**Response Letter**

Paper Title: News ‘Attraction’ and Digital Inequalities: Incidental News Exposure and the Equalization or Stratification of Political Information

ID: 223259527

**Editor’s Comments**

**Comment:** We have now received the reviewers’ reports on your paper “News ‘Attraction’ and Digital Inequalities: Incidental News Exposure and the Equalization or Stratification of Political Information." Your paper has more reviews than the usual two. We initially received divergent reviews, which led as to reach out another two reviewers to make a robust assessment of your piece. That said, we apologize for the long wait this paper had to go through.  
  
In general, the reviewers had positive Comments about the manuscript, but there is still work to do for this piece to be ready for publication. As such, we are offering you the opportunity to revise and resubmit your paper addressing the suggestions made by the reviewers as well as the Digital Journalism editorial team.

Reviewer 1 invites you to better visualize your results, while Reviewer 2 would like you to reflect on some concepts and measures in the study. Reviewer 3 requests a better justification for presenting competing hypotheses, as well as stronger arguments to back up some conceptual definitions (such as news engagement or news attraction). This reviewer also has many methodological observations. Reviewer 4 would like you to discuss the context of your study, and invites you to provide more information for the methods section regarding sampling, variables, and models. This reviewer also provides helpful insight to strengthen your discussion section.

**Response**: Thank you for facilitating the review process and for providing your own critical feedback on the paper. We’ve made changes through the paper in response to the reviewers’ suggestions. Most notably, we’ve refined conceptualization and strengthened arguments in the front end of the paper, we’ve clarified methodological choices, and we’ve streamlined the analysis and results, and we’ve updated the discussion to integrate important issues raised by the reviewers. We think these changes have substantively improved the paper. Thank you again for your time and attention to this manuscript. We’re very excited about the possibility of published with Digital Journalism.

**Comment:** In addition, the Digital Journalism editorial team would like you to respond to the following:

Sample and Data

The study relies on a cross-sectional survey of adult social media users in the US, but there is no explanation as of how these users were recruited. Did you hire a polling company, perhaps?

**Response**: Thank you for catching this; it was an oversight on our part. We used Qualtrics, and the sample was balanced for age, race, gender, and census region. We added this important detail to the manuscript under the Survey Design and Sample sections on pages 10-11.

**Comment:** Also, the paper indicates that “survey responses were linked with social media content collected via Brandwatch (formerly Crimson Hexagon) and then validated by crosschecking content lists with CrowdTangle.” It’s not clear why survey responses needed to be validated, and how (what is that you validated, exactly?).

**Response**: This comment from the editors aligns with similar concerns from reviewers. Our description of key elements of the study design wasn’t entirely clear, and this lack of clarity lead to some confusion about the nature of the rolling cross-section (RCS) survey and the ‘linkage’ between respondents and social media content.

We have clarified our explanation of the study design in the Methods section to address these points on pgs. XX-XX.

The survey responses themselves are not ‘validated.’ Rather, we implemented a multi-step strategy to identify and verify the top news circulating on Facebook at the time of data collection.

We then embedded the top two news articles circulating on Facebook at the time (ranked by engagement metrics) into the survey as a cue for attention to news during the previous news cycle. The assumption underlying this approach is that posts with higher engagement are more likely to show up in people’s feeds, and therefore respondents who follow news only passively are more likely to have seen it. To determine the ‘top’ news organizations, as well as the most widely circulated/top stories, we employed public and private ranking systems from two third-party data brokers.

First, we used NewsWhip to identify the top 25 U.S. news organizations on Facebook (NewsWhip regularly releases reports to the public). Next, we used Brandwatch (formerly Crimson Hexagon) to identify the top posts on Facebook based on engagement from those 25 outlets, and finally, we cross-checked these stories on Facebook’s proprietary platform, CrowdTangle, in order to verify that those stories were indeed the most widely circulated on Facebook during four days prior to data collection.

The main benefit of this approach is that we can a) identify the top stories widely circulating on Facebook with a higher level of certainty than simply relying on a single data provider or external source, and b) we can better take advantage of the RCS design. RCS designs are useful, because the survey responses can be more closely tied to media content. We followed the lead of previous published work in this area (see, De Vreese et al., 2017 for an overview and past uses of RCS to link media content with survey data).

**Comment:** In addition, validating survey responses with social media data requires knowing the respondents’ usernames. Does this violate IRB concerns about anonymity/confidentiality? Please explain.

**Response**: No personally identifying information was collected from individual participants and we did not have access to respondents’ social media accounts. As outlined in the previous response, our discussion of the RCS and nature of the linkage study have been revised for clarity. We specifically address this concern in the revised manuscript. We note that no personally indefinable information was collected on page XX.

**Comment:** Is there any reason to measure age as intervals instead of a continuous ratio variable?

**Response**: We made a conscious decision to capture age in this manner. Ultimately, we priorities ease of response over precision, as the demographic questions appeared at the end of the survey questionnaire when survey fatigue may affect response quality. In designing the questionnaire, we ultimately decided that ratio-level measurement was not required to answer our research questions. While we agree that ordinal variables can make interpretation of statistical results less straight-forward, we think the measure adequately controls for differences in age groups.

**Comment:** Why did you impute missing values in a 2,000-case sample? Did you have too many missing values? How big was your sample if you didn’t input missing data?

**Response**: Thank you for raising this question. The sample size for our models before imputation is N = 1,731. We’ve included this information in the manuscript on pages 11-12.

We decided to impute data not for the purpose of increasing statistical power via a larger N, but rather to protect against one of the drawbacks of the RCS design.

RCS designs typically have more waves than a typical survey (17 in our case), but there are fewer respondents per wave. Thus, small variations in sample size could potentially bias statistical results via sampling bias (i.e., underrepresentation on one or more of the quotas). The imputation technique minimizes this bias. By using the partial responses filled in with imputed data, we minimize study mortality and the threat of sampling bias.

Based on an analysis of missingness patterns, we did not see any issue with data quality in this regard. The missingness ranges from 6.3% (Age) to < 1% (Ideology and Party ID), and only 6 of the 28 modelled variables showed missing cases.

**Comment:** Exposure and Engagement; The methods section describes ‘trait-like’ and ‘state-like’ properties of news exposure, but none of this is explained in the front end, let alone the most adequate way to measure these properties. For instance, total exposure to political information was measured with Weeks et al., 2017’s selective exposure items, but it’s not clear why ‘trait-like’ properties should be measured as selective exposure. Please address the ‘trait-like’ and ‘state-like’ concepts in the lit review and explain the rationale to measure them.

**Response**: Thank you for raising this issue, as we view our theorizing and measurement of incidental exposure on both trait- and state-levels to be an important contribution. We have now offered clearer definitions of trait and state incidental exposure early in the paper (p. 4) and more consistently integrated this distinction throughout the introduction. These changes offer a stronger rationale for why we measure incidental exposure on these two levels (see p. 9).

**Comment:** Please indicate if you ran a factor analysis to create the high-effort engagement variable, to make sure the items loaded together.

**Response**: We had conceived of the high-effort variable as a conceptual robustness check, and therefore did not previously perform a factor analysis. We did perform an EFA for the revision, and we found that the items load onto a single factor. After much deliberation, we decided to drop the high-effort variable from the paper entirely. We acknowledge that there are multiple possible solutions for this issue, and we would be happy to approach it differently at the suggestion of the editorial team.

**Comment:** Controls; What are the theoretical reasons to control for political ideology, party identity, and identity strength? These variables are significant in most of the models, so there is clearly something going on (especially if you are studying exposure to political information) but these significant findings are not discussed in the paper. You might want to give it some thought.

**Response**: We feel that these are important variables to include as controls, because in the U.S. they tend to be associated with patterns of digital news consumption. For example, party identity has been linked to selective exposure and homophilous engagement on social media platforms. Party identity is also featured in incidental exposure studies via the news attraction concept introduced by Thorson (as partisans may be more interested in news, as cited in the paper), as well as incidental exposure to attitude-consistent political information. In the U.S., party identity and ideology tend to be closely aligned, and there is ambiguity in the literature about which variable is more important and which should feature more prominently in theories and models of news consumption. To be safe, we included both variables (we tested for multicollinearity and found no evidence of variance inflation). That said, we would be happy to remove one variable or the other at the suggestion of the editorial team.

To address this comment, we added a justification for including these variables to the manuscript on page 16. In addition, we noted these findings in the Discussion section on pages 26 and 27.   
  
**Comment:** Limitations; Finally, you mention self-reported measures as a limitation, pointing out this is an endemic issue to survey research. However, some of the self-reported measures in this study are particularly problematic. For instance, how reliable is asking respondents whether they clicked on a story, scanned the headline, or read it entirely, Commented, discussed, etc.? These actions are not easy to remember. The same might happen with network size (is this something you could cross validate with Brandwatch or CrowdTangle?). Please elaborate on this issue.

**Response**: We agree that survey measures of digital news consumption suffer from measurement error arising from poor recall and/or impression-based response processes. But this is precisely why employed the RCS design, which assures temporal proximity between event and the reporting of the event (i.e., between exposure and survey response). This approach affords the possibility of asking about specific content, as the likelihood that respondents will remember that content is much higher than in the typical survey. This method offers more specificity and a closer connection to real-world content than typical survey measures. Respondents answered questions about a specific story known to be circulating immediately prior to data collection. Those who did not recall whether they had been exposed to the story were filtered out of the analyses. The approach minimizes measurement error related to inaccurate recall, and provides a level of external validity not found in the general measures because of its closer connection to real-world content.

Notably, the study also includes the general measures. Following King et al.’s (1994) advice to increase the observable implications of the theory in order to maximize leverage over a research problem, we intentionally included both sets of measures in the hopes that their comparative advantages would compensate for their comparative disadvantages. Specifically, the trait-like measures are more likely to be limited by poor recall (these measures typically underestimate news exposure; see Gonzalez-Bailon & Xenos, 2021), but they are better for capturing broad habits of consumption, at least as they are perceived by the individual. On the other hand, the state-like measures are unable to capture broad habits and patterns, but they are also less likely to be limited by poor recall. Thus, the disadvantages of the measures are, to some extent, cancelled out by examining both indicators.

Thank you again for your constructive feedback. Although RCS linkage studies are becoming more widely used in the field, this is a relatively recent development. We ourselves are still learning how to effectively communication about the important details, and we apologize for any confusion about the design in the previous version. We’d be happy to further clarify these design-related questions, as well as our decision-making process, in future rounds of review.

**Reviewer 1 Comments**

**Comment:** The way in which the findings are presented in this paper will contribute significantly to the development of the field of incidental news exposure. The results are well structured, and the methodology used is clear. The limitations of the study are also declared. It is suggested to incorporate tables that allow faster visualization of the results in the different variables used.

**Response**: Thank you for your time in reading and responding to the manuscript.

**Reviewer: 2**

**Comment:** This manuscript investigates whether the use of digital media reduces or exacerbates inequalities in news exposure and engagement based on survey data. The paper is very well-written and deals with an important topic. The literature review section is thorough and makes logical sense. The analyses are rigorous. Limitations (e.g., cross-sectional nature of the data) are also well-noted. Overall, I highly value this paper and recommend this paper be published in this journal after addressing the following concerns.  
  
**Response**: Thank you for the close reading of the paper and we appreciate your helpful comments.

**Comment:** While the author's way of measuring IE is more sophisticated compared to the previous way of measuring this concept which relied on a single item, the author(s) still need to acknowledge that it is still very difficult to measure IE with the survey. Survey respondents are not good at distinguishing to what extent their exposure was "accidental" or "purposeful." The author(s) can mention this point in the limitation section.

**Response**: This is an excellent point. We made several direct efforts to both employ and build upon existing measures of IE (Nanz & Matthes, 2022; Weeks & Lane, 2020). However, we agree that these concepts are imperfect and inherently difficult to capture. We added this note to the limitations section on page 25.

**Comment:** I don't think the author(s) explained what story they chose and why they chose this topic. I need more information.

**Response:** This comment relates to concerns raised by other reviewers as well as the editors. To address this point, we will simply paste the comment made to the editors above here:

“We lacked precision in our reporting of the study design. In turn, this lack of clarity lead to some ambiguity about other aspects of the study, including the nature of the rolling cross-section (RCS) survey and the ‘linkage’ between respondents and social media content.

We have heavily revised our explanation of the study design in the Methods section to address these points on pages 10-12.

To be clear, the survey responses are not validated. Rather, we implemented a multi-step strategy to identify and verify the top news circulating on Facebook at the time of data collection.

We then embedded the top two news articles circulating on Facebook at the time (ranked by engagement metrics) into the survey as a cue for attention to news during the previous news cycle. The assumption is that posts with higher engagement metrics would be more likely to show up in people’s feeds, and therefore a respondent who only passively follows the news should have seen it.

To decide what counted as ‘top’ news organization, as well as the most widely circulated/top stories, we developed a criterion that employs third-party data brokers’ public and private ranking systems.

First, we used NewsWhip to identify the top 25 news organizations on Facebook (NewsWhip regularly release reports to the public). Next, we used Brandwatch (formerly Crimson Hexagon) to identify the top posts on Facebook based on engagement from those 25 outlets, and finally, CrowdTangle was referenced to verify that those top stories were indeed widely circulating on Facebook during four days prior to data collection. The author has accounts with Brandwatch and CrowdTangle, so, while the news posts were publicly available, the rankings for news engagement are only available to researchers.

The main benefit of this approach is that we can a) ID the top stories widely circulating on Facebook with a higher level of certainty than simply relying on the researcher’s intuition, and b) we can better take advantage of the RCS design. RCS designs are useful, because survey responses can be more closely tied media campaigns and election cycles. We followed the lead of previous published work in this area (see, De Vreese et al., 2017 for an overview and past uses of RCS to link media content with survey data). This approach is now rather common in election studies.”

**Comment:** When measuring social media use, the author(s) used a single item. It is much more desirable to use multiple items when measuring this concept since a) different social media platforms have different functionalities and b) survey respondents may not take the term "social media" in the same way. For instance, some may think "YouTube" is not social media. Likewise, some may think of WhatsApp as social media (while some may not). So, the authors better use multiple items, or even if the authors end up using a single item, the authors need to at least add examples in the parenthesis such as "social media (e.g., x,y,z)."

**Response:** We agree with the premise of this comment and indeed rely on multiple-item measures, as well as more advanced reliability scales and techniques where appropriate (as evidenced by our approach to measurement of key independent and dependent variables).

Since this study was focused on answering a research question related IE and associated behavioral phenomena, we were not particularly concerned with controlling for overall social media use across multiple platforms.

But we were cognizant of the limitations of a single-item measure used in this way. This particularly survey item was adapted from a paper that recommends a similar measure for time spent on Facebook. Ernala and colleagues (2020) validity checked various measures of Facebook use in a large N sample from Facebook Research (15 countries and N = 49,934) that included both self-report and observed, live social media usage data. The single-item correlates best with real time usage. To our surprise, we omitted this reference.

To address this point, we included the citation for this measure in the manuscript on page 16.

**Reviewer: 3**  
  
**Comment:** This paper addresses how incidental news exposure and news engagement relate to the concept of “news attraction.” The authors attempt to explicate and measure this concept and then test it to better understand if incidental exposure can help reduce informational inequalities. My primary concerns lie in the presentation of the hypotheses and in the operationalization of the “news attraction” variable. I have presented my concerns in the order they appeared in the paper:

**Response**: Thank you for the careful reading of the manuscript. These comments forced us to clarify key concepts and provide deeper rationale for our modelling and reporting decisions. We welcomed these critiques and hope the current version of the manuscript appropriately addresses these concerns.

**Comment:** The quotes on page 6 need correcting to capture which portions of the sentence are directly quoted.

**Response**: Thanks for catching this. We updated this sentence for the proper quotation placement.

**Comment:** I’m not following the paragraph that leads to H1a and H1b. Why would overall exposure be equal for those who are high and low in news attraction? It seems by definition that those high in news attraction would have greater levels of overall exposure.  
There’s a typo in H1a and H1b as well as H2a and H2b (“and” should be “in”).

I generally don’t like competing hypotheses. In some instances, competing hypotheses are used to ensure a supported hypothesis is possible no matter the results. I would prefer a stronger theoretical argument that leads to a specified prediction. Given that it’s difficult to do this ethically after data has been analyzed, I would like to see the authors do a better job justifying the decision to present competing hypotheses.

I would like to see a stronger argument for why “news engagement” matters in this context. For example, is engagement necessary for someone to learn from news content? It seems that simple exposure could prompt some knowledge gain. What exactly does engagement add to people’s knowledge gain process that is not available through exposure? Perhaps a stronger argument could be made for more in-depth processing (e.g., central route) occurring during “news engagement.”

**Comment:** The paragraph before H2a and H2b suggests interaction effects, but the hypotheses are not written to predict interaction effects. Please consider revising.

**Response**: We’ve taken your suggestion and revised the H2a and H2b to more explicitly predict interaction effects (pg. XX).

**Comment:** Was the trait-like scale for total exposure specific to social media use? It seems that specifying this information should come from social media would be important to this measure.  
Asking people to know if they came across information accidentally is asking a lot of their memory and accuracy. Are there prior studies that validate the use of this measure as an accurate way to know that someone encountered information accidentally? (The limitations of this self-report measure should be more thoroughly addressed in the paper – not just briefly touched on in the limitations paragraph.)

**Response**: Yes, the questionnaire did specify “information on social media.” Thank you for pointing out this oversight. We have clarified the measure on pg. 13.

We agree that self-reported measures are limited. However, we would offer three defenses of the approach.

First, the measures are indeed based on prior literature (Weeks et al., 2017), and this paper is citied in the measures section.

Second, the primary limitation of self-reported measures is inaccurate recall. Prior research shows that survey respondents tend to underestimate their exposure, producing point estimates that are systematically lower than the “true” population parameters (see González-Bailón & Xenos, 2020). This is a form of systematic measurement error. According to King et al. (1994), systematic measurement error will bias descriptive inferences. However, it does not necessarily bias causal inferences if all variables are affected in the same way. That is, if both the independent and dependent variables are systematically underreported, the estimates of the strength of the relationship will not be biased. Considering the nature of the variables in our study (attraction, exposure, engagement), there is a high likelihood that they are all biased in the same way, which counterintuitively means that the effects estimates are likely *not biased.*

We have expanded our discussion of these limitations in both the measures (pg. 11) and limitations sections (pgs. 25-26).

González-Bailón, S., & Xenos, M. (2020). Surveys underestimate online news exposure: a comparison of self-reported and observational data in nine countries. *SSRN Electronic Journal*.

King, G., Keohane, R. O., & Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press.

Third, despite their limitations, we respectfully argue that self-reported measures are still the best way to measure news exposure in the general population. While advances in computational methods allow for direct observation of exposure, it is difficult for researchers to obtain reliable and externally valid data from social media companies (notably, Twitter is now charging $45k per month for their research API). This has been a particularly important issue for Facebook and other Meta platforms, as researchers have had a difficult time obtaining privacy-protected data directly from these companies. Web scraping techniques and third-party apps violate their terms of service. Collaborative research initiatives (i.e., Social Science One) ultimately failed, and research awardees never received the data they were promised due to a reversal from Meta. Finally, Meta’s proprietary platform (CrowdTangle) does not provide individual-level data. Rather, it only provides aggregate engagement metrics for particular posts. Thus, despite the fact that direct news exposure data exists, it is not readily available to academic researchers (including us).

Web-tracking software provides another approach that researchers can employ. However, participation rates in web-tracking studies are generally low, and combining survey data with web-tracking data results in study attrition that alters the contours of the sample. These studies are best treated as “small-N” or “phenomenon-driven” studies rather than as studies of the general population. Therefore, we respectfully assert that while self-reported measures are indeed limited by poor recall and critiquing them has become common in our field, they still provide valuable insights into the relationships between exposure and other variables of interest in the general population.

**Comment:** Why was total exposure multiplied by incidental exposure? Would the results change if these scales were not combined in this way (i.e., if the models were rerun with these variables assessed independently)?

**Response**: The incidental exposure question was designed as a corollary to the total exposure battery. That is, it’s placement in the survey (directly after the total exposure battery), means that respondents should have their previous responses at the top of their minds when answering the incidental question. Thus, the incidental question makes an indirect reference to the total exposure question and was not intended to be a stand-alone measure.

For that reason, we were not particular interested in modeling the interaction between the composite dimensions of our incidental exposure measure. Rather, our aim was to create a measure that is both internally and externally valid, and we think that our multiplicative measure accomplished that aim.

**Comment:** Why was a separate “high-effort engagement” variable created? This needs more justification both theoretically and operationally.

**Response**: Thank you for raising this issue. We had conceived of the high-effort variable as a conceptual robustness check. However, we performed a factor analysis at the request of the editors and found that the items load onto a single factor. Therefore, we decided to drop the high-effort variable from the paper completely.

**Comment:** It’s not clear to me how the second measure of news attraction, “self-reported interest,” is a news-related variable. Is political interest inherently part of news attraction? This measurement seems to suggest that someone who is interested in news and politics would necessarily get news on social media. Including interest as a covariate in the model makes sense, but it’s not clear to me why it is part of a measure of “news attraction.”  
Given the context of the study, I would encourage the authors to consider relabeling “news attraction” to “social media news attraction.” The crux of the argument surrounds exposure to news on social media.

**Response**: This is a valid concern, and we agree that it is important to be specific when measuring key variables. However, the three items are highly correlated and form a reliable scale, which reduces random measurement error in comparison to single-item measures. Additionally, the models are slightly less efficient (i.e., there is slightly more noise) if only the single item (i.e., the news interest item) is used, which is to say that the slope estimates are not substantively different, but the standard errors are slightly larger (but not large enough to alter the statistical significance of results). Therefore, we would prefer to leave the variable as is, and we have made specific mention of the measure in the limitations on pg. 25. That said, we would be willing to alter the variable if you believe this is the best approach.

**Comment:** Per my previous Comment, it seems entirely possible that someone could have high levels of political interest but only get news from non-social media sources.  
You address this previous point in the “regression analyses: exposure” section of the paper. I would like to see the overlap of non-social media news use and social media news use parsed out better in the literature review, specifically as it relates to “news attraction.” It wasn’t clear to me that “news attraction” would apply both to non-social media news as well as social media news, in part due to the way the variables in the news attraction variable were operationalized. I would like to see these decisions better justified (both conceptually and operationally).

**Reviewer: 4**  
  
**Comment:** This paper elaborates on the concept of ‘news attraction’ to examine the extent to which digital media users are exposed to news intentionally or accidentally, and how incidental exposure is related to engagement with news. It uses data collected through a cross-sectional online survey fielded in the USA during the 2020 elections. The results provide evidence of a paradox: while the online environment can be an equalizer of exposure to news across social groups, it can also stratify engagement with the news.  
  
There are many things to like about the paper. It tackles a relevant topic, namely, whether digital media reduces or exacerbates inequalities in passive and active news use. It is theoretically rich, covering the key works in the literature on incidental exposure. The statistical analysis is more sophisticated than what is typical for papers using cross-sectional data, as it includes a latent class analysis, estimates hierarchical models, etc. The appendix is helpful, too, as it enables readers to assess the representativeness of the sample and the robustness of some findings.  
  
Having said that, I see several areas for improvement, especially regarding the context of the study, the methods used, and discussion sections. Let me elaborate on these shortcomings.

**Response**: Thank you for the supportive comments. We also appreciate you taking the time to outline your concerns. Your comments (along with those of the editors) made it clear to us on rereading the manuscript that some of the methodological reporting was either vague or incomplete. We revised those sections. We also made changes to the literature review to better explicate core concepts. We hope the current version of the manuscript sufficiently addresses these concerns.   
  
**Comment:** Somewhat ironically (as the authors make highlight the importance of studying the context of media exposure), I missed a discussion on the context of the study. Empirical findings are always bounded by cultural, temporal, and other forces. The polarized American media and political systems are rather unique in the world. I’m sure incidental exposure to political news and current events differs between election and nonelection years. All this is to say that it will greatly benefit the international audience of the journal if the authors include one or two paragraphs in the methods section about the particularities of the US case that are relevant for this particular study.

**Response**: You’re absolutely right on this point. We’ve taken your suggestion and added a “Context of Study” section after the hypotheses and before the Methods (pgs. 10-11).   
  
**Comment:** The methods section is lacking important pieces of information, and many of the authors’ choices are not justified or explained. This has the unfortunate consequence of making the statistical models less parsimonious and the results harder to follow.

**Response**: Thank you for raising this legitimate concern. As discussed elsewhere in this letter, we heavily revised the Methods section (pages 10-17) to better justify our research design. We also further explicated core concepts related to the trait/state distinction (page 4). Details are included in the responses below.

**Comment:** Why were the data collected using a rolling cross-sectional design (RCSD) instead of a single one-shot design? I’m asking considering that the longitudinal aspect of the survey does not seem of any relevance to this study.

**Response**: This comment echoed concerns raised above. For convenience, we will include previous answers here:

“We chose the RCS design to better capture news attention and engagement behaviors. The benefit of this design is that we are cuing respondents by embedding actual news stories circulating on social media during the data collection waves. Each wave we adjusted the stories, based on third-party ranking metrics, to ensure that respondents received stories that were tied to that news cycle.

This practice is not uncommon in election studies (De Vreese et al., 2017) and we feel that it offers a higher level of external and internal validity.”

And, from above:

“We lacked precision in our reporting of the study design. In turn, this lack of clarity lead to some ambiguity about other aspects of the study, including the nature of the rolling cross-section (RCS) survey and the ‘linkage’ between respondents and social media content.

We have heavily revised our explanation of the study design in the Methods section to address these points on pages 10-12.

To be clear, the survey responses are not validated. Rather, we implemented a multi-step strategy to identify and verify the top news circulating on Facebook at the time of data collection.

We then embedded the top two news articles circulating on Facebook at the time (ranked by engagement metrics) into the survey as a cue for attention to news during the previous news cycle. The assumption is that posts with higher engagement metrics would be more likely to show up in people’s feeds, and therefore a respondent who only passively follows the news should have seen it.

To decide what counted as ‘top’ news organization, as well as the most widely circulated/top stories, we developed a criterion that employs third-party data brokers’ public and private ranking systems.

First, we used NewsWhip to identify the top 25 news organizations on Facebook (NewsWhip regularly release reports to the public). Next, we used Brandwatch (formerly Crimson Hexagon) to identify the top posts on Facebook based on engagement from those 25 outlets, and finally, CrowdTangle was referenced to verify that those top stories were indeed widely circulating on Facebook during four days prior to data collection. The author has accounts with Brandwatch and CrowdTangle, so, while the news posts were publicly available, the rankings for news engagement are only available to researchers.

The main benefit of this approach is that we can a) ID the top stories widely circulating on Facebook with a higher level of certainty than simply relying on the researcher’s intuition, and b) we can better take advantage of the RCS design. RCS designs are useful, because survey responses can be more closely tied media campaigns and election cycles. We followed the lead of previous published work in this area (see, De Vreese et al., 2017 for an overview and past uses of RCS to link media content with survey data). This approach is now rather common in election studies.”

**Comment:** The data was analyzed using multilevel modeling. According to the authors, this was justified: it helps to control for measurement invariance across the 17 sampling frames employed. But the authors should be aware that MLM are harder to interpret for the lay reader than a simple, straightforward OLS. Perhaps comparing the robustness of results across using different estimators would help answer the question of how dependent the results on MLM are.

**Response**: Thank you for raising this concern, as it gives us a chance to stress-test our analysis. We spot-checked robustness using Model 1 (DV: Incidental Exposure) and Model 3 (DV: Total Exposure) in Table 1. In both cases, model comparisons show that the MLM approach fits the data better than the OLS approach, with lower AICs and statistically significant log-likelihood tests (for Model 1, chi-square = 5.16, p = .023, indicating the MLM is a slightly better fit; for Model 3, chi-square = 29.17, p < .001, indicating MLM is a much better fit).

Furthermore, the coefficient estimates are similar in both cases. For Model 1 the key estimates are Low = 1.15 (MLM) vs. 1.14 (OLS); Medium—Unmotivated = 0.62 (MLM) vs 0.62 (OLS); Medium—Motivated = 0.36 (MLM) vs. 0.37 (OLS); and High = -0.16 (MLM) vs. -0.14 (OLS). The differences are slightly bigger for Model 3, but not big enough to alter the paper’s substantive conclusions: Low = 1.12 (MLM) vs. 1.09 (OLS); Medium—Unmotivated = 0.79 (MLM) vs. 0.82 (OLS); Medium—Motivated = 1.22 (MLM) vs. 1.30 (OLS); High = 1.59 (MLM) vs. 1.67 (OLS).

We would be happy to add a footnote to the manuscript describing these robustness checks if the reviewer feels this is appropriate and would clarify the findings for the readers.

**Comment:** How much data was missing to justify the use of multiple imputation using chained equations? Do the results change if missing data is not imputed?

**Response**: This is an important detail that we omitted to report in the previous version of the manuscript. There are two answers here. The first is related to our rationale to impute (along with a report of the missingness by variable) and the second is related to our approach to data treatment in general.

First, to repeat a response to the editors:

Thank you for catching this. Indeed, readers may want to know about the nature of missingness. The sample size for our models before imputation is N = 1,731. We revised the sample reporting to include this information on pages 11-12.

We decided to impute data not for reasons of smaller sample sizes, but rather to protect against one of the drawbacks of the RCS design. That is, RCS typically means more waves (17 in our case) but fewer respondents per wave. Thus, there is an inherent potential for sample size variation between waves, which could theoretically bias statistical results (sampling bias). We took two steps to minimize this bias: multilevel modelling (as is common in designs that rely on multiple data collection points) and imputation of missing cases.

Based on analysis of missing patterns, we did not see any issue with data quality in this regard. The missingness ranges from 6.3% (Age) to 1% (Strength of Party) and only 9 of the 28 modelled variables showed missing cases. Here is the report (which we will omit from the manuscript for space considerations:

|  |  |  |
| --- | --- | --- |
| variable | number missing | % |
| age | 126 | 6.2749004 |
| ine2 | 106 | 5.27888446 |
| alg | 59 | 2.93824701 |
| eff.int | 22 | 1.09561753 |
| int | 22 | 1.09561753 |
| ideo | 19 | 0.94621514 |
| pid | 19 | 0.94621514 |
| ideo.ext | 19 | 0.94621514 |
| pid.str | 19 | 0.94621514 |
| ine1 | 0 | 0 |
| smnews | 0 | 0 |
| engage | 0 | 0 |
| know | 0 | 0 |
| nml | 0 | 0 |
| sml | 0 | 0 |
| int.freq | 0 | 0 |
| sm.freq | 0 | 0 |
| female | 0 | 0 |
| poc | 0 | 0 |
| edu | 0 | 0 |
| inc | 0 | 0 |
| sm.newsintent | 0 | 0 |
| smfollow | 0 | 0 |
| avoid | 0 | 0 |
| sm.ns | 0 | 0 |
| sm.div.occ | 0 | 0 |
| sm.div.tie | 0 | 0 |
| grp | 0 | 0 |

Second, we used a multiple imputation algorithm (predictive mean matching) which uses full information (drawing from a group of similar, but complete cases) and thus should reduce bias as much as possible (see Van Buuren, 2018).

As a matter of data treatment, we did not run alternative models without the imputations. The overall number missing is rather low. This is also a matter of research ethics, as of course we would not choose results based on a favorable outcome relative to imputing missing data. Our decision is based solely on the nature of data collection, and we prefer to use the imputed datasets.

For the reviewer’s reference, here are the model statistics for the imputed versus non-imputed datasets:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1 Models** | | | | |
|  | M1 | M2 | M3 | M4 |
| **Fit Statistics Imputed** |  |  |  |  |
| ICC | .01 | .02 | .03 | .02 |
| LL | -3,137.86 | -1,051.30 | -3,088.25 | -1,339.30 |
| Pseudo-*R*2 | .17 | .11 | .44 | .11 |
| **Fit Statistics List-Wise Deletion** | | | | |
| ICC |  |  |  |  |
| LL |  |  |  |  |
| Pseudo-*R*2 |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **Table 2 Models** | | |
|  | M1 | M2 |
| **Fit Statistics Imputed** |  |  |
| ICC | 01 | .02 |
| LL | -1,634.20. | -1,285.53 |
| Pseudo-*R*2 | .45 | 0.46 |
| **Fit Statistics List-Wise Deletion** | | |
| ICC |  |  |
| LL |  |  |
| Pseudo-*R*2 |  |  |

**Comment:** Do results change when using unweighted data? I’m asking because weights, while helping to address deviations from the population distribution, inflate standard errors, too.

**Response:** In this particular case, the weights do not inflate the standard errors. We spot-checked robustness using the same two models before, and the unweighted standard error estimates are very similar to the weighted estimates. We prefer to use the weighted models given the nature of the survey sample, and we would be happy to add a footnote explaining these robustness checks. Briefly, the key standard error estimates are (order = Low, Medium—Unmot, Medium—Mot, High) are:

Model 1 weighted: 0.05, 0.06, 0.08, 0.11

Model 1 unweighted: 0.05, 0.06, 0.08, 0.10

Model 3 weighted: 0.06, 0.05, 0.07, 0.10

Model 3 unweighted: 0.05, 0.06, 0.07, 0.10

**Comment:** More importantly, how were survey responses linked with social media content, exactly? And the validation of content lists with CrowdTangle, what was that and why was it necessary? And if most stories come from Fox News, what does this say about the representativeness of the media stimuli?

**Response:** We pasted the comments related to social media content above several times. Briefly, we relied on third-party rankings of top news circulating on Facebook at the time of data collection to ensure that respondents received stories that were tied to that news cycle. We looked at several ranking lists to validate that indeed, the stories we chose to embed in the survey were actually the ‘top’ posts that week. The assumption is that posts with higher engagement metrics would be more likely to show up in people’s feeds, and therefore a respondent who only passively follows the news should have seen it.

Regarding Fox News, they dominated the rankings for engagement metrics according to all three tracking companies we referenced (NewsWhip, Brandwatch, and CrowdTangle). This is why we chose two stories to embed and then randomize (each respondent saw only one story), one for the top story (which was always Fox) and the first non-Fox story.

In this sense, we were not aiming for a ‘representative’ sample of news that circulated during a given survey wave, but rather as a point of external validity, we wanted to capture the news that was being pushed to most people based on the algorithms that favor engagement rankings.

This is an important limitation worth discussing, and if we had to make difficult choices about what stories to embed in the survey, other scholars may want to learn from our experiences. Thus, we added a paragraph to the limitations section on page 26.

**Comment:** I find it somewhat confusing why some IVs are described as covariates and others as control variables. Covariates predict variance in the DV that is clearly not attributable to the IVs of interest. That’s why socio-demographics are usually covariates. Control variables, however, are included to eliminate spurious relationships between the IV of interest and the DV that might otherwise be thought to be causal. Again, more explanation would be helpful.

**Response:** You’re right. Our terminology was needlessly confusing. We cut all mentions of covariates and now refer to all variables as controls.  
  
**Comment:** In the concluding section, I missed two central aspects. First, a discussion of how the study findings relate to the most important function of news: to produce informed readers. There is a larger debate on the consequences of the digitization of news and the rise of social media on citizen competence. Questions such as: what do people learn from they news? Why is social media news use related (or not) to misinformation? These are important questions, and I would like to know the implications of this study on those issues.

**Response:** These are good questions. Certainly, there is room for follow-up work on the role of news attraction in stimulating or hampering the normative functions of political learning and spread of misinformation.

In this paper, we were interested in how the emerging media reality (where algorithmic news flows alter who gets political information and who pays attention to it) shape information inequality. Thus, the results are inherently normative in that we assume that these inequalities will equate to stratified news audiences with radical hierarchies that are typical of so-called information deserts.

Thus, it follows, that the high news attraction group will disproportionally reap the benefits of a range of pro-social outcomes related to active attention, including higher levels of learning, and a reduced likelihood of falling for misinformation. We tried to make these connections in the Discussion section, but your comment had us revise this section to specifically mention these potential outcomes typical of information hierarchies.

We integrated these thoughts into the discussion on page 26-27.   
  
**Comment:** Second, for a paper that mentions repeatedly concepts such as datafication and algorithmic categorization, I expect some a more thorough discussion of what the study findings mean for the debate on the platformitization of news in the current media environment. The last paragraph of the paper hints at this. I’m sure the authors can elaborate more.

**Response:** We may need to reconsider or be more careful about this language.  
  
**Comment:** Last but not least, please proofread the manuscript. Some propositions are missing, there are spelling problems, etc.

**Response**: We have proofread the manuscript. Thank you again for you time and attention to this paper. We feel it is stronger after responding to your comments.

**References**

Bennett, W. L., Wells, C., & Rank, A. (2009). Young citizens and civic learning: Two paradigms of citizenship in the digital age. *Citizenship studies*, *13*(2), 105-120.

de Vreese, C. H., Boukes, M., Schuck, A., Vliegenthart, R., Bos, L., & Lelkes, Y. (2017). Linking Survey and Media Content Data: Opportunities, Considerations, and Pitfalls. *Communication Methods and Measures*, *11*(4), 221–244

Ernala, S. K., Burke, M., Leavitt, A., & Ellison, N. B. (2020). How Well Do People Report Time Spent on Facebook? An Evaluation of Established Survey Questions with Recommendations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14.

Holbert, R. L., Hmielowski, J. D., & Weeks, B. E. (2012). Clarifying relationships between ideology and ideologically oriented cable TV news use: A case of suppression. *Communication Research*, *39*(2), 194-216.

Lee, J. K., & Kim, E. (2017). Incidental exposure to news: Predictors in the social media setting and effects on information gain online. *Computers in human behavior*, *75*, 1008-1015

Shehata, A., Ekström, M., & Olsson, T. (2016). Developing self-actualizing and dutiful citizens: Testing the AC-DC model using panel data among adolescents. *Communication Research*, *43*(8), 1141-1169.

Stroud, N. J. (2011). *Niche news: The politics of news choice*. Oxford University Press on Demand.

Van Buuren, S. (2018). *Flexible imputation of missing data*. CRC press. https://stefvanbuuren.name/fimd/

Zaller, J. (2003). A new standard of news quality: Burglar alarms for the monitorial citizen. *Political Communication*, *20*(2), 109-130.