (see Figure 4). The SVT side x Twitter polarity x attitudes interaction did not approach significance (F = 1.06, p = .35).

Participants who held negative attitudes towards unions recalled less article content, as indicated by both relatively shorter recalls and lower SVT scores, after they were primed with pro-union Twitter messages. Conversely, participants who held positive attitudes recalled less article content after they were primed with anti-union Twitter messages. Thus, the polarized messages in the Twitter thread may have primed participants’ attitude about the topic, which then “spilled over” and biased processing of the subsequent article in favor of attitude-consistent information. This biased processing may have led to poorer integration of information into the participants’ mental representations, thereby reducing recall.

The finding that participants constructed inferior memory representations after they were primed with attitude-inconsistent information has implications for social media contexts. Namely, the finding suggests that users’ understanding of controversial topics may not benefit from exposure to messages outside their filter bubble. Indeed, as the current experiment shows, being primed with attitude-inconsistent social media content may be relatively detrimental to their memory for related information.

Additionally, the extent to which participants recalled pro-union and anti-union article information, as measured by performance on SVT items corresponding to pro-union or anti-union article content, depended on their attitudes. Specifically, participants with more positive attitudes toward unions recalled more pro-union information compared to participants with more negative attitudes, who performed slightly better on the anti-union questions. This finding supports the idea that participants’ attitudes biased their processing of the article content in favor of attitude-consistent information. Twitter polarity did not influence the effect of participants’ attitudes on their recognition of pro-union or anti-union information. Thus, it may be the case that Twitter polarity may have influenced overall performance, whereas participants’ attitudes biased processing of specific article content.

We also found that participants who did not hold particularly strong pro-union or anti-union attitudes had similar recall of article content regardless of the Twitter messages with which they were primed. It is possible that these participants were not as susceptible to priming from the Twitter messages simply because they lacked strong attitudes in the first place. Any reactivation of attitudes that may have occurred in the context of reading the Twitter messages did

**Discussion**

The current study examined the extent to which the consistency between social media content and participants’ attitudes influenced their memory for subsequent article content. We were particularly interested in the effects of encountering information outside of one’s filter bubble (i.e., attitude-inconsistent information). We hypothesized that participants with strong attitudes would recall less article content when they were primed with attitude-inconsistent Twitter messages (corresponding to information outside their filter bubble) compared to when they were primed with attitude-consistent or attitude-neutral Twitter messages (corresponding to information inside their filter bubble—the current norm on social media). We also hypothesized that participants with relatively neutral attitudes would show similar memory for article content regardless of Twitter condition. The results are consistent with both of these hypotheses across multiple indicators of memory for article content.

![Figure 4: Memory for pro-union and anti-union article content as a function of attitudes](image-url)

**Figure 4:** Memory for pro-union and anti-union article content as a function of attitudes
not appear to spill over into processing of the subsequent article. It may also be the case that social media users who do not have strong attitudes about controversial topics may be less likely to sequester themselves in filter bubbles. Thus, encountering information that represents both sides of a controversy may have little effect. Taken together, the results suggest that the effects of exposure to information outside of users’ filter bubbles on participants’ memory for information may depend on the strength of their attitudes towards the topic.

It is important to note that the current experiment cannot directly speak to whether priming participants with polarized Twitter messages influenced their understanding of the article content. This is because our dependent variables targeted memory for content rather than understanding. Although memory for information and understanding are related (Kintsch, 1988), subsequent work must examine the extent to which the current conditions influence participants’ deeper understanding of the article content.

One limitation of the current experiment is that the current work did not measure participants’ prior knowledge about unions. It is possible that participants who held strong attitudes towards unions also possessed a more coherent prior knowledge base. A more coherent knowledge base facilitates both learning from text and subsequent recall (Haenggi & Perfetti, 1994). The extent to which the Twitter messages would reactivate prior knowledge and, in turn, facilitate understanding of the article content over and above attitudes is unclear. Thus, future work should examine how attitudes and knowledge interact to influence understanding in the context of social media. Doing so would strengthen the contribution of subsequent work to existing discourse theory.

A second limitation of the current work is the rather small sample size. Consequently, our ability to detect small effects and complex interactions was limited due to low statistical power. To address this, it is critical to replicate these effects with a larger sample size based on an a priori power analysis.

A third limitation is that the current Twitter environment was relatively inauthentic. Namely, the Tweet thread consisted of only six short messages in a hypothetical Twitter thread. Although a thread of six messages is hardly akin to the filter bubbles that occur in authentic social media contexts, this experiment represents a first step towards understanding the consequences of filter bubbles on users’ memory and learning. One could expect the effects observed in the current experiment to be drastically amplified by more substantial simulations of filter bubbles in the context of more hotly contested topics on social media (e.g., abortion, drug legalization, gun control). Also, future research should leverage additional features of Twitter with which users routinely engage. For example, follow-up studies could examine the influence of attitudes and Twitter polarity on participants’ decisions to retweet, like, and respond to Tweets they encounter.

A fourth limitation is that the pro-union and anti-union article content shared extremely high semantic similarity (latent semantic analysis cosine = .96; Mikolov et al., 2013). Due to the semantic relatedness, it was not feasible to examine whether biased processing of the article content manifested in participants’ free recall responses based on assessments of overlap with the pro-union and anti-union article content. The high degree of semantic overlap may be a consequence of carefully controlling the article content to evenly present pro-union and anti-union content. This was necessary to examine the extent to which polarized attitudes influenced recall of content. However, the content that users access from mainstream media sources is unlikely to provide such an even presentation. Using more authentic article content may allow for natural variability that improves ability to detect differences in recall across Twitter conditions.

Overall, future work should examine how attitudes and social media influence learning of authentic, biased media content from sources that vary in credibility and partisanship. To further aid generalizability, the limitations of the current work should also be replicated in other controversial domains that permeate social media, such as gun control, police reform, and vaccine mandates.

Conclusion

Social media users are inundated with vast amounts of information on numerous complex and controversial topics. In an effort to personalize the flow of information, they often produce (or become victims of) filter bubbles that contain only content that is consistent with their attitudes about the topic. The priming that occurs from a users’ filter bubbles may affect processing of subsequent information. There is currently a lack of research investigating how this social media phenomenon impacts knowledge construction.

Results of the current experiment indicate that even brief exposure to attitude-inconsistent social media content can negatively influence individuals’ memory for related content. This is critical because the consequences of social media on users’ information streams do not stop when they close social media. Thus, future work must continue to examine interactions amongst attitudes and social media information exposure on learning and memory in everyday contexts.

Acknowledgments

The research reported here was supported by the Office of Naval Research, through Grant N00014-19-1-2424. The opinions expressed are those of the authors and do not represent views of the Office of Naval Research.

References


semantic similarity and text overlap. *Behavior Research Methods, 51*(1), 14-27.


