

UCLA

UCLA Electronic Theses and Dissertations

Title

Behind the Screens: News Selectivity in Algorithmic Digital Environments

Permalink

<https://escholarship.org/uc/item/9qm9g7v5>

ISBN

9798247989318

Author

Noh, Seonhye

Publication Date

2026-06-08

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Behind the Screens: News Selectivity in Algorithmic Digital Environments

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Communication

by

Seonhye Noh

2026

© Copyright by

Seonhye Noh

2026

ABSTRACT OF THE DISSERTATION

Behind the Screens: News Selectivity in Algorithmic Digital Environments

by

Seonhye Noh

Doctor of Philosophy in Communication

University of California, Los Angeles, 2026

Professor Stuart Soroka, Chair

This dissertation argues that dominant narratives of digital media fragmentation—echo chambers and filter bubbles—present a picture that is considerably more complicated than commonly assumed. The discrepant findings in the selective exposure literature reflect three recurring analytical tendencies: (a) measuring expressed preferences in place of actual behavior; (b) treating news engagement as a single undifferentiated act rather than a layered sequence of distinct behavioral stages; and (c) insufficient attention to disentangling what algorithmic systems make available from what users themselves choose to engage with. The three studies in this dissertation address each of these tendencies in turn, drawing on controlled experiments and large-scale digital trace data to examine political news engagement. What emerges is not a refutation of selective exposure, but a more precise and nuanced account of when, how, and why it occurs—and what it means for democratic politics in the algorithmic age.

The dissertation of Seonhye Noh is approved.

Peter John Lamberson

Jungseock Joo

Michelle Torres

Stuart Soroka, Committee Chair

University of California, Los Angeles

2026

Table of Contents

Abstract	ii
Committee Page	iii
Acknowledgments	xii
Vita	xiii
1 Introduction	1
1.1 The Paradox of Digital Information Environment	1
1.2 The Theoretical Landscape: Selective Exposure and Its Complications	2
1.3 Overview of Three Studies	5
2 Chapter 1. Expressive News Preferences: Identity-Signaling in News Selection	9
2.1 Introduction	9
2.2 Partisan Expressive Responding	11
2.3 Expressive News Preferences in Polarized Media Environment	14
2.4 Research Design	18
2.4.1 Procedures and Materials	20
2.4.2 Measures	21
2.5 Results	23
2.6 Discussion	28

3	Chapter 2. Engaging to Oppose: Cross-Cutting Patterns in Hostile News Commentary	31
3.1	Introduction	31
3.2	Media Selectivity and News Engagement Types	32
3.3	Cross-Cutting Potential in Active News Engagement	34
3.4	Data	37
3.5	Study 1: Content-Level Engagement	39
3.5.1	Methods	40
3.5.2	Results	42
3.6	Study 2: Source-Level Engagement	43
3.6.1	Methods	43
3.6.2	Results	47
3.7	Discussion	50
4	Chapter 3. Algorithmic Exposure, User Choice: Engagement Segregation in Mobile News Feeds	55
4.1	Introduction	55
4.2	Algorithmic Segregation in News Engagement	57
4.2.1	Algorithmic Segregation	57
4.2.2	Individual Selectivity vs. Algorithmic Personalization	59
4.3	Beyond Partisanship: Negativity and Topical Interest as Dimensions of Engagement Segregation	63
4.4	Data	65
4.4.1	The SmartNews Platform	65
4.4.2	Data	66
4.5	Methods	67
4.5.1	Headline Attribute Estimation	67
4.5.2	Co-Engagement Network	69

4.5.3	Models of Engagement Segregation	71
4.6	Results	73
4.6.1	Descriptive Statistics	73
4.6.2	Engagement Segregation from Algorithmic Exposure and User-Driven Stages	74
4.7	Predictors of Engagement Segregation	75
4.8	Discussion	78
5	Appendix for Chapter 1	83
5.1	Experiment Materials	83
5.1.1	Hypothetical Scenarios and News Selection Instructions	83
5.1.2	Twelve Gun Control Arguments	85
5.1.3	Headlines and Lead Texts	87
5.1.4	News Selection Task Interface	89
5.1.5	Pre- and Post-Treatment Survey Measures	90
5.2	Supplemental ACNS	92
5.3	Additional Preregistered Hypotheses	93
5.3.1	OLS Regressions Predicting ACNS (Excluding “Keep the Rules About the Same” Responses)	96
5.3.2	OLS Regressions Predicting ACNS (Including “Keep the Rules About the Same” as Anti-Gun Control)	97
6	Appendix for Chapter 2	98
6.1	Monthly Averages of Hostility and News Engagement by News Domain . . .	98
6.2	Monthly Trends in Confrontation and News Engagement	99
6.2.1	Angry Reaction Over Time	99
6.2.2	Hostile Comments Over Time	100
6.2.3	Modularity Over Time	100

6.2.4	Selected Media Proportion Over Time	101
6.3	Distributions of Monthly Confrontation	102
6.3.1	Distribution of Monthly Average Number of Angry Reaction	102
6.3.2	Distribution of Monthly Percentage of Hostile Comments	103
6.3.3	Distribution of Monthly Percentage of Hostile Users	103
6.4	An Alternate Measure of Modularity	104
6.4.1	Centrality by News Domain	104
6.4.2	Centrality Over Time by News Domain	105
6.4.3	Centrality of Hostile versus Non-Hostile Users	105
6.5	Additional Aggregate-level Analyses of Hostility and Source Diversity	106
6.5.1	Multilevel Regression Result Graph	107
6.5.2	Multilevel Regression Result Table	108
6.5.3	User Hostility and Selected Media Sources	108
6.6	Results from the Generalized Linear Mixed Model (GLMM)	110
6.6.1	Generalized Linear Mixed Model (GLMM) Result	110
6.6.2	Bootstrap Result: Confidence Intervals for GLMM Estimates	111
6.6.3	Bootstrap Result: Distributions of GLMM Estimates	112
6.7	Case Study 1: Nuclear Phase-Out Policy	113
6.7.1	Context	113
6.7.2	Data and Classification	114
6.8	Case Study 2: THAAD Deployment	116
6.8.1	Context	116
6.8.2	Data and Classification	116
6.8.3	Results	117
7	Appendix for Chapter 3	120
7.1	Full Regression Results	121
7.2	Topical Categories	122

7.2.1	Platform-Assigned Categories: Examples	122
7.2.2	Uncategorized Articles	123
7.2.3	Topical Composition of Uncategorized Articles	124
7.3	Partisan Slant Estimation	127
7.3.1	Estimation and Validation	127
7.3.2	Examples by Category	128
7.4	Co-Engagement Network	129
7.4.1	Basic Network Characteristics	129
7.4.2	Degree-Corrected Stochastic Block Model	129
7.4.3	Distribution of Block Sizes	130
7.5	Additional Segregation Metrics	131
	References	132

List of Figures

2.1	Mean ACNS by news engagement modes	25
2.2	Mean ACNS by partisanship and selection publicity	26
3.1	Monthly number of news comments and unique commenters on Naver News	39
3.2	Commenting patterns on nuclear power phaseout policy articles	44
3.3	Illustrative examples of media network fragmentation	47
3.4	Monthly modularity scores for hostile and non-hostile users	48
3.5	Media engagement networks of hostile and non-hostile users in politics domain	49
3.6	Average angry reactions and hostile comments across six news domains . . .	53
3.7	Average modularity scores across six news domains	54
3.8	GLMM results for hostility, news domains, and source diversity	54
4.1	Engagement segregation according to user selectivity and algorithmic adaptation	62
4.2	Engagement segregation in algorithmic exposure and user-driven stages . . .	75
4.3	Engagement segregation in algorithmic exposure and user-driven stages: po- litical interest	77
4.4	Predicted probability of segregation by political interest and partisan distance	78
5.1	News selection task screen for selection-focused and consumption-focused con- ditions	90
5.2	Mean ACNS by condition	93
6.1	Monthly trends in angry reactions over time by news domain	99
6.2	Monthly trends in hostile comments over time by news domain	100

6.3	Monthly trends in modularity over time by news domain	100
6.4	Monthly trends in selected media proportion over time by news domain . . .	101
6.5	Distribution of monthly average number of angry reactions by news domain .	102
6.6	Distribution of monthly percentage of hostile comments by news domain . .	103
6.7	Distribution of monthly percentage of hostile users by news domain	103
6.8	Centrality by news domain	104
6.9	Centrality over time by news domain	105
6.10	Centrality of hostile versus non-hostile users	105
6.11	Multilevel regression results predicting modularity	107
6.12	User hostility and selected media sources.	109
6.13	Bootstrap distributions of GLMM estimates	112
6.14	Commenting patterns across news stance and user orientation for the THAAD deployment case	119
7.1	Distribution of uncategorized articles across users	125
7.2	Topic distribution of categorized and uncategorized articles	126
7.3	Distribution of block sizes across three engagement stages	131

List of Tables

4.1	Descriptive Statistics across Viewing, Clicking, and Reading Phases	74
4.2	Summary Statistics of DCSBM-Fitted Co-Engagement Networks	74
5.1	Hypothetical scenarios and images	84
5.2	News selection task instructions	85
5.3	OLS Regressions Predicting ACNS (Excluding “Keep the Rules About the Same” Responses)	96
5.4	OLS Regressions Predicting ACNS (Including “Keep the Rules About the Same” as Anti-Gun Control)	97
6.1	Monthly Averages of Hostility and News Engagement by News Domain . . .	98
6.2	Multilevel Regression Results Predicting Modularity	108
6.3	Generalized Linear Mixed Model (GLMM) Result	110
6.4	Bootstrap Result: Confidence Intervals for GLMM Estimates	111
7.1	Logistic Regression Models Predicting Engagement Segregation	121
7.2	Example Headlines by Platform-Assigned Topical Category	122
7.3	Example Headlines by Partisan Slant Category	128
7.4	Average network characteristics of the co-engagement networks across 30 sampled days, by engagement stage.	129
7.5	Block assortativity (weighted) and mean conductance across engagement stages.	131

Acknowledgments

I am grateful to my committee members for their guidance throughout this dissertation. Stuart Soroka provided invaluable intellectual direction and mentorship at every stage of this project. PJ Lamberson, Jungseock Joo, and Michelle Torres offered generous feedback and support that shaped this work in important ways.

I also thank the members of the Political Communication and Behavior Lab at UCLA for their collegiality and intellectual community throughout my time as a graduate student.

Chapter 2 is a version of Noh, S., & Soroka, S. (2026). Engaging to Oppose: Cross-Cutting Patterns in Hostile News Commentary. *The International Journal of Press/Politics*. The chapter has been slightly revised from the published version to match the formatting and numbering conventions of this dissertation. Stuart Soroka contributed to the conceptual development of the study, provided guidance on the empirical approach, and offered feedback on successive drafts.

Chapter 3 is a version of Noh, S. (2026). Expressive News Preferences: Identity-Signaling in News Selection. *Journalism Studies*. The chapter has been slightly revised from the published version to match the formatting and numbering conventions of this dissertation.

The data analyzed in Chapter 4 were provided by SmartNews under a non-disclosure agreement. The analyses and conclusions presented are those of the author and do not represent the views of SmartNews.

Vita

Education

Bachelor of Arts in Communication, Pusan National University, South Korea, 2017

Master of Arts in Communication, Seoul National University, South Korea, 2021

Publications

Noh, S., & Soroka, S. (2026). Engaging to Oppose: Cross-Cutting Patterns in Hostile News Commentary. *The International Journal of Press/Politics*.

Noh, S. (2026). Expressive News Preferences: Identity-Signaling in News Selection. *Journalism Studies*.

Kernell, G., & Noh, S. (forthcoming). The AI Referee: How Online Interventions Shape Incivility and User Engagement in News Discussions. *Social Media + Society*.

Kim, S., & Noh, S. (2026). Disproportionate Voices: Participation Inequality and Hostile Engagement in News Comments. *International AAAI Conference on Web and Social Media (ICWSM 2026)*.

Noh, S., & Soroka, S. (2025). Negativity Biases Online: The Interplay of Individuals and Algorithms in News Consumption. *Journal of Media Psychology*.

Akcakir, G., Jiang, J., Luo, J., & Noh, S. (2023). Validating a Mixed Approach for Multilingual News Framing Analysis: A Case Study of COVID-19. *Computational Communication Research*.

Selected Awards and Fellowships

UCLA Departmental Dissertation Proposal Recognition Award, 2026

Research Grant, UCLA Bedrai Kindness Institute, 2025

UCLA Political Psychology Fellowship, 2025

Graduate Summer Research Mentorship, 2024

Research Grant, UCLA Initiative to Study Hate, 2022–2023

UCLA Graduate Fellowship and Assistantship, 2021–2026

Introduction

1.1 The Paradox of Digital Information Environment

In the decades since digital platforms became the dominant infrastructure of public information, a particular diagnosis has taken hold in academic research, in policy circles, and in popular discourse alike. That diagnosis holds that the contemporary media environment has become deeply and dangerously fragmented: citizens increasingly inhabit ideological enclaves, consuming news that confirms what they already believe, avoiding content that challenges them, and growing ever more insulated from the perspectives of those who think differently. The metaphors have become familiar almost to the point of cliché — the "echo chamber," in which like-minded voices amplify one another toward increasing extremity (Sunstein 2001); the "filter bubble," in which algorithmic personalization silently narrows the informational world each user inhabits (Pariser 2011).

The implications for democratic life are serious. Democratic theory across its many traditions—from deliberative accounts that emphasize reasoned public exchange (Habermas 1989) to more minimal accounts centered on informed electoral choice (Zaller 1992)—presupposes that citizens have access to a range of views and information. A media environment that systematically forecloses such access would represent a genuine threat to those conditions. It is no surprise, then, that these concerns have shaped not only academic debate but also regulatory discussions and platform design decisions on both sides of the Atlantic.

This dissertation argues that the picture is considerably more complicated than this diagnosis allows, and that the complication is not incidental but systematic.

A growing body of empirical research has begun to trouble the received account. Studies using large-scale behavioral data and web-tracking find that most users, regardless of political orientation, consume a more moderate and cross-cutting mix of news sources than

echo chamber models would predict (Guess 2021; Tyler et al. 2022). Experimental research reveals that individuals do not uniformly avoid counter-attitudinal information, and may in fact engage with it more than they report (Garrett 2009; Garrett and Stroud 2014). Evidence from social media platforms suggests that the strength of algorithmic filtering may be more modest than commonly assumed (Bakshy et al. 2015; González-Bailón et al. 2023). Yet the fragmentation narrative retains a powerful grip on both scholarly and policy debates, in part because the empirical challenges to it have been piecemeal, and in part because it remains unclear why studies frequently arrive at different conclusions.

This dissertation offers a systematic answer to that question. The discrepant findings in the literature, I argue, are not merely the product of different samples, platforms, or political contexts. They reflect three deeper, recurring analytical tendencies: a tendency to measure expressed preferences in place of actual behavior; a habit of treating news engagement as a single, undifferentiated act when it is in fact a layered sequence of distinct behavioral stages; and insufficient attention to disentangling what algorithmic systems make available to users from what users themselves choose to engage with. Each of these tendencies may contribute to an incomplete picture of fragmentation—one that risks overstating its strength or mislocating its sources.

The three studies in this dissertation address each of these tendencies in turn, drawing on controlled experiments, large-scale digital trace data, and computational methods to examine political news engagement across its full complexity. What emerges is not a refutation of selective exposure, but a more nuanced account of when, how, and why it occurs, and what it means for democratic politics in the algorithmic age.

1.2 The Theoretical Landscape: Selective Exposure and Its Complications

The concern that citizens avoid information challenging their existing views has deep roots in political communication research. Early work by Lazarsfeld et al. (1944) and later by Klapper (1960) documented a tendency toward selective exposure to attitude-consistent

media, though they also noted that this tendency was far from absolute. Later theoretical developments helped explain the psychological mechanisms behind it. Research on motivated reasoning showed that people process information in ways that protect prior beliefs, evaluating congenial information more favorably and counter-attitudinal information more skeptically (Lodge and Taber 2013). This body of work provided a cognitive foundation for understanding why individuals might gravitate toward news that confirms what they already think.

As media environments diversified and digital platforms emerged, these tendencies took on new significance. Studies using survey data and experimental designs found that individuals prefer ideologically aligned news sources and are more likely to select attitude-consistent content (Iyengar and Hahn 2009; Stroud 2008, 2011). These findings, combined with the rapid growth of online news consumption, gave rise to widespread concern about echo chambers and filter bubbles—the idea that digital media environments not only reflect but actively reinforce ideological segregation at scale (Pariser 2011; Sunstein 2001, 2018).

Yet as the empirical record has grown, a more complicated picture has emerged. Studies using experimental or behavioral data have found that most users, regardless of political affiliation, consume a more moderate and varied mix of news than strong selective exposure accounts would predict (Guess 2021), and that individuals do not systematically avoid counter-attitudinal information when it is made available to them (Garrett 2009; Garrett and Stroud 2014). Other work finds that ideological segregation online is often less pronounced than offline, and that exposure to cross-cutting content is more common than echo chamber models imply (Gentzkow and Shapiro 2011; Guess et al. 2018). The empirical picture, in short, is unsettled. And this unsettledness is not simply a matter of conflicting findings—it points to deeper questions about how we conceptualize and measure political information engagement in the first place.

Three analytical gaps in particular stand out. The first concerns the gap between expressed preferences and actual behavior. Much of the evidence for selective exposure rests

on measures that ask people which news they would choose to read, or that observe headline selections in controlled settings. These measures may not fully capture what people actually read in everyday life. Research on expressive responding shows that individuals sometimes misreport their attitudes or preferences in ways that align with their political identity because expressing congenial preferences carries social and psychological value in itself (Bullock et al. 2015; Malka and Adelman 2023; Prior et al. 2015; Schaffner and Luks 2018). In news selection contexts, this raises the possibility that observed ideological alignment in preference tasks reflects, at least in part, identity signaling rather than genuine reading intent. When people select headlines without expecting to actually read the articles, their choices may be more expressive than informational. This distinction has rarely been tested directly, leaving open the question of how much apparent selectivity in news preferences reflects what people would truly choose to read.

The second gap concerns the diversity of engagement behaviors that the selective exposure literature has tended to treat as equivalent. Research in this tradition has focused predominantly on passive forms of news engagement, such as which outlets people visit or which headlines they click. These are important behaviors, but they do not exhaust the range of ways people interact with political news in digital environments. Commenting, sharing, and other more active forms of participation may involve different motivational logics than passive reading or selection. Active engagement with news, particularly commenting, is often characterized by emotional reactions such as anger or hostility rather than by information-seeking (Coe et al. 2014; Rowe 2015). Research on the psychology of anger suggests that it is associated with a motivation to confront and oppose rather than to seek out congenial content (Carver and Harmon-Jones 2009; Young et al. 2011), and that it carries a mobilizing effect on political participation more broadly (Valentino et al. 2011). If this is right, then active engagement may follow quite different patterns than passive consumption, potentially producing more cross-cutting rather than pro-attitudinal behavior. Whether this is empirically the case remains underexplored.

The third gap concerns the relationship between what platforms make available and what users choose to engage with. The selective exposure literature has largely focused on user demand, meaning the choices people make, without fully accounting for algorithmic supply, meaning the content platforms present to users in the first place. This distinction matters enormously for how we interpret observed patterns of engagement. If a user’s news feed is already structured in a way that reflects their past behavior, then what looks like a voluntary preference may partly be a response to a curated choice set. Several studies have begun to examine this question, finding that individual choice tends to drive segregation more strongly than algorithmic filtering alone (Bakshy et al. 2015; González-Bailón et al. 2023), while others find evidence of ideologically amplifying effects in algorithmic feeds (Huszár et al. 2022). But because most existing studies either examine the algorithm or the user in isolation, the joint contribution of both forces across different stages of engagement remains poorly understood. Without separating algorithmic provision from user choice within the same observational framework, it is difficult to determine how much of observed fragmentation originates on each side.

These three gaps—in measurement, in engagement-stage differentiation, and in the supply-demand distinction—motivate the three studies that follow. While each study was conducted independently and addresses its own specific research questions, together they shed light on analytical challenges that shape what we can and cannot conclude about selective exposure in digital news environments. Collectively, they offer a more layered account of how political information engagement unfolds across individuals, platforms, and behavioral stages.

1.3 Overview of Three Studies

This dissertation brings together three empirical studies that each address one of the gaps identified above. The studies were conducted independently, and each stands as a self-contained contribution with its own literature review, data, and findings. Taken together,

however, they offer a more layered account of political news engagement than any single study could provide, examining how selectivity manifests differently depending on how it is measured, what form of engagement is observed, and whether one accounts for the role of algorithmic curation.

The first study, "Expressive News Preferences: Identity-Signaling in News Selection," published in *Journalism Studies* (2026), examines whether commonly used measures of news preferences actually capture what people intend to read. Using a preregistered experiment with a national U.S. sample, the study asks whether people select different news content depending on whether they expect to actually read the articles afterward. The findings show that participants selected significantly more attitude-consistent news when they did not anticipate reading the content, suggesting that headline selection tasks—widely used in survey and experimental research—may capture not only genuine reading intent, but also identity-signaling behavior. This has implications for how we interpret experimental findings on selective exposure, and for how we understand the gap between stated news preferences and actual media diets observed in behavioral studies.

The second study, "Engaging to Oppose: Cross-Cutting Patterns in Hostile News Commentary," co-authored with Stuart Soroka and published in *The International Journal of Press/Politics* (2026), shifts from passive news selection to active participation. Using ten years of commenting data from South Korea's largest news aggregator, amounting to over 250 million comments from approximately six million users, the study examines whether hostile commenters engage primarily with attitude-consistent content or whether they are more likely to cross partisan lines. The findings suggest that hostile commentary is characterized by cross-cutting rather than pro-attitudinal engagement: users are more likely to comment on counter-attitudinal news, and hostile commenters show weaker echo chamber structures in co-engagement networks than less hostile users. This study points to the importance of distinguishing between passive and active forms of news engagement, as the two appear to follow quite different logics.

The third study, "Algorithmic Exposure, User Choice: Engagement Segregation in Mobile News Feeds," examines audience segregation across three sequential stages of news engagement—algorithmic exposure (viewing), and two user-driven stages (clicking and reading)—using large-scale behavioral trace data from SmartNews, a mobile news aggregator app dedicated to providing politically balanced news coverage. The data capture over 22 million user-article interactions from 5,000 U.S. users, and include records of what the algorithm surfaces to users as well as what users choose to engage with on their own. The study finds that segregation intensifies as users move from algorithmic exposure to more choice-driven stages of engagement, and that partisan distance is associated with segregation primarily at the user-driven stages rather than at the algorithmic exposure stage. These findings suggest that, at least in this platform context, user selectivity may contribute more to engagement segregation than algorithmic curation does. This study has not yet been published as of the time of this dissertation's submission.

While each study addresses its own specific research questions, a common thread runs through all three. Each one involves disaggregating something that prior research has tended to treat as a single, uniform phenomenon. The first study separates expressed preferences from behavioral intent. The second separates passive from active engagement. The third separates algorithmic provision from user choice. In each case, making these distinctions changes the conclusions one draws about the extent and nature of selective exposure. Taken together, the three studies suggest that selective exposure is real but context-dependent— that its strength and character vary meaningfully depending on which aspect of news engagement one is looking at. This does not overturn the existing literature, but it does call for greater precision in how the concept is operationalized, measured, and interpreted.

The remainder of this dissertation proceeds as follows. Chapter 1 presents Study 1, an experimental study of expressive news preferences and identity signaling in news selection. Chapter 2 presents Study 2, an observational study of hostile news commentary and cross-

cutting engagement patterns. Chapter 3 presents Study 3, a trace data study of engagement segregation across algorithmic and user-driven stages of news consumption.

Chapter 1. Expressive News Preferences: Identity-Signaling in News Selection

2.1 Introduction

The polarized media environment, where media consumption is sharply divided along partisan lines, has become a defining feature of today’s information landscape (Lelkes et al. 2017; Peterson and Kagalwala 2021). Research consistently shows that individuals are drawn to politically consistent content, reinforcing ideological divisions and fostering so-called echo chambers (Benkler et al. 2018; Tokita et al. 2021). This divide can be further reflected in survey responses, where partisan biases often influence how individuals express their preferences or beliefs. Studies on partisan cheerleading in fact demonstrate that survey respondents frequently align their answers with their political identities, even when those responses may not fully reflect their actual views (e.g., Bullock et al. 2015; Prior et al. 2015; Schaffner and Luks 2018; Yair and Huber 2020). These trends highlight a critical question for researchers interested in news consumption: do individuals’ expressed news preferences actually capture their genuine media consumption, or are they shaped by broader expressive motives tied to identity signaling?

This study considers the possibility that individuals’ expressed news preferences may diverge from their actual intention to read the news. In highly polarized media environments, where media outlets and their content are closely tied to partisan identities, individuals may express news preferences that serve to signal in-group alignment regardless of what they would actually choose to read. These expressive tendencies, when not followed by actual reading behavior, may lead to overstated selectivity in news diets.

To examine this possibility, this study employs a preregistered web-based experiment

testing whether people’s news preferences (i.e., news they say they are interested in reading) differ depending on whether they do or do not expect to actually read the articles afterward. Participants were asked to select news headlines on a politically divisive issue (i.e., gun control) under different engagement conditions—focused on either selection or consumption—and different levels of selection publicity. Across all conditions, participants selected the stories they were interested in reading, but only those in the consumption-focused condition were told they will be redirected to the full articles to consume after each selection, similar to browsing news online in real-worlds. This design isolates the effect of consumption expectation—an indicator of behavioral intent—on ideological news selectivity, separating it from selection behavior made without that expectation, a distinction that previous studies have not tested in a controlled setting.

Results reveal a clear expressive bias: participants in the selection condition (indicating their news preferences without expecting to read the articles) were significantly more likely to select attitude-consistent news than participants in the consumption condition (indicating their news preferences with the expectation of actually reading the articles). Participants in the latter condition showed more diverse, cross-cutting news choices. This gap in ideological selectivity persisted regardless of partisanship strength—though moderate partisans showed greater bias—and regardless of whether selections were made in public or private settings. These findings reflect what this study refers to as expressive news preferences: the tendency for individuals to select news that aligns with their political attitudes but is not necessarily intended to be read. In such cases, news preferences reflect more identity signaling than is likely evident in real-world news environment where individuals intend to actually engage with the content.

These findings have two key implications. First, they help explain why experimental studies often report stronger ideological selectivity than large-scale behavioral studies using web-tracking data (Tyler et al. 2022): the difference may lie, at least in part, in whether consumption is expected. Second, the findings suggest that headline selection tasks, which are

commonly used to measure news preferences, may not only reflect respondents’ actual reading preferences but also identity-signaling motives that exaggerate those preferences. The implications extend beyond measurement bias within survey settings, since identity-signaling motives may also shape media engagement behaviors observed in real-world contexts, particularly in today’s polarized information environment.

2.2 Partisan Expressive Responding

Prior research identifies a phenomenon known as expressive responding, in which individuals intentionally misreport their beliefs in responses to factual political questions in order to signal partisan identity and sentiment. In such cases, individuals align their answers with their partisan affiliations even when those responses do not reflect their actual beliefs or knowledge, using their responses as a way to signal support for their political camp.

Seminal studies demonstrate this pattern by showing that incentives designed to increase response accuracy reduce the partisan alignment that appears under unincentivized conditions (Bullock et al. 2015; Prior et al. 2015). For example, Bullock et al. (2015) found that even minimal monetary incentives could reduce partisan differences in beliefs about factual issues such as economic conditions during the Clinton and Reagan administrations. They also observed that a larger proportion of respondents chose to answer “don’t know” to factual questions when offered financial incentives. Prior et al. (2015) similarly found that monetary incentives and on-screen appeals for accuracy significantly diminished partisan biases in responses concerning objective economic conditions.

Although earlier seminal studies indicate expressive responding, the effects of financial incentives have not been consistently observed (Flynn et al. 2017), and many studies have found that financial incentives have little or no effect in reducing expressive responding (Malka and Adelman 2023). For instance, (Allcott et al. 2020) found that the partisan gap in predicting COVID-19 case numbers did not shrink when financial incentives were offered, suggesting no expressive responding in reporting their predictions. Peterson and Iyengar

(2021) likewise found no evidence that financial incentives produced meaningful changes in news headline choices—an area they argued may also be susceptible to partisan cheerleading.

Even with non-financial incentives, studies have found little or no evidence of expressive responding. Berinsky (2018) consistently found no expressive responding regarding political rumors across several interventions intended to increase response accuracy, such as asking respondents to ignore their feelings, encouraging honesty, or incentivizing them with a shorter survey. Graham and Yair (2024) similarly found no or only limited expressive responding regarding claims of 2020 election fraud when using honesty encouragement, response substitution (i.e., providing a prior question to channel expressive endorsement), list experiments (i.e., offering an anonymous channel for endorsement), as well as financial incentives. Fahey (2023) also found no evidence of expressive responding in endorsements of electoral and public health misperceptions when respondents faced accuracy pressure or when response substitution was available.

Nevertheless, even if most responses are truthful rather than expressive, prior studies suggest some portion may still reflect expressive bias that cheerleads partisan affiliations (Malka and Adelman 2023), thereby potentially distorting our understanding of partisan bias. Even Berinsky (2018), whose study reported no expressive responding, estimated that up to 15% of respondents endorsed a political rumor—an upper-bound estimate of possible expressive effects. Peterson and Iyengar (2021) also found no significant effect of financial incentives on news headline choices, yet the correlation between experimental headline choices and real-world news traces remained modest ($r = 0.58$), leaving open the possibility of expressive bias.

In addition, the findings of no expressive responding are by no means conclusive. Null or limited effects of interventions designed to increase response accuracy do not necessarily rule out expressive responding; rather, they may indicate that the interventions were not strong enough to reduce it, rather than demonstrating that expressive bias does not exist.

A lack of effects from financial incentives, for example, does not necessarily mean

that responses given without incentives reflect true beliefs (Flynn et al. 2017). There is no definitive way to determine which responses more accurately capture individuals’ genuine perceptions, as individuals may adjust their answers depending on the considerations that become salient across contexts (Zaller 1992), or may simply provide whichever answer secures the incentive rather than the one that reflects their actual beliefs. The same reasoning applies to non-financial incentives: responses provided under these conditions cannot be assumed to be free of expressive influence.

It is also difficult to identify respondents’ true beliefs when the political factual questions used to test expressive responding are so challenging that people cannot realistically provide a correct answer from memory, regardless of whether accuracy incentives are offered—for instance, the exact change in unemployment from 2009 to 2012 or the percentage of TARP funds repaid by March 2012 (Bullock et al. 2015), or the precise number of COVID-19 cases in April 2020 (Allcott et al. 2020).

Studies that rely on neither financial incentives nor difficult factual questions have, in fact, found significant expressive bias favoring respondents’ political identities (Schaffner and Luks 2018; Yair and Huber 2020), for example, compared responses to visually obvious photographic evidence of the crowd sizes at the inaugurations of Donald Trump and Barack Obama and found that respondents’ answers favored their preferred president despite clear visual evidence to the contrary. Yair and Huber (2020) observed a similar pattern when participants rated the attractiveness of target profiles that signaled partisan affiliations—an evaluation that requires no political knowledge.

The discussion above suggests that expressive responding may not always be detected under existing interventions and therefore warrants further examination using different approaches. My study accordingly introduces a novel experimental design that compares self-reported responses with actual behavioral indicators in a controlled setting, focusing on news preferences—an area closely tied to partisan identity and thus potentially prone to expressive responding (as elaborated in the next section).

This design avoids relying on incentive-based methods or difficult factual questions and instead directly compares isolated self-reports with their behavioral counterparts. Such direct comparisons have rarely been examined, yet they are important because expressive responding likely entails a divergence between expressed reports and corresponding behavior—what people say they would do or think versus what they actually do or think—beyond the shifts in survey responses observed across different conditions. If, as prior studies suggest, expressive responding occurs when individuals align their answers with their political identities in ways that do not accurately reflect their sincere beliefs or intentions, then these expressed preferences should diverge from behavior. This design therefore tests an important yet underexplored aspect of expressive responding in news selection tasks: if expressive responding is present, self-reported news selection made in isolation should exhibit stronger partisan alignment than the corresponding behavioral choices.

2.3 Expressive News Preferences in Polarized Media Environment

Although previous studies on expressive responding have focused primarily on beliefs or attitudes about factual questions, expressive bias may extend to any domain associated with political affiliations and capable of signaling a desired political image.

One domain where expressive responding may plausibly emerge is individuals’ news preferences in today’s polarized media environment. A substantial body of research documents this environment, in which media readerships are sharply divided along partisan lines (e.g., Benkler et al. 2018; Flaxman et al. 2016; Iyengar and Hahn 2009; Peterson et al. 2021; Stroud 2008, 2011; Tokita et al. 2021). In this context, news preferences are often interpreted as cues to political identity. Choosing certain media outlets over others therefore signals more than personal taste; it frequently reflects partisan alignment. As Prior (2013) noted, partisans may “forget, underestimate, or fail to admit” their exposure to counter-attitudinal content while readily reporting engagement with their own side, thereby producing self-reports that overstate selective exposure. Individuals may also experience discomfort when

deviating from their desired social identity or image (Keltner and Haidt 1999; Suhay 2015), which may further push them toward expressive responding. In this sense, today’s polarized media landscape may create conditions in which expressing news preferences becomes politically charged, prompting individuals to align their stated preferences with partisan expectations—even when these preferences do not reflect what they actually want to read. This tendency reflects what I refer to as expressive news preferences.

To examine whether expressive tendencies influence how people indicate their news preferences, an important question is whether these expressed preferences correspond to the news they actually intend to consume. If the former exhibit stronger partisan alignment, it would suggest the presence of expressive news preferences. Although not always directly compared, prior studies suggest different tendencies in news preferences between controlled setups and real-world tracking. As Tyler et al. (2022) point out, there is a general mismatch between findings from experimental and webtracking studies: experiments often reveal strong partisan selectivity (e.g., Iyengar and Hahn 2009; Stroud 2008), whereas large-scale behavioral data from web-tracking studies show more ideologically diverse news consumption (e.g., Gentzkow and Shapiro 2011; Guess 2021).

Peterson and Iyengar (2021) directly examined this issue by comparing respondents’ chosen headlines in a survey experiment with their real-world web-tracking behavior. They found that the partisanship of survey-based source choices and real-world browsing was not substantially divergent ($r = 0.58$), though not strongly aligned either, and that the extent of partisan divide—calculated as the difference between the average media choices of Democrats and Republicans—was similar across the two measures.

Because their setup does not allow direct comparison between survey responses and actual behavior using identical news options, however, we still have little insight into what the same individuals would have chosen to read in the absence of expressive influence, relative to their survey choices. This gap underscores the need for a more controlled design to test these differences directly.

My study therefore directly compares the news people say they want to consume with the news they actually intend to consume in a controlled experimental setup. Although this approach limits the broad range of options available in the real-world environment examined by (Peterson and Iyengar 2021), it enables a direct comparison of news choices under identical choice sets. To isolate what people report from their actual informational interests (i.e., the intention to consume the content), this study compares news choices made without the expectation of consumption and those made with the expectation of reading the content afterward. The first setup reflects an expressive condition —similar to survey selections—in which participants choose headlines without expecting to read them, a context more conducive to symbolic or expressive motives than informational ones. The second condition mirrors a behavioral condition—akin to real-world browsing—in which clicking a headline directs individuals to the full article, thereby creating an immediate expectation of engagement. If expressive responding is at play, selections made without the expectation of consumption should exhibit stronger partisan alignment than selections made with the expectation of actually reading the content. This yields the following hypothesis:

H1: *Individuals will select attitude-consistent news more frequently when selecting without the expectation of consumption than when selecting with the expectation of consumption.*

In addition, as suggested by research on partisan cheerleading in survey responses, those with stronger partisan attachments may be more motivated to engage in expressive responding (Malka and Adelman 2023) and thus more likely to express identity-aligned news preferences when selection carries no behavioral indicators (i.e., actual reading). This study therefore hypothesizes that the gap in attitude-consistent news selections across these two conditions will be greater among individuals with stronger partisan identities.

H2: *The difference in attitude-consistent news selection between the two conditions will be greater for strong partisan than weak partisans or independents.*

Beyond individuals’ partisan attachments, situational contexts may also shape incen-

tives for expressive responding. Public situations, in which news choices are visible to others, can alter how individuals make those choices. According to impression management theory, individuals seek to control how they are perceived (Goffman 1959; Leary and Kowalski 1990) by engaging in strategic self-presentation in both online (Marder et al. 2016; Rosenberg and Egbert 2011) and offline contexts (Hart et al. 2020). In such settings, news choices may become more susceptible to expressive bias, as people align their public selections with their political identity.

Empirical evidence supports this possibility. Hart et al. (2020) distinguish information display (i.e., displaying one’s information preferences) from information processing (i.e., reading) and show that information display is more ideologically congruent. They found that displaying content in public, such as holding a book visibly without reading it, can function as a signal of one’s identity, and that doing so with identity-incongruent content produces psychological discomfort, making individuals less likely to display such content. They also showed that people express greater intent to read identity-congruent content, though this measure captures expressed intent and may still be shaped by identity-signaling rather than informational interests. In a related online experiment, Coppini et al. (2017) found that the visibility of news choices activates both identity-management motivations (using news to express identity or manage one’s public image) and opinion management motivations (using news to shape one’s own or others’ views). They further showed that public conditions—where participants were told their choices would be visible—made politically congruent news more likely to be shared, particularly among conservatives, compared to private conditions in which choices were described as invisible to others.

Prior research also shows that identity signaling in public situations occurs particularly through the sharing of news stories, a behavior often referred to as selective sharing (Shin and Thorson 2017). Bergamaschi Ganapini (2023) argues that sharing fake news in adversarial contexts can serve a signaling function that demonstrates commitment to one’s side, facilitates cooperation, and reinforces identity-oriented narratives. Examining this be-

havior on Twitter, Shin and Thorson (2017) found that fact-checks favorable to Democrats or Republicans were retweeted predominantly by co-partisans. In an experiment replicating Facebook, Johnson et al. (2020) found that selective sharing (i.e., the intention to share congenial news) exhibited even stronger partisan alignment than selective exposure (i.e., reading time). Selective sharing also appears to be the dominant pattern on Twitter, despite a meaningful amount of cross-cutting sharing (Liang 2018).

Given findings that displaying information in public can serve as a signal of identity, and that selective sharing occurs in contexts where shared content is visible to others, the visibility of news choices may play an important role in shaping expressive bias when individuals express or display their news preferences. This leads to the following hypothesis:

H3: *The difference in attitude-consistent news selection between the two conditions will be greater in public settings than in private settings.*

2.4 Research Design

The experiment that follows is designed to test whether individuals are more likely to select attitude-consistent news when they indicate their preferences without the expectation of reading the articles, compared to when they select news knowing they will consume the content. If an expressive bias exists, participants are expected to choose more attitude-consistent news when there is no expectation of consumption—a context more conducive to identity signaling (H1). This tendency is also expected to be greater for stronger partisans (H2) and in public settings (H3).

The three hypotheses outlined above were preregistered before conducting the analysis, along with additional hypotheses not discussed in the present article for the sake of brevity. These omitted hypotheses tested the moderating effects of partisanship and impression management motivations, but none produced significant results. Full details on these additional hypotheses and their tests are provided in Appendix 5.3.¹

¹This study was preregistered on the Open Science Framework and is available at:

To examine the hypotheses, a web-based experiment was conducted in which participants reported their news preferences on a politically divisive issue (gun control) under varying conditions. The study was fielded through the online survey company Verasight from August 22 to August 27, 2024 ($N = 1,280$ U.S. adults aged 18 and older). Respondents were recruited through Verasight’s verified nonprobability panel, which uses multiple recruitment methods, including address-based sampling, person-to-person text messaging, and dynamic online targeting. All respondents underwent multi-step identity verification, including SMS authentication with U.S.-based mobile carriers, IP verification, bot-screening, and survey invitations. The survey took approximately 10-15 minutes to complete, and participants received a \$5 incentive upon completion. To ensure data quality, respondents who failed attention checks, exhibited straight-lining, or completed the survey in less than 30% of the median completion time were removed.

In the final sample, 55.6% of respondents identify as male and 44.1% as female. Age distribution is as follows: 9.8% are in their 20s or younger, 26.2% in their 30s, 17.3% in their 40s, 16.9% in their 50s, 16.6% in their 60s, and 13.2% are aged 70 or older. The racial composition of the sample is 72.6% White, 11.6% Black, 7.7% Hispanic, 4.7% Asian, and 1.7% (22) Native American.

Using a two-by-two experimental design, participants are randomly assigned to one of four conditions that combine (a) two news engagement modes (selection-focused or consumption-focused) and (b) two levels of selection “publicity” (public or private).

For news engagement modes, participants are asked to select five headlines on gun control that they would be interested in reading. In the selection-focused condition, participants are told they will not read the full articles after making selections and that the task will end immediately afterward—allowing them to indicate interest without any expectation of actual consumption. In the consumption-focused condition, participants are informed that they will be automatically directed to the full article after each selection and

<https://doi.org/10.17605/OSF.IO/J8FG9>.

can spend as much time reading as they wish and freely return back to the news homepage at any point—closely mirroring real-world digital browsing, where clicking on a headline leads directly to the full story. They continue this process until they select and view five articles.

It is important to note that in both conditions, participants are instructed to select news headlines they are interested in reading. The key difference lies in whether the selection is made with or without the expectation of actually engaging with the full content. Comparing these two conditions allows for an assessment of whether expressive bias is more likely to occur when individuals select news without expecting to consume the content.

For publicity (or “public-ness”), participants are presented with hypothetical scenarios describing how visible their news selections would be to others. In the public condition, participants are asked to carefully read a scenario in which their news selections are visible to others on a social media platform. Conversely, in the private condition, the scenario describes a situation in which their news selections are entirely private and not visible to others while browsing social media. To reinforce these contexts, participants are shown corresponding images: a person browsing online while connected to multiple users for the public condition and a person browsing alone for the private condition. These images remain visible at the top of the screen throughout the news selection task, reminding participants of their assigned scenario. Participants are instructed to imagine themselves in the described situation. The scenarios and images are provided in Appendix 5.1.

2.4.1 Procedures and Materials

Participants first complete pre-treatment questions about their attitudes toward various topics, including the topic of interest, gun control. They are then randomly assigned to one of four experimental conditions: selection-focused public, consumption-focused public, selection-focused private, and consumption-focused private. In all conditions, participants are presented with the same set of twelve news stories about gun control—half pro-gun control and half anti-gun control—and are asked to select five stories they are interested in

reading.

The twelve gun control news stories (headlines, lead texts, and full articles) were adapted from real-world news sources and align with different gun control arguments identified in previous studies (Cragun 2022; Kiil 2024). The stories are displayed in randomized order, and all headlines, lead texts, and full articles are kept at similar lengths.

To provide clear partisan cues, the story headlines are highlighted in distinct colors—red for anti-gun control news and blue for pro-gun control news—and participants are informed which color represents each stance. In addition to the partisan framing embedded in the news content, these visual cues are expected to reinforce ideological positioning of the news stories. It is worth noting that this approach differs from that of (Peterson and Iyengar 2021), who used news sources to signal the partisan orientation of headlines, as well as from earlier studies that relied on outlet cues in headline selection tasks to examine selective exposure (e.g., Iyengar and Hahn 2009). Recent research, however, suggests that people may align more strongly with the content of news stories than with their sources (Green et al. 2025). This study accordingly employs story-level argumentative cues to signal political alignment. Moreover, to reduce potential bias arising from variation in participants’ political sophistication or knowledge (Lodge and Taber 2013; Prior 2007), the partisan orientation of each headline is made visually explicit through color, ensuring that all participants can easily identify which articles align with their political identity without needing to infer this from the headlines.

After completing the news selection task, participants answer questions about their party identification, partisan strength and demographic information. Full details of the experimental materials and survey questions are provided in Appendix 5.1.

2.4.2 Measures

This study manipulates two aspects of news engagement: the mode of engagement (selection vs. consumption) and the level of selection publicity (public vs. private). Ac-

cordingly, there are two binary manipulated variables. Selection is a binary variable coded as 1 for participants assigned to the selection-focused condition and 0 for those assigned to the consumption-focused condition. Public is also a binary variable indicating the level of publicity in news engagement, coded as 1 for participants assigned to the public condition and 0 for those assigned to the private condition.

The primary outcome variable is *attitude-consistent news selections (ACNS)*, measured as the number of selected news articles that align with a participant’s attitude toward gun control. Participants’ attitudes toward gun control are assessed using the question drawn from the American National Election Studies (ANES): "Do you think the federal government should make it more difficult for people to buy a gun than it is now, make it easier for people to buy a gun, or keep the rules about the same as they are now?"

The response “Make it more difficult” is classified as pro-gun control, and “Make it easier” is classified as anti-gun control. In line with the preregistration, the middle response, “Keep the rules about the same,” is coded in two ways. First, given the relatively low level of gun control regulation in the United States, these respondents are coded as anti-gun control. Second, because these respondents are less anti-gun control than those who respond “Make it easier,” and it may therefore be inaccurate to assume that anti-gun control articles are attitude-consistent for them, models are also estimated excluding these respondents. In all cases, participants’ news selections are coded as attitude-consistent if they match their position on gun control, while other choices are coded as attitude-inconsistent. Since both treatments of “Keep the rules about the same” produce similar patterns in subsequent analyses, the results reported use the measure that includes all respondents, categorizing these responses as anti-gun control. Statistics based on the alternative measure are provided in Appendix 5.2.

Partisan strength captures the strength of participants’ political identification. Participants are asked the standard two-party party identification question drawn from the ANES. The first question is: “Generally speaking, do you usually think of yourself as a Democrat, a

Republican, an independent, or something else?” Those who identify as either Democrat or Republican are then asked a follow-up question: “Would you call yourself a strong [Democrat/Republican] or a not very strong [Democrat/Republican]?” Participants who identify as independent or “something else” are asked whether they lean toward either the Democratic or Republican party. Based on these responses, participants are classified into four categories: strong Republican/Democrat, weak Republican/Democrat, leaning Republican/Democrat, or independent. Partisan strength is then coded on a scale from 0 (independent) to 3 (strong partisan), with 1 indicating leaning partisan and 2 indicating weak partisan.

2.5 Results

The primary outcome variable in this study is the number of attitude-consistent news selections (ACNS). The analysis focuses on whether ACNS differs between the consumption-focused condition, in which participants select headlines with the expectation of reading the full articles after each selection, and the selection-focused condition, in which participants select headlines without any expectation of reading the articles. If expressive bias is present, ACNS should be higher in the selection-focused condition than in the consumption-focused condition. The analysis then assesses whether this difference between conditions is moderated by partisan strength and selection publicity. Based on the preceding discussion, individuals with stronger partisan identities, as well as those in the public (versus private) condition, are expected to exhibit a larger gap in ACNS between the consumption-focused and selection-focused conditions.

To provide context for the analyses that follow, it is useful to first examine overall patterns in the outcome variable and news selection across the two gun-control stances in the experiment. On average, respondents selected 3.17 ($SD = 1.29$) attitude-consistent headlines. Because each respondent selected five stories, values above three indicate that individuals tended to choose slightly more attitude-consistent news overall. Across all conditions, participants also selected pro-gun-control articles more frequently than anti-gun-control ones

(55% vs. 45%), a statistically significant difference ($\chi^2(1) = 57.38, p < .001$). Selection frequencies also differ across the twelve headlines ($\chi^2(11) = 120.95, p < .001$). The most frequently selected headline was a pro-gun-control story about a mass school shooting (chosen by 51% of participants), whereas the least selected was an anti-gun-control story about constitutional gun rights (chosen by 33% of participants). Note that these differences in headline popularity do not undermine the betweensubject comparisons that follow, as all participants, regardless of condition, were exposed to the same set of headlines in randomized order. The key test is whether the relative tendency to select attitude-consistent news differs across conditions.

To test the hypotheses, this study estimate an ANOVA model with ACNS as the dependent variable. The model includes the main effect of selection mode, along with two interaction terms: the interaction between selection mode and partisan strength and the interaction between selection mode and selection publicity. The main effect of selection tests H1, while the interaction terms test H2 and H3, respectively. Figures 1 and 2 present raw group means with 95% confidence intervals for descriptive interpretation, whereas all hypothesis tests are based on the ANOVA model.

Figure 1 displays average ACNS for the consumption-focused and selection-focused conditions, averaged across partisan strength and selection publicity. The pattern is clear: consistent with H1, participants in the selection-focused condition exhibit significantly higher levels of ACNS ($M = 3.32, SD = 1.27$) compared to those in the consumption-focused condition ($M = 3.03, SD = 1.29$) ($F = 18.510, p < .001$).

Note that ACNS ranges from 0 to 5, so the difference of 0.30 represents 6% of the scale and roughly a 10% increase from the consumption-focused to the selection-focused condition. Because baseline selectivity is already high (3.03–3.32 ACNS on average), there is limited room for treatment-induced change. The presence of expressive bias on top of this already strong baseline of attitude-consistent selection therefore makes the observed shift substantively meaningful.

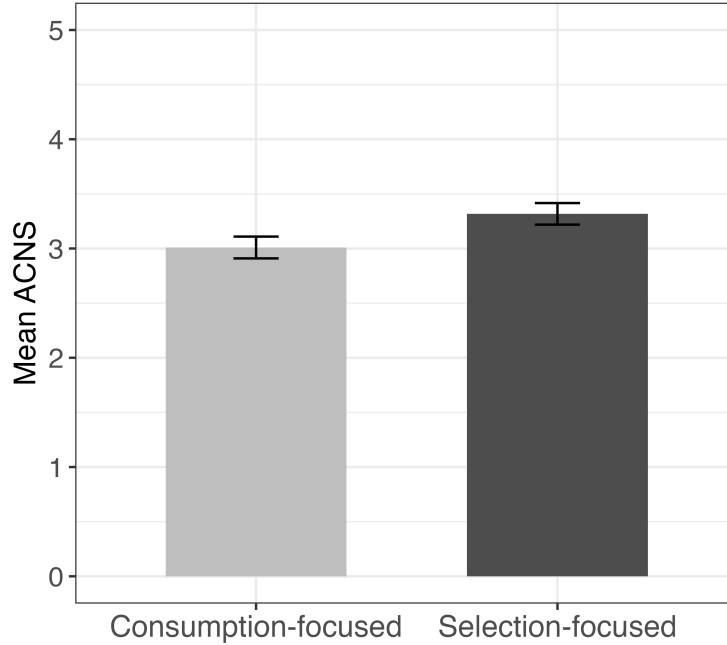


Figure 2.1: Mean ACNS by news engagement modes

Figure 2 presents average ACNS by partisan strength (left panel) and by selection publicity (right panel). As shown in the left panel, all partisan groups select more attitude-consistent news when selecting headlines without the expectation of reading the articles than when they expect to read the articles. However, the magnitude of this difference does not vary significantly across partisan-strength groups; the interaction between selection mode and partisan strength is not statistically significant ($F = 1.30$, $p = .27$). H2 is accordingly not supported.

For exploratory purposes—beyond the preregistered hypotheses—simple comparisons based on estimated marginal means are examined within each partisan group. These comparisons reveal that the difference between consumption- and selection-focused conditions is statistically significant and most pronounced among weak partisans and leaning partisans. Weak partisans exhibit the largest gap (diff = 0.50, $M = 2.89$ in consumption-focused vs. $M = 3.40$ in selection-focused, $t = 2.18$, $p < .05$), followed by leaning partisans (diff = 0.47, $M = 3.02$ vs. $M = 3.49$, $t = 2.54$, $p < .05$). Strong partisans also show a significant difference between conditions, though the magnitude of the gap is smaller (diff = 0.22, $M = 3.07$ vs.

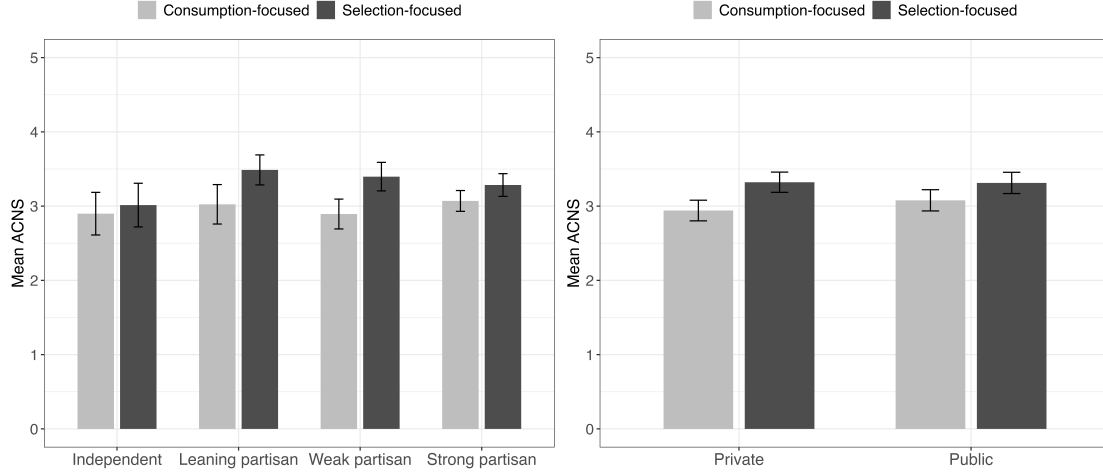


Figure 2.2: Mean ACNS by partisanship and selection publicity

$M = 3.28$, $t = 2.18$, $p < .05$). Independents, while selecting less attitude-consistent news overall, exhibit only a modest and insignificant difference between conditions (diff = 0.11, $M = 2.90$ vs. $M = 3.01$, $t = -0.48$, $p = .63$), suggesting that they have weaker political identities and are therefore less motivated by partisan signaling.

The larger gap observed among moderate partisans (weak and leaning partisans), relative to strong partisans or independents, suggests that they may be particularly susceptible to expressive motives, expressing more partisan-aligned preferences when selection carries no behavioral cost but making more balanced choices when consumption is expected. Consistent with this pattern, an exploratory analysis—conducted beyond preregistered hypotheses—confirms a significant moderating effect for moderate partisans compared to the combined group of strong partisans and independents ($F = 4.23$, $p < .05$).

While the expectation was that stronger partisans would exhibit more consistently biased news preferences, the results suggest a more nuanced pattern. As shown above, it is moderate partisans—those who identify as weak partisans or leaning partisans—who exhibit the greater divergence between the selection-focused and consumption-focused conditions. Strong partisans, although showing significant difference between conditions, exhibit relatively high levels of attitude-consistent selection in both conditions, suggesting that their

preferences may be more stable and habitual. Independents, meanwhile, show overall lower levels of attitude-consistent selection and no significant difference between conditions—likely reflecting their weaker partisan identities and thus a weaker motivation for expressive signaling. These patterns help explain why the linear measure of partisan strength does not significantly moderate the effect of consumption expectation. Expressive bias appears to be most pronounced among individuals whose partisan identities are salient enough to motivate signaling but not so entrenched as to drive habitual consumption.

The right panel of Figure 2 then presents average ACNS by selection publicity (public vs. private). The average ACNS values are quite similar across conditions (selection-focused: $M = 3.31$ public vs. $M = 3.32$ private; consumption-focused: $M = 3.08$ public vs. $M = 2.94$ private), and the moderating effect of publicity is not statistically significant ($F = 1.33$, $p = .25$). Accordingly, H3 is not supported. Although the pattern is partially in the expected direction—with slightly higher average ACNS in the public condition ($M = 3.20$, $SD = 1.31$) than in the private condition ($M = 3.13$, $SD = 1.27$)—this difference is also not statistically significant ($t = -0.89$, $p = .37$).

One possible explanation for this result may lie in the relatively weak publicity stimulus—although Coppini et al. (2017) used a similar imagined-publicity stimulus and reported partial effects. Participants in the present study were asked only to imagine selecting news in public or private contexts, rather than experiencing actual social exposure. Indeed, when asked whether they felt discomfort in these imagined contexts, a majority of participants in both conditions reported no discomfort (52% in the public condition; 63% in the private condition). Given that Hart et al. (2020) identify psychological discomfort over displaying identity-incongruent content as a key mechanism driving attitude-consistent display in public, the absence of strongly induced discomfort in this study may have limited the effect of the publicity treatment.

2.6 Discussion

The findings reveal a clear pattern of expressive responding in news preferences: individuals select more attitude-consistent news when they do not expect to consume the content than when they do. Participants in the no-consumption condition selected roughly 10% more attitude-consistent news than those who expected to read the articles. This suggests that when news selection carries no expectation of actual engagement, individuals are more likely to choose content that reflects symbolic or identity-driven preferences. Their actual reading intentions, however, appear more ideologically diverse.

It is important to note that news selections were still skewed toward attitude-consistent content across all conditions, indicating that most choices remain ideologically congruent. Prior work focused on factual beliefs finds a sizeable (sometimes 50–60%) reduction in partisan cheerleading when monetary incentives are provided (e.g., Bullock et al. 2015; Prior et al. 2015). To be clear, this study focuses on a substantively different outcome (i.e., news preferences), so the direct comparison with effect sizes from these studies may not always be appropriate. It also does not dispute the well-documented tendency toward attitude-consistent news consumption, nor the possibility that properly incentivized survey responses can still reflect sincere behavioral inclinations.

Even so, the preceding findings suggest that expressive responding to signal political identity can inflate the apparent strength of ideological alignment under circumstances where individuals do not anticipate actual information engagement.

The expressive bias observed in this study appears consistent across conditions, with no significant moderating effects of partisan strength or selection publicity. However, a closer look at the results reveals a more nuanced pattern. While strong partisans consistently selected politically congruent news across both conditions, it was moderate partisans—those who lean toward or identify with a party but less strongly—who exhibited the greater difference between selection-focused and consumption-focused settings. Strong partisans and

independents, by comparison, showed smaller differences across conditions, probably for different reasons: the former likely due to deeply internalized political commitments, and the latter due to weaker partisan identity and correspondingly weaker motivation to signal ideological alignment. These patterns suggest that expressive motives may be most influential when political identities are present but not firmly anchored.

The null effect of selection publicity may reflect limitations in the experimental manipulation. Participants were asked to imagine hypothetical public or private selection environments, which may not have produced the psychological discomfort associated with real-world public exposure. Consistent with this possibility, most participants reported little or no discomfort in these imagined scenarios. This aligns with prior work suggesting that psychological discomfort over displaying identity-incongruent content is a key mechanism driving public identity expression (Hart et al. 2020). Future research could therefore employ more immersive or ecologically grounded designs to more fully assess the role of publicity in expressive news preferences.

These findings also shed light on the limitations of selection-only measures in capturing everyday news behavior. While prior studies have noted potential gaps between stated preferences and actual engagement (e.g., Nelson and Webster 2017; Prior 2009; Tyler et al. 2022), this study directly isolates one possible mechanism—expectation of consumption—and shows how it shapes ideological selectivity. When people do not anticipate reading the articles they choose, their selections may reflect identity signaling more than genuine reading intent. This suggests that selection tasks used in surveys or experiments may overstate ideological polarization by capturing preferences that are partly performative.

The implications extend beyond measurement bias. According to impression management theory, individuals often seek to present themselves in ways that align with their social identities (Goffman 1959; Marder et al. 2016; Rosenberg and Egbert 2011). These expressive tendencies may be particularly visible in digital environments through behaviors such as sharing, liking, or curating content, where users publicly signal alignment with po-

litical groups. This highlights the importance of considering how identity signaling operates not only in survey responses but also across broader forms of digital engagement. Future studies should therefore examine how user engagement on digital platforms interacts with expressive motives in contemporary online information environments.

Several limitations of this study should nevertheless be acknowledged. First, the use of hypothetical scenarios to manipulate publicity may have limited the realism and effectiveness of the treatment. Future work could incorporate real-time visibility or social exposure to test the effects of public versus private contexts more rigorously. Second, although gun control is a highly salient and polarized issue, this study focuses on a single topic. The patterns of expressive bias observed here may therefore not fully generalize to issues with different moral, political, or social dynamics. Future research should examine whether expressive news preferences vary across topics and identity domains.

Despite these limitations, this study offers new insight into how identity-signaling motives shape news preferences. While political alignment characterizes news selection in general, expressive bias appears to emerge in low-cost, symbolic contexts. These findings highlight the importance of distinguishing between genuine informational interest and identity signaling when interpreting selection behaviors. As political polarization continues to shape the media landscape, understanding these expressive dynamics will be essential for interpreting observed media behaviors and the informational patterns they collectively produce.

Chapter 2. Engaging to Oppose: Cross-Cutting Patterns in Hostile News Commentary

3.1 Introduction

The theory of selective exposure posits that individuals prefer to consume information aligned with their existing beliefs. A large body of work has shown that people tend to select or consume pro-attitudinal content, thereby reinforcing ideological segregation and contributing to the formation of echo chambers (e.g. Iyengar and Hahn 2009; Peterson et al. 2021).

Emerging evidence suggests, however, that this pattern may not hold uniformly across all forms of engagement, especially in today’s digital media environment. Digital news engagement unfolds in phases—exposure, selection, consumption, and active response—that are conceptually and behaviorally distinct. Users may be exposed to stories they never select, select content they do not read, or comment on stories they did not seek out or read. While more passive behaviors such as selection and consumption tend to reflect information-seeking behavior consistent with users’ reading preferences, active engagement may reflect different motives, namely, a desire to express, confront, or reject.

Most research on user selectivity has focused on passive forms of news engagement. Far less is known about how more active forms fit into the broader picture of selective or cross-cutting news use. Commenting, as an active form of engagement, represents the most public and expressive mode of news engagement. It requires effort and time (Krebs and Lischka 2019) and is thus distinguished from other activities such as sharing or liking (Guo and Sun 2020). It is also often a venue for aggressive or uncivil expression (Coe et al. 2014; Masullo et al. 2021). This raises the possibility that comments may not occur in echo

chambers but rather in response to counter-attitudinal content, motivated by disagreement and hostility.

To examine this possibility, we analyze ten years of user commenting behavior on Naver News, South Korea’s largest news aggregation platform. Naver enables commenting across nearly all major outlets without paywalls, offering a unique opportunity to assess how hostile users engage with news content at both the story and source level. We ask whether users are more likely to comment on counter-attitudinal content, whether hostile users are less confined to outlet-based echo chambers, and whether cross-cutting engagement is more common in domains with higher hostility.

Our findings suggest that hostile commentary often reflects boundary-crossing engagement. Hostile users are more likely to target content they disagree with and exhibit weaker echo chamber structures within engagement networks. In today’s media environment, hostile commentary may thus signal not a deepening of ideological silos but rather an active disruption of what might otherwise be predominantly pro-attitudinal news engagement.

3.2 Media Selectivity and News Engagement Types

A substantial body of literature has documented individuals’ tendency toward selective news engagement, that is, a preference for pro-attitudinal information or ideologically aligned news outlets (Prior 2013). This pattern has been repeatedly demonstrated in empirical research, particularly through studies showing that people are more likely to seek out or engage with information that reinforces their existing views (e.g. Iyengar and Hahn 2009; Peterson et al. 2021; Stroud 2008, 2011). Selective exposure has also been linked to the formation of echo chambers, informational environments in which individuals are surrounded primarily by like-minded perspectives (e.g., Benkler et al. 2018; Flaxman et al. 2016; Tokita et al. 2021), further reinforcing attitudinal consistency and limiting exposure to cross-cutting content.

Much of the empirical evidence for selective exposure comes from studies that focus

on more passive types of news engagement. Studies of selection behavior have typically examined headline choices or click-through decisions in experimental settings (e.g., Iyengar and Hahn 2009; Jang 2014; Kim and Lu 2020), for instance, or have tracked domain-level browsing patterns using web-tracking data (e.g., Flaxman et al. 2016; Peterson et al. 2021; Tyler et al. 2022). Similarly, consumption, another relatively passive form of news engagement, has been analyzed through self-reported measures of newspaper reading (e.g., Stroud 2008, 2010) or through behavioral data such as time spent reading articles in experimental environments (e.g., Jang 2014; Knobloch-Westerwick 2012a,b).

Although many studies confirm the general pattern of media selectivity, there is still debate regarding its strength and scope. When relying on web visit data, for instance, Guess (2021) found that most individuals, regardless of political orientation, tend to consume a relatively moderate mix of media sources, whereas Tyler et al. (2022) and Peterson et al. (2021) observed clear patterns of partisan selectivity, particularly when attention was narrowed to political news content. Experimental studies have produced similarly complex findings. Several experiments have shown a nonavoidance tendency, where individuals do not systematically steer clear of counterattitudinal political information (Garrett 2009; Garrett and Stroud 2014), while others find that although people may not actively avoid opposing views, they still exhibit stronger tendencies to approach pro-attitudinal information (Jang 2014; Kim and Lu 2020).

Prior studies also suggest that story- and source-level news engagement can reveal different patterns of audience selectivity. For example, previous work has shown that echo chambers are often more pronounced at the story level than at the source level (González-Bailón et al. 2023), particularly on digital media platforms where user curators play an active role in organizing content that aligns with their predisposition (Green et al. 2025). Green et al. (2025) further theorize the process of “curation bubble,” in which user curation involves unbundling discrete pieces of information from their parent sources and re-bundling them into individualized streams of content. This process produces patterns of engagement that

appear integrated at the source level but are, in fact, highly segregated at the story level.

More active forms of engagement, such as sharing or commenting, nevertheless remain relatively underexplored in the context of media selectivity. There is reason to believe that active forms of engagement may follow different patterns and deserve closer conceptual and empirical attention. Specifically, exposure, selection, consumption, and participation (e.g., commenting or sharing) may vary in how intentional they are and what they reveal about user intent and behavior. One may be exposed to information without choosing it, select information without fully consuming it, or react to content they never actively sought out or consumed. These distinctions may matter, particularly in today’s digital media environment where users are constantly presented with a high amount of information but ultimately engage with only a small portion of it. Digital platforms allow, and may even encourage, these phases to unfold independently of one another. Research on media selectivity, however, has often assumed behavioral consistency across these phases, treating selection or consumption as sufficient to infer broader engagement patterns. This risks overlooking meaningful variation, particularly in active engagement, where expressive, emotional, or oppositional motives may underlie responses that diverge from more passive behaviors. Disentangling these phases is therefore critical to understanding the layered dynamics of audience selectivity in digital information behavior.

3.3 Cross-Cutting Potential in Active News Engagement

Compared to passive engagement, active engagement provides greater potential for animosity or hostility to emerge. This is because, whereas passive engagement primarily reflects quiet or implicit preferences, active engagement involves, by definition, users’ explicit expression of opinions and reactions; and individuals’ attitudes are more likely to be expressed when they are hostile or aggressive, as such negative emotional states tend to provoke greater engagement (Huddy et al. 2021; Valentino et al. 2011).

Importantly, hostility is associated with an action-oriented tendency to oppose. It is

defined as negative or antagonistic attitudes or beliefs toward others (Barefoot et al. 1994; Ermakov et al. 2016), reflecting a general disposition of mistrust, resentment, or ill will (Powell and Williams 2007), and it is often accompanied by the emotion of anger (Fabiansson and Denson 2016). Anger, an emotional state closely associated with hostility, is known for its high arousal potential (Russell 1980) and its strong association with increased engagement and participation (Berger 2011; Brady et al. 2017; Crockett 2017). It also carries an antagonistic impulse—a motivation to confront or fight against an object or target of hostility (Carver and Harmon-Jones 2009). Young et al. (2011) found that this emotional state may even lead users to deliberately seek out counter-attitudinal information to challenge it, due to its “move against” tendency. This suggests that anger or hostility may influence not only how people engage with news, but also what they choose to engage with in the first place.

Active forms of engagement may consequently have a high potential to reveal cross-cutting news engagement because of their association with hostility. Hostile commentary—the focus of this study—is a verbal, behavioral manifestation of hostility. It seems likely that hostile commentary will be directed at counter-rather than proattitudinal information—although this expectation warrants empirical testing. In short: when users encounter counter-attitudinal information, it may trigger emotional reactions such as anger, leading them not to avoid it but instead to respond to it—not to endorse, but to oppose or disconfirm it. Active engagement, therefore, may exhibit cross-cutting patterns shaped not only by what users support but also by what they reject or oppose.¹

Indeed, prior work finds evidence of cross-cutting behavior in active forms of news engagement. The literature has mostly focused on sharing rather than commenting, but even so, it finds evidence of cross-cutting patterns that appear to be driven by nonpreferences or animosity. For instance, Eady et al. (2019) found that retweets are more ideologically balanced than authored tweets, suggesting that sharing can involve not only pro-attitudinal

¹Note that the presence of uncivil commentary, which is typically hostile in tone (Coe et al. 2014; Masullo et al. 2021; Muddiman and Stroud 2017; Santana 2014; Stroud et al. 2016; Wulczyn et al. 2017), suggests that active forms of news engagement may indeed exhibit cross-cutting patterns since they are likely to be directed at counter-attitudinal information.

but also counter-attitudinal content. Similarly, Rathje et al. (2021) showed that Facebook and Twitter posts about political out-groups were shared and retweeted approximately twice as often as those about in-groups, and these posts tended to provoke more “angry” reactions. Yu et al. (2024) similarly found that Twitter users are more likely to share negative tweets about an out-party than positive tweets about their in-party.

Commenting behavior is likely to be at least as expressive as sharing, and likely more so. Commenting allows users to directly express their perceptions rather than merely forwarding content. Commenting may also provide a clearer window into users’ attitudes than more passive forms of news engagement, where underlying motivations often remain ambiguous. These passive behaviors may reflect partisan selective exposure to ideologically aligned content (Iyengar and Hahn 2009; Stroud 2008, 2011), information utility (Knobloch-Westerwick and Kleinman 2012; Valentino et al. 2009), or biased self-reports (Prior 2013). Given the expressive and relatively unambiguous nature of news commentary, cross-cutting patterns may be more readily apparent.

The discussion thus far suggests that commentary may be closely associated with cross-cutting engagement, as hostile responses are often directed at counter-attitudinal content. We nevertheless lack clear evidence about the structure of this hostility in news commentary: Does it typically occur within ideologically selective echo chambers, where users criticize opposing views without ever leaving their in-group media environments? Or do users frequently cross ideological or outlet lines to leave hostile commentary on content they disagree with? These are open empirical questions—and ones this study seeks to address.

To do so, we draw on large-scale digital trace data to examine selective or cross-cutting patterns of news commentary at both the story and source levels, in two respective studies. At the story level, we examine whether users are more likely to comment on counter-attitudinal news stories than on pro-attitudinal ones (RQ1). The sourcelevel analysis then assesses whether individuals who post hostile comments are less likely to form echo chambers (RQ2), and whether cross-cutting commentary is more prevalent in news domains where

hostility occurs more frequently (RQ3).

This approach helps fill a gap in the literature by focusing on the most active and expressive forms of news engagement—largely underexplored in the context of media selectivity—and by engaging with ongoing debates about how selectivity may differ across levels of analysis.

3.4 Data

Both of the studies that follow rely on ten years of news comment data, from January 2011 to September 2020, collected from Naver News, the largest online news aggregating platform in South Korea. South Korea is known for high levels of digital news consumption, and Naver News stands as the leading online news aggregation platform, handling most of the country’s digital news traffic. As of 2021, approximately 87% of portal site users accessed news through Naver (Korea Press Foundation, 2021, p. 87).

This dataset offers two key benefits for studying news engagement. First, the platform gathers articles from virtually every major national outlet and allows users to read, react to, and comment on news content directly within the platform, without being redirected to external sites. To post a comment, users click on a headline, scroll to the bottom of the story, and write and submit their comment after signing in. This open-access structure enables engagement across outlet boundaries, making Naver an ideal setting for studying large-scale public news commentary. Second, because Naver assigns identification codes to each user (distinct from their actual account IDs), we are able to track individual-level news engagement over time. This granularity, combined with the platform’s scale and open structure, allows for a rare examination of cross-cutting or selective engagement patterns in online news commentary within a national media ecosystem.

Note that, unlike social media platforms where algorithmic systems actively curate content that aligns with users’ preferences, news aggregators are often designed to compile stories from a wide range of outlets. This facilitates at least some cross-cutting exposure.

News commentary still requires a series of deliberate user actions, however: clicking on an article, scrolling through it, writing a comment, and posting it. Initial exposure to opposing headlines may thus be incidental or involuntary, but users are under no obligation to engage further. They can simply ignore incongruent headlines, as assumed in prior selective exposure studies in which individuals are expected to avoid clicking on dissonant news (e.g., Iyengar and Hahn 2009). If our data reveal cross-cutting patterns in news commentary, then, our inference is that such behavior reflects deliberate user decisions rather than outcomes driven merely by the platform’s design.

During the study period, Naver News published a daily list of the thirty most-read articles (“Ranking News”) in each of six news domains (defined by Naver News): Politics, Society, Economy, World, IT/Science, and Life/Culture. These lists, including 180 articles each day, were compiled from January 2011 to September 2020. Every news comment for all listed articles was then collected, including the user identification code of the commenters, as well as information on the news sources, articles, and the six news domains.

While we recognize the interactive nature of comment sections—where user comments, rather than news itself, can trigger subsequent replies—our dataset includes only parent comments (those posted directly under news stories) and not child comments (replies to other users). On this platform, child comments are automatically hidden and remain collapsed unless a user actively clicks to view them. Because our focus is on the interaction between news stories and commenters, rather than commenter-to-commenter exchanges, we analyze only parent comments. These are more likely prompted by the news content itself, whereas users responding to others’ remarks would typically post replies within subthreads rather than as new parent comments.

Also, while most comments were posted on the day the news articles appeared on the most-read list, some comments were posted much later. These late comments may not share the same context as those made on the day of peak attention for the news stories. Comments posted more than three days after the article’s listing date (less than .03% of the

sample) are accordingly excluded from the analyses that follow.²

Naver News introduced a feature in March 2017 allowing users to express various reactions to news articles, including an “angry” reaction. In April 2016, Naver News also began disclosing story-level aggregate demographic information about commenters, such as generation and gender, once an article received at least one hundred comments. Our data include counts of angry reactions and aggregate demographic information from the time that each was available.

Our final working dataset includes 251,876,173 comments posted by 5,818,356 unique users over a period of ten years. There are more comments and users from later in the time period. Figure 3.1 shows the number of comments and unique commenters over time.

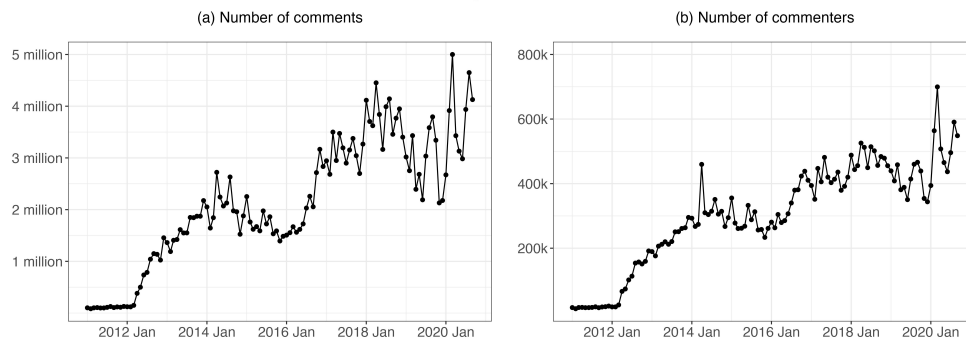


Figure 3.1: This figure shows the monthly number of news comments (panel a) and unique commenters (panel b) on Naver News from January 2011 to September 2020. Both measures increase steadily over time, indicating the platform’s growing scale of user participation in public news commentary. (a) Number of comments. (b) Number of commenters.

3.5 Study 1: Content-Level Engagement

Examining cross-cutting patterns across news content requires identifying both the stance of the news and that of the users. To do this, we focus on users’ commenting behavior on a subset of politically salient news stories on a politically divisive policy issue in the country: the phasing out of nuclear power. This policy was implemented by a liberal government

²As news tends to be timely, few users likely seek out and comment on older popular stories. Therefore, removing these late comments is not expected to significantly affect the overall dataset.

in South Korea in 2017 and sparked opposition from conservatives, who favored the cost-effective electricity supply provided by nuclear power plants. Liberals, in contrast, exhibited strong support for the policy, viewing it as a step toward transitioning the country’s energy supply system to renewable energy sources (For further background on this issue, see, e.g., Nam et al. 2022).

3.5.1 Methods

To measure content-level news engagement, we first identify articles on the issue by searching for article titles containing relevant nuclear power keywords published during the peak of policy debates in 2017. There are 108 such articles. We then examine the 102,256 comments attached to those articles.

Each article and comment is categorized as either (1) opposing the policy, (2) supporting the policy, or (3) taking a neutral stance towards the policy. The articles were manually coded, resulting in thirty-two policy-opposing articles, thirty policy-supportive articles, and forty-six policy-neutral articles. Comments were then classified as policy-supporting or policy-opposing using a transformer-based model trained on a manually coded sample comprising 5 percent of the comments.³ This process yielded 66,747 policy-opposing comments and 35,509 policy-supportive comments.⁴ Individual users were categorized as either proponents or opponents of the policy, depending on the balance of policy positions evident in their comments. If the majority of an individual’s comments supported the policy, for instance, that individual was categorized as a proponent (16,126 users); if a majority of their comments opposed the policy, they were categorized as an opponent (21,671 users).⁵ If cross-cutting engagement dominates, proponents should comment more on policy-opposing news stories, while opponents should comment more on policy-supportive ones.⁶

³We adopted kcELECTRA-base-v2022, pre-trained on Korean comments.

⁴To test the validity the author’s coding, a research assistant coded the same data. Inter-coder reliability was high, with Krippendorff $\alpha = .87$ for the articles and $.94$ for comments.

⁵We excluded 1,369 users (4 percent) without a clear stance on the policy. We similarly do not present articles classified as neutral.

⁶Additional methodological details are provided in Appendix 6.7.

We measure hostility in news commentary using the same transformer-based model, trained in this instance to replicate Cleanbot, an automated moderation system developed by Naver. The Cleanbot system was introduced in June 2020 to detect and flag offensive or inappropriate comments. Our model allows us to extend the detection of hostile comments to the full ten-year dataset, including periods before the official launch of Cleanbot. The model was trained on a sample of 61,440 comments labeled by Cleanbot from its initial implementation and the four months following. Model performance was validated against 7,680 randomly sampled comments from the same period, achieving 95.1 percent classification accuracy. Using this classifier, each comment in our dataset was categorized as either hostile or non-hostile. The monthly percentage of hostile comments was calculated for aggregate-level data. For user-level data, the total number of hostile comments by each individual in a given month was counted.⁷

Note that we rely on Cleanbot-labeled data in part because of its ready availability in our corpus, even as we recognize the categorizations made by Cleanbot may capture comment hostility imperfectly. That said, Naver’s description of Cleanbot is largely in line with our conception of overtly hostile commentary. The feature was intended to automatically hide potentially offensive comments, creating a more civil and less hostile news engagement environment for any user who activated it. According to Naver, it uses deep-learning techniques to detect six types of hateful and denigrating speech in the news comment section: (1) common swear words from a predefined dictionary, (2) offensive vulgar language, (3) sexually provocative expressions, (4) language implying physical threats, (5) language that discriminates based on region, race, nationality, etc., and (6) humiliating and deprecating language.

While there may be concerns about repetitive commenting behavior—where the same users repeatedly post to support one side or intentionally disrupt the other, potentially

⁷We measured each individual’s behavior by count rather than percentage, since comment frequency varies widely across users; using proportions would understate the hostility among frequent commenters and overstate it among non-frequent ones.

biasing the estimation of cross-cutting engagement—such behavior is negligible in our study. Duplicated comments (those that appear more than once during the Study 1 period) account for only 3.6 percent of all comments.

3.5.2 Results

To assess content-level news engagement, we examine whether users commented more on policy-supportive or policy-opposing articles. Figure 3.2 displays the average number of total comments (top panels) and hostile comments (bottom panels) posted by opponents (left panels) and proponents (right panels) of phasing out nuclear power. In each cell, the average number of comments is presented separately for articles categorized as either opposing or supporting the policy. And within each article category, the comments themselves are separated by whether they were supportive or opposing the policy.

The pattern in Figure 3.2 is relatively clear: users are more likely to comment on counter-attitudinal news articles than on pro-attitudinal ones. Both opponents and supporters of the nuclear phaseout policy disproportionately commented on stories that opposed their own stance—often to challenge the article. Notably, this pattern holds not only in total comments, but also in the subset of comments identified as hostile.

Individuals who opposed the policy (top-left panel) left comments that were nearly universally opposed to the policy. Our categorization of these policy-opposing users is defined by their comments, of course—so we should expect there to be a greater number of opposing than supporting comments in this panel. Even so, the fact that there are nearly no supporting comments here (and nearly no opposing comments in the right panel) makes clear the degree of polarization on the issue. Note in the top-left panel that policy-opposing comments were attached to both opposing and supporting articles, but more frequently to policy-supporting articles ($p < .001$). The top-right panel shows the same data for supporters of the policy. Supporters left fewer comments than did users who opposed the policy; however, supporters also made more policy-supporting comments on articles that oppose the policy than on

articles that support the policy. (This difference is barely perceptible in the figure, but it is statistically significant at $p < .001$). The same pattern is observed when the results are restricted to hostile comments (bottom panel).

To confirm that results for nuclear power are not anomalous, we conducted a second case study focused on the Terminal High Altitude Area Defense missile system—a contested policy that arose under a conservative rather than liberal government. Due to space constraints, results are reported in detail in Appendix 6.8. Those results reflect very similar patterns to Figure 3.2.

In sum, our findings suggest that users are not merely passive when encountering opposing viewpoints: they respond to such content with a greater frequency than they do when they encounter content reflecting similar viewpoints. News consumers’ greater tendency to engage with counter-attitudinal stories indicates cross-cutting engagement, reflecting the fact that commenting behavior extends beyond users’ ideologically congruent spaces.

3.6 Study 2: Source-Level Engagement

Study 1 was focused on analyses of story-level engagement; Study 2 shifts the focus to source-level analyses. Using the full dataset, this study examines whether hostile users are less confined to fragmented outlet clusters and whether domains with higher hostility exhibit more cross-cutting engagement.

3.6.1 Methods

We assess news engagement at the source level using network analysis to capture the structural patterns of news commentary. To do so, we construct a co-engagement media network that captures the structure of user engagement across news outlets. Each node in the network represents a media outlet, and weighted edges between nodes reflect the number of unique users who commented on both outlets during a given time window. In this co-engagement network, higher edge weights indicate greater audience overlap between

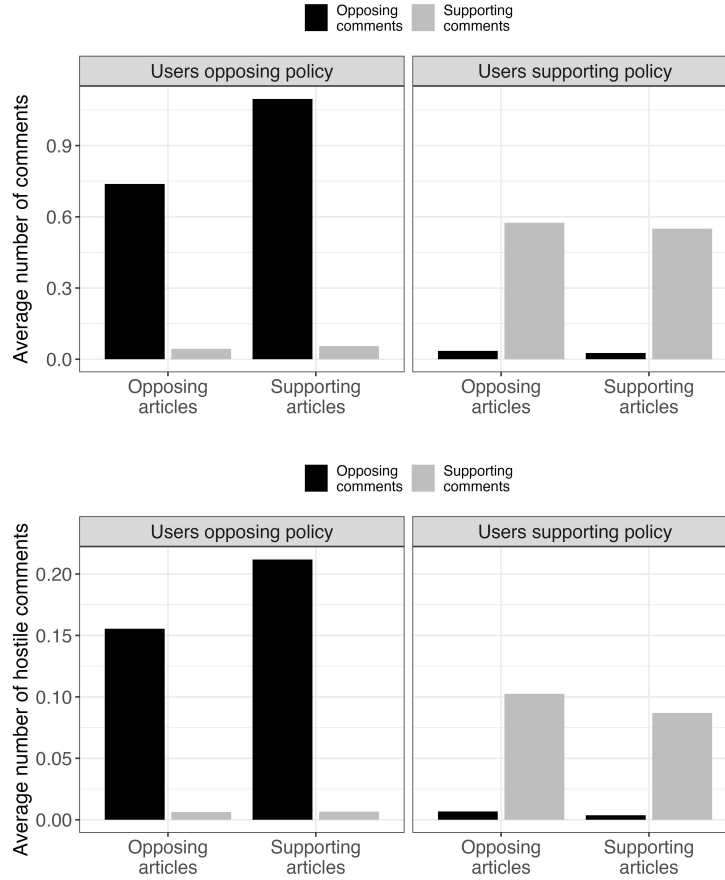


Figure 3.2: This figure compares commenting patterns on news articles about the nuclear power phaseout policy. The top panels show the average number of total comments, and the bottom panels show the average number of hostile comments, made by users opposing (left) and supporting (right) the policy.

outlets.

We use this network to calculate modularity scores, a widely used measure of community structure that reflects the degree of audience fragmentation. Prior studies suggest that selective news engagement tends to produce a fragmented media network, where outlets or their audiences cluster based on shared viewpoints, forming distinct and insulated communities (e.g., Barnidge et al. 2021; González-Bailón et al. 2023; Majó-Vázquez et al. 2019; Mukerjee et al. 2018; Webster 2014). In this context, high modularity suggests that users primarily comment within tightly connected clusters of outlets—what can be considered echo chambers (Majó-Vázquez et al. 2019). Low modularity, in contrast, indicates a more

integrated information ecosystem, where users regularly engage across outlet boundaries.⁸

To build these networks, we first created adjacency matrices for pairs of media outlets by month. Each matrix cell was filled with the count of co-commenters for the pairs, and commenters were identified using unique user codes. After constructing these audience overlap networks (referred to as “media networks” hereafter), we then extracted the network’s backbone, following the approach suggested by Mukerjee et al. (2018). This step eliminates any links that do not reach the significance level for a probability value of $p < .01$. Since not all the links are equally important for understanding the media consumption pattern, removing non-significant links can help avoid overestimating the audience overlap between media outlets.

Network fragmentation is then calculated using modularity scores for each monthly media network. We adopt a community detection method that identifies the optimal media clusters, maximizing modularity across all possible partitions (Brandes et al. 2007). Modularity is a metric that reflects the ratio of connections within a cluster to total connections, in comparison to what would be expected from a random distribution of connections. This score thus controls for the number of commenters (which varies over time), allowing for comparisons across different networks.

Unlike simple source-count metrics (e.g., the number of outlets visited), modularity captures the structural organization of media use by showing how tightly users are grouped and how distinct those groups are from one another. This makes modularity a more meaningful indicator of echo chambers, especially in large-scale, multisource environments. Figure 3.3 illustrates this distinction: Panel a shows a relatively less fragmented network (cross-cutting engagement), while panel b depicts a more fragmented network (selective engagement).

Importantly, this modularity-based approach does not require predefined ideological

⁸Network centrality can also indicate media selection diversity, as prior studies show that a few key outlets capture most audience attention (Barnidge et al. 2021; Mukerjee et al. 2018). However, centrality reflects audience hierarchy rather than fragmentation across media clusters. Because this study focuses on fragmentation rather than outlet dominance, modularity was used instead. Centrality scores showed similar trends (see Appendix 6.4).

labels for news outlets. Rather, it reflects latent clustering patterns based on actual user behavior. Existing studies suggest that such clusters are likely to align with partisanship, given evidence of growing ideological polarization in South Korea’s media landscape (Lee et al. 2024) as well as among political elites and the public (Cheong and Haggard 2023). Regardless of the underlying dimension, however, the more users cross-cut between outlet clusters, the lower the modularity score.

The same measure of hostile commentary used in Study 1 is applied here. In addition, Study 2 incorporates users’ use of the “angry” reaction feature as a secondary indicator. Introduced by Naver in March 2017, this feature allows users to express anger, an emotional state closely associated with hostility, toward news articles.

To address whether hostile users exhibit less fragmented engagement patterns and whether higher hostility corresponds to greater cross-cutting interaction across news outlets, we first compare hostility and modularity scores across news domains using t-tests, Kruskal–Wallis tests, and post hoc Dunn’s tests for pairwise significance. To further examine these relationships at the individual level, we estimate a Generalized Linear Mixed Model (GLMM). The dependent variable is source diversity, measured as the percentage of media outlets commented on (out of all available in a given domain and month), and the key predictor is the number of hostile comments posted. Given the considerable size of the dataset, we randomly sample 1,000 cases where each case is a unique user-month combination. This results in a dataset of 11,656 comments over ten years. Since users can appear more than once when they post comments over multiple months, random slopes for users and months are included to account for repeated observations. The model uses a gamma distribution to account for the positive, right-skewed distribution of the dependent variable, while controlling for differences across issue domains. To assess the robustness of the fixed-effect estimates from this model, we conduct a parametric bootstrap with 1,000 iterations; the results are reported in Appendix 6.6.

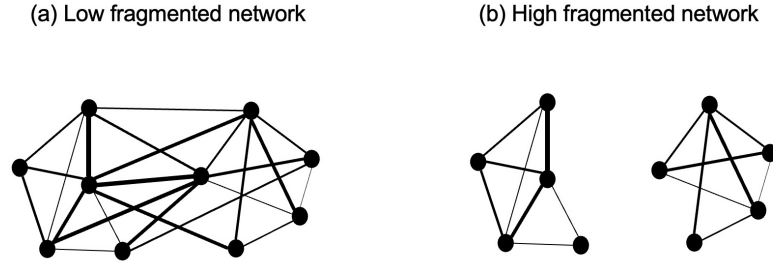


Figure 3.3: Illustrative examples of media network fragmentation. Panel (a) shows a low fragmentation co-engagement network, representing cross-cutting news engagement in which users comment across outlets. Panel (b) shows a high-fragmentation network, reflecting selective news engagement in which users remain concentrated within distinct outlet clusters, forming echo chambers.

3.6.2 Results

Across the ten-year dataset, the overall level of comment hostility on Naver seems relatively high. From April 2017 onward, each article across all issue domains received an average of 975 angry reactions per month. Over the entire period, 19.4 percent of monthly comments were classified as hostile, and 31.6 percent of users posted at least one hostile comment a month.⁹ On average, there are 44,382 hostile and 49,520 nonhostile users commenting each month.

Figure 3.5 shows the monthly average modularity scores of co-engagement networks built separately for hostile and nonhostile users. Hostile users, defined as those who posted at least one hostile comment in a given month, consistently exhibit lower modularity scores than nonhostile users ($p < .001$), indicating that their engagement spans more outlets and is less confined to clustered, like-minded environments. Figure 3.5 illustrates this difference during a high-modularity month (November 2015) in the Politics domain, using identical node positions for easy comparison: the network of hostile users is visibly more interconnected and less fragmented (modularity = .004 for 35,663 users) than that of nonhostile users (modularity score = .266 for 46,833 users).

⁹Detailed distributions can be found in Appendix 6.3. As depicted in the distribution, Politics and Society domains are marked by the highest degree of angry reactions, hostile comments, and percentage of hostile users relative to other domains.

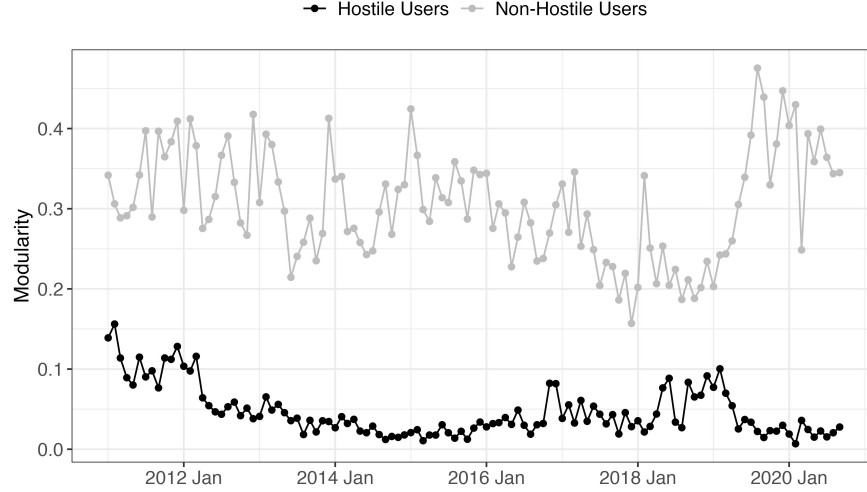


Figure 3.4: The figure shows monthly modularity scores for hostile and non-hostile users from 2011 to 2020. Lower modularity among hostile users indicates that their engagement spans a wider range of outlets, suggesting less fragmented and more cross-cutting patterns of news engagement compared to non-hostile users.

Figures 3.6 and 3.7 present domain-level comparisons. Figure 6 shows the Politics and Society domains have the highest average number of angry reactions per article (2,117 and 1,662, respectively) and the highest percentage of hostile comments from audiences (22.2 and 20.2 percent). The same pattern appears in the proportion of hostile users, with Politics and Society showing the highest monthly averages (32.4 and 29.8 percent). Importantly, Figure 3.7 reveals that these domains have the lowest modularity scores (.017 and .024), indicating less fragmented and more cross-cutting engagement.¹⁰ These findings consequently suggest an inverse relationship between hostility and modularity (RQ3): news domains with higher hostility tend to exhibit more crosscutting interactions in news commentary. In other words, reacting angrily to political or societal content is associated with active engagement across outlet boundaries.¹¹

To validate these findings at the individual level, we estimate a GLMM with the

¹⁰Descriptive statistics for the aggregate variables of hostility, angry reactions, and modularity across six news domains, presented as monthly average values, are available in the Appendix 6.1

¹¹The observed trend remains consistent over time, with the domains of Politics and Society consistently exhibiting the highest levels of comment hostility and angry reactions compared to other news domains over the ten years. See Appendix 6.2.

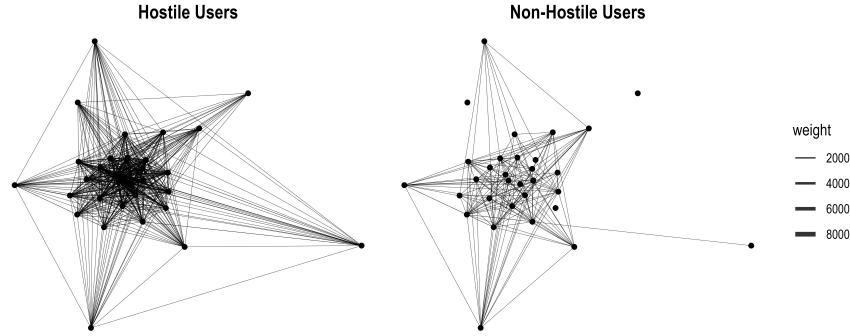


Figure 3.5: This figure shows a visualization of media engagement networks of hostile and non-hostile users in the politics domain during the month with the highest modularity. The network for hostile users (left) appears denser and more interconnected, reflecting crosscutting engagement across outlets, whereas the network for non-hostile users (right) shows more distinct clusters, indicating selective engagement within fragmented communities.

sampled data described in the Methods section. Figure 3.8 presents the fixed-effect estimates from this model. Estimates show differences across five issue domains relative to the reference domain, Life/Culture. Although these results cannot establish causality, hostile commentary and source diversity are quite clearly related. Even after accounting for differences across news domains, hostility is a strong positive predictor of source diversity ($p < .001$). A one-unit increase in individual-level hostility is associated with a .16-unit increase in the diversity of outlets a user comments on in a given period (bootstrapped mean coefficient = .162, 95% CI [0.150, 0.177]).¹²

To further assess the robustness of these findings, we conducted two supplemental analyses at different levels of aggregation. First, a multilevel model regressed monthly mod-ularity scores on hostility and angry reactions while controlling for issue domain and time (see Appendix 6.5.1-6.5.2). Results confirm the negative association between hostility and source-level fragmentation, reinforcing the link between hostile engagement and lower mod-ularity. Second, analysis of monthly user-level averages shows a strong positive correlation between hostility (number of hostile comments) and media diversity (percentage of media

¹²The full table for the GLMM result and distributions of the bootstrap estimates can be found in Appendix 6.6.

outlets a user commented on; $r = .715$), suggesting that more hostile users engage with a broader set of news sources (see Appendix 6.5.3.).

These findings consistently indicate that hostile engagement is associated with more cross-cutting patterns of news commentary. At both aggregate and individual levels, hostile users are less likely to be confined to echo chambers and more likely to engage across clusters of news outlets. We consider the implications of these findings below.

3.7 Discussion

Drawing on ten years of large-scale digital trace data from Naver, this study examined whether news commentary—an active form of news engagement—reflects cross-cutting rather than selective patterns of audience behavior across both content and source levels.

Study 1 indicated that users frequently crossed ideological boundaries at the content level by posting confrontational comments on counter-attitudinal news stories rather than supportive comments on pro-attitudinal ones. Interestingly, this pattern holds for commenting behavior in general, regardless of tone. Even non-hostile users appeared motivated to express disagreement, albeit in more civil language, when issues were salient or highly polarizing. This suggests that, for widely publicized and divisive issues, cross-cutting engagement may not always stem from emotional activation but may also arise from reasoned or deliberative expression. Such expressions nevertheless tend to occur disproportionately when users disagree rather than agree—and, importantly, on counter-attitudinal rather than pro-attitudinal news.

Study 2 extended the analysis to the broader media ecosystem and found evidence that users who posted hostile comments engaged across clusters of news outlets, exhibiting weaker structural fragmentation (and thus more cross-cutting engagement). Cross-cutting patterns in news commentary were most pronounced in the Politics and Society domains, where hostility was most prevalent.

These findings complicate the common and often implicit assumption that online hos-

tility reflects ideologically insular spaces that reinforce selective news engagement. Instead, they suggest that expressive and oppositional motives in active engagement may broaden, rather than narrow, users’ interactions with diverse perspectives.

These findings have implications for how we understand digital news engagement. Much of the literature, shaped by theories of confirmation bias and motivated reasoning, emphasizes alignment between preferences and behaviors. Our findings show that active engagement, especially commenting, does not always follow this logic. Hostile users often engage with content they reject, not to affirm it but rather to challenge it.

From a behavioral standpoint, this suggests that news engagement is shaped not just by what users support, but also by what they oppose. Hostile commentary may reflect oppositional attention, where disagreement becomes a motivator for interaction. In high-choice digital environments, such behavior may be a marker of crosscutting engagement, not ideological isolation.

The normative implications of these findings are mixed. While cross-cutting engagement is often seen as democratically beneficial, engagement driven by hostility may amplify animosity rather than promote understanding. Hostile engagement may serve expressive or identity-signaling functions without promoting deliberation or mutual understanding. As such, the presence of cross-cutting engagement does not necessarily imply constructive discourse.

Empirically, our results build on a growing literature showing that echo chambers are not as impermeable as once feared. Prior studies have found that users often encounter diverse perspectives through platform architecture, interpersonal networks, or incidental exposure (e.g., Bakshy et al. 2015; Flaxman et al. 2016). Our findings contribute to this literature by showing that users themselves may play an active role in breaking echo chamber boundaries, even when that engagement is characterized by opposition and hostility rather than curiosity or open-mindedness.

Still, several limitations remain. The data are observational, and while the patterns

are robust, they do not allow us to make strong causal claims. While Naver News offers a rich and unique environment for studying large-scale news engagement, it remains an open question whether similar patterns hold in other media systems or platform contexts. Moreover, while we observe behavior, we cannot directly measure users' underlying motivations. Future work should explore whether hostile engagement reflects deliberate confrontation or other intentions not captured in this study.

In sum, this study suggests that hostility does not always reinforce selective engagement—it may challenge it. Far from being confined to ideologically aligned spaces, hostile users engage across boundaries, reacting to content they disagree with. This behavioral tendency points to an important, if uncomfortable, possibility: that hostility and opposition may drive some of the most cross-cutting forms of digital news engagement in today's media environment.

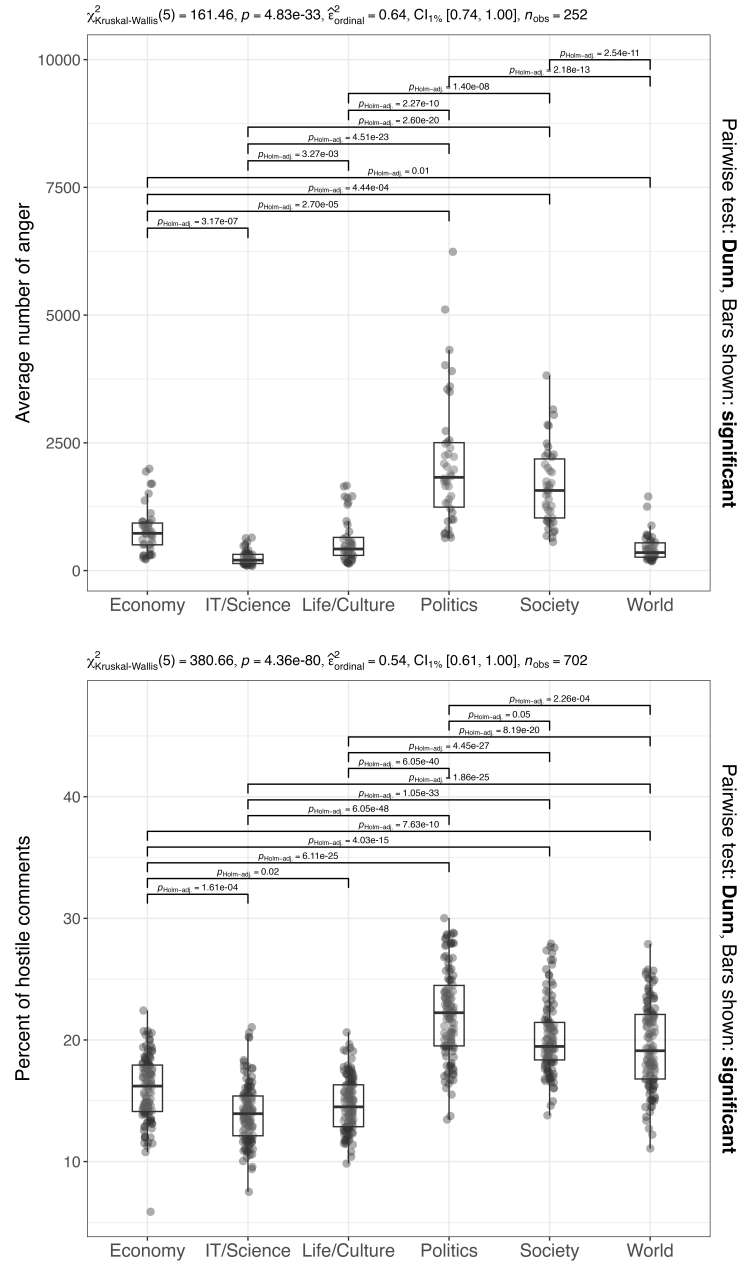


Figure 3.6: This figure compares the average number of angry reactions (top panel) and the percent of hostile comments (bottom panel) across six news domains. Bars indicate statistically significant pairwise differences based on Dunn's tests. It shows that political and societal news elicits the highest levels of anger and hostile engagement.

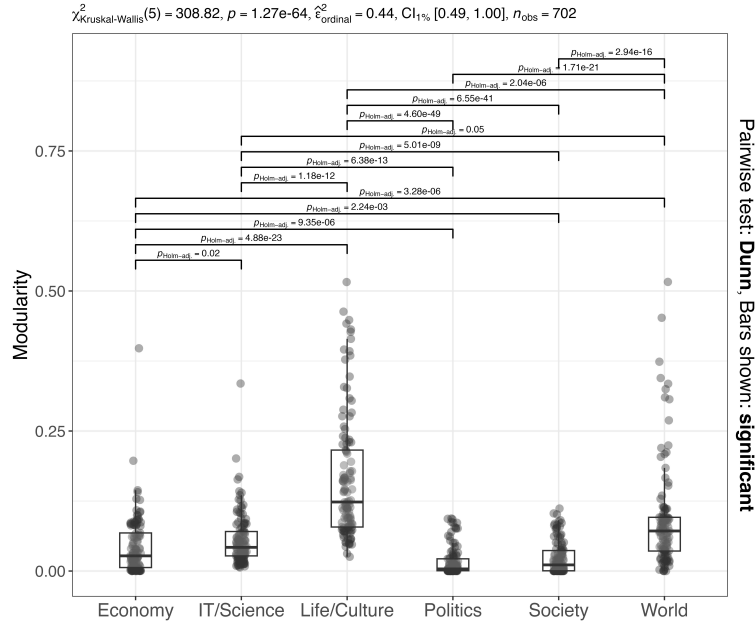


Figure 3.7: This figure compares average modularity scores across six news domains. Bars indicate statistically significant pairwise differences based on Dunn’s tests. Lower modularity in politics and society suggests less fragmented, more cross-cutting engagement, whereas higher modularity in life/culture indicates more selective engagement confined within outlet clusters.

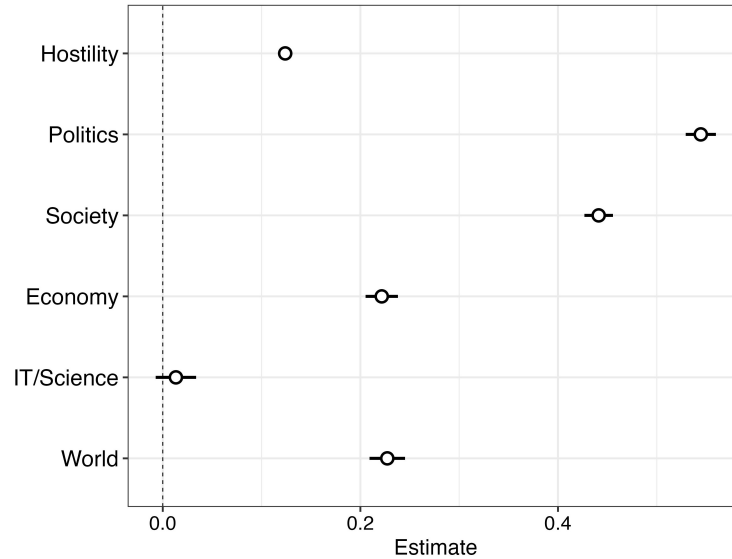


Figure 3.8: This figure presents results from the Generalized Linear Mixed Model (GLMM) estimating the relationship between hostility, news domains, and source diversity. Positive coefficients indicate that higher values of the corresponding variable are associated with greater diversity in the outlets users engage with.

Chapter 3. Algorithmic Exposure, User Choice: Engagement

Segregation in Mobile News Feeds

4.1 Introduction

The relationship between digital media and political polarization has become one of the most debated questions in contemporary social science. As news consumption increasingly migrates to algorithmic platforms, concerns have grown about whether personalized recommendation systems narrow the informational environments individuals encounter, reinforcing existing beliefs and deepening societal divides (Pariser 2011; Sunstein 2001, 2018). These phenomena are known as filter bubbles or echo chambers, in which news engagement is segregated among consumers.

Yet the mechanisms underlying these phenomena remain unclear. Engagement segregation—the tendency for media audiences to cluster into distinct informational niches—is widely documented, but its origins are contested. Does it arise primarily from individuals’ own selective tendencies, as decades of research on selective exposure suggest? Or is it produced, at least in part, by the recommendation algorithms that curate what users see in the first place? In practice, the answer almost certainly lies somewhere in between, as the two forces are deeply entangled: user behavior shapes algorithmic outputs, and algorithmic outputs in turn shape user behavior, creating feedback loops that are difficult to disentangle with conventional research designs (Matias 2023; Pedreschi et al. 2025).

This entanglement has important implications for how we study and interpret engagement segregation in the digital era. Much existing research on selective exposure has focused on users’ own choices—what they click on, share, or read—without fully accounting for what the platform made available to them in the first place. Many studies on algorithmic

mic filtering, meanwhile, have largely examined effects on downstream political attitudes rather than on the content users actually choose to engage with. We thus have a limited understanding of how much of the observed segregation in digital news environments reflects individual selectivity, how much reflects algorithmic provision, and how these forces combine across different stages of the engagement process.

This study addresses these gaps using a uniquely rich behavioral dataset from SmartNews, a mobile news aggregator app, which captures over 22 million user–article interactions from 5,000 U.S. users in the second half of 2023. Crucially, the data allow us to identify three distinct stages of engagement: viewing (algorithmically surfaced headlines in the feed), clicking (user-initiated selection), and reading (sustained consumption of at least five seconds). This sequential structure allows us to treat the viewing stage as a representation of algorithmic exposure and the clicking and reading stages as reflections of user-driven selectivity on top of that baseline, providing a partial window into how the two forces jointly produce engagement segregation.

To measure segregation, daily co-engagement networks were constructed in which articles are connected by users who engage with both on the same day, and degree-corrected stochastic block models were applied to assess the strength of clustering at each stage. Logistic regression on article pairs was then conducted to examine which content-level features—partisan slant, negativity, and topical category—are associated with segregation at each stage, and whether these associations differ between the algorithmic exposure stage and the user-driven engagement stages.

The results show that engagement segregation intensifies as users move from algorithmic exposure to active selection and sustained reading, and that the content-level predictors of segregation differ across stages in meaningful ways. Partisan distance is significant only at the user-driven stages, indicating that ideological selectivity emerges primarily from users’ own choices rather than from algorithmic provision even among users with low political interest. Negativity is by contrast associated with segregation only at the algorithmic exposure

stage, suggesting it is a dimension along which platform curation already structures content.

Together, these findings highlight the multivariate and sequential nature of engagement segregation, and underscore the importance of considering both algorithmic provision and user behavior when studying audience fragmentation in digital news environments. Engagement segregation is not reducible to a single content dimension—partisan alignment, affective cues, and topical categories each play meaningful roles, and their relative importance shifts across different stages of news consumption. These findings also carry practical implications for platform governance. SmartNews’s bipartisan curation appears to be at least partly undone by users’ own selective tendencies, suggesting that simple algorithmic changes may be insufficient to meaningfully reduce segregation overall. While the specific dynamics observed here may vary across platforms with different recommendation logics and user bases, the broader insight holds: understanding how users respond to algorithmic provision and how the two forces jointly shape information environments are essential for accurately diagnosing the sources of audience fragmentation and for designing interventions that can effectively address them.

4.2 Algorithmic Segregation in News Engagement

4.2.1 Algorithmic Segregation

A substantial body of literature has documented individuals’ tendency toward selective news engagement—that is, a preference for pro-attitudinal information or ideologically aligned news sources. This pattern has been repeatedly demonstrated in both experimental and observational research, particularly through studies showing that people are more likely to seek out or engage with information that reinforces their existing views (e.g., Iyengar and Hahn 2009; Peterson et al. 2021; Stroud 2008, 2011).

Empirical studies have supported these patterns using a variety of designs. Click-through experiments reveal that individuals are more likely to select news headlines that reinforce their prior attitudes (Iyengar and Hahn 2009; Stroud 2008, 2011), while web-tracking

studies show that users often visit partisan-aligned domains (Flaxman et al., 2016; Peterson et al., 2021). Other studies infer consumption patterns based on article dwell time, again finding ideological biases in what people choose to read (Jang 2014; Knobloch-Westerwick 2012a).

This tendency toward selective exposure has been closely linked to engagement segregation—a pattern in which individuals are predominantly surrounded by like-minded perspectives and content that aligns with their prior preferences, while largely avoiding content that does not. Such segregation is relatedly captured by the concepts of *echo chambers* (Sunstein 2001, 2018) and *filter bubbles* (Pariser 2011)—both describing how digital media environments can narrow the range of information individuals encounter, reinforcing existing viewpoints while limiting exposure to cross-cutting content. Sunstein’s echo chamber describes a collective phenomenon in which like-minded individuals, insulated from dissenting views, reinforce one another’s positions and drift toward more extreme opinions. Pariser’s filter bubble relatedly focuses on individual-level isolation, describing how algorithmic personalization creates epistemic bubbles in which personal viewpoints are rarely exposed to challenge or contradiction (Kitchens et al. 2020).

Importantly, algorithmic personalization is identified as the key mechanism underlying both phenomena, driving group polarization at the collective level (Sunstein 2001, 2018) and intellectual isolation at the individual level (Pariser 2011) by carefully selecting content to match inferred user preferences.

This points to a crucial insight: the premise of engagement segregation rests not on one force but two. On one side is individuals’ well-documented tendency to select content that aligns with their predispositions, as the theory of selective exposure suggests. On the other is platform recommendation algorithms that curate content in ways that may reproduce or further amplify this tendency.

4.2.2 Individual Selectivity vs. Algorithmic Personalization

While both forces are important, research on media selectivity has largely focused on users’ own selective tendencies, and while this body of work has made significant contributions to the field, it has often done so without accounting for the supply side: what users are provided by recommendation algorithms in the first place. Many influential studies on engagement selectivity have examined user demand by tracking, for example, what people share, follow, click on, or otherwise choose to engage with, and have found consistent evidence of selective patterns and audience segregation (e.g., Benkler et al. 2018; Flaxman et al. 2016; Iyengar and Hahn 2009; Peterson et al. 2021; Stroud 2008, 2011; Tokita et al. 2021). Platform recommendation algorithms and platform design, meanwhile, have emerged as an increasingly important issue, not only within media and communication research but across many sectors of society (Pedreschi et al. 2025; Wagner et al. 2021). Yet many studies on user selectivity have rarely accounted for the likelihood that the baseline choice sets provided by algorithms may have shaped the very choices they observe. What looks like a user’s preference may partly reflect what the platform made available to them in the first place.

On the other hand, studies examining the supply side of news consumption have largely focused on the effects of platform recommendation algorithms on users’ attitudes and opinions, with far less attention given to how users interact with and select from algorithmically curated feeds—obscuring the mechanisms through which algorithmic influences shape downstream attitudes. A number of studies, for example, have found that exposure to algorithmically filtered feeds can produce polarizing effects on political attitudes (Gauthier et al. 2026; Huszár et al. 2022; Ludwig et al. 2025), while others find little to no such effects (Cho et al. 2020; Guess et al. 2023; Ludwig et al. 2023, 2025). Yet across both sets of findings, a critical question remains largely unaddressed: what do users actually choose to engage with when interacting with algorithmically curated content, and how does this dynamic contribute to collective information patterns such as echo chambers and filter bubbles,

which are arguably related to polarizing effects?

Several studies have begun to address this gap. In an experimental setting replicating a YouTube-like interface, Liu et al. (2025) found that users are more likely to select ideologically aligned videos when algorithms present them with a higher proportion of slanted content. This study makes a valuable contribution by placing both algorithmic provision and user choice within the same analytical framework and tracing a causal chain between the two. Nevertheless, because the ratio of ideologically slanted content in each experimental condition remains fixed throughout (e.g., 3:1 or 1:1), the design does not capture how real-world algorithms progressively personalize content over time based on user behavior, leaving open the question of how users would respond to increasingly tailored choice sets.

Other studies have examined algorithmic filtering through the lens of echo chambers and filter bubbles. On Facebook, Bakshy et al. (2015) found that individual choice limits exposure to ideologically cross-cutting content more than the algorithm itself does, and González-Bailón et al. (2023) showed that users become increasingly ideologically segregated as they move from potential exposure to actual exposure and active engagement. However, Facebook’s social network features mean that homophily among users, in addition to or even more so than content-based algorithmic filtering, may be a significant confounding factor in these results (Kitchens et al. 2020), underscoring the need to examine platforms where algorithmic recommendations are driven more by content rather than social connections.

A related gap concerns the algorithmic feeds themselves. Relatively little attention has been paid to whether recommendation algorithms already reflect filter bubbles in what they serve to users. Although some studies have examined algorithmic composition directly—finding, for example, no evidence of filter bubbles in Google Search (Courtois et al. 2018) or less ideologically congruent content on Twitter (Wang et al. 2024)—most studies offer limited insight into how imbalanced or segregated algorithmically curated feeds actually are in the first place, and, importantly, into how users respond to specific algorithmic content compositions. While there have been efforts to categorize algorithms into broad types based

on content sorting, such as reverse chronological versus algorithmic feeds, and to compare their effects (e.g., Gauthier et al. 2026; Guess et al. 2023; Huszár et al. 2022; Wang et al. 2024), we still have a very limited picture of what content those feeds actually consist of.

This gap matters because without knowing the degree to which algorithmic provision is itself segregated, it becomes difficult to determine whether observed filter bubbles stem from (a) individuals’ tendency to selectively engage with content that aligns with their pre-dispositions, regardless of what they are shown, or (b) their conformity to what algorithms recommend based on implicit or explicit estimates of user preferences. The answer almost certainly lies somewhere between the two, since user selectivity and algorithmic recommendations are deeply entangled through feedback loops in which each continuously shapes the other (Matias 2023; Pedreschi et al. 2025).

Despite these difficulties, disentangling the contributions of algorithmic adaptation and user selectivity remains crucial for accurately characterizing the sources of filter bubbles and for better informing platform governance. The overall degree of engagement segregation observed in scholarly work most likely reflects their joint contribution rather than either force alone.

Figure 1 illustrates this theoretical framework, showing how engagement segregation varies across the full space of possible combinations of algorithmic adaptation and user selectivity. When both forces are high—that is, when users strongly favor pro-attitudinal content and algorithms are highly responsive to those preferences—engagement segregation is expected to be most pronounced (top-right). When user selectivity is high but algorithmic adaptation is low, however, the algorithm fails to pick up on user preferences when curating subsequent feeds, meaning that observed segregation may still be elevated but falls short of the top-right scenario (top-left). When user selectivity is low to begin with—meaning users are not particularly selective and occasionally engage with content outside their preferences—segregation is expected to be low regardless of how adaptive the algorithm is. In this case, a non-adaptive algorithm produces a largely random feed with no systematic

match to user preferences (bottom-left), while a highly adaptive algorithm would still curate a relatively non-selective feed, simply because there are no strong preferences to reflect (bottom-right).

These dynamics carry direct implications for the present study. Given SmartNews’s arguable commitment to bipartisan content curation, its recommendation algorithm is likely to surface a relatively diverse mix of content—positioning it toward the left side of the figure. Whether engagement segregation nonetheless emerges on this platform may therefore depend largely on the degree to which users exercise their own selective tendencies on top of that algorithmically balanced baseline.

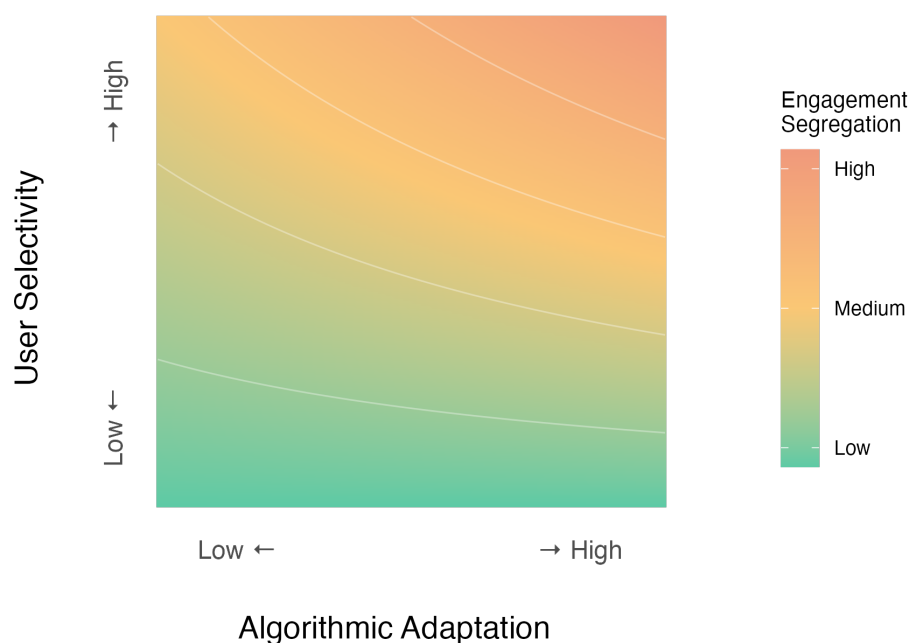


Figure 4.1: Engagement Segregation According to User Selectivity and Algorithmic Adaptation

4.3 Beyond Partisanship: Negativity and Topical Interest as Dimensions of Engagement Segregation

Existing research on selective exposure, engagement segregation, and algorithmic filtering has largely emphasized partisan alignment as the primary axis along which these dynamics operate. Yet news selectivity is likely to manifest across multiple dimensions beyond ideology, and little attention has been paid to how other content attributes, such as emotional valence and topical interest, may also shape patterns of engagement segregation.

There is, for instance, a large literature suggesting that negative valence is associated with heightened attention and news engagement across cultural contexts. Negative content is more likely to be read, remembered, and shared (Robertson et al. 2023; Soroka et al. 2019; Soroka 2014). One possibility is that negativity biases function as a force that de-segregates news consumers across all phases, rather than segregates them, since individuals are broadly drawn to affectively charged, especially negative, stories. That said, studies also suggest that subsets of users are attracted to positive content (Soroka and Krupnikov 2021), and that individuals exhibit variations in stable trait-level negativity biases over time (Bachleda et al. 2020; Noh and Soroka 2025). Some people may be less inclined than others to seek out negatively toned content, potentially contributing to engagement fragmentation. Sentiment-based news preferences may therefore be an important but underexamined factor in user segregation. Accounting for sentiment is also important when studying algorithmic influences, given that recommendation algorithms are likely sensitive to it as a cue that platforms may act on (Hilbert et al. 2018; Ludwig et al. 2023) and that incorporating sentiment has been shown to improve the quality of recommendation systems (Dang et al. 2021; Osman and Noah 2018).

Topical focus is another potentially salient dimension. People prioritize news that speaks to their interests, concerns, or perceived relevance, which can vary considerably by topic area (Kulshrestha et al. 2015; Tewksbury 2003). Political news may drive greater

engagement among users who are interested in politics, while others may gravitate toward entertainment, lifestyle, or health content. Mummolo (2016) demonstrates that topic relevance can even outweigh political source reputation when individuals decide what to read. Relatedly, Krupnikov and Ryan (2022) argue that the central divide among Americans is not ideological but rooted in political interest: while a minority of highly engaged citizens focus on politics, most people show limited interest and instead prioritize other topics. Content topic is also among the most fundamental attributes recommendation algorithms rely on to predict and optimize user engagement (Narayanan 2023). Topical interests, however, are rarely examined alongside partisanship or emotional tone as a correlate of engagement segregation.

This study addresses these gaps using a uniquely rich behavioral dataset that captures both algorithmic exposure and individual selectivity in mobile news feeds. While the observational nature of the data precludes direct manipulation of algorithmic feeds, it offers a valuable opportunity to examine how engagement segregation differs between the algorithmically curated feed and the subsequent stages of active user choice. By comparing segregation at the viewing stage, which reflects the composition of algorithmically provided content, against segregation at the clicking and reading stages, which reflect users’ own selective engagement on top of that algorithmic baseline, the study provides a partial window into the relative contributions of these two forces. It does so, moreover, not only along the partisan dimension that has dominated prior research, but also across the affective and topical dimensions outlined above. This leads to the following research questions:

RQ1. *How does engagement segregation at the algorithmic exposure stage compare to that observed at user-driven stages of clicking and reading?*

RQ2. *How are content-level attributes—partisan slant, negativity, and topical category—associated with engagement segregation at each stage, and do these associations differ between the algorithmically driven exposure stage and the user-driven engagement stages?*

4.4 Data

4.4.1 The SmartNews Platform

SmartNews is a mobile news aggregation app that was launched in Japan in 2012 and has since expanded to other countries, including the United States. Powered by machine learning, the platform evaluates tens of millions of articles, behaviors, and social signals to surface what it identifies as the most interesting and important stories, and has been downloaded by more than 50 million people worldwide. The app draws content from a large network of publishing partners spanning major national outlets across the political spectrum—including ABC News, CNN, Fox News, NPR, Reuters, and Bloomberg, among many others—as well as local sources, covering a broad range of topical categories including politics, business, technology, entertainment, sports, and lifestyle.

A central part of SmartNews’s product strategy, particularly in the U.S. market, has been its commitment to politically balanced news curation, positioning the platform as designed to bridge the partisan divide by surfacing news from across the political spectrum (Rest of World 2023). The company describes itself as a machine-learning company deeply committed to helping users find quality news beyond the filter bubble.

SmartNews presents content through a horizontally scrollable set of topical channels—such as Top News, Politics, Entertainment, and Sports—with each channel displaying a vertically scrollable feed of article headlines and thumbnail images. Users can view headlines directly within the feed, tap to open an article, and read it either within the app’s built-in reader or via the publisher’s original page. The data described below record the three behavioral stages central to this study: algorithmic exposure to a headline in the feed, the decision to click on it, and the subsequent act of reading.

Though the data used in this study do not include user demographics, SmartNews’s U.S. user base is known to skew toward highly educated adults with above-average household incomes compared to the general U.S. population (SmartNews 2022), and the platform’s

audience is roughly evenly split by gender, with the largest segment of users falling in the 35–44 age range (Similarweb 2026). These demographic characteristics suggest that the sample analyzed in this study may not be representative of the broader U.S. population, warranting caution when interpreting the findings.

4.4.2 Data

This study draws on a large-scale behavioral trace dataset provided by SmartNews under a nondisclosure agreement. The dataset captures the full range of in-app activity for 25,171 U.S.-based users over a six-month period, from July to December 2023, containing 113,157,655 engagement events across 4,494,669 unique news stories. These records include user-level behaviors such as viewing and clicking on headlines, as well as the time spent reading each article after clicking. Each activity is timestamped in UTC and accompanied by article-level metadata including the headline, URL, and SmartNews-assigned topical category.

This dataset offers a rare opportunity to analyze digital news engagement at the individual level across two analytically distinct stages: algorithmic exposure (viewing) and user-driven choices (clicking and reading). The viewing data captures the set of headlines that appear in an individual’s feed—that is, the content curated by the platform algorithms. While viewing may not be a pure measure of algorithmic output, as the feed also reflects users’ prior engagement history and navigation choices, it nonetheless captures what the platform has decided to surface to each user at a given moment, since what appears in a user’s feed is ultimately an algorithmic decision. Even when the platform incorporates user preferences, it is the algorithm that determines how much weight to assign those preferences, which articles to retrieve, and how to rank and present them. In this sense, treating viewing as a representation of algorithmic provision is consistent with how prior research has approached feed-level exposure data (Bakshy et al. 2015; González-Bailón et al. 2023), and it provides a meaningful baseline against which user-driven engagement at the clicking and reading stages

can be compared.

For the analysis, a random sample of 5,000 users was drawn and all of their activities retained, yielding a subsample of 22,171,412 engagement events. To account for meaningful behavioral sequences, session-level filtering was applied. Sessions were defined as continuous stretches of activity separated by no more than five minutes of inactivity, such that each user could have multiple sessions throughout the day depending on how often they opened and closed the app. To reduce noise from idle or anomalously long sessions, the top 5% of sessions by duration, measured in number of articles viewed, were excluded. In rare cases, users appear to have viewed tens of thousands of articles within a single session, likely reflecting background app activity, automated behavior, or data recording errors rather than genuine news engagement. This exclusion step is intended to reduce the influence of passive app behavior and erroneously inflated activity on the analyses.

Following session filtering, datasets were constructed for each of the three engagement phases. The viewing phase includes all instances where users were actually exposed to headlines in their feed. The clicking phase includes cases in which the user clicked on a headline, indicating a decision to engage. The reading phase includes only those cases where users spent at least five seconds on the article page, signaling relatively meaningful attention to the content. Notably, the most frequent reading time after clicking on an article was zero seconds, with five seconds falling around the 25th percentile of reading time, indicating that users often do not read the articles they click on and simply return to the home screen immediately. The final dataset includes 22,171,412 viewing events, 3,992,364 clicking events, and 3,141,502 reading events.

4.5 Methods

4.5.1 Headline Attribute Estimation

Headline features were estimated to predict engagement segregation patterns across co-engagement networks. First, the partisan slant of each article was estimated using *GPT*-

5-mini, which classified each headline on a three-point scale based on its ideological framing and partisan language: -1 (liberal), 0 (neutral), and 1 (conservative). To validate this measure, a random sample of 100 headlines was independently coded by a human coder, yielding a Cohen’s kappa of 0.96. Average partisan scores remain close to zero across all phases (viewing: $M = -0.057$, $SD = 0.353$; clicking: $M = -0.063$, $SD = 0.370$; reading: $M = -0.064$, $SD = 0.375$), suggesting a modest liberal lean in the content engaged with, though the overall distribution remains largely centered. Further details on the estimation procedure, validation, and example headlines by partisan category are provided in Appendix 7.3.

Second, each headline’s negativity was computed using the Lexicoder Sentiment Dictionary (LSD) (Young and Soroka 2012). The headline text was tokenized and the number of positive and negative words was counted using the dictionary. A normalized negativity score was calculated by subtracting the number of positive words from the number of negative words and dividing by the total number of words in the headline, yielding scores that range from -1 (greatest positivity) to 1 (greatest negativity). Resulting negativity scores are modest but show a slight increase from viewing ($M = 0.017$, $SD = 0.107$) to clicking ($M = 0.032$, $SD = 0.102$) and reading ($M = 0.033$, $SD = 0.102$).

Each article was assigned a single topical label by SmartNews based on its internal classification system. Binary variables were created for each of the most frequent topical categories in the dataset: politics (11.1%), entertainment (15.6%), lifestyle (0.09%), sports (0.09%), national news (0.05%), world news (0.05%), business (0.02%), technology (0.01%), and science (0.01%). Example headlines for each platform-assigned topical category are provided in Appendix 7.2.1.

It is worth noting that the platform-assigned topical categories are incomplete, with a large portion of articles labeled as unknown or uncategorized (41.9%). These articles were excluded from the regression analyses, and as a result, the current analysis captures only a subset of topical variation. Platform-assigned categories nevertheless carry an important

conceptual advantage: they directly reflect the taxonomic logic the recommendation engine itself operates on. SmartNews organizes content into topical channels using its own internal classification system, meaning that its clustering and curation logic operates on these categories rather than on externally derived topic labels. Topical categories identified by an external classification model may thus not correspond to how the platform actually groups and serves content to users. In this sense, platform-assigned categories are conceptually closer to the mechanism of interest. Furthermore, supplemental analyses (available in Appendix 7.2.2) suggest that the excluded uncategorized articles do not differ meaningfully from categorized articles along the key dimensions examined in this study (partisan slant and negativity) and that the unknown rate is distributed roughly uniformly across users rather than concentrated among particular user types. Topical composition estimated using an external classification model also shows broadly similar distributions across categorized and uncategorized articles, though political and sports content are somewhat better represented among categorized articles (see Appendix 7.2.3 for details).

4.5.2 Co-Engagement Network

To assess the extent to which engagement segregation varies across different phases of digital news use, the analysis constructs co-engagement networks that capture the structure of user engagement across news articles. Examining networks of engagement is an effective way to summarize the complex structure of news use, and prior studies suggest that selective news engagement tends to produce fragmented networks, forming distinct and insulated communities (e.g., Barnidge et al. 2021; González-Bailón et al. 2023; Mukerjee et al. 2018).

Daily co-engagement networks are constructed in which nodes represent unique news articles and edges represent the number of unique users who engaged with pairs of articles on the same day. Pairs of articles that share similar attributes are more likely to be connected, as users tend to engage with both when articles are similar along latent dimensions. Engagement is defined differently across phases: viewing refers to being exposed to the article in the feed,

selection to clicking on it, and consumption to reading it for at least five seconds.

To reduce spurious overlaps, a hypergeometric filtering procedure is applied to article-article edges. Extracting the backbone of a network by removing insignificant ties is particularly important in noisy online engagement data, where co-engagement can occur incidentally and blur the underlying structure (Mukerjee et al. 2018). The procedure compares the observed number of shared users between two articles (x) with the expected overlap under a hypergeometric null model, given the total number of users that day (N) and each article’s audience size (K_i, K_j). The resulting p -value indicates the probability of observing at least x shared users by chance. Edges are retained only if this probability falls below a threshold ($p < .05$), ensuring that the resulting networks reflect meaningful co-engagement rather than random overlap.

To assess the extent of engagement segregation, degree-corrected stochastic block modeling (DCSBM) is applied to the daily filtered co-engagement networks. The DCSBM is a generative network model that groups nodes into blocks based on shared patterns of connectivity, while explicitly accounting for overall network density and heterogeneous degree distributions (Peixoto 2019). Articles that share many users are more likely to be placed in the same block, and the network can be considered more segregated when articles fall into distinct blocks with limited audience overlap. For each daily co-engagement network at each stage, the model is fit using the *graph-tool* package in Python.

From these partitions, a segregation index is calculated as the ratio of within-block to between-block connectivity:

$$\text{Segregation Index} = \frac{\sum_{(i,j) \in \text{same block}} w_{ij}}{\sum_{(i,j) \in \text{different blocks}} w_{ij}} \quad (4.1)$$

where w_{ij} denotes the edge weight between articles i and j , defined as the number of users who engaged with both articles on the same day. The index thus captures how much more densely articles within the same block are co-engaged with relative to articles across different blocks. A value of 1 would indicate that within-block and between-block connectivity are

equal, reflecting no meaningful segregation. Values greater than 1 indicate that users disproportionately co-engage with articles within the same block, with higher values reflecting greater segregation. A value of 10, for instance, would mean that within-block co-engagement is ten times as dense as between-block co-engagement.

The degree to which engagement segregation differs across algorithmic exposure and user-driven stages is an open empirical question. As illustrated in Figure 1, the overall level of segregation reflects the joint contribution of algorithmic adaptation and user selectivity. Segregation at the user-driven stages may exceed that of algorithmic exposure if users selectively engage with a narrower subset of what they are shown, or if platform recommendation algorithms surface a relatively diverse mix of content to begin with. Alternatively, the two stages may exhibit similar levels of segregation if algorithmic personalization already clusters content into distinct audience communities before any active user choice occurs.

Basic network characteristics, block size distributions, and additional segregation metric (i.e., block assortativity and conductance) are reported in Appendix 7.4-7.5, and show patterns consistent with the main segregation index reported in the results.

4.5.3 Models of Engagement Segregation

To examine which content-level features are associated with the clustering of articles into distinct blocks, logistic regression models are estimated that predict whether a pair of articles belongs to different blocks derived from the block modeling described above. The model answers the question of which content-level attributes make any two given articles more or less likely to be assigned to different blocks. Note that both measures—the network-level segregation index and the pairwise outcome—draw on the same block partitions produced by the DCSBM: the network-level index summarizes the overall degree of clustering across all article pairs, while the pairwise models examine the content-level correlates of that clustering at the level of individual article pairs. To do this, 1,000 articles are randomly selected from the filtered networks across thirty randomly sampled days for each

engagement phase, and all possible pairs of these articles are generated, producing datasets in which each observation represents an article pair.

The model is specified as follows. For a pair of articles i and j observed on day d , the probability that they are assigned to different blocks is modeled as:

$$\begin{aligned} \Pr(Y_{ijd} = 1) = \text{logit}^{-1} & \left(\beta_0 + \beta_1 |\Delta \text{Neg}_{ij}| + \beta_2 |\Delta \text{Part}_{ij}| \right. \\ & + \sum_k \gamma_k \text{TopicMismatch}_{ijk} \\ & + \delta \cdot \text{HighInterest}_{ij} \\ & \left. + \lambda (|\Delta \text{Part}_{ij}| \times \text{HighInterest}_{ij}) \right) \end{aligned} \quad (4.2)$$

where $Y_{ijd} = 1$ indicates that articles i and j belong to different blocks, and $Y_{ijd} = 0$ indicates they belong to the same block. $|\Delta \text{Neg}_{ij}|$ denotes the absolute difference in negativity scores between the two articles (range 0–2), and $|\Delta \text{Part}_{ij}|$ denotes the absolute difference in partisan slant (where 0 = same stance, 1 = liberal vs. neutral or conservative vs. neutral, and 2 = full ideological opposition). $\text{TopicMismatch}_{ijk}$ is a set of binary indicators for whether the two articles differ in topical category k , including politics, entertainment, lifestyle, sports, national news, world news, business, and technology, with science serving as the reference category. Positive coefficients indicate that the predictor is associated with a higher probability of two articles being assigned to different blocks, interpreted as greater engagement segregation, while negative coefficients indicate the reverse. HighInterest_{ij} denotes the share of high political-interest users among those who engaged with the pair, where political interest for each user is measured as the proportion of political articles among all articles they engaged with, and high-interest users are defined as those falling in the top 25% of this distribution. The interaction term $|\Delta \text{Part}_{ij}| \times \text{HighInterest}_{ij}$ tests whether the association between partisan distance and segregation is moderated by the political interest of the audience. Models are estimated with cluster-robust standard errors at the day level

to account for the non-independence of article pairs within the same daily network. The full model results are reported in Supplemental Information Section 1.

4.6 Results

4.6.1 Descriptive Statistics

Table 1 summarizes key descriptive statistics for unique articles and users across the three engagement phases: algorithmic exposure (viewing), selection (clicking), and consumption (reading). As expected, the number of users who engaged in each phase declines as they move down the funnel—5,000 users in the viewing phase, 4,689 in the clicking phase, and 4,431 in the reading phase—indicating that not every user clicks on or reads articles from their feed. The number of unique articles also decreases sharply across stages, since each subsequent phase is a subset of the previous one: over 22 million unique articles appeared in users’ feeds during the six-month observation period, 691,832 unique articles were clicked, and 561,076 unique articles were read for at least five seconds. Of all viewed articles, only 18.0% were clicked, and of those clicked, 78.7% were read for at least five seconds.

Average partisan scores and negativity shift subtly across phases. Partisan scores remain slightly negative across all phases (viewing: $M = -0.057$, $SD = 0.353$, clicking: $M = -0.063$, $SD = 0.370$, reading: $M = -0.064$, $SD = 0.375$), suggesting a modest liberal lean in the content engaged with at each stage, though the overall distribution remains largely centered. Negativity scores show a modest increase across phases (viewing: $M = 0.017$, $SD = 0.107$; clicking: $M = 0.032$, $SD = 0.102$; reading: $M = 0.033$, $SD = 0.102$), rising from algorithmic exposure to user-driven stages.

The topical composition of engaged content also varies across phases. The proportion of unique articles categorized as politics roughly doubles from 4.1% in viewing to 8.2% in clicking and 8.8% in reading, indicating that although political content represents a small share of what is algorithmically surfaced, it is disproportionately likely to be selected and read. Entertainment content also increases, from 11.8% in viewing to 14.3% in clicking and

14.8% in reading. Lifestyle content remains relatively flat across phases.

These descriptive patterns suggest that user engagement narrows in scope as individuals move down the funnel of news use, with some content features becoming more prominent. The following analyses examine whether this narrowing process produces meaningful engagement segregation across users.

Table 4.1: Descriptive Statistics across Viewing, Clicking, and Reading Phases

Phase	Users	Articles	Partisan Score (<i>SD</i>)	Negativity (<i>SD</i>)	Entertainment (%)	Politics (%)	Lifestyle (%)
Viewing	5,000	22,171,412	−0.057 (0.353)	0.017 (0.107)	11.8	4.1	5.0
Clicking	4,689	691,832	−0.063 (0.370)	0.032 (0.102)	14.3	8.2	5.8
Reading	4,431	561,076	−0.064 (0.375)	0.033 (0.102)	14.8	8.8	6.1

4.6.2 Engagement Segregation from Algorithmic Exposure and User-Driven Stages

To evaluate whether engagement becomes more segmented as users move from algorithmically curated exposure to active user-driven phases, this section analyzes daily co-engagement networks constructed separately for each stage. Table 2 reports mean network characteristics after fitting DCSBM across the three stages. The algorithmic exposure networks are the largest, with an average of about 9,617 nodes and 3.57 million edges, while clicking and reading networks are smaller and denser subsets. The number of active blocks estimated by the models decreases as engagement deepens, from an average of 355 blocks in viewing to 271 in clicking and 224 in reading.

Table 4.2: Summary Statistics of DCSBM-Fitted Co-Engagement Networks

Phase	# Blocks	# Nodes	# Edges	Segregation Index
Viewing	355	9,617	3,573,320	17.73
Clicking	271	5,878	311,836	76.23
Reading	224	4,733	175,613	93.52

The mean segregation index rises across the three engagement stages: 17.73 for viewing ($SD = 2.60$), 76.23 for clicking ($SD = 9.76$), and 93.52 for reading ($SD = 11.99$), with relatively small variation across days. This indicates that articles in the clicking and reading

networks are far more concentrated within blocks relative to the viewing network, pointing to substantially greater engagement segregation at user-driven stages. Figure 2 visualizes the daily time series of the segregation index, showing a clear separation between stages: algorithmic exposure networks consistently exhibit the lowest segregation, while clicking and especially reading networks display substantially higher within-block clustering. This pattern indicates that engagement becomes more segregated as users move from algorithmic exposure to more active and sustained interaction with news content.

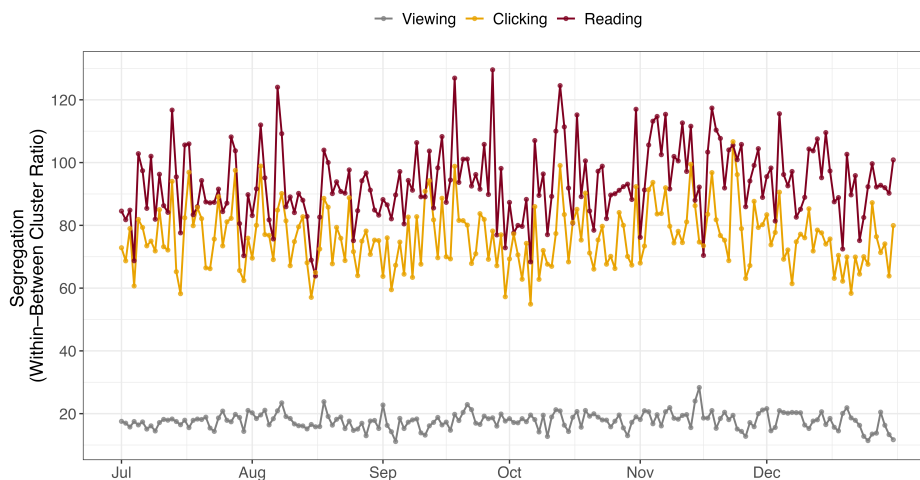


Figure 4.2: Engagement Segregation in Algorithmic Exposure and User-Driven Stages

4.7 Predictors of Engagement Segregation

To assess which content-level features are associated with the clustering of articles into distinct blocks, Figure 3 displays logistic regression point estimates with 95% confidence intervals for each predictor across the three engagement stages. The left panel shows negativity difference and partisan distance, while the right panel shows topical mismatch indicators. Full model results are reported in Appendix 7.1.

Partisanship. Partisan distance is not significantly associated with segregation at the algorithmic exposure stage ($B = 0.066$, $SE = 0.042$), but becomes a significant positive predictor in both clicking ($B = 0.305$, $SE = 0.048$, $p < .01$) and reading ($B = 0.195$, $SE =$

0.047, $p < .01$). It is worth noting that these are marginal effects conditional on the share of high political-interest users being zero. As At the algorithmic exposure stage, this means that the SmartNews algorithm does not appear to differentially cluster liberal and conservative articles—consistent with the platform’s stated mission of providing politically balanced news. For the clicking and reading stages, however, even users who are not particularly interested in politics exhibit partisan selectivity when actively engaging with content, likely reflecting underlying political predispositions that shape clicking and reading choices even among those who do not explicitly follow political news.

Negativity. Negativity shows the opposite pattern: negativity difference is positively associated with engagement segregation at the algorithmic exposure stage ($B = 0.304$, $SE = 0.165$, $p < .1$), but this association is not significant in clicking ($B = 0.097$, $SE = 0.119$) or reading ($B = 0.177$, $SE = 0.118$). This suggests that sentiment is a dimension along which algorithmic curation already structures content into distinct clusters—consistent with evidence that recommendation algorithms are sensitive to sentiment as a signal (Hilbert et al. 2018; Noh and Soroka 2025)—but that when user agency comes into play in the selection and consumption phases, negativity differences no longer play a significant role in segregating engagement.

Topics. Among topical mismatches, politics is the strongest and most consistent predictor of engagement segregation across all three stages ($B = 0.455$, 0.738 , and 0.720 for viewing, clicking, and reading, respectively; all $p < .01$), and its magnitude is considerably larger than that of any other topical predictors. This suggests that the primary topical divide structuring engagement patterns is between political and non-political content. National news mismatch, by contrast, shows a small negative and non-significant association at the viewing stage ($B = -0.054$, $SE = 0.040$), suggesting that national news content (such as severe weather or major public events of broad interest) may bridge otherwise distinct audience clusters in the algorithmically curated feed. Other topical mismatches, including entertainment, lifestyle, sports, and world news, are also positively associated with segrega-

tion across stages, though their magnitudes are consistently smaller than politics.

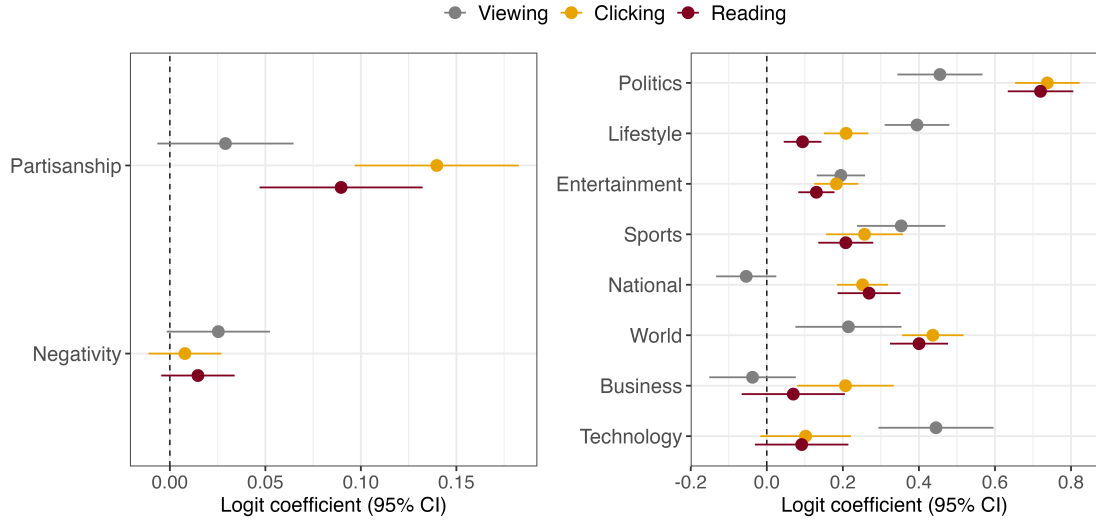


Figure 4.3: Engagement Segregation in Algorithmic Exposure and User-Driven Stages

Political Interest. Figure 4 displays predicted probabilities of segregation as a function of the share of high political-interest users among co-engaged users, separately for each level of partisan distance between articles (same stance, one apart, opposite stances) and each engagement stage. Across all three stages, a higher share of politically interested users is associated with a lower predicted probability of article-pair segregation, though this trend does not reach conventional levels of statistical significance (viewing: $B = -0.029$, $SE = 0.052$; clicking: $B = -0.053$, $SE = 0.071$; reading: $B = -0.050$, $SE = 0.046$). The direction of these estimates suggests that politically interested users are somewhat more likely to engage with articles across the partisan spectrum—consuming both liberal and conservative content—which would reduce rather than increase segregation among the pairs they co-engage with. While this pattern does not reach statistical significance, it is substantively noteworthy as it runs counter to the intuition that political interest amplifies partisan segregation.

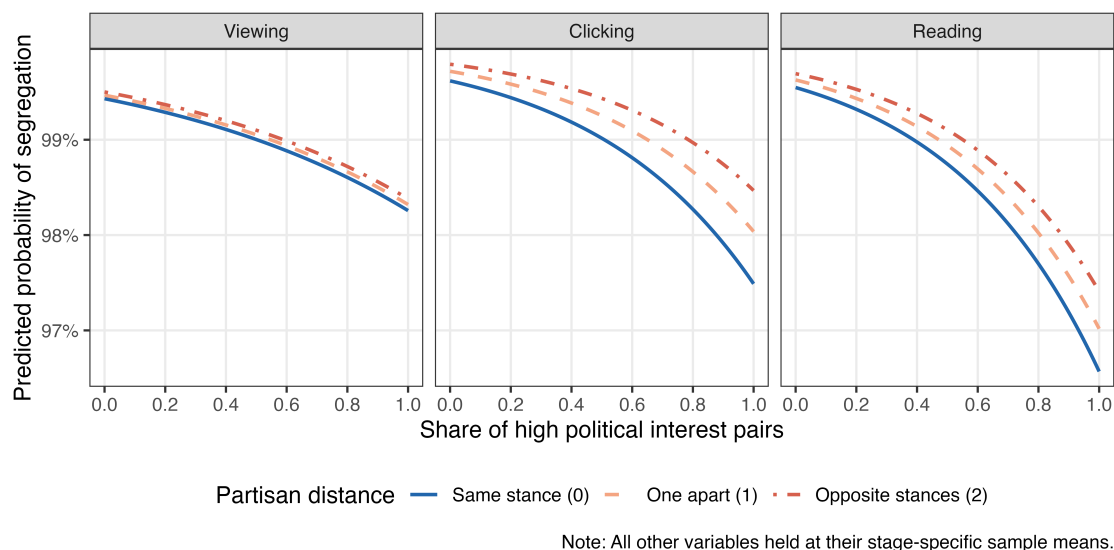


Figure 4.4: Predicted Probability of Segregation by Political Interest and Partisan Distance

4.8 Discussion

This study set out to examine how engagement segregation evolves across different stages of digital news use and to identify which content-level attributes are associated with these patterns—attending specifically to whether those associations differ between the algorithmically curated exposure stage and the user-driven stages of clicking and reading. Unlike studies conducted on social media platforms such as Facebook or Twitter, where social network homophily can confound the effects of content-based algorithmic filtering (Kitchens et al., 2020), this study examines a mobile news aggregator whose recommendation system is driven primarily by content rather than social connections, allowing for a cleaner assessment of how algorithmic provision and user selectivity jointly shape engagement segregation.

Findings addressing the first research question confirm that segregation intensifies as users move from algorithmic exposure to active selection and sustained reading. The SmartNews algorithm appears to surface a relatively diverse mix of content—consistent with the platform’s stated mission of bridging the partisan divide—yet users systematically engage with a more segregated subset of what they are shown. This progressive narrowing of the informational environment highlights that while algorithmic provision may not itself

be a primary source of partisan segregation in this context, user agency amplifies it at each subsequent stage.

Findings addressing the second research question reveal meaningful differences in content-level predictors across stages. Among topical mismatches, the divide between political and non-political content is the most consistent predictor of segregation, present across all three stages and increasing from the algorithmic exposure stage to the user-driven stages. This aligns with the argument of Krupnikov and Ryan (2022) that the central divide among Americans is rooted in political interest, and suggests that topical segmentation is a fundamental feature of both algorithmic provision and user selectivity.

The patterns for partisanship and negativity are also illuminating in light of the algorithmic-versus-user-choice distinction central to this paper. Partisan distance is not associated with segregation at the algorithmic exposure stage but becomes a significant predictor at the clicking and reading stages. This suggests that the SmartNews algorithm does not differentially cluster liberal and conservative content in its feed—producing a relatively ideologically balanced mix—but that users, when exercising their own agency, selectively engage with content that aligns with their partisan predispositions. Importantly, this partisan selectivity is present even among users with low political interest, suggesting that political predispositions shape news engagement at a level that operates below explicit political attention.

Negativity shows the opposite pattern, being associated with segregation at the algorithmic exposure stage but not at the user-driven stages. This suggests that sentiment is a dimension along which algorithmic curation already structures content into distinct clusters—consistent with evidence that recommendation systems are sensitive to emotional tone (Hilbert et al., 2018)—but that when users exercise active choice, negativity differences do not seem to systematically drive them into distinct informational niches. This finding extends the well-documented role of negativity in attracting individual attention (Robertson et al., 2023; Soroka, 2014) by showing that at the collective level, affective tone structures

segregation in algorithmic provision more than in user selectivity.

The interaction between partisan distance and political interest yields a non-significant but substantively interesting pattern. The association between partisan distance and article-pair segregation does not vary significantly with the share of politically interested users in the audience. Yet the share of high political-interest users is itself significantly and negatively associated with overall engagement segregation across all three stages (viewing: $B = -1.131$, $SE = 0.057$; clicking: $B = -1.904$, $SE = 0.046$; reading: $B = -2.056$, $SE = 0.039$; all $p < .01$), suggesting that politically interested users tend to co-engage with a broader range of articles, including those that are politically distant. While the absence of a significant interaction with partisan distance does not allow us to interpret this as indicative of cross-partisan engagement, the overall pattern runs counter to the intuition that political interest amplifies partisan segregation.

Taken together, these findings have several implications for how we understand selective exposure and media fragmentation. First, they underscore that both algorithmic personalization and user behavior must be considered jointly when examining engagement segregation. The results show that even when a platform actively curates a relatively balanced content mix—as SmartNews appears to do—users systematically narrow their informational environment through their own choices. This suggests that simple algorithmic changes alone may be insufficient to meaningfully reduce segregation overall, as user selectivity will continue to operate on top of whatever content is provided. These findings also position SmartNews closer to the top-left of Figure 1, a platform whose algorithmic provision is relatively balanced yet whose users nonetheless exercise considerable selective tendencies, pushing observed engagement segregation upward. Whether other platforms with different recommendation logics and user bases would fall elsewhere in this space, particularly toward the top-right where both algorithmic adaptation and user selectivity are high, remains an important question for future research.

Second, the findings highlight the importance of distinguishing between different

phases of engagement. Even among user-driven stages, the magnitude of segregation varies across viewing, clicking, and reading, suggesting that treating these phases as interchangeable risks obscuring meaningful differences in how deeply audiences are fragmented at each stage of news engagement.

Third, debates about echo chambers and filter bubbles in today’s information landscape should attend to multiple content dimensions beyond partisanship. Topical distance, particularly the divide between political and non-political content, and affective cues such as negativity each emerge as meaningful correlates of engagement segregation at various stages of news engagement, suggesting that partisan alignment is one of several dimensions along which audience fragmentation operates.

Finally, these findings carry direct implications for platform governance. Interventions designed to reduce partisan segregation may therefore consider targeting user behavior rather than algorithmic feed composition alone, while efforts to address sentiment-based clustering may be more productively directed at recommendation systems themselves. More broadly, effective governance of digital news environments requires distinguishing between what algorithms provide and what users choose and designing interventions accordingly.

This study has several limitations worth acknowledging. The observational nature of the data precludes causal claims about the effects of algorithmic provision on user behavior; experimental designs that manipulate algorithmic composition would allow stronger causal inference about the relative contributions of algorithmic provision and user selectivity to engagement segregation. Analyses incorporating a wider range of content dimensions, including framing, source credibility, and emotional valence beyond negativity, would also offer a more comprehensive picture of the multidimensional nature of audience segmentation.

Perhaps most importantly, the findings are specific to SmartNews, a platform whose recommendation system is seemingly designed to surface politically balanced content and whose user base is known to skew toward educated, higher-income adults. The specific content dimensions that shape segregation, and the relative contributions of algorithmic

provision and user selectivity, may differ across platforms with different recommendation logics, user demographics, and content ecosystems. Future work should examine whether and how these patterns replicate across a broader range of digital news environments.

Nevertheless, the central insight of this study—that understanding engagement segregation requires attending to the interaction between algorithmic provision and user behavior—is likely to hold across platforms. Even as the specific content dimensions that structure segregation may vary, and even as the relative weight of algorithmic versus user-driven forces may differ, the two forces are deeply entangled in any algorithmically mediated news environment. Disentangling their contributions, and examining how segregation accumulates across distinct stages of engagement, remains essential for accurately characterizing the sources of audience fragmentation and for designing effective interventions in today’s digital news landscape.

Appendix for Chapter 1

5.1 Experiment Materials

This section provides detailed information on the experimental materials used in the study. First, it provides the instructions and materials given to participants for each of the four conditions (selection-focused public, consumption-focused public, selection-focused private, and consumption-focused private). The section then presents twelve gun control arguments and the corresponding news headlines used in the news selection task. Lastly, it includes a screenshot of the actual news selection task interface.

5.1.1 Hypothetical Scenarios and News Selection Instructions

Using a two-by-two experimental design, participants are randomly assigned to one of four conditions that combine (a) two news engagement modes (selection-focused or consumption-focused) and (b) two levels of selection “publicity” (public or private).

Before starting the news selection task, participants were first presented with a hypothetical scenario about selection publicity based on their assigned condition. They were also shown a corresponding image for the scenario, which remained visible at the top of the screen throughout the subsequent news selection task. Table 5.1 displays the scenarios along with the provided images.

Table 5.1: Hypothetical scenarios and images

<i>Public</i>	<i>Private</i>
 <i>Please carefully read the following scenario and imagine yourself in that situation:</i> You are at home, comfortably browsing online news on your personal computer. You are logged into a popular social media platform known for its vibrant community and active engagement. This platform is highly interactive, and your actions, including the news articles you select or read, are visible to others. You have a large network here—hundreds of followers including close friends and acquaintances. Every article you click on sends a notification to your network and will appear on the feed of others. People can also see the time you spend reading each post or news articles. This platform prides itself on fostering an environment where users’ choices are part of a broader social conversation.	 <i>Please carefully read the following scenario and imagine yourself in that situation:</i> You are at home, comfortably browsing some news stories on social media. You are freely exploring and consuming news online without anyone observing your choices. Your actions, including the news articles you select or read, are completely private and not visible to others. You can focus solely on the content that interests you the most, without any social pressure or external judgment.

After being presented with the hypothetical scenarios, participants were directed to the news selection task and given corresponding instructions based on their assigned conditions (*selection-focused* or *consumption-focused*). Table 5.2 presents the instructions provided to participants for the news selection task in each condition.

Table 5.2: News selection task instructions

<i>Selection-focused</i>	<i>Consumption-focused</i>
<i>Selection-focused public:</i> While you are on this social media platform, some news articles about gun control pop up, and you want to read some of them. Please select six headlines from the list provided on the following page that you would be interested in reading. You are NOT required to read the full articles — just indicate the ones you are most interested in reading. Note that on this platform, your selections will be PUBLIC on your feed and accordingly may be seen by others on the platform.	<i>Consumption-focused public:</i> While you are on this social media platform, some news articles about gun control pop up, and you want to read some of them. Please select AND read six full articles that you would be interested in on the following page. After you select each story, you will have time to read through the entire article. Note that on this platform, your news consumption will be PUBLIC on your feed and accordingly may be seen by others on the platform.
<i>Selection-focused private:</i> While you are browsing online news at home, some news articles about gun control pop up, and you want to read some of them. Please select six headlines from the list provided on the following page that you would be interested in reading. You are NOT required to read the full articles — just indicate the ones you are most interested in reading. Note that in this private setting, your selections are completely PRIVATE and will not be seen by others online.	<i>Consumption-focused private:</i> While you are browsing online news at home, some news articles about gun control pop up, and you want to read some of them. Please select AND read six full articles that you would be interested in on the following page. After you select each story, you will have time to read through the entire article. Note that in this private setting, your news consumption is completely PRIVATE and will not be seen by others online.

5.1.2 Twelve Gun Control Arguments

For the news selection task, twelve gun control news stories (headlines, lead texts, and full articles) were adapted from real-world news sources, each aligned with different gun control arguments identified in previous studies by Kiil (2024) and Cragun (2022). Of the 16 arguments about gun control used in their studies, this study selected the following twelve:

Pro-gun control arguments:

- A study in a prominent medical journal found that you or a member of your family are 43 times more likely to be killed by your own gun than by an intruder's. Guns aren't the protection many people think they are. We need stricter gun control.
- In one poll of imprisoned felons, only 27% report buying guns on the black market; the rest got their weapons through legal channels. Obviously, tougher gun controls are needed to keep these 'legal' guns out of criminal hands.
- A study of 743 gunshot deaths reports that 398 occurred in a home where a gun was kept. Only 9 of the 743 were deemed to be justified by the police. It follows that gun owners are not as responsible as they claim to be.
- A gun should be fired only if one's life is in danger and all other options have been exhausted. Most 'self-defense' shootings do not meet these criteria. Thus use of guns in self-defense only contributes to the crime rate.
- Several recent school tragedies highlight the fact that guns have become a menace to our children. It's very simple: our schoolyards should not be battlefields. We need to reduce access to guns; we need stricter gun control.
- Recent trials against gun manufacturers have consistently found them guilty, and have forced the gun industry to pay out huge sums of money. If the courts can find good reason to rein in the gun industry, then it is high time for Congress to follow suit.

Anti-gun control arguments:

- The Bill of Rights guarantees the right of all citizens to bear arms. Quite simply, gun control measures are unconstitutional infringements on a basic right of citizenship.
- Most privately-owned guns in America are owned by sportsmen and are used for completely peaceful purposes. These guns pose no risk to society, but they are unfairly targeted by gun control legislation.

- A national council reported in a recent year that handgun accidents killed less than 15 children under the age of 6. This number is minuscule when compared to the total number of accidental deaths of young children. It simply is not worth outlawing guns to save just a handful of lives.
- The liberal media distort gun issues: they only talk about tragedies involving guns. Yet guns were used defensively 2.5 million times last year. The real tragedy would be to outlaw guns – crime would spiral out of control.
- A main reason why our murder rate is so high is that most crime victims do not resist. These victims are twice as likely to be injured compared to those who defend themselves. Carrying a gun is thus one’s ultimate protection against violent crime.
- Laws that require guns to be locked up defeat the purpose of gun ownership: how can I protect my family if I must first retrieve my gun from its locker? We thus need to repeal laws regulating guns in private homes.

To generate news stories based on these gun control arguments, real news articles that can capture each argument were identified and adapted to ensure consistent length across the stories (i.e., 10–15 words for headlines, 20–30 words for lead text, and 400–500 words for the main text). The selected articles were among those published close to the experiment dates (August 22–27, 2024).

5.1.3 Headlines and Lead Texts

For each of the gun control arguments listed in Section 1.2, a headline and lead text were adapted from real-world news stories. These appeared on the main news screen during the selection task. Below, each headline is followed by its corresponding lead text, aligned with the relevant argument.

Pro-gun control articles

- *Handguns at Home Increase Fatal Risks, Not Safety*

Recent findings reveal living with handgun owners raises homicide risks by over twice, challenging beliefs that guns provide safety against intruders.

- *Most Crime Guns Sourced from Legal Sales, Urging Stricter Gun Control Measures*

Study reveals the majority of crime guns in California originate from legal purchases, calling for stricter regulations on legal gun sales.

- *Unlocked Guns at Home Proven Lethal for Children, Calls for Gun Control*

Recent study shows most accidental shootings involve guns kept at home, often unlocked and loaded, posing significant risks.

- *Defensive Gun Use Rarely Prevents Crime and Often Increases Risk*

Analyses debunk myths of defensive gun use, highlighting its rarity and potential societal harm compared to its purported benefits.

- *Urgent Call for Gun Control After Surge in U.S. School Shootings*

Following devastating years marked by frequent school shootings, experts demand immediate, stringent gun control to protect American youth.

- *Gun Manufacturers Can't Avoid Responsibility for Their Role in Gun Violence*

Recent lawsuits reveal gun manufacturers' liabilities in gun violence, urging Congress to enforce stricter gun control measures.

Anti-gun control articles

- *Gun Control is Unconstitutional: Courts Invalidate Maryland and Oregon Gun Restrictions*

Recent rulings from U.S. courts declare Maryland and Oregon gun control measures unconstitutional, affirming the Second Amendment rights.

- *Gun Owners Criticize Unfair Targeting by Gun Control Overlooking Peaceful and Responsible Use*

Amid rising gun crime, sportsmen argue that new firearms legislation unfairly targets peaceful, law-abiding gun owners, not the criminals.

- *Are Firearms Truly the Leading Cause of Death in Children? Critics Question Data*

Recent statistics label firearms as the top cause of death among children, but critics challenge these findings based on age definitions and context.

- *Survey Reveals Defensive Gun Uses Common, Challenges Gun Control*

New survey finds millions use firearms defensively each year, casting doubt on the effectiveness of strict gun control measures.

- *Good Guys With Guns Save Lives: Armed Citizens Play a Key Role in Preventing Crimes*

Data challenges gun control narratives, showing that armed citizens frequently prevent crimes and stop potential mass shootings.

- *Locked Guns Limit Self-Defense Capabilities, Gun Owners Argue for Accessibility*

Amid debates over gun storage laws, gun owners assert that locking up firearms compromises their ability to quickly defend their families.

5.1.4 News Selection Task Interface

In the news selection task, the twelve news headlines, along with their lead texts, are presented in color (red for anti-gun control news and blue for pro-gun control news) to provide clear partisan cues for each story. The stories are displayed in a randomized order. A partial screenshot of the news selection interface is shown below. Since there are twelve news stories in total, participants needed to scroll down to view all available options.

Please **select five headlines** you would like to read on this platform...

Pro-gun control articles are highlighted in blue, while **anti-gun control articles** are highlighted in red.

Gun Owners Criticize Unfair Targeting by Gun Control Overlooking Peaceful and Responsible Use

☐ Amid rising gun crime, sportsmen argue that new firearms legislation unfairly targets peaceful, law-abiding gun owners, not the criminals.

Handguns at Home Increase Fatal Risks, Not Safety

☐ Recent findings reveal living with handgun owners raises homicide risks by over twice, challenging beliefs that guns provide safety against intruders.

Most Crime Guns Sourced from Legal Sales, Urging Stricter Gun Control Measures

☐ Study reveals the majority of crime guns in California originate from legal purchases, calling for stricter regulations on legal gun sales.

Gun Manufacturers Can't Avoid Responsibility for Their Role in Gun Violence

☐ Recent lawsuits reveal gun manufacturers' liabilities in gun violence, urging Congress to enforce stricter gun control measures.

Urgent Call for Gun Control After Surge in U.S. School Shootings

☐ Following devastating years marked by frequent school shootings, experts demand immediate, stringent gun control to protect American youth.

Unlocked Guns at Home Proven Lethal for Children, Calls for Gun Control

☐ Recent study shows most accidental shootings involve guns kept at home, often unlocked and loaded, posing significant risks.

Please **select and read five full news articles** you would be interested in on this platform...

To read a full article, select one headline and then click on "See the full article" at the bottom of the screen. You will be directed to the selected article. When you are done reading, you will be able to make another selection from the remaining articles. You will repeat this process until you have read six articles in total.

Pro-gun control articles are highlighted in blue, while **anti-gun control articles** are highlighted in red.

Gun Control is Unconstitutional: Courts Invalidate Maryland and Oregon Gun Restrictions

☐ Recent rulings from U.S. courts declare Maryland and Oregon gun control measures unconstitutional, affirming the Second Amendment rights.

Survey Reveals Defensive Gun Uses Common, Challenges Gun Control

☐ New survey finds millions use firearms defensively each year, casting doubt on the effectiveness of strict gun control measures.

Urgent Call for Gun Control After Surge in U.S. School Shootings

☐ Following devastating years marked by frequent school shootings, experts demand immediate, stringent gun control to protect American youth.

Gun Owners Criticize Unfair Targeting by Gun Control Overlooking Peaceful and Responsible Use

☐ Amid rising gun crime, sportsmen argue that new firearms legislation unfairly targets peaceful, law-abiding gun owners, not the criminals.

Gun Manufacturers Can't Avoid Responsibility for Their Role in Gun Violence

☐ Recent lawsuits reveal gun manufacturers' liabilities in gun violence, urging Congress to enforce stricter gun control measures.

Good Guys With Guns Save Lives: Armed Citizens Play a Key Role in Preventing Crimes

☐ Data challenges gun control narratives, showing that armed citizens frequently prevent crimes and stop potential mass shootings.

Figure 5.1: News selection task screen for selection-focused (left panel) and consumption-focused (right panel) conditions

5.1.5 Pre- and Post-Treatment Survey Measures

Pre-treatment survey on issue attitudes (All items adapted from ANES)

- *Death penalty*

Do you favor, oppose, or neither favor nor oppose the death penalty for persons convicted of murder?

Response options: Favor, Oppose, Neither favor nor oppose

- *Gun control*

Do you think the federal government should make it more difficult for people to buy a gun than it is now, make it easier for people to buy a gun, or keep the rules about the same as they are now?

Response options: Easier, Keep the rules about the same, Don't know

- *Income taxes*

Do you favor, oppose, or neither favor nor oppose increasing income taxes on people making over one million dollars per year?

Response options: Favor, Oppose, Neither favor nor oppose, Don't know

- *Federal spending*

Should federal spending on welfare programs be increased, decreased, or kept the same?

Response options: Increased, Decreased, Kept the same, Don't know

Post-treatment survey (All items adapted from ANES)

- *Party identification*

Generally speaking, do you usually think of yourself as a Democrat, a Republican, an independent, or what?

Response options: Democrat, Republican, Independent, Something else

- *Partisan strength* (for participants identifying as Democrat or Republican)

Would you call yourself a strong [Democrat / Republican] or a not very strong [Democrat / Republican]?

Response options: Strong, Not very strong

- *Partisan strength* (for participants identifying as Independent or Something else)

Do you think of yourself as close to the Democratic Party or to the Republican Party?

Response options: Closer to Democratic Party, Closer to Republican Party, Neither

- *Ideology*

Where would you place yourself on this scale, or haven't you thought much about this?

Response options: Extremely liberal, Liberal, Slightly liberal, Moderate, Slightly conservative, Conservative, Extremely conservative, Haven't thought much about this

- *Birth year*

In what year were you born?

- *Race*

What racial or ethnic group or groups best describes you?

Response options: Black, Asian, Native American, Hispanic or Latino, White, Other, Don't know

- *Education*

What is the highest level of school you have completed or the highest degree you have received?

Response options: Less than high school, High school diploma or equivalent, Bachelor's degree, Graduate or professional degree, Other, Prefer not to say

- *Gender*

What is your gender?

Response options: Male, Female, Non-binary/Other, Prefer not to say

5.2 Supplemental ACNS

This section presents an alternative measure of the primary dependent variable, ACNS (attitude-consistent news selections), excluding participants who selected “Keep the rules about the same” as their gun-control attitude. The number of participants included in this supplemental analysis is 954 (excluding 326 participants who selected the middle response on gun control). Note that this exclusion did not create a noticeable imbalance between conditions: 485 (50.8%) in the consumption-focused condition and 469 (49.2%) in the selection-focused condition. The results provide a complementary perspective to the analysis in the main text.

The results based on this supplemental variable closely mirror those reported in the main text, where the same responses were classified under the anti-gun control category.

This consistency suggests that the overall pattern of attitude-consistent news selection is robust to how these middle-category responses are coded. The two figures presented below show the average supplemental ACNS across conditions. Specifically, regarding the main effect, participants in the selection-focused condition again exhibit significantly higher ACNS scores ($M = 3.32$, $SD = 1.27$) than those in the consumption-focused condition ($M = 3.03$, $SD = 1.29$) (diff = 0.29, $F = 12.163$, $p < .01$).

Consistent with the results in the main text, the magnitude of this difference does not vary significantly across partisan-strength groups. The interaction between selection mode and partisan strength is not statistically significant ($F = 1.263$, $p = .29$). The moderating effect of publicity is also not statistically significant ($F = 0.729$, $p = .39$), with similar average ACNS across conditions.

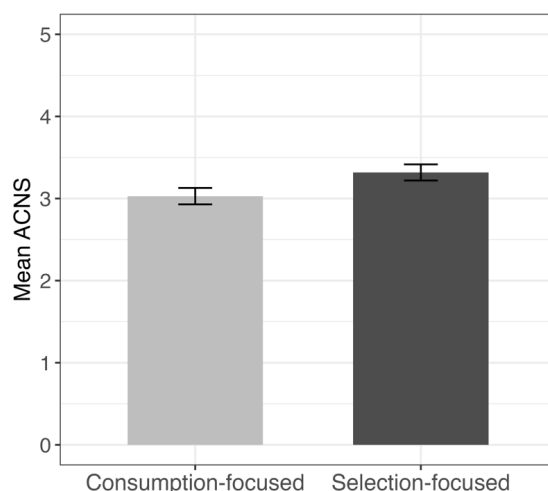


Figure 5.2: Mean Attitude Consistency in News Selection (ACNS) by condition

5.3 Additional Preregistered Hypotheses

There are additional preregistered hypotheses involving moderators beyond the three tested in the main study. As noted in the paper, none of these preregistered hypotheses involving moderators were statistically supported, and including them in the main text does not add value to the study's findings. This section presents these hypotheses along with their

brief background and corresponding statistical tests. Note that the terminology and numbering/order of the hypotheses in this section differ slightly from those in the preregistration, though the substantive content remains the same.

The first set of hypotheses examines whether *selection publicity* (*public* vs. *private*) has an independent effect on ACNS, and whether this effect—or its interaction with the selection/consumption condition—is moderated by partisan strength. As suggested by Hart et al. (2020), the public or private nature of information selection may significantly influence individuals’ choices, particularly when selections are made in public. Publicity may thus exert an independent effect on the primary dependent variable (ACNS). Moreover, as outlined in the main-text hypotheses, partisan strength may moderate these relationships, including the interaction between publicity and selection mode. These hypotheses are as follows:

H4a: ACNS will be greater in the public condition than in the private condition.

H4b: The difference in ACNS between public and private conditions will be greater for strong partisans than for weak partisans.

H4c: The interaction between selection/consumption and public/private conditions will be stronger for strong partisans than for weak partisans.

The second set of hypotheses explores a potential mechanism underlying expressive bias: *impression management motivation*. Prior research suggests that individuals often manage how they are perceived by others by selectively displaying information aligned with their identity, even when that information does not reflect their actual consumption habits. Impression management theory holds that people strategically present themselves in ways that shape others’ perceptions (Goffman 1959; Leary and Kowalski 1990)—a tendency observed both online (Marder et al. 2016; Rosenberg and Egbert 2011) and offline (Hart et al. 2020). This motivation is particularly relevant to the influence of selection publicity, which may activate impression management concerns.

The following hypotheses therefore test whether impression management motivation

moderates expressive bias:

H5a: The difference in ACNS between selection and consumption conditions will be greater for participants with higher impression management motivation.

H5b: The difference in ACNS between public and private conditions will be greater for participants with higher impression management motivation.

H5c: The interaction between selection/consumption and public/private conditions will be stronger for participants with higher impression management motivation.

Impression management motivation was measured via a survey item administered after the selection task: “How much do you agree with the following statement? — I usually worry about the impression I make on others.” Participants responded on a 5-point scale, from Strongly Agree (5) to Strongly Disagree (1). Responses were treated as a continuous measure, with higher values indicating stronger impression management motivation.

As noted in the main text, selection publicity had no significant effect on the dependent variable (ACNS). Therefore, it was expected that neither the first set of hypotheses, which tested moderators of the publicity effect, nor the second set, which tested impression management as a possible mechanism, would be supported.

The following tables present the results of the ACNS analyses based on two different codings of gun-control attitudes: excluding participants who selected “Keep the rules about the same” or including them under the anti-gun control category. As expected, none of the hypotheses in this section were statistically supported. For both coding approaches, selection publicity did not significantly influence ACNS, nor did its effect vary by partisan strength. While impression management motivation was negatively associated with ACNS overall—suggesting that participants concerned with self-presentation tended to select less attitude-consistent news—there was no evidence that this motivation moderated the effects of selection mode or publicity.

5.3.1 OLS Regressions Predicting ACNS (Excluding “Keep the Rules About the Same” Responses)

Table 5.3: OLS Regressions Predicting ACNS (Excluding “Keep the Rules About the Same” Responses)

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	3.157*** (0.060)	3.063*** (0.138)	2.875*** (0.200)	3.463*** (0.142)	3.598*** (0.143)	3.582*** (0.196)
Public	0.027 (0.083)	0.040 (0.198)	-0.042 (0.287)		-0.121 (0.206)	-0.254 (0.285)
Partisan Strength		0.043 (0.057)	0.049 (0.082)			
Public $\times$ Partisan Strength		-0.006 (0.082)	0.054 (0.116)			
Selection			0.359 (0.275)	0.133 (0.204)		0.024 (0.285)
Selection $\times$ Public			0.122 (0.394)			0.242 (0.410)
Selection $\times$ Partisan Strength			-0.001 (0.114)			
Selection $\times$ Public $\times$ Partisan Strength			-0.108 (0.163)			
Impression				-0.146*** (0.044)	-0.150*** (0.044)	-0.205*** (0.061)
Selection $\times$ Impression				0.050 (0.064)		0.114 (0.088)
Public $\times$ Impression					0.050 (0.064)	0.120 (0.087)
Selection $\times$ Public $\times$ Impression						-0.132 (0.128)
Num.Obs.	954	954	954	954	954	954
R^2	0.000	0.001	0.016	0.028	0.017	0.031
<i>Note.</i> * $p < .05$, ** $p < .01$, *** $p < .001$						

5.3.2 OLS Regressions Predicting ACNS (Including “Keep the Rules About the Same” as Anti-Gun Control)

Table 5.4: OLS Regressions Predicting ACNS (Including “Keep the Rules About the Same” as Anti-Gun Control)

Predictor	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	3.131*** (0.051)	3.041*** (0.116)	2.867*** (0.172)	3.427*** (0.126)	3.538*** (0.123)	3.462*** (0.175)
Public	0.064 (0.072)	0.159 (0.167)	0.060 (0.250)		-0.070 (0.177)	-0.078 (0.252)
Partisan Strength		0.042 (0.049)	0.033 (0.071)			
Public $\times$ Partisan Strength		-0.044 (0.070)	0.032 (0.102)			
Selection			0.316 (0.232)	0.121 (0.176)		0.122 (0.245)
Selection $\times$ Public			0.143 (0.335)			0.003 (0.353)
Selection $\times$ Partisan Strength			0.033 (0.097)			
Selection $\times$ Public $\times$ Partisan Strength			-0.135 (0.140)			
Impression				-0.141*** (0.039)	-0.140*** (0.039)	-0.176** (0.054)
Selection $\times$ Impression				0.059 (0.056)		0.085 (0.077)
Public $\times$ Impression					0.045 (0.056)	0.074 (0.078)
Selection $\times$ Public $\times$ Impression						-0.053 (0.112)
Num.Obs.	1280	1280	1280	1280	1280	1280
R^2	0.001	0.001	0.017	0.027	0.015	0.030
<i>Note.</i> * $p < .05$, ** $p < .01$, *** $p < .001$						

Appendix for Chapter 2

6.1 Monthly Averages of Hostility and News Engagement by News Domain

The table that follows shows average values for measures of hostility and news engagement, monthly, across the six issue domains.

Table 6.1: Monthly Averages of Hostility and News Engagement by News Domain

News domains	# of angry	% of hostile comments	% of hostile users	Modularity	% of selected media sources
Politics	2,117	22.2	34.2	0.017	7.0
Society	1,662	20.2	29.8	0.024	5.5
Economy	793	16.0	23.5	0.044	4.8
IT/Science	252	14.0	19.2	0.056	3.9
World	439	19.4	27.3	0.092	4.8
Life/Culture	586	14.7	19.8	0.164	3.5
Across domains	975	19.4	31.6	0.07	5.3

6.2 Monthly Trends in Confrontation and News Engagement

The following figures show trends in monthly averages over time, across the six issue domains, for two measures of confrontation: angry reactions (Figure 6.1), and hostile comments (Figure 6.2). Figures 6.3 and 6.4 then show trends in monthly average, across the six issue domains, for two measures of news engagement: modularity and percent of media sources.

6.2.1 Angry Reaction Over Time

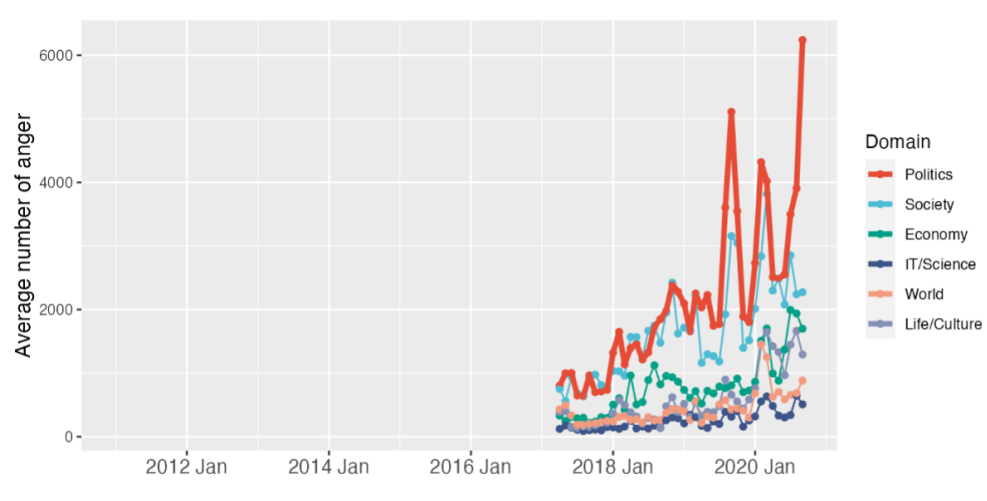


Figure 6.1: Monthly trends in angry reactions over time by news domain

6.2.2 Hostile Comments Over Time

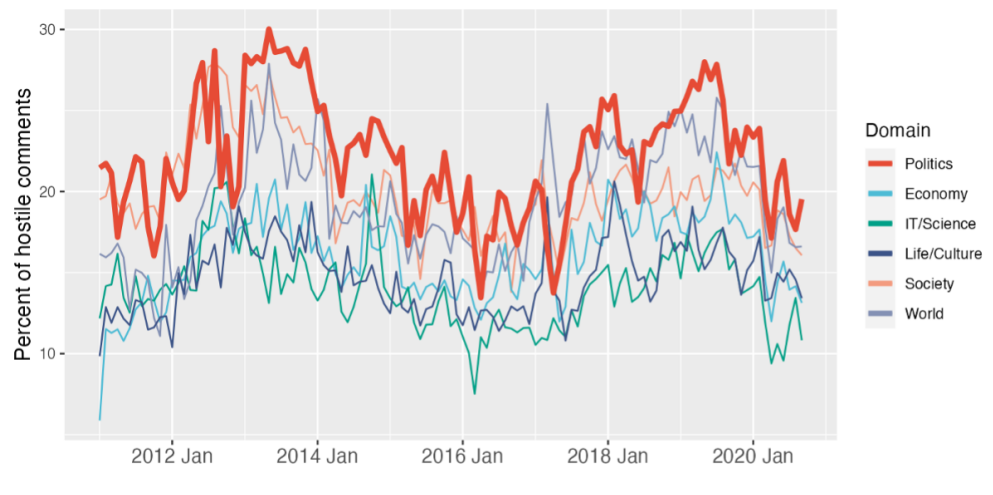


Figure 6.2: Monthly trends in hostile comments over time by news domain

6.2.3 Modularity Over Time

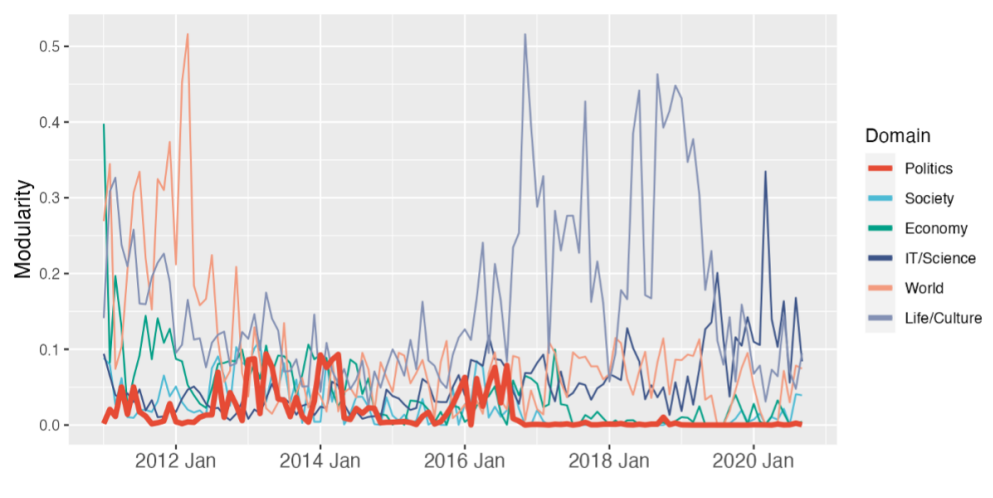


Figure 6.3: Monthly trends in modularity over time by news domain

6.2.4 Selected Media Proportion Over Time

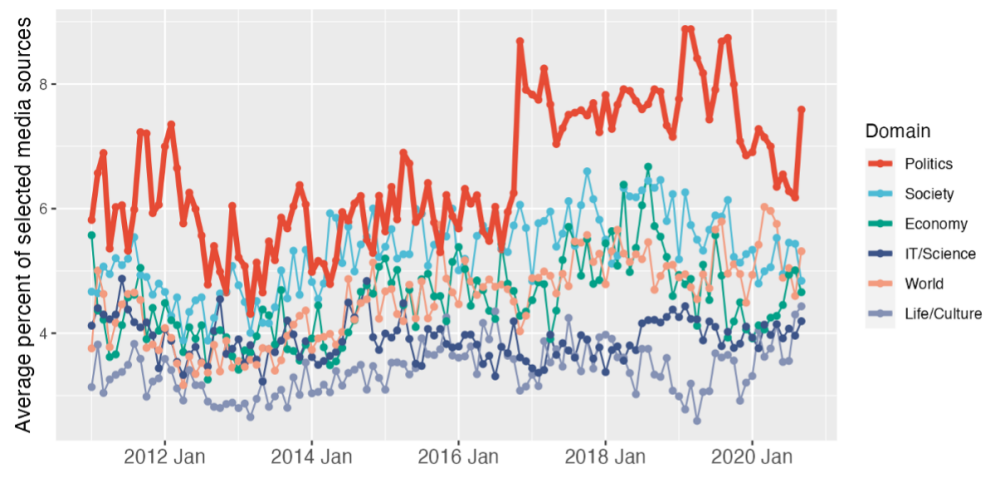


Figure 6.4: Monthly trends in selected media proportion over time by news domain

6.3 Distributions of Monthly Confrontation

The following figures show the distribution of angry reactions, by month, across the six issue domains (Appendix 6.3.1), the distribution of hostile comments, by month, across the six issue domains (Appendix 6.3.2) and the distribution of hostile users (users who post at least one hostile comment), by month, across the six issue domains (Appendix 6.3.3).

6.3.1 Distribution of Monthly Average Number of Angry Reaction

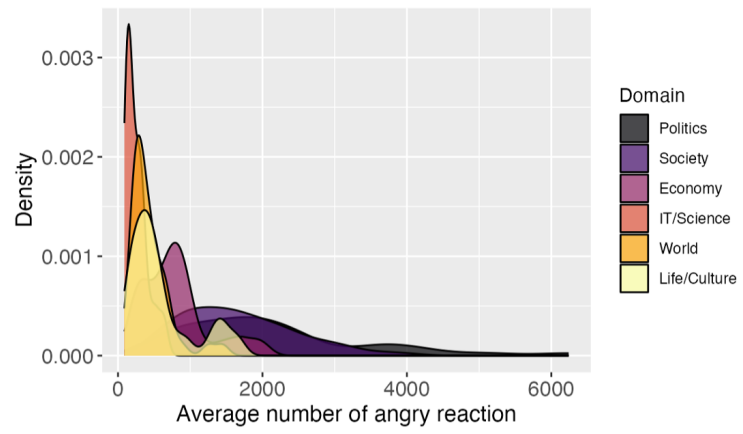


Figure 6.5: Distribution of monthly average number of angry reactions by news domain

6.3.2 Distribution of Monthly Percentage of Hostile Comments

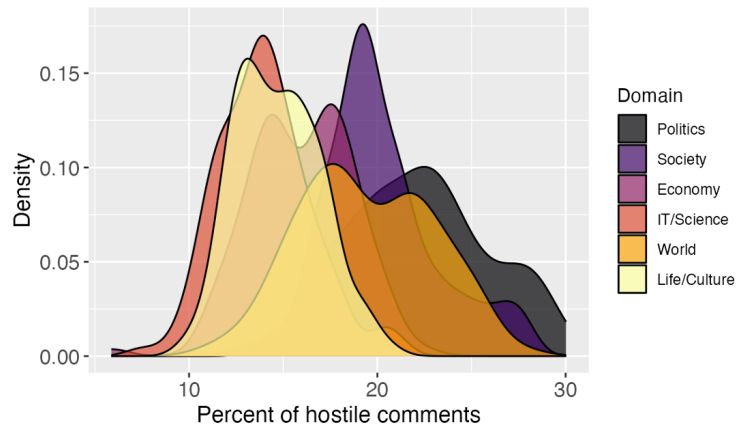


Figure 6.6: Distribution of monthly percentage of hostile comments by news domain

6.3.3 Distribution of Monthly Percentage of Hostile Users

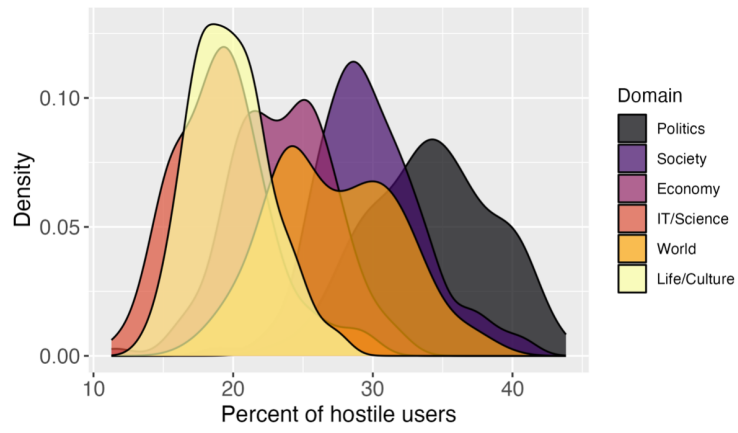


Figure 6.7: Distribution of monthly percentage of hostile users by news domain

6.4 An Alternate Measure of Modularity

Figure 6.8 replicates the analysis in Figure 4 using a different measure of modularity, namely, centrality. Figure 6.9 replicates the trend examined in Appendix 6.2.3 using the same centrality measure.

6.4.1 Centrality by News Domain

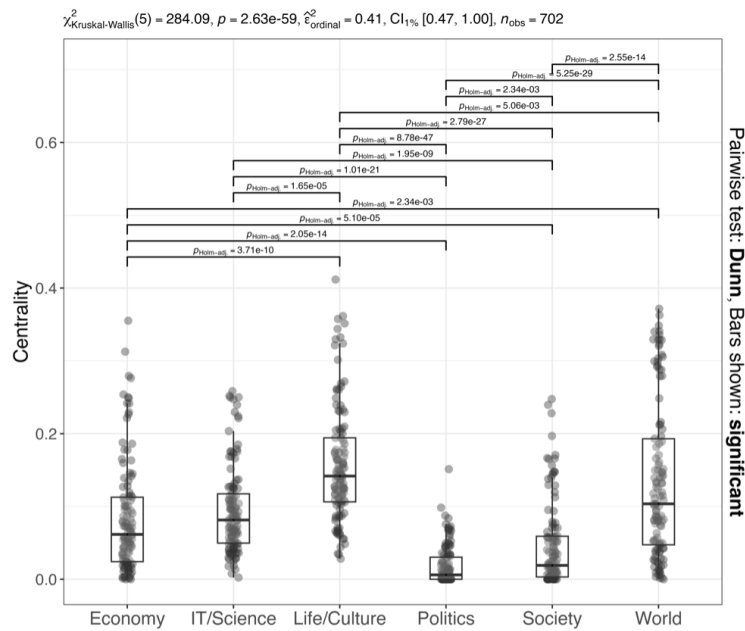


Figure 6.8: Centrality by news domain

6.4.2 Centrality Over Time by News Domain

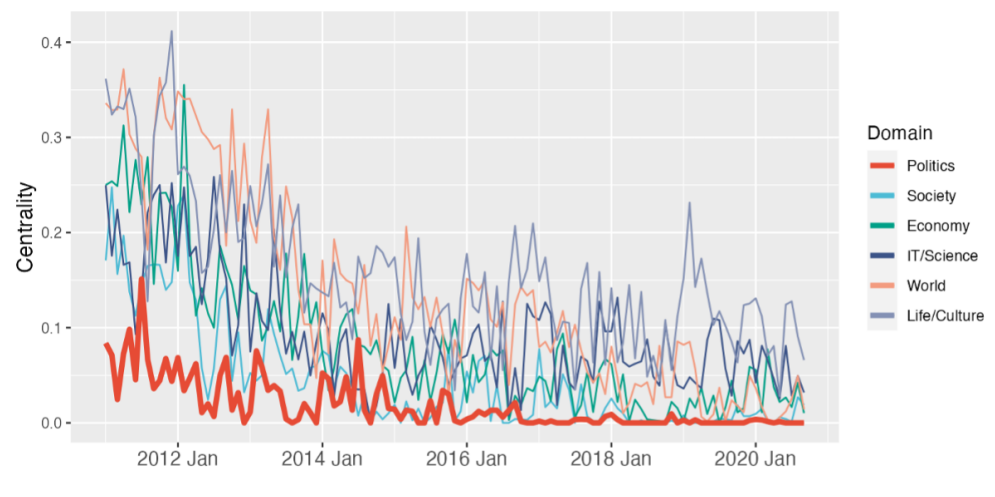


Figure 6.9: Centrality over time by news domain

6.4.3 Centrality of Hostile versus Non-Hostile Users

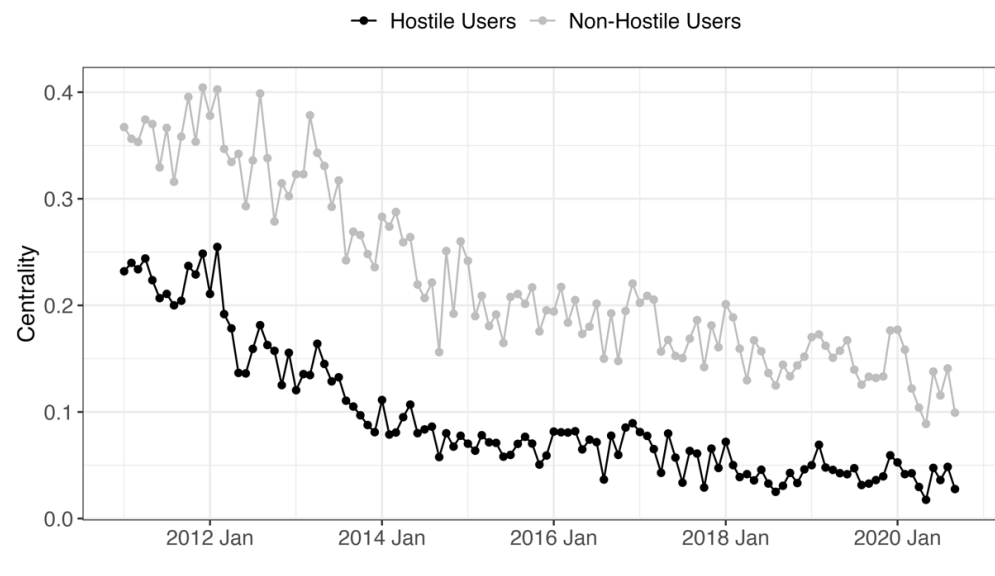


Figure 6.10: Centrality of hostile versus non-hostile users

6.5 Additional Aggregate-level Analyses of Hostility and Source Diversity

In the multilevel regression model, the primary independent variables are *hostility* and *anger*, and the dependent variable is *modularity*. We included a month indicator as a random slope. All key variables – hostility, anger, and modularity – were standardized with a mean of 0.

The following graph (Figure 6.11) and table (Table 6.2) shows the results from four models: Model 1 includes just hostility alongside news domains. This first model relies on data across the entire decade. Model 2 includes anger as dependent variable, and is constrained to data from April 2017 onwards. Again, it includes news domains. Models 3 and 4 then add aggregate-level demographic controls, provided at the story level by Naver. When accounting for demographics, the data is confined to the period starting April 2016, when the disclosure of the demographic composition within each article’s comment section began.

In every case, hostility is significantly negatively correlated with modularity. In a model excluding control variables, an increase in comment hostility by one standard unit is linked to a decrease of 0.19 standard units in modularity. Figure 6.11 does not suggest any major change when controls are added. Anger is also significantly negatively correlated. An increase in anger by one standard unit is associated with a 0.14 standard unit decrease in modularity without controls; the effect increases to 0.25 with controls.

6.5.1 Multilevel Regression Result Graph

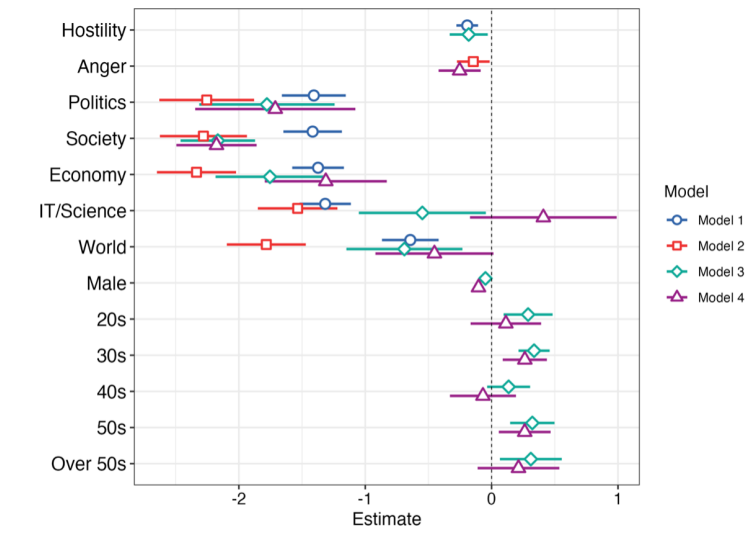


Figure 6.11: Multilevel regression results predicting modularity

6.5.2 Multilevel Regression Result Table

Table 6.2: Multilevel Regression Results Predicting Modularity

Predictor	Model 1 Modularity	Model 2 Modularity	Model 3 Modularity	Model 4 Modularity
(Intercept)	1.03*** (0.08)	1.66*** (0.12)	-20.55** (6.97)	-3.97 (10.59)
Hostility	-0.19*** (0.04)		-0.18* (0.08)	
Anger		-0.14* (0.07)		-0.25** (0.07)
Politics	-1.41*** (0.13)	-2.55*** (0.19)	-1.78*** (0.27)	-1.71*** (0.32)
Society	-1.42*** (0.12)	-2.28*** (0.18)	-2.17*** (0.15)	-2.18*** (0.16)
Economy	-1.37*** (0.10)	-2.34*** (0.16)	-1.75*** (0.22)	-1.31*** (0.25)
IT/Science	-1.32*** (0.10)	-1.53*** (0.16)	-0.55* (0.26)	0.41 (0.30)
World	-0.64*** (0.11)	-1.78*** (0.16)	-0.69** (0.23)	-0.45 (0.24)
Male			-0.05*** (0.01)	-0.10*** (0.02)
20s			0.29** (0.10)	0.11 (0.14)
30s			0.34*** (0.06)	0.26** (0.09)
40s			0.13 (0.09)	-0.07 (0.13)
50s			0.32*** (0.09)	0.26* (0.10)
Over 50s			0.31 (0.12)	0.21 (0.16)
<i>Random Effects</i>				
σ^2	0.63	0.53	0.45	0.40
τ_{00}	0.01	0.00	0.00	0.01
ICC	0.01		0.01	0.02
N	117	42	54	42
Observations	702	252	324	252
Marginal R^2 / Conditional R^2	0.371 / 0.378	0.592 / NA	0.655 / NA	0.691 / 0.696
<i>Note.</i> * $p < .05$, ** $p < .01$, *** $p < .001$				

6.5.3 User Hostility and Selected Media Sources

Figure 6.12 provides an additional overview of the relationship between hostility and user engagement with media outlets, this time focusing on monthly averages of (a) user hostility, measured as the number of hostile comments left by individual users, and (b) diversity in users' engagement across media sources, measured as the percent of media sources on which users made comments out of all available sources in a given news domain and time. The figure illustrates a strong positive correlation (0.715) between hostility and news source diversity: the more hostile comments individuals leave, the greater the variety of media outlets they choose from the available media pool.

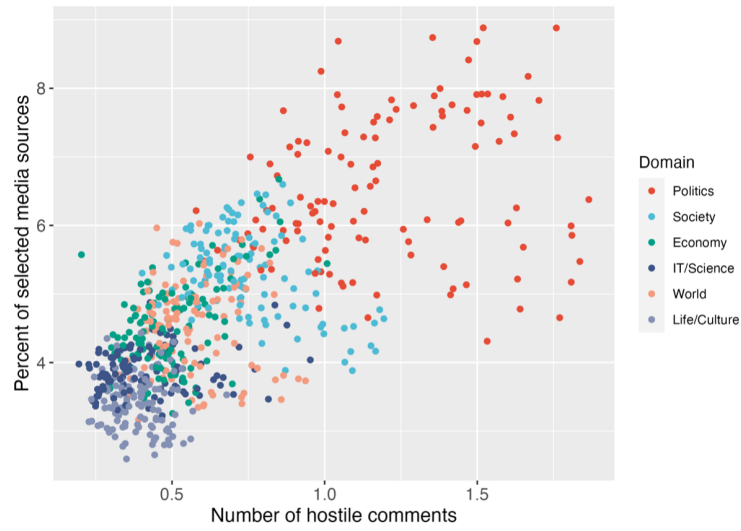


Figure 6.12: User hostility and selected media sources.

6.6 Results from the Generalized Linear Mixed Model (GLMM)

Table 6.3 shows results from the GLMM estimation that informs Figure 3.8 in the paper. Table 6.4 shows bootstrapped confidence intervals for the GLMM estimates, and Figure 6.13 illustrates the distributions of the bootstrapped estimates.

6.6.1 Generalized Linear Mixed Model (GLMM) Result

Table 6.3: Generalized Linear Mixed Model (GLMM) Result

Predictor	Percent of selected media outlets
(Intercept)	0.72*** (0.02)
Hostility	0.16*** (0.00)
Politics	0.55*** (0.02)
Society	0.44*** (0.02)
Economy	0.32*** (0.02)
IT/Science	−0.00 (0.02)
World	0.23*** (0.02)
<i>Random Effects</i>	
σ^2	0.34
τ_{00} userIdNo	0.05
τ_{00} yearmonth_id	0.00
ICC	0.13
N yearmonth_id	117
N userIdNo	1000
Observations	11656
Marginal R^2 / Conditional R^2	0.274 / 0.369
<i>Note.</i> * $p < .05$, ** $p < .01$, *** $p < .001$	

6.6.2 Bootstrap Result: Confidence Intervals for GLMM Estimates

Table 6.4: Bootstrap Result: Confidence Intervals for GLMM Estimates

Variable	Mean Estimate	Lower CI	Upper CI
(Intercept)	0.72	0.62	0.79
Hostility	0.16	0.15	0.18
Politics	0.55	0.47	0.64
Society	0.44	0.36	0.53
Economy	0.32	0.23	0.42
IT/Science	0.01	−0.11	0.12
World	0.24	0.13	0.34

Note. 95% confidence intervals calculated from bootstrap distributions using percentiles.

6.6.3 Bootstrap Result: Distributions of GLMM Estimates

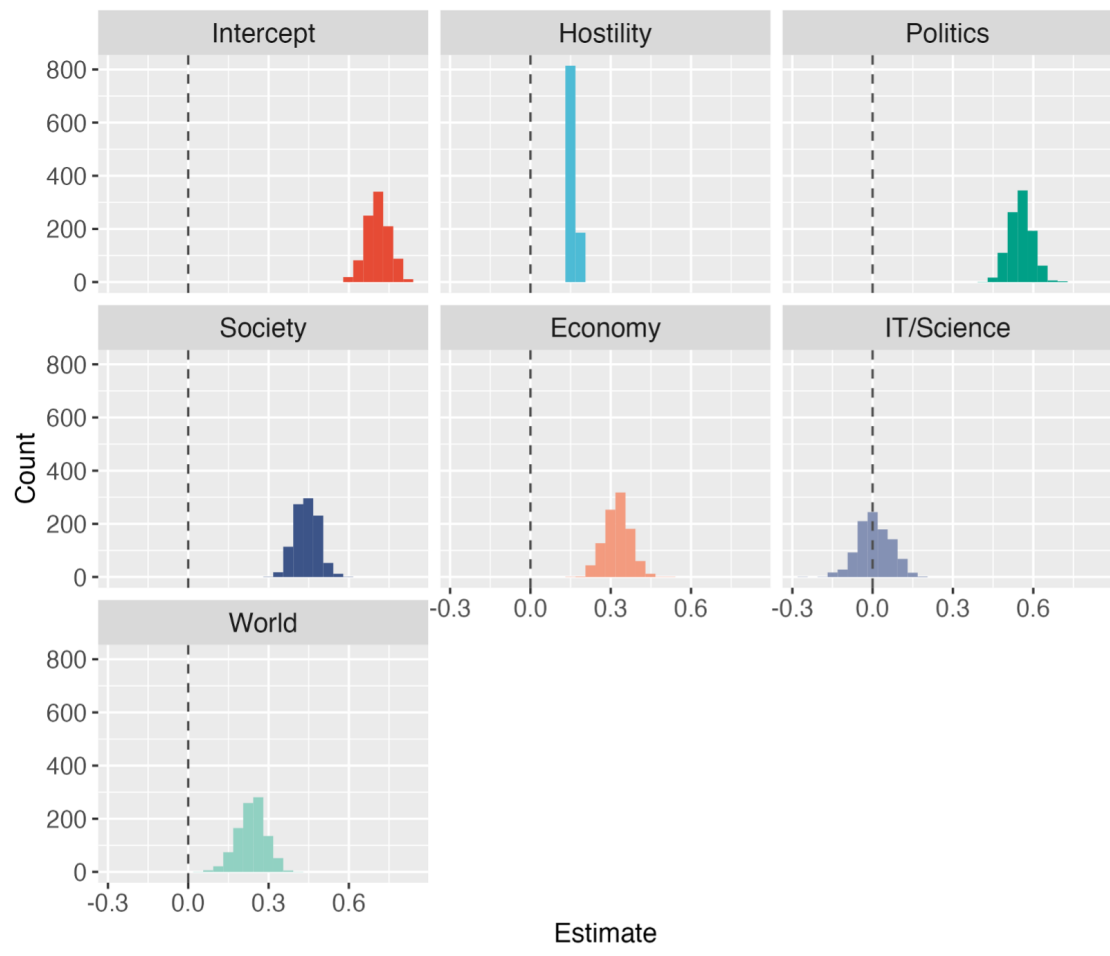


Figure 6.13: Bootstrap distributions of GLMM estimates

6.7 Case Study 1: Nuclear Phase-Out Policy

6.7.1 Context

In 2017, South Korea announced a policy to gradually reduce and eventually phase out its reliance on nuclear power as a source of energy. This policy was introduced by President Moon Jae-in's administration. The primary goal of the policy was to shift the country's energy mix towards more renewable sources, such as solar and wind power, while reducing the reliance on nuclear energy over time. The policy aimed to address safety concerns following the Fukushima disaster in Japan and to respond to public sentiment regarding nuclear energy's potential risks. In line with this policy direction, President Moon made a commitment to halt the construction of Shin-Gori Nuclear Reactors No. 5 & 6, as well as the decommissioning of one of Korea's oldest nuclear reactors, Gori No. 1.

On May 18th, Korea Hydro & Nuclear Power (KHNP) announced their adherence to the administration's policy, generating significant attention from the media and the general public. This led to heightened polarization between the policy's supporters, often referred to as liberals, and its opponents, commonly referred to as conservatives. While supporters favored the policy, arguing that transitioning to renewable energy sources would promote environmental sustainability and long-term viability, opponents expressed concerns about potential electricity cost increases and the risk of energy shortages. They asserted that alternatives to nuclear power might not offer the same level of reliability. The policy debate encompassed a range of issues, from specific nuclear reactors such as Shin-Gori No. 5 & 6, Gori No. 1, and Wolsong, to broader considerations encompassing energy security, economic impact, environmental sustainability, public safety, and technological progress.

As public controversies escalated, the Moon administration declared its intention to pursue societal agreement on the matter and to adhere to the results of public deliberation. Consequently, the administration officially established the Public Deliberation Committee for this policy on July 24th. The extensive public deliberation took place amid considerable

media attention in October. During this process, representative citizens were provided with scientific and technical insights by experts from both pro-nuclear and anti-nuclear perspectives (Nam et al. 2022). These citizens then voted on the policy based on the knowledge they had gained. After the citizens’ deliberative decision, which opposed halting the construction of Shin-Gori No. 5 & 6, the administration indicated its intention to align with this decision.

6.7.2 Data and Classification

The case study focuses on news articles and their comments published from May 18th, when the policy began gaining momentum and attracting media and public attention, to July 23rd, just before the establishment of the Public Deliberation Committee when the issue started shifting from the policy to the deliberation process itself. This timeframe belongs to the transition phase as described by Nam et al.’s (2022) study, during which South Korea shifted from its previous nuclear-friendly policy to the implementation of the nuclear phase-out policy.

We initially identified Korean domestic news articles containing specific keywords in their titles from the Naver Ranking News list within the specified timeframe. The search keywords used were as follows: (1) Nuclear, (2) Nuclear Power, (3) Nuclear Power Plant, (4) Nuclear Fuel, (5) Nuclear phase-out, (6) Fukushima, (7) IAEA, (8) Shin-Gori, (9) Wolseong, (10) No. 1 (Unit 1), (11) No. 5 & 6 (Unit 5 & 6). All keywords were utilized in the Korean language. Subsequently, we gathered all comments posted on these identified articles. This process resulted in a total of 108 news articles and 102,256 comments.

The collected articles and comments were subsequently categorized into three groups: (1) those opposing the policy, (2) those supporting the policy, and (3) those adopting a neutral perspective towards the policy. In the case of articles, a manual coding approach was employed. The coding process took into account both article titles and content. Articles that framed the policy in a negative light (e.g., citing high costs, governmental haste, etc.) were classified as policy-opposing articles. Conversely, articles that portrayed the policy in a pos-

itive manner (e.g., highlighting its promising future, affordability of renewable energy, etc.), were categorized as policy-supportive articles. Articles that presented objective information or balanced viewpoints were designated as policy-neutral articles. This procedure yielded 32 policy-opposing articles, 30 policy-supportive articles, and 46 policy-neutral articles.

For comment classification, we employed kcELECTRA-base-v2022, a Korean pre-trained language model based on the ELECTRA architecture.¹ To train this model, 5% of the total comments were randomly selected and manually labeled into the same three categories. The categorization process was relatively straightforward due to the limited scope of the issue during the target timeframe and the clear division between policy supporters and opponents (Nam et al. 2022). Comments were labeled as policy-opposing if they explicitly criticized the policy or the Moon administration. Conversely, comments were labeled as policy-supportive if they expressed direct endorsement for the administration. Comments that did not clearly take a stance on the policy or the government were classified as policy-neutral. The trained model achieved an accuracy level of 85.1%. Using this model, all comments were automatically categorized, resulting in 66,747 policy-opposing comments and 35,509 policy-supportive comments. Our model did not identify any neutral comments.

¹kcELECTRA-base-v2022 is a pre-trained transformer model optimized for Korean text classification tasks.

6.8 Case Study 2: THAAD Deployment

6.8.1 Context

The Terminal High Altitude Area Defense (THAAD) missile system was deployed in South Korea under the conservative Park Geun-hye administration in 2016 in response to escalating threats from North Korea. While intended as a national security measure, the policy quickly became one of the most polarized issues in South Korean politics. Conservatives largely supported THAAD deployment, emphasizing its role in protecting the nation from missile attacks and strengthening the U.S.–Korea alliance. Liberals and progressives, by contrast, opposed the deployment, citing environmental and health risks, the lack of democratic deliberation, and diplomatic backlash from China and Russia.

The public controversy intensified after the government’s announcement of the deployment site in Seongju in July 2016. Residents and civic groups organized large-scale protests over the perceived procedural unfairness and potential harm to local communities. The conservative side framed the protesters as “pro-North” or “anti-alliance,” while the liberal side highlighted concerns about sovereignty and regional peace. The THAAD issue thus mirrored the polarization observed in the 2017 nuclear-phaseout case examined in Study 1, representing the opposite ideological direction but a similar degree of public division.

6.8.2 Data and Classification

We analyzed news coverage and public commentary surrounding the THAAD policy using the same procedures applied in Study 1. News articles were collected from Naver News between July 8 and July 31, 2016, the period spanning the government’s official announcement and the subsequent escalation of protests and diplomatic tension. Using the keywords “THAAD” (both in English and Korean) in article titles, we identified 213 articles accompanied by 166,330 comments.

Each article was manually coded into one of three categories based on tone and

framing toward the policy: policy-supportive, policy-opposing, or policy-neutral. Articles emphasizing North Korean missile threats, national security, or the effectiveness of U.S.-Korea defense cooperation were coded as policy-supportive (86 articles). Articles highlighting protest movements, environmental risks, Chinese retaliation, or procedural illegitimacy were coded as policy-opposing (82 articles). Articles presenting largely factual or balanced coverage were coded as policy-neutral (45 articles).

A random sample of 3,000 comments was then manually labeled into the same three categories. Comments supporting the policy typically endorsed government decisions or criticized protesters and liberal parties; those opposing the policy criticized the Park administration, the U.S. alliance, or the decision-making process; and neutral comments lacked a clear stance.

The labeled data—comprising 1,453 policy-supportive, 1,146 policy-opposing, and 566 neutral comments—were then used to train the same transformer-based model employed in Case Study 1, achieving an overall accuracy of approximately 75%. Despite lower performance in detecting neutral comments, the classifier showed strong accuracy in distinguishing the two polarized sides, correctly identifying policy-supportive comments at 84.6% and policy-opposing comments at 86.6%. The classifier was then applied to all 166,330 comments to categorize them into policy-supportive, policy-opposing, or neutral positions. This yielded 75,696 policy-supportive, 67,576 policy-opposing, and 23,058 policy-neutral comments. Users’ political orientations were then determined by the dominant stance in their comments, resulting in 22,996 policy supporters, 30,962 policy opponents, and 3,801 neutral users. The results are presented excluding these neutral comments and users.

6.8.3 Results

The figure below illustrates commenting patterns across news stance and user orientation, parallel to Figure 3.2 in the main text. Results closely replicate those from the 2017 nuclear-phaseout case.

Users on both sides of the THAAD debate were significantly more likely to post comments on counter-attitudinal news articles than on congruent ones. Supporters of THAAD frequently commented on articles critical of the policy — often to challenge the coverage or denounce protesters—while opponents more actively engaged with articles supportive of the deployment, criticizing the government or the perceived militarization of national policy. This asymmetry is evident for both total and hostile comments.

The persistence of this pattern across ideologically opposing policy contexts—a liberal-led issue (nuclear phaseout) and a conservative-led one (THAAD deployment)—demonstrates the robustness of the main finding: active and hostile commentary tends to cross attitudinal boundaries rather than remain confined within them. Even in a security-framed, conservative-initiated policy debate, commenting behavior reflects confrontation with counter-attitudinal content rather than confinement within like-minded narratives.

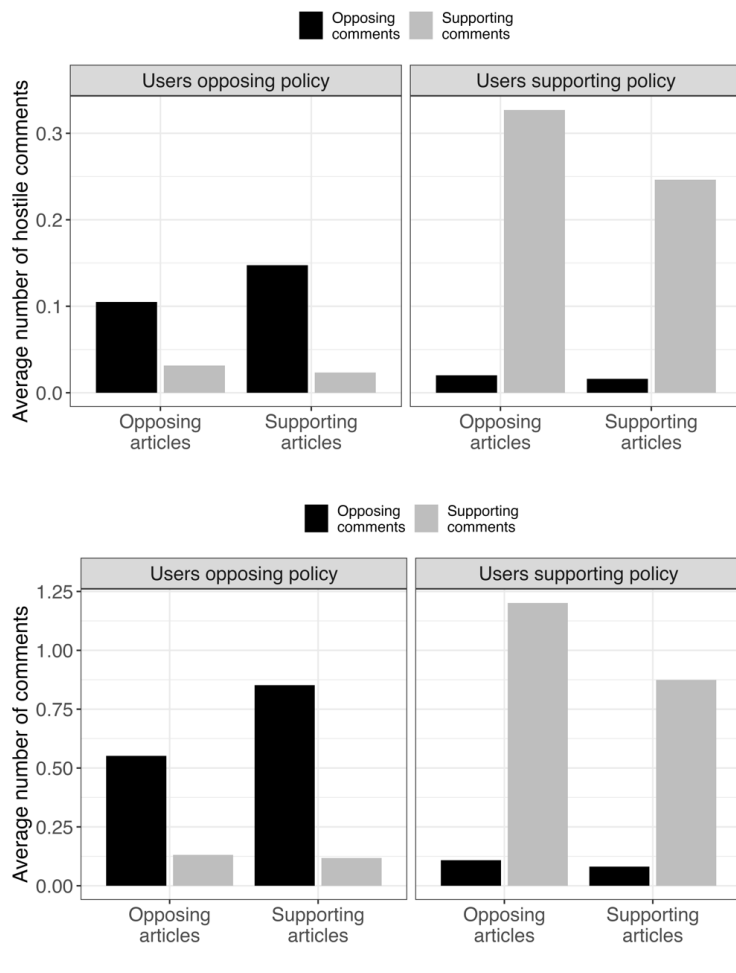


Figure 6.14: Commenting patterns across news stance and user orientation for the THAAD deployment case

Appendix for Chapter 3

7.1 Full Regression Results

Table 7.1: Logistic Regression Models Predicting Engagement Segregation

Predictor	Viewing (1)	Clicking (2)	Reading (3)
<i>Dependent variable: Different clusters (label = 1)</i>			
Constant	4.698*** (0.058)	5.066*** (0.050)	4.971*** (0.041)
Negativity difference	0.304* (0.165)	0.097 (0.119)	0.177 (0.118)
Partisan distance	0.066 (0.042)	0.305*** (0.048)	0.195*** (0.047)
Politics mismatch	0.455*** (0.057)	0.738*** (0.043)	0.720*** (0.044)
Entertainment mismatch	0.195*** (0.032)	0.183*** (0.030)	0.130*** (0.024)
Lifestyle mismatch	0.395*** (0.043)	0.208*** (0.030)	0.094*** (0.025)
Sports mismatch	0.353*** (0.059)	0.257*** (0.052)	0.208*** (0.037)
National news mismatch	−0.054 (0.040)	0.252*** (0.034)	0.269*** (0.042)
World news mismatch	0.215*** (0.071)	0.437*** (0.041)	0.400*** (0.039)
Business mismatch	−0.038 (0.058)	0.207*** (0.065)	0.070 (0.069)
Technology mismatch	0.445*** (0.077)	0.102* (0.061)	0.092 (0.063)
Partisan distance × High-interest share	−0.029 (0.052)	−0.053 (0.071)	−0.050 (0.046)
Share of high-interest pairs	−1.131*** (0.057)	−1.904*** (0.046)	−2.056*** (0.039)
Family SEs	Binomial (logit)	Binomial (logit) Cluster-robust by date	Binomial (logit)
Observations	14,970,230	14,965,103	14,961,840

*p<0.1; **p<0.05; ***p<0.01

7.2 Topical Categories

7.2.1 Platform-Assigned Categories: Examples

Table 7.2 presents three randomly selected headlines from each platform-assigned topical category, ordered by their frequency in the dataset.

Table 7.2: Example Headlines by Platform-Assigned Topical Category

Category	Example Headlines
Entertainment	1. Hollywood star, 80, falls off stage after getting out of wheelchair 2. ‘Family Crisis’ Finale: Mama June breaks down in tears as she recalls her ‘struggles’ during wedding vows 3. Royal Divorce: Prince Harry And Meghan Markle Attempting To Save Marriage
Politics	1. When an Angry Jill Biden Said Kamala Harris Should ‘Go F–K Herself’ for Debate Attack on Joe Biden 2. Trump team changes obscure GOP rules in hopes of clinching presidential nomination early 3. Sen. Bernie Sanders opposes giving Israel \$10 billion in aid
Lifestyle	1. Couple Spends Years Saving Up For First Child-Free Vacation, It Gets Ruined By Women Next Door 2. The Absolute Worst Things You Can Wear To An Airport 3. The Best And Worst Frozen Egg Products To Buy At The Grocery Store
Sports	1. College Football’s Fastest Recruit ‘Shocks The World’ With His Commitment 2. Chase Elliott confronted Mike Rockenfeller post-race for blocking him during Verizon 200 3. Jordan Poole Threw An Off-Glass Alley-Oop To Kyle Kuzma Down 21 To The Hawks
National	1. 2 injured, thousands of dollars in damage: Drone used to spot teenage suspects throwing rocks onto Wisconsin highway 2. Police found 115 bodies at Colorado ‘green’ funeral home while investigating putrid smells 3. Alabama police investigating after doorbell video captures deaf dog being stolen from inside home
World	1. Hamas rejects Egyptian proposal to end war with Israel, refuses free elections in Gaza 2. I’m an American who’s been living in Paris for 5 years. Here are 8 things that have most surprised me about living in the French capital.

Continued on next page

Table 7.2 (continued)

Category	Example Headlines
	3. Ukraine’s ‘deep strike battle’ in Black Sea deals biggest blow so far in war to Putin fleet, says UK
Business	1. Bank of America to pay \$250M in fines, customer refunds over junk fees, fake accounts 2. Dollar Tree Is Dropping Prices Back to \$1 — Here’s What Could Be Impacted 3. Coca-Cola follows rival Pepsi’s lead and makes major pricing change as CEO says people are ‘looking for value’
Technology	1. Urgent warning for Android owners to ‘delete app now’ – it steals your texts and everything you type 2. Microsoft just made Paint useful again 3. Users ‘stop going on YouTube’ after streaming site’s new rule against ad blockers
Science	1. The Truth Behind The “Aztec” Crystal Skulls Continues To Fascinate 2. Einstein, Newton could have been wrong about gravity, Gaia telescope data 3. Algae blooms, record heat: Florida climate change puts us all in movie with bad ending

7.2.2 Uncategorized Articles

A substantial share of articles in the dataset (41.9%) were assigned an “unknown” topical label by the platform and consequently excluded from the regression analyses. This raises the question of whether the uncategorized articles differ systematically from categorized articles along the key dimensions examined in this study—namely, negativity and partisan slant—in ways that might bias the reported results. This section presents three complementary checks suggesting that uncategorized articles are not meaningfully different from categorized ones.

Comparison of Key Variables Uncategorized articles show nearly identical negativity ($M = 0.04$, $SD = 0.09$) compared to categorized articles ($M = 0.04$, $SD = 0.09$) in the regression sample, with a difference of 0.002 and a Cohen’s d of 0.022. Similarly, partisan slant differs only marginally between uncategorized ($M = -0.04$, $SD = 0.33$) and categorized ar-

ticles ($M = -0.06$, $SD = 0.35$), with a difference of -0.022 and a Cohen’s d of -0.064 . Both effect sizes—reported here in lieu of significance tests, which would flag even trivially small differences as significant given the large sample—fall well below the conventional threshold of $|d| = 0.20$ for a small effect (Cohen 1988), suggesting that excluding uncategorized articles is unlikely to introduce systematic bias into the regression estimates along these dimensions.

User-Level Distribution of Uncategorized Articles A second concern is whether uncategorized articles are concentrated among particular users, which could introduce user-level bias even if the articles themselves are similar on average. To address this, the analysis is restricted to active users who viewed more than 1,000 articles ($n = 1,503$). Figure 7.1 addresses this concern in two ways. The left panel shows the distribution of the proportion of uncategorized articles across these active users. The distribution is roughly unimodal and approximately normal, centered around 40% ($M = 39.97\%$, $SD = 10.64\%$), suggesting that the unknown rate reflects a stable property of the platform’s classification system rather than anything systematic about user type—every active user encounters roughly the same proportion of uncategorized content, give or take. The right panel presents a scatter plot of each user’s proportion of uncategorized articles against their total engagement volume. The unknown rate stabilizes quickly as activity increases, converging to a 30–50% band with no discernible upward or downward trend as total article volume grows. Together, these patterns are consistent with missingness reflecting platform-level classification practice rather than user-level selection, suggesting that the exclusion of uncategorized articles does not disproportionately affect any particular type of user.

7.2.3 Topical Composition of Uncategorized Articles

A third check examines whether uncategorized articles are topically distinct from categorized ones—that is, whether they cluster into topic areas that might not be well represented in the categorized sample. To assess this, I estimate the topical composition

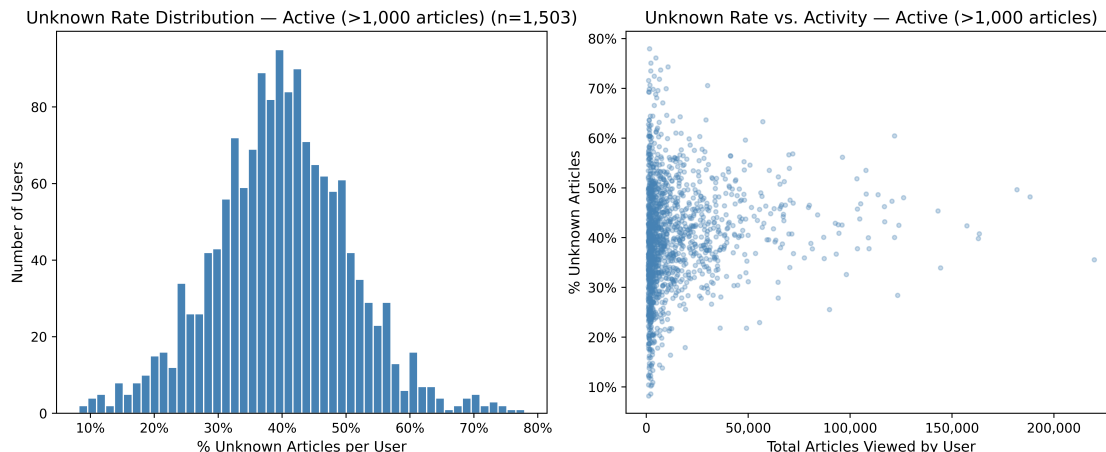


Figure 7.1: Distribution of uncategorized articles across users. Left panel: histogram of the proportion of uncategorized articles per user. Right panel: scatter plot of uncategorized article proportion against total user engagement volume.

of both categorized and uncategorized articles using a zero-shot text classification model (facebook/bart-large-mnli; Yin et al. 2019), which assigns topic probabilities to each article independent of the platform-assigned labels. Note that the topic labels were manually specified and differ slightly from the platform-assigned categories, though all major topical domains are retained. Figure 7.2 plots the resulting topic distributions for categorized and uncategorized articles across eight topic categories.

The two distributions are broadly similar, though some differences are worth noting. Political and sports content are somewhat better represented among categorized articles (politics: 14.0% vs. 8.4%; sports: 14.8% vs. 7.6%), while domestic, entertainment, technology/science, lifestyle, and business/economy content are slightly more prevalent among uncategorized articles. International content is nearly identical across the two groups (1.8% vs. 1.8%).

The exclusion of uncategorized articles therefore does not appear to be entirely random with respect to topic. However, since the theoretical interest of this study lies precisely in how topical boundaries—particularly the political/non-political divide—structure engagement segregation, the fact that political content is more completely captured in the regression sample means that the core theoretical comparisons of interest are well-represented in the

analysis. If anything, the underrepresentation of political content among uncategorized articles suggests that the regression sample is well-suited to detect the politics mismatch effect. Nevertheless, the direction and magnitude of any bias introduced by this non-random exclusion remains difficult to determine with precision, and this is acknowledged as a limitation of the study.

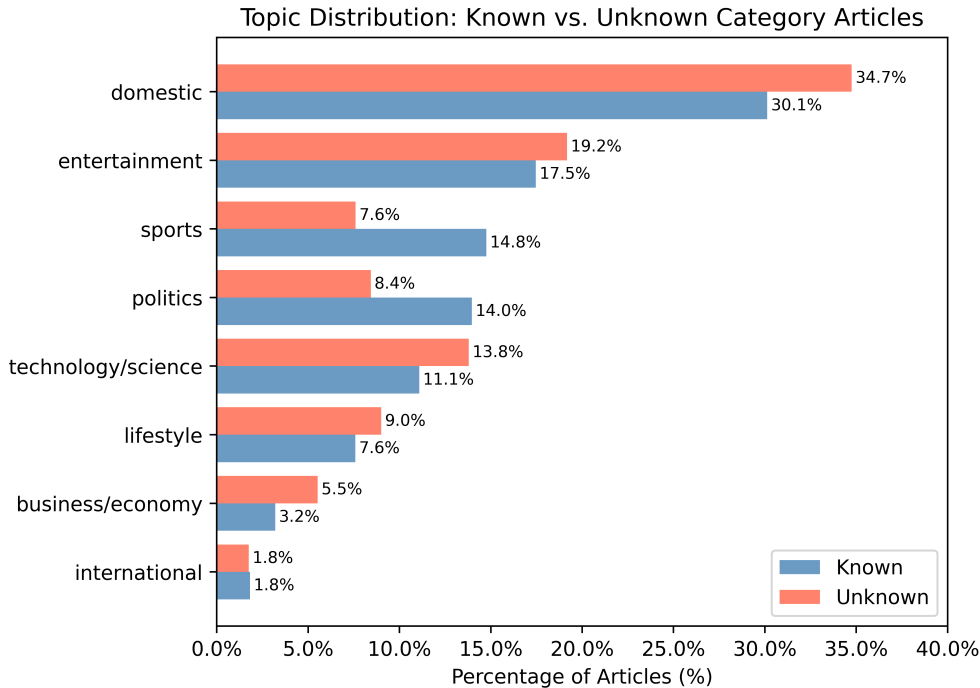


Figure 7.2: Topic distribution of categorized and uncategorized articles estimated using a zero-shot text classification model (Yin et al. 2019). Blue bars represent articles with platform-assigned topical labels (known); red bars represent articles without platform-assigned labels (unknown).

7.3 Partisan Slant Estimation

7.3.1 Estimation and Validation

Partisan slant was estimated for all unique headlines in the regression sample—covering randomly selected 30 dates from the full dataset, restricted to articles with known platform-assigned topical categories—using GPT-5-mini. Each headline was classified into one of three categories: Liberal, Conservative, or Neutral, based on the following prompt:

You are classifying the political orientation of news headlines. Determine whether the headline expresses a liberal, conservative, or neutral political stance. Definitions: Liberal: The wording or framing reflects positions, priorities, or perspectives commonly associated with liberal or left-leaning politics. Conservative: The wording or framing reflects positions, priorities, or perspectives commonly associated with conservative or right-leaning politics. Neutral: The headline reports information without clear ideological framing or partisan language. Return ONLY one word and nothing else: Liberal, Conservative, Neutral.

The resulting distribution is heavily skewed toward neutral across all three engagement stages. In the viewing dataset, 87.2% of headlines were classified as Neutral, 9.3% as Liberal, and 3.5% as Conservative ($n = 5,372,027$). The distribution is similar in the clicking dataset (Neutral: 85.9%, Liberal: 10.2%, Conservative: 3.9%; $n = 4,053,609$) and the reading dataset (Neutral: 85.6%, Liberal: 10.4%, Conservative: 4.0%; $n = 3,710,895$). The consistent liberal lean in the non-neutral content across all stages is reflected in the modest negative mean partisan scores reported in the main text.

To validate the model’s classifications, a random sample of 150 headlines was drawn (50 Conservative, 50 Liberal, and 50 Neutral) and independently coded by a research assistant. Raw agreement between the model and human coder was 96.7% (145 out of 150), with a Cohen’s κ of 0.95.

7.3.2 Examples by Category

Table 7.3 presents five randomly selected headlines from each partisan category to illustrate the types of content classified as Liberal, Conservative, and Neutral.

Table 7.3: Example Headlines by Partisan Slant Category

Category	Example Headlines
Conservative	1. Illegal Immigration Surges Months After Biden, Trudeau Make Deal To Crack Down On Northern Border Influx 2. ‘I WAS ANGRY’: Donald Trump Fires Back After ‘Crazy’ Liz Cheney Claims Ex-prez Was ‘Depressed and Not Eating’ After Jan 6 3. Lessons from Vietnam: Israel must destroy Gaza tunnels, Hamas — Washington Times 4. Austin crime victim’s mother rips city council ‘cowards’ for defunding police, turning citizens into ‘targets’ 5. Soros-Backed California Lieutenant Governor Calls on Elections Official To Block Trump From Ballot
Liberal	1. Left-wing writer accuses Israel’s Netanyahu of ‘toxic masculinity’ 2. Trump’s supporters protesting outside courthouse have no idea what’s actually going on: legal experts 3. Deputy’s Brutal Assault and Kneeing of Elderly Black Woman Spurs Protest by LA Activists 4. Trump Is Hosting \$100,000-Per-Plate Dinner For Cash-Strapped Rudy Giuliani. Snark Is Served. 5. A 2% global tax on all billionaires could raise \$250 billion annually, a research report backed by a Nobel Prize-winning economist says
Neutral	1. Governor awards \$1.2 million to install EV charging stations 2. Ezra Miller’s The Flash Is on Track to Be One of the Worst-Performing Superhero Films Ever 3. ‘Fancy’: Iggy Azalea On Pace to Join Nicki Minaj & Cardi B As Only Female Rappers with Diamond Singles 4. Hunky Beyoncé Security Guard Goes Viral After His Reaction To Crowd’s Silence—And The Thirst Is Real 5. Nebraska Volleyball Player Making Headlines Ahead Of Huge Match

7.4 Co-Engagement Network

7.4.1 Basic Network Characteristics

Daily co-engagement networks were constructed for each of the three engagement stages across 30 randomly sampled dates. As described in the main text, nodes represent unique news articles and edges connect article pairs that share at least one common user on the same day, with edge weights reflecting the number of shared users. Table 7.4 summarizes the average network characteristics after hypergeometric backbone extraction, separately for each engagement stage.

Network size and density decline sharply across stages, reflecting the progressive narrowing of co-engagement as users move from passive exposure to active reading. To characterize community structure independently of the degree-corrected stochastic block model, modularity was also computed on the filtered networks using the Louvain algorithm (Blondel et al. 2008). Modularity scores increase consistently across stages, from 0.277 at viewing to 0.407 at clicking and 0.432 at reading, consistent with the pattern of intensifying community structure reported in the main text.

Table 7.4: Average network characteristics of the co-engagement networks across 30 sampled days, by engagement stage.

Stage	Users	Articles	Edges	Density	Modularity
Viewing	1,295	22,981	3,573,320	0.076	0.277
Clicking	1,235	5,964	311,888	0.018	0.407
Reading	1,170	4,806	175,643	0.015	0.432

7.4.2 Degree-Corrected Stochastic Block Model

Community detection was performed using the degree-corrected stochastic block model (DCSBM; Karrer and Newman 2011), as implemented in the `graph-tool` library (Peixoto 2017). The DCSBM extends the standard stochastic block model by introducing

a per-node degree parameter, allowing the model to account for the highly heterogeneous degree distributions characteristic of co-engagement networks—where a small number of articles attract very large shared audiences while most attract few. Without this correction, high-degree nodes tend to dominate block assignments, producing spurious community structure. The number of blocks is determined automatically using the minimum description length (MDL) criterion, which selects the partition that minimizes the total description length of the network—that is, the number of bits required to encode both the model and the data given the model—thereby avoiding overfitting without requiring the number of communities to be specified in advance.

Average MDL per edge is comparable across the three stages: 3.57 bits per edge at the provision stage, 3.88 at selection, and 4.08 at consumption, indicating that the DCSBM provides consistent and reliable partitions across all three engagement stages.

7.4.3 Distribution of Block Sizes

Figure 7.3 displays the distribution of block sizes, defined as the number of articles assigned to each non-empty block, pooled across all 30 sampled days and presented separately for each engagement stage. At all three stages, block sizes follow a strongly right-skewed distribution: the large majority of blocks are small, with a median of 18 articles at the viewing stage, 16 at clicking, and 17 at reading. A long tail of larger blocks is present in each network, reflecting a small number of high-traffic clusters such as breaking news or major political stories that attract much broader shared readership. The maximum observed block size declines across stages, from approximately 500 articles in viewing networks to around 200 in both clicking and reading networks. This pattern is substantively meaningful: the algorithm surfaces a wide and diverse set of content producing some very large co-engagement clusters, whereas user-driven engagement narrows and fragments the landscape into smaller and more specialized communities, consistent with the intensifying segregation reported in the main text.

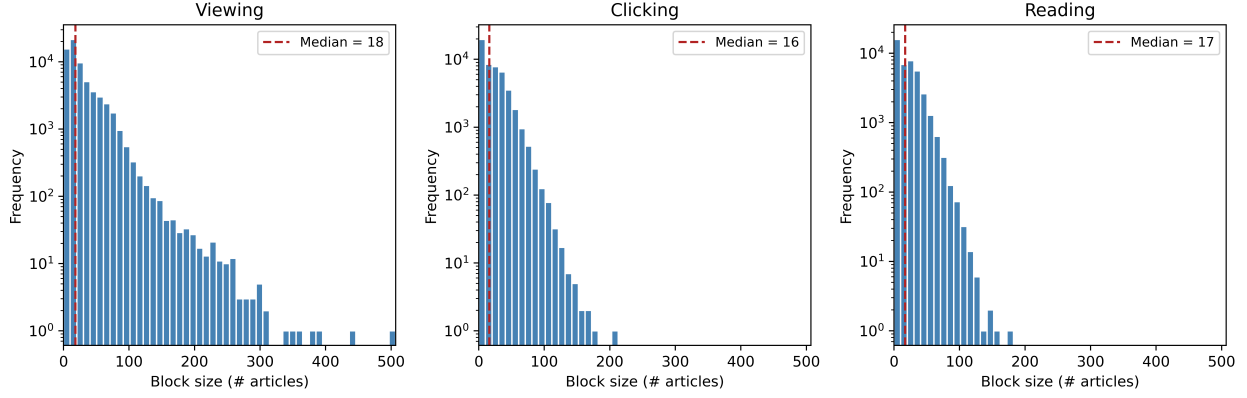


Figure 7.3: Distribution of block sizes across the three engagement stages, pooled across 30 sampled days. Each bar represents the frequency of blocks containing a given number of articles (log scale). Dashed red lines indicate the median block size at each stage (viewing: 18, clicking: 16, reading: 17).

7.5 Additional Segregation Metrics

Table 7.5 reports two additional metrics of network segregation across the three engagement stages. Block assortativity measures the tendency of articles to co-engage with articles in the same block, ranging from 0 (no preference) to 1 (perfectly assortative). Conductance measures the proportion of each community’s total co-engagement weight that crosses block boundaries, where lower values indicate more cohesive and isolated communities. Both metrics are consistent with the segregation index reported in the main text: assortativity increases across stages while conductance decreases, indicating that communities become progressively more internally cohesive and boundary-bound as users move from algorithmic exposure to active reading.

Table 7.5: Block assortativity (weighted) and mean conductance across engagement stages.

Stage	Block Assortativity	Mean Conductance
Viewing	0.038	0.959
Clicking	0.156	0.838
Reading	0.195	0.805

References

- Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., and Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of Public Economics*, 191:104254.
- Bachleda, S., Neuner, F. G., Soroka, S., Guggenheim, L., Fournier, P., and Naurin, E. (2020). Individual-level differences in negativity biases in news selection. *Personality and Individual Differences*, 155:109675.
- Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132.
- Barefoot, J. C., Patterson, J. C., Haney, T. L., Cayton, T. G., Hickman, J. R., and Williams, R. B. (1994). Hostility in asymptomatic men with angiographically confirmed coronary artery disease. *The American Journal of Cardiology*, 74(5):439–442.
- Barnidge, M., Diehl, T., Sherrill, L. A., and Zhang, J. (2021). Attention centrality and audience fragmentation: An approach for bridging the gap between selective exposure and audience overlap. *Journal of Communication*, 71(6):898–921.
- Benkler, Y., Faris, R., and Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.
- Bergamaschi Ganapini, M. (2023). The signaling function of sharing fake stories. *Mind & Language*, 38(1):64–80.
- Berger, J. (2011). Arousal increases social transmission of information. *Psychological Science*, 22(7):891–893.
- Berinsky, A. J. (2018). Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey responding. *The Journal of Politics*, 80(1):211–224.

- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., and Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28):7313–7318.
- Bullock, J. G., Gerber, A. S., Hill, S. J., and Huber, G. A. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, 10(4):519–578.
- Carver, C. S. and Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin*, 135:183–204.
- Cheong, Y. and Haggard, S. (2023). Political polarization in Korea. *Democratization*, 30(7):1215–1239.
- Cho, J., Ahmed, S., Hilbert, M., Liu, B., and Luu, J. (2020). Do search algorithms endanger democracy? an experimental investigation of algorithm effects on political polarization. *Journal of broadcasting & Electronic media*, 64(2):150–172.
- Coe, K., Kenski, K., and Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4):658–679.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 2 edition.
- Coppini, D., Duncan, M. A., McLeod, D. M., Wise, D. A., Bialik, K. E., and Wu, Y. (2017). When the whole world is watching: A motivations-based account of selective expression and exposure. *Computers in Human Behavior*, 75:766–774.
- Courtois, C., Slechten, L., and Coenen, L. (2018). Challenging google search filter bubbles

- in social and political information: Disconforming evidence from a digital methods case study. *Telematics and Informatics*, 35(7):2006–2015.
- Cragun, J. (2022). Rigid religious faith promotes selective exposure to attitude-congruent political information. *Political Behavior*, 44(3):1175–1201.
- Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1(11):769–771.
- Dang, C. N., Moreno-García, M. N., and Prieta, F. D. l. (2021). An approach to integrating sentiment analysis into recommender systems. *Sensors*, 21(16):5666.
- Eady, G., Nagler, J., Guess, A., Zilinsky, J., and Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *Sage Open*, 9(1):2158244019832705.
- Ermakov, P. N., Abakumova, I. V., Fedotova, O., and Shchetinina, D. P. (2016). Hostility as a psychological phenomenon and object of scientific research. *International Journal of Environmental and Science Education*, 11(18):10829–10837.
- Fabiansson, E. C. and Denson, T. F. (2016). Anger, hostility and anger management. In *Encyclopedia of Mental Health*, pages 64–67.
- Fahey, J. J. (2023). The big lie: Expressive responding and misperceptions in the United States. *Journal of Experimental Political Science*, 10(2):267–278.
- Flaxman, S., Goel, S., and Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1):298–320.
- Flynn, D. J., Nyhan, B., and Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38(S1):127–150.

- Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication*, 14(2):265–285.
- Garrett, R. K. and Stroud, N. J. (2014). Partisan paths to exposure diversity: Differences in pro- and counterattitudinal news consumption. *Journal of Communication*, 64(4):680–701.
- Gauthier, G., Hodler, R., Widmer, P., and Zhuravskaya, E. (2026). The political effects of x’s feed algorithm. *Nature*, pages 1–8.
- Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Doubleday, Garden City, New York.
- González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A. M., Iyengar, S., Kim, Y. M., Malhotra, N., Moehler, D., Nyhan, B., Pan, J., Rivera, C. V., Settle, J., Thorson, E., and Tucker, J. A. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, 381(6656):392–398.
- Graham, M. H. and Yair, O. (2024). Expressive responding and belief in 2020 election fraud. *Political Behavior*, 46(2):1349–1374.
- Green, J., McCabe, S., Shugars, S., Chwe, H., Horgan, L., Cao, S., and Lazer, D. (2025). Curation bubbles. *American Political Science Review*, pages 1–19.
- Guess, A., Nyhan, B., Lyons, B., and Reifler, J. (2018). Avoiding the echo chamber about echo chambers. *Knight Foundation*, 2(1):1–25.
- Guess, A. M. (2021). (almost) everything in moderation: New evidence on Americans’ online media diets. *American Journal of Political Science*, 65(4):1007–1022.

- Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656):398–404.
- Guo, M. and Sun, F.-S. (2020). Like, comment, or share? Exploring the effects of local television news Facebook posts on news engagement. *Journal of Broadcasting & Electronic Media*, 64(5):736–755.
- Habermas, J. (1989). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press.
- Hart, W., Richardson, K., Tortoriello, G. K., and Earl, A. (2020). ‘You Are What You Read.’ Is selective exposure a way people tell us who they are? *British Journal of Psychology*, 111(3):417–442.
- Hilbert, M., Ahmed, S., Cho, J., Liu, B., and Luu, J. (2018). Communicating with algorithms: A transfer entropy analysis of emotions-based escapes from online echo chambers. *Communication Methods and Measures*, 12(4):260–275.
- Huddy, L., Smirnov, O., Snider, K. L. G., and Perliger, A. (2021). Anger, anxiety, and selective exposure to terrorist violence. *Journal of Conflict Resolution*, 65(10):1764–1790.
- Huszár, F., Ktena, S. I., O’Brien, C., Belli, L., Schlaikjer, A., and Hardt, M. (2022). Algorithmic amplification of politics on twitter. *Proceedings of the national academy of sciences*, 119(1):e2025334119.
- Iyengar, S. and Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1):19–39.
- Jang, S. M. (2014). Challenges to selective exposure: Selective seeking and avoidance in a multitasking media environment. *Mass Communication and Society*, 17(5):665–688.

- Johnson, B. K., Neo, R. L., Heijnen, M. E., Smits, L., and van Veen, C. (2020). Issues, involvement, and influence: Effects of selective exposure and sharing on polarization and participation. *Computers in Human Behavior*, 104:106155.
- Karrer, B. and Newman, M. E. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 83(1):016107.
- Keltner, D. and Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition and Emotion*, 13(5):505–521.
- Kiil, F. (2024). Looking for relief: Developing and testing the emotion-regulation explanation of selective exposure to political information. *Political Psychology*, 45(2):279–297.
- Kim, M. and Lu, Y. (2020). Testing partisan selective exposure in a multidimensional choice context: Evidence from a conjoint experiment. *Mass Communication and Society*, 23(1):107–127.
- Kitchens, B., Johnson, S. L., and Gray, P. (2020). Understanding echo chambers and filter bubbles: The impact of social media on diversification and partisan shifts in news consumption. *MIS quarterly*, 44(4):1619–1649.
- Klapper, J. T. (1960). The effects of mass communication.
- Knobloch-Westerwick, S. (2012a). Selective exposure and reinforcement of attitudes and partisanship before a presidential election. *Journal of Communication*, 62(4):628–642.
- Knobloch-Westerwick, S. (2012b). Selective exposure and reinforcement of attitudes and partisanship before a presidential election. *Journal of Communication*, 62(4):628–642.
- Knobloch-Westerwick, S. and Kleinman, S. B. (2012). Preelection selective exposure: Confirmation bias versus informational utility. *Communication Research*, 39(2):170–193.

- Krupnikov, Y. and Ryan, J. B. (2022). *The other divide*. Cambridge University Press.
- Kulshrestha, J., Zafar, M., Noboa, L., Gummadi, K., and Ghosh, S. (2015). Characterizing information diets of social media users. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 9, pages 218–227.
- Lazarsfeld, P. F., Berelson, B., and Gaudet, H. (1944). *The People’s Choice: How the Voter Makes Up His Mind in a Presidential Campaign*. Columbia University Press.
- Leary, M. R. and Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107(1):34–47.
- Lee, B., Ryu, H., Lee, J. K., Jeong, H., and Kim, B. J. (2024). Network analysis reveals news press landscape and asymmetric user polarization. *arXiv*. arXiv:2408.07900.
- Lelkes, Y., Sood, G., and Iyengar, S. (2017). The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, 61(1):5–20.
- Liang, H. (2018). Broadcast versus viral spreading: The structure of diffusion cascades and selective sharing on social media. *Journal of Communication*, 68(3):525–546.
- Liu, N., Hu, X. E., Savas, Y., Baum, M. A., Berinsky, A. J., Chaney, A. J., Lucas, C., Mariman, R., de Benedictis-Kessner, J., Guess, A. M., et al. (2025). Short-term exposure to filter-bubble recommendation systems has limited polarization effects: Naturalistic experiments on youtube. *Proceedings of the National Academy of Sciences*, 122(8):e2318127122.
- Lodge, M. and Taber, C. S. (2013). *The Rationalizing Voter*. Cambridge University Press, New York.
- Ludwig, K., Grote, A., Iana, A., Alam, M., Paulheim, H., Sack, H., Weinhardt, C., and Müller, P. (2023). Divided by the algorithm? the (limited) effects of content-and sentiment-based news recommendation on affective, ideological, and perceived polarization. *Social Science Computer Review*, 41(6):2188–2210.

- Ludwig, K., Müller, P., Nikolajevic, N., and Grote, A. (2025). Putting ‘filter bubble’ effects to the test: evidence on the polarizing impact of ideology-based news recommendation from two experiments in germany and the us. *Information, Communication & Society*, 28(13):2321–2340.
- Majó-Vázquez, S., Nielsen, R. K., and González-Bailón, S. (2019). The backbone structure of audience networks: A new approach to comparing online news consumption across countries. *Political Communication*, 36(2):227–240.
- Malka, A. and Adelman, M. (2023). Expressive survey responding: A closer look at the evidence and its implications for American democracy. *Perspectives on Politics*, 21(4):1198–1209.
- Marder, B., Slade, E., Houghton, D., and Archer-Brown, C. (2016). ‘I Like Them, But Won’t ‘Like’ Them’: An examination of impression management associated with visible political party affiliation on Facebook. *Computers in Human Behavior*, 61:280–287.
- Masullo, G. M., Lu, S., and Fadnis, D. (2021). Does online incivility cancel out the spiral of silence? A moderated mediation model of willingness to speak out. *New Media & Society*, 23(11):3391–3414.
- Matias, J. N. (2023). Humans and algorithms work together—so study them together. *Nature*, 617(7960):248–251.
- Muddiman, A. and Stroud, N. J. (2017). News values, cognitive biases, and partisan incivility in comment sections. *Journal of Communication*, 67(4):586–609.
- Mukerjee, S., Majó-Vázquez, S., and González-Bailón, S. (2018). Networks of audience overlap in the consumption of digital news. *Journal of Communication*, 68(1):26–50.
- Mummolo, J. (2016). News from the other side: How topic relevance limits the prevalence of partisan selective exposure. *The Journal of Politics*, 78(3):763–773.

- Nam, A., Weible, C. M., and Park, K. (2022). Polarization and frames of advocacy coalitions in south korea’s nuclear energy policy. *Review of Policy Research*, 39(4):387–410.
- Narayanan, A. (2023). Understanding social media recommendation algorithms.
- Nelson, J. L. and Webster, J. G. (2017). The myth of partisan selective exposure: A portrait of the online political news audience. *Social Media + Society*, 3(3):205630511772931.
- Noh, S. and Soroka, S. (2025). Negativity biases online: The interplay of individuals and algorithms in news consumption. *Journal of Media Psychology*, 37(5):296–306.
- Osman, N. A. and Noah, S. A. M. (2018). Sentiment-based model for recommender systems. In *2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, pages 1–6. IEEE.
- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. penguin UK.
- Pedreschi, D., Pappalardo, L., Ferragina, E., Baeza-Yates, R., Barabási, A.-L., Dignum, F., Dignum, V., Eliassi-Rad, T., Giannotti, F., Kertész, J., et al. (2025). Human-ai coevolution. *Artificial Intelligence*, 339:104244.
- Peixoto, T. P. (2017). The graph-tool python library. *figshare*.
- Peixoto, T. P. (2019). Bayesian stochastic blockmodeling. In Doreian, P., Batagelj, V., and Ferligoj, A., editors, *Advances in Network Clustering and Blockmodeling*, pages 289–332. Wiley, 1st edition.
- Peterson, E., Goel, S., and Iyengar, S. (2021). Partisan selective exposure in online news consumption: Evidence from the 2016 presidential campaign. *Political Science Research and Methods*, 9(2):242–258.
- Peterson, E. and Iyengar, S. (2021). Partisan gaps in political information and information-seeking behavior: Motivated reasoning or cheerleading? *American Journal of Political Science*, 65(1):133–147.

- Peterson, E. and Kagalwala, A. (2021). When unfamiliarity breeds contempt: How partisan selective exposure sustains oppositional media hostility. *American Political Science Review*, 115(2):585–598.
- Powell, L. H. and Williams, K. (2007). Hostility. In *Encyclopedia of Stress*, pages 354–358. Elsevier.
- Prior, M. (2007). *Post-broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge University Press, Cambridge.
- Prior, M. (2009). Improving media effects research through better measurement of news exposure. *The Journal of Politics*, 71(3):893–908.
- Prior, M. (2013). The challenge of measuring media exposure: Reply to dilliplane, goldman, and mutz. *Political Communication*, 30(4):620–634.
- Prior, M., Sood, G., and Khanna, K. (2015). You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*, 10(4):489–518.
- Rathje, S., Van Bavel, J. J., and van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26):e2024292118.
- Rest of World (2023). Smartnews was ready to take on the U.S. market. where did it go wrong? <https://restofworld.org/2023/smartnews-japan-unicorn-layoffs-media/>.
- Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., and Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, pages 1–11.

- Rosenberg, J. and Egbert, N. (2011). Online impression management: Personality traits and concerns for secondary goals as predictors of self-presentation tactics on Facebook. *Journal of Computer-Mediated Communication*, 17(1):1–18.
- Rowe, I. (2015). Civility 2.0: A comparative analysis of incivility in online political discussion. *Information, communication & society*, 18(2):121–138.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161.
- Santana, A. D. (2014). Virtuous or vitriolic: The effect of anonymity on civility in online newspaper reader comment boards. *Journalism Practice*, 8(1):18–33.
- Schaffner, B. F. and Luks, S. (2018). Misinformation or expressive responding? What an inauguration crowd can tell us about the source of political misinformation in surveys. *Public Opinion Quarterly*, 82(1):135–147.
- Shin, J. and Thorson, K. (2017). Partisan selective sharing: The biased diffusion of fact-checking messages on social media. *Journal of Communication*, 67(2):233–255.
- Similarweb (2026). Smartnews.com traffic analytics, ranking & audience. <https://www.similarweb.com/website/smartnews.com/>. Retrieved April 2026.
- SmartNews (2022). SmartNews launches U.S. advertising business, expanding footprint in U.S. tech market. Press release. PR Newswire. <https://www.prnewswire.com/news-releases/smartnews-launches-us-advertising-business-expanding-footprint-in-u-s-tech-market-301650928.html>.
- Soroka, S., Fournier, P., and Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, 116(38):18888–18892.

- Soroka, S. and Krupnikov, Y. (2021). *The increasing viability of good news*. Cambridge University Press.
- Soroka, S. N. (2014). *Negativity in democratic politics: Causes and consequences*. Cambridge University Press.
- Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, 30(3):341–366.
- Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3):556–576.
- Stroud, N. J. (2011). *Niche News: The Politics of News Choice*. Oxford University Press.
- Stroud, N. J., Van Duyn, E., and Peacock, C. (2016). News commenters and news comment readers. *Engaging News Project*, 21:1–21.
- Suhay, E. (2015). Explaining group influence: The role of identity and emotion in political conformity and polarization. *Political Behavior*, 37:221–251.
- Sunstein, C. (2001). Republic. com princeton, nj: Princeton univ.
- Sunstein, C. R. (2018). Republic: Divided democracy in the age of social media.
- Tewksbury, D. (2003). What do Americans really want to know? tracking the behavior of news readers on the Internet. *Journal of Communication*, 53(4):694–710.
- Tokita, C. K., Guess, A. M., and Tarnita, C. E. (2021). Polarized information ecosystems can reorganize social networks via information cascades. *Proceedings of the National Academy of Sciences*, 118(50):e2102147118.
- Tyler, M., Grimmer, J., and Iyengar, S. (2022). Partisan enclaves and information bazaars: Mapping selective exposure to online news. *The Journal of Politics*, 84(2):1057–1073.

- Valentino, N. A., Banks, A. J., Hutchings, V. L., and Davis, A. K. (2009). Selective exposure in the internet age: The interaction between anxiety and information utility. *Political Psychology*, 30(4):591–613.
- Valentino, N. A., Brader, T., Groenendyk, E. W., Gregorowicz, K., and Hutchings, V. L. (2011). Election night’s alright for fighting: The role of emotions in political participation. *The Journal of Politics*, 73(1):156–170.
- Wagner, C., Strohmaier, M., Olteanu, A., Kiciman, E., Contractor, N., and Eliassi-Rad, T. (2021). Measuring algorithmically infused societies. *Nature*, 595(7866):197–204.
- Wang, S., Huang, S., Zhou, A., and Metaxa, D. (2024). Lower quantity, higher quality: auditing news content and user perceptions on twitter/x algorithmic versus chronological timelines. *Proceedings of the ACM on human-computer interaction*, 8(CSCW2):1–25.
- Webster, J. G. (2014). *The marketplace of attention: How audiences take shape in a digital age*. MIT Press.
- Wulczyn, E., Thain, N., and Dixon, L. (2017). Ex Machina: Personal attacks seen at scale. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1391–1399.
- Yair, O. and Huber, G. A. (2020). How robust is evidence of partisan perceptual bias in survey responses? A new approach for studying expressive responding. *Public Opinion Quarterly*, 84(2):469–492.
- Yin, W., Hay, J., and Roth, D. (2019). Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. *arXiv preprint arXiv:1909.00161*.
- Young, L. and Soroka, S. (2012). Affective news: The automated coding of sentiment in political texts. *Political communication*, 29(2):205–231.
- Young, M. J., Tiedens, L. Z., Jung, H., and Tsai, M.-H. (2011). Mad enough to see the

other side: Anger and the search for disconfirming information. *Cognition and Emotion*, 25(1):10–21.

Yu, X., Wojcieszak, M., and Casas, A. (2024). Partisanship on social media: In-party love among American politicians, greater engagement with out-party hate among ordinary users. *Political Behavior*, 46(2):799–824.

Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge University Press, Cambridge.