CURIOSITYBITS LIVING LAB BOSTON · STUDY #07 RESEARCH ARTICLE (V1.0 — VIRALITY, CROWD GATEKEEPING & AGONISTIC TRENCH DUELS)

Clicks Without Conversation? Negative Valence, Moral-Conflict Framing, and the Structure of Public Deliberation in Digital Journalism

CuriosityBits Living Lab — Communication and Media Science Series (Study 7, Edition v1.0)

Abstract

Digital communication scholarship widely concludes that negative emotional tone and moralized conflict amplify online engagement. Yet treating clicks, shares, upvotes, and raw comment totals as a single behavioral quantity obscures a foundational question in democratic theory: does affective news framing foster broad, evidence-grounded citizen dialogue, or does it merely trigger shallow reactions and polarized exchanges? Using a pre-registered two-stage hurdle design across M = 85 non-financial digital journalism outlets from January 2021 through December 2025, we separate initial crowd selection (N = 3653 outcome-blind submissions drawn from N = 220696 journalism posts) from 48-hour recursive conversation geometry (N = 2114 nested comment trees containing $226317 valid comments). Four open-weight language model families blindly scored 5398 unique headlines (21592 live evaluations) on Headline Negativity and Moral-Conflict and Blame Framing, achieving two-way absolute-agreement intraclass correlations of 0.9665 and 0.9650 in Stage 1 and 0.9638 and 0.9623 in Stage 2. Absorbing outlet, quarter, submission slot, and beat fixed effects alongside cubic splines of comment volume, we establish four findings. First, although negativity and moral conflict ignite initial first-comment reactions (+3.06 and +3.00 percentage points, p = 0.0129 and p = 0.0211) and crowd upvotes, neither predicts crossing the 15-comment front-page discussion hurdle (t = +1.52 and t = -0.51, Holm p ≥ 0.3946). Second, within active discussions, non-conflictual negative valence is dead zero across all five deliberative metrics (Holm p = 1.0000), whereas moral-conflict framing shifts all five past multiplicity control (p ≤ 0.0177). Third, exact additive tree decomposition proves that moral-conflict headlines induce *agonistic dyadic closure*: 80.44% (+1.09 percentage points, t = +3.13, p = 0.0024) of conflict's +1.35 percentage-point reciprocity gain (t = +2.82, Holm p = 0.0240) reflects repeat two-person ping-pong within locked-in pairs, while unique first-time dyad breadth stalls (+0.26 percentage points, t = +1.05), speakers per comment contract (-1.74 percentage points, t = -3.34), Shannon entropy drops (-0.0029, t = -2.42), and external documentary links fall (-0.73 percentage points, t = -2.95, Holm p = 0.0125). Fourth, readers meet news paywalls with a 5.38-fold surge in archive mirrors (1.089% versus 0.202%, t = +5.73), neutralizing paywall moderation (Holm p = 1.0000). Chronological 2024–2025 forecasts confirm that multi-judge semantic scores outperform structural controls and bag-of-words dictionaries (p ≤ 0.0342$).

1. Introduction

Few empirical regularities in contemporary communication science appear more settled than the dominance of negativity and moral conflict in digital attention markets. Psychophysiological and experimental studies show that citizens monitor threatening or adverse news more attentively than positive reporting (Soroka, 2014; Soroka & McAdams, 2015; Soroka, Fournier, & Nir, 2019; Trussler & Soroka, 2014), while large-scale field experiments demonstrate that negative headline words systematically elevate click-through rates (Robertson et al., 2023). Simultaneously, social-diffusion studies rooted in framing theory (Entman, 1993; Scheufele, 1999; Semetko & Valkenburg, 2000) show that moral-emotional expressions and out-group animosity substantially boost sharing and reaction totals on digital platforms (Brady et al., 2017; Rathje, Van Bavel, & van der Linden, 2021; Iyengar et al., 2018). Together, these literatures have fostered a *single-metric engagement paradigm* in which negative valence and conflictual framing are treated as unified accelerators of online civic participation.

From the perspective of public-sphere theory, however, collapsing clicks, upvotes, shares, and raw comment volume into an undifferentiated "engagement" index conflates low-cost attentional capture with democratic deliberation. Journalism scholars have long documented persistent "choice gaps" and "social news gaps" separating what editors prioritize, what readers click, and what audiences amplify socially (Boczkowski & Peer, 2011; Bright, 2016; Van Aelst et al., 2017). Recent feed-ranking experiments similarly reveal a wedge between behavioral interaction volume and democratic quality: divisive posts inflate clicks and comments while lowering user satisfaction and heightening partisan hostility (Milli et al., 2025; Piccardi et al., 2025). Normative accounts of the public sphere insist that the democratic value of citizen talk depends not on crowd noise, but on conversational structure and epistemic grounding—whether participants engage in reciprocal multi-turn dialogue, maintain broad speaker inclusion rather than two-person shouting matches, and substantiate contested claims with verifiable external sources (Habermas, 2006; Graham & Witschge, 2003; Stromer-Galley, 2007; Frieß & Eilders, 2015).

Once shallow crowd amplification is distinguished from downstream conversation architecture, core communication theories generate competing predictions. If negative and conflictual headlines primarily elicit expressive signaling or aversive withdrawal (Noëlle-Neumann, 1974; Mutz, 2002), they should inflate shallow reactions while crowding out reciprocal, evidence-based dialogue. Conversely, theories of affective intelligence (Marcus, MacKuen, & Neuman, 2011) and agonistic pluralism (Papacharissi, 2004; Karppinen, Moe, & Svensson, 2008; Muddiman & Stroud, 2017; Rossini, 2020) contend that conflict disrupts passive consensus and mobilizes deeper counter-argumentation. A third perspective—forum-design theory (Wright & Street, 2007; Halpern & Gibbs, 2012; Jaidka et al., 2021)—predicts that moderated, non-personalized forum architecture decouples downstream conversation topology from headline framing altogether ($\beta_2 = 0$).

Testing these rival mechanisms requires overcoming four pervasive empirical pitfalls: (1) viral-survival collider bias when sampling only active discussion threads (Heckman, 1979); (2) construct conflation between non-conflictual tragedy or threat (Soroka, Fournier, & Nir, 2019) and adversarial blame attribution (Entman, 1993); (3) mechanical scaling of tree depth and turn counts with comment volume, compounded by author self-threads

and paywall-mirror hyperlinks; and (4) institutional access barriers from digital news paywalls (Myllylahti, 2013; Sjøvaag, 2015; Benson, 2006).

We address all four challenges via a pre-registered two-stage investigation strictly within communication and media science—excluding all financial press, corporate earnings stories, and market variables—across M = 85 journalism outlets over 2021–2025. Pairing an outcome-blind Stage-1 submission panel (N = 3653 from N = 220696 clean journalism submissions) with full 48-hour recursive discussion trees in Stage 2 (N = 2114 threads containing $226317 comments), and scoring headlines blindly across four independent open-weight language model families (21592 live ballots), we uncover a sharp structural dissociation. Headline negativity and conflict spark initial zero-to-one reactions and upvotes but stall at the 15-comment community gatekeeping hurdle. Inside active discussions, pure negative valence is completely inert across all five deliberative metrics, whereas moral-conflict framing drives *agonistic dyadic closure*—elevating multi-turn reciprocity almost entirely (80.44%) through repeat two-person ping-pong among fewer speakers while eroding external documentary evidence citing. Meanwhile, a 5.38$-fold surge in communal archive mirrors neutralizes paywall information starvation, demonstrating grassroots *networked epistemic repair* (Barzilai-Nahon, 2008).

---

2. Theoretical Framework and Competing Hypotheses

2.1 Psychophysiological Negativity Versus Moral-Conflict Framing

Two distinct mechanisms explain why adverse and contentious news captures audience attention, yet empirical studies routinely merge them into a single sentiment dimension. The first is *psychophysiological negativity bias*. Cross-national physiological experiments show that human autonomic arousal responds more strongly to negative news—disease outbreaks, accidents, environmental or civic deterioration—than to positive news, regardless of self-reported preferences (Soroka, 2014; Soroka & McAdams, 2015; Soroka, Fournier, & Nir, 2019; Trussler & Soroka, 2014). Analyzing randomized headline experiments, Robertson et al. (2023) confirmed that negative words significantly raise click-through rates. Crucially, threat surveillance requires no human culprit: a natural disaster or medical hazard carries high negative valence without interpersonal blame.

The second mechanism is *moral-conflict framing and blame attribution*. Entman (1993) conceptualized framing around problem definition, causal attribution, moral evaluation, and treatment recommendation (Scheufele, 1999), and newsroom analyses identify conflict and responsibility attribution as pervasive journalistic templates (Semetko & Valkenburg, 2000). In networked environments, moralized conflict engages group identity and moral indignation: moral-emotional words and out-group animosity capture visual attention and drive social sharing (Brady et al., 2017; Rathje, Van Bavel, & van der Linden, 2021; Iyengar et al., 2018). Yet because clicking, sharing, and deliberating obey distinct logics (Boczkowski & Peer, 2011; Bright, 2016; Neuman et al., 2014), whether valence negativity and moral conflict enrich or degrade downstream conversation remains an open theoretical clash.

2.2 Public-Sphere Crowding-Out, Agonistic Mobilization, and Forum Affordances

Three theoretical traditions yield rival predictions for Stage-2 conversation geometry.

Habermasian Crowding-Out (H_1). Normative deliberative theory grounds democratic legitimacy in reciprocal listening, inclusive speaker participation, and reasoned documentary justification (Habermas, 2006). Empirical deliberation schemes operationalize these norms via multi-turn interpersonal reciprocity (A \to B \to A), speaker diversity, and external evidence citing (Graham & Witschge, 2003; Stromer-Galley, 2007; Frieß & Eilders, 2015). Combined with spiral-of-silence and cross-cutting exposure research showing that adversarial climates chill broader participation (Noëlle-Neumann, 1974; Mutz, 2002), affective and conflictual headlines should boost Stage-1 volume ($\beta_1 > 0$) while crowding out Stage-2 reciprocity, nesting, speaker breadth, and documentary citing into shallow depth-1 venting ($\beta_2 < 0$; Milli et al., 2025).

Agonistic Deliberative Mobilization (H_2). Critics of consensus deliberation argue that impassioned contestation and sharp disagreement are catalysts rather than enemies of democratic talk (Papacharissi, 2004; Karppinen, Moe, & Svensson, 2008). Empirical work separating substantive partisan or issue conflict from personal intolerance shows that conflictual cues stimulate active rebuttals (Muddiman & Stroud, 2017; Rossini, 2020), while affective intelligence and polarized-crowd theories hold that threat and conflict motivate scrutiny and counter-evidence search (Marcus, MacKuen, & Neuman, 2011; Shi et al., 2019). Under H_2, framing elevates both Stage-1 entry ($\beta_1 > 0$) and Stage-2 reciprocal argumentation and evidence citing ($\beta_2 > 0$).

Valence-versus-Conflict Bifurcation (H_3) and Platform Decoupling (H_0). Because uncontested tragedy elicits consensual sympathy whereas blame attribution posits competing sides (Entman, 1993; Semetko & Valkenburg, 2000), H_3 predicts systematic divergence between Moral-Conflict Framing and pure Headline Negativity ($\beta_{2,\text{conf}} \neq \beta_{2,\text{neg}}$). Conversely, institutional design theory contends that threaded architecture, community moderation, and journalistic beat norms govern how citizens converse once inside a forum, rendering Stage-2 geometry invariant to headline framing ($\beta_2 \approx 0$; Wright & Street, 2007; Halpern & Gibbs, 2012; Jaidka et al., 2021).

2.3 Journalistic Fields, Paywalls, and Strong-Inference Matrix

Journalistic framing operates across differentiated fields—from national public-affairs wires to specialist science, technology, and culture beats (Benson, 2006)—and across open-access versus subscription business models (Myllylahti, 2013; Sjøvaag, 2015). Under our paywall information-starvation hypothesis (H_4), restricting article text behind paywalls forces non-subscribers to react to headline cues alone, steepening framing-induced deliberative erosion ($\gamma_{\text{pw}} < 0$) unless offset by networked counter-gatekeeping (Barzilai-Nahon, 2008; Carlson, 2015).

We therefore test four research questions—Stage-1 crowd breakthrough (RQ_1), Stage-2 deliberative geometry (RQ_2), valence-versus-conflict bifurcation (RQ_3), and paywall moderation (RQ_4)—summarized in Table 1.

Hypothesis	Canonical Lineage	Stage-1 Breakthrough (β_1)	Stage-2 Size-Adjusted Deliberation & Evidence (β_2)	Predicted Equilibrium
H_1: Crowding-Out	Robertson et al. (2023); Brady et al. (2017); Habermas (2006); Milli et al. (2025)	$\beta_1 > 0$	$\beta_2 < 0$ (reciprocity, nesting, & evidence fall)	*Clicks Without Conversation*: Shallow volume crowds out reciprocal, evidence-based talk.
H_2: Agonistic Mobilization	Marcus et al. (2011); Papacharissi (2004); Muddiman & Stroud (2017); Rossini (2020)	$\beta_1 > 0$	$\beta_2 > 0$ (reciprocity, depth, & evidence rise)	*Productive Contestation*: Conflict mobilizes multi-turn rebuttals and counter-evidence.
H_3: Valence-Conflict Split	Soroka et al. (2019) vs. Entman (1993); Semetko & Valkenburg (2000)	$\beta_{1,conf} \neq \beta_{1,neg}$	$\beta_{2,conf} \neq \beta_{2,neg}$	*Valence Is Not Conflict*: Blame/ contestability diverges sharply from pure tragedy.
H_0: Affordance Decoupling	Wright & Street (2007); Halpern & Gibbs (2012); Frieß & Eilders (2015); Jaidka et al. (2021)	$\beta_1 \geq 0$	$\beta_2 \approx 0$ (Holm $p \geq 0.05$)	*Architecture Over Headlines*: Forum design insulates talk from headline framing.
H_4: Paywall Starvation	Myllylahti (2013); Sjøvaag (2015); Benson (2006); Barzilai-Nahon (2008)	Entry friction	$\gamma_{pw} < 0$ on paywalled domains	*Headline-Only Reacting*: Paywalls worsen distortion unless repaired by communal mirrors.

3. Pre-Registered Two-Stage Hurdle Data and Research Design

3.1 Curated Journalism Universe and Pre-Data Selection Protocol

Before locking our confirmatory design in a standalone cryptographic commit prior to full-panel ingestion, we audited five pure-communication designs using live archival probes. Four alternatives were rejected pre-data due to construct censoring or selection bias: within-thread incivility testing suffered $80.00% floor-censoring ($\leq 0.05$) under anti-insult moderation; primary-versus-legacy news dyads yielded only a 17.24% match rate under platform duplicate suppression; encyclopedia talk-page gatekeeping exhibited 54.17% acute zero-

censoring; and preprint abstract hedging-versus-hype faced viral-survival collider bias because over 90%$ of raw preprints draw zero to two comments.

Locking the digital journalism design on Hacker News enabled a clean *two-stage hurdle architecture* across complete, timestamped conversation trees without personalized feed confounding. Enforcing $100% communication-science purity, we excluded all financial and business outlets and applied a deterministic headline filter removing corporate earnings, share-price, initial public offering, securities-filing, and cryptocurrency market stories. Our curated registry comprises M = 85 qualifying non-financial journalism outlets (58 open-access, 27$ paywalled) across five beats: (1) national, international, and public-service wires/broadcasters (*Reuters*, *Associated Press*, *BBC*, *NPR*, *The Guardian*, *The New York Times*, *The Washington Post*, *Politico*); (2) investigative, legal, and policy reporting (*ProPublica*, *The Markup*, *SCOTUSblog*, *Lawfare*, *Bellingcat*); (3) science, environment, and health journalism (*Nature*, *Science*, *Quanta Magazine*, *STAT*, *Scientific American*); (4) technology-in-society journalism (*Ars Technica*, *Wired*, *The Verge*, *MIT Technology Review*, *404 Media*); and (5) culture and ideas magazines (*The Atlantic*, *The New Yorker*, *Harper's*, *Aeon*, *Vox*, *Slate*).

Adaptive time-sliced queries across January 1, 2021 through December 31, 2025 (< 900 hits per slice, > 10 months matured so right-censoring equals $0%) scanned N = 221497 raw submissions. Dropping 294 host mismatches, 14 platform meta-posts, and 493 market headlines yielded N = 220696 clean journalism submissions across all M = 85 outlets (N = 23453 reaching ≥ 15$ comments), from which we drew two panels:

1. **Stage-1 Unconditional Panel (N = 3653, G = 85 outlets, G = 262 weeks):** A seeded stratified random draw across outlet-by-year cells over all submissions (≥ 0 comments), with inverse-probability weights summing exactly to $220696.00 (398 stories reaching ≥ 15$ comments automatically entered Stage 2).
2. **Stage-2 Deliberative Tree Panel (N = 2114, G = 85 outlets, G = 261 weeks):** Combining those $398 stories with stratified discussed draws yielded 2251 candidate trees (100% retrieval; weights summing to 23453). Retaining threads with ≥ 10 valid comments inside 48 hours produced N = 2114 recursive trees containing 226317 valid 48-hour comments (235510$ over 14 days).

3.2 Blind Four-Family Semantic Annotation and Shrout–Fleiss Reliability

All $5398 unique headlines were scored blindly at temperature 0.0 (seeing *only* headline text, stripped of outlet, date, points, and comments) by four independent open-weight model families—Llama 3.3 70B, Qwen 2.5 72B, DeepSeek V3, and Gemma 3 27B—producing 21592 live JSON evaluations (0 fallbacks) on [0.00, 1.00] for **Headline Negativity** (N_i: harm, crisis, loss, threat, or adverse outcomes without requiring blame) and **Moral-Conflict and Blame Framing** (C_i$: interpersonal, institutional, or political conflict, accusation, or moral wrongdoing).

Table 2 reports textbook two-way Shrout and Fleiss (1979) reliability. Four-judge absolute agreement ICC(2,4) reaches $0.9665 and 0.9650 in Stage 1 and 0.9638 and 0.9623 in Stage 2 (single-rater ICC(2,1) ≥ 0.8645, pairwise r ∈ [0.8684, 0.8886]). Negativity and Conflict correlate at r = +0.7533 in Stage 1 and r = +0.7176 in

Stage 2, yielding modest variance inflation factors (2.3527 and 2.4105 in Stage 1; 2.1736 and 2.2687 in Stage 2). Pre-registered 0.50-midpoint typology counts in Stage 2 are 1198 Neutral/Constructive (C_0), 315 Pure Threat/Tragedy (C_1), 412 Moral-Conflict/Blame (C_2), and 189 Non-Negative Debate (C_3; Stage-1 counts: 2342, 441, 568, and 302$).

Panel & Construct	N	ICC(1,4)	ICC(2,4)	ICC(3,4)	ICC(2,1)	Pairwise r	Mean (SD)	Exact Zero
Stage 1: Headline Negativity	$3653	0.9664	0.9665	0.9682	0.8782	0.8886	0.2944 (0.2734)	20.67% $
Stage 1: Moral-Conflict & Blame	$3653	0.9649	0.9650	0.9655	0.8732	0.8768	0.2465 (0.2741)	31.37% $
Stage 2: Headline Negativity	$2114	0.9637	0.9638	0.9659	0.8694	0.8806	0.3460 (0.2750)	15.14% $
Stage 2: Moral-Conflict & Blame	$2114	0.9623	0.9623	0.9629	0.8645	0.8684	0.2903 (0.2817)	23.37% $

3.3 Conversation-Tree Outcomes and Econometric Specifications

Stage-1 outcomes are Discussed Breakthrough (≥ 15 comments, base rate $10.90%), Log Upvote Points (mean +1.8228, SD 1.1932), and Log Comment Volume (mean +0.8259, SD 1.3806). In Stage 2, we retain valid comments posted within 48 hours (172800 seconds; 6016 later comments pruned) and exclude 562 author self-replies (0.25%$) from interpersonal metrics across five outcomes: 1. **Dyadic Reciprocity Rate (48h, Co-Primary):** Share of interpersonal replies (d ≥ 2, distinct parent–child authors) where child author equals grandparent author (d ≥ 3) across two distinct citizens (A \to B \to A; mean +0.2117, SD 0.1053). 2. **Interpersonal Nested Reply Share (48h):** Share of valid 48-hour comments that are interpersonal nested replies rather than depth-1 root statements or self-appends (mean +0.7982, SD 0.1111). 3. **90th-Percentile Branch Depth (48h):** Upper-decile tree depth (mean +5.2674, SD 1.6128). 4. **External Evidence Link Rate (48h):** Share of HTML-unescaped 48-hour comments citing external URLs after stripping archive/ paywall mirrors, platform self-links, and article self-domain links (mean +0.0991, SD 0.0650), logging **Paywall Mirror Link Rate** separately. 5. **Speaker Shannon Entropy (48h):** Normalized turn-share entropy [0, 1] across unique commenters (mean +0.9548, SD 0.0269).

All models absorb outlet domain ($85), year-quarter, UTC day-by-six-hour slot, and URL beat fixed effects plus structural headline features and deterministic lexical negativity (2487 words) and conflict (199 words) dictionary rates. Stage-2 models include four-knot restricted cubic splines of \log(V_{48}) (knots 2.772589,

3.663562, 4.543295, 5.834811). Inference uses two-way `(domain, week)` clustering (Cameron, Gelbach, & Miller, 2011) and Holm (1979) multiplicity adjustments (K = 4 co-primary slopes and K = 5$ Stage-2 outcomes).

---

4. Empirical Results

4.1 Stage-1 Crowd Gatekeeping: Upvotes Without Front-Page Breakthrough

Table 3 reports Stage-1 estimates (N = 3653). On Discussed Breakthrough (≥ 15 comments), neither Headline Negativity (+0.0158, SE = 0.0104, t = +1.52, p = 0.1315, K = 4 Holm p = 0.3946) nor Moral-Conflict Framing (-0.0059, SE = 0.0115, t = -0.51, p = 0.6079, Holm p = 0.6199) is statistically significant. Standalone models (t = +1.65 and +0.70), inverse-probability universe-weighted models (t = +0.21 and +0.66), and logit estimates (z = +1.52 and +0.10) confirm this front-page breakthrough null (omitting lexical controls yields only a faint surface-lexical negativity slope of +0.0203, t = +2.01, p = 0.0473).

Stage-1 Outcome (N = 3653)	Joint Neg (z)	Joint Conflict (z)	Standalone Neg	Standalone Conf	Pure Threat (C_1)	Moral Blame (C_2)	Debate (C_3)
Breakthrough (≥ 15)	+0.0158 (+1.52)	-0.0059 (-0.51)	+0.0120 (+1.65)	+0.0057 (+0.70)	+0.0231 (+1.24)	+0.0016 (+0.09)	+0.0257 (+1.03)
Log Upvote Points	+0.0579 (+1.66)	+0.0872^* (+2.29)	+0.1129** (+4.67)	+0.1299** (+4.88)	+0.1277^* (+2.29)	+0.2387** (+3.89)	+0.2675** (+2.69)
Log Comment Volume	+0.0693 (+1.66)	+0.0459 (+1.14)	+0.0983** (+3.38)	+0.0971** (+3.46)	+0.1535^* (+2.06)	+0.1237^* (+2.00)	+0.2516^* (+2.25)

Notes: Four-way fixed effects plus structural/lexical controls; two-way clustered t-statistics in parentheses (G = 85, G = 262). ^*p < 0.05*, *p < 0.01 (unadjusted; K = 4 Holm p = 0.3946 and 0.6199 on Breakthrough).

By contrast, continuous crowd volume responds strongly at lower margins: standalone Negativity and Conflict elevate Log Upvote Points (+0.1129, t = +4.67, p = 0.000011; +0.1299, t = +4.88, p = 0.000005) and Log Comments (+0.0983, t = +3.38, p = 0.0011; +0.0971, t = +3.46, p = 0.0008). Jointly, Moral-Conflict Framing significantly lifts Log Upvote Points (+0.0872, SE = 0.0381, t = +2.29, p = 0.0246) and Poisson comment counts (+0.2869, z = +2.16, p = 0.0311), with Moral Blame (+0.2387, p = 0.0002) and Non-Negative Debate (+0.2675, p = 0.0086) leading neutral headlines.

4.2 Stage-2 Deliberative Architecture: Valence Null Versus Five-Dimension Conflict Shift

Table 4 presents Model 2 and Model 3 results across N = 2114 recursive 48-hour conversation trees ($226317 comments), controlling for four-knot cubic splines of \log(V_{48})$.

First, **Headline Negativity conditional on Moral Conflict is precisely null across all five deliberative outcomes (K = 5 Holm p = 1.0000 on five of five)**: Dyadic Reciprocity (+0.0035, SE = 0.0034, t = +1.02, p = 0.3100, K = 4 Holm p = 0.6199), Interpersonal Nested Reply Share (-0.0018, t = -0.60), 90th-Percentile Branch Depth (+0.0117, t = +0.27), External Evidence Link Rate (+0.0017, t = +0.74), and Speaker Shannon Entropy (-0.0004, t = -0.39), plus Mean Depth (-0.0012, t = -0.06) and Log Comment Words (-0.0110, t = -0.95). Standalone negativity correlations with reciprocity (+0.0118, t = +4.51) vanish completely once moral blame is separated.

Stage-2 48h Outcome (N = 2114)	Joint Neg (z)	Joint Conflict (z)	K = 5 Holm p	Conflict \beta$_{std}$	Wald (Conf - Neg)	Moral Blame (C_2)	Debate (C_3)
Dyadic A \to B \to A Reciprocity	+0.0035 (+1.02)	+0.0135** (+2.82)	$0.0125	+0.1286 SD	+0.0101 (0.1829)	+0.0246^{**} (+2.73)	+0.0212^* (+2.20$)
Interpersonal Nested Share	-0.0018 (-0.60)	+0.0127** (+3.76)	$0.0013	+0.1145 SD	+0.0146^* (0.0158)	+0.0204^{**} (+3.16)	+0.0179^{**} (+2.70$)
90th-Pct Branch Depth	+0.0117 (+0.27)	+0.1812** (+3.80)	$0.0013	+0.1123 SD	+0.1695^* (0.0395)	+0.3487^{**} (+3.36)	+0.2694^{**} (+2.99$)
External Evidence Link Rate	+0.0017 (+0.74)	-0.0073** (-2.95)	$0.0125	-0.1121 SD	-0.0090^* (0.0500)	-0.0031 (-0.91)	-0.0054 (-1.18$)
Speaker Shannon Entropy	-0.0004 (-0.39)	-0.0029^* (-2.42)	$0.0177	-0.1063 SD	-0.0025 (0.2048)	-0.0040 (-1.84)	-0.0028 (-0.97$)

Notes: Four-way fixed effects, four-knot \log(V_{48}) splines, and structural/lexical controls; two-way `(domain, week)` t-statistics (G = 85, G = 261) or Wald p-values in parentheses. Primary K = 4 Holm p for Reciprocity Conflict equals $0.0240; all five Negativity Holm p-values equal 1.0000$.

Second, **Moral-Conflict and Blame Framing shifts all five deliberative dimensions past K = 5 Holm multiplicity control (p ≤ 0.0177), splitting structural engagement against epistemic quality**. A one-SD increase in Moral Conflict *increases* Dyadic Reciprocity Rate by +0.0135 (+4.81 percentage points across [0, 1], SE = 0.0048, t = +2.82, p = 0.0060, K = 4 Holm p = 0.0240, K = 5 Holm p = 0.0125; total-comment-denominator reciprocity +0.0137, t = +3.28), *increases* Interpersonal Nested Reply Share by +0.0127 (SE = 0.0034, t = +3.76, Holm p = 0.0013), and *deepens* 90th-Percentile Branch Depth by +0.1812 levels (SE = 0.0476, t = +3.80, Holm p = 0.0013; Mean Depth +0.1010, t = +4.27, p = 0.000050). Yet Moral Conflict simultaneously *reduces* External Evidence Link Rate by -0.0073 per SD (-2.59 percentage points across [0, 1] off the +0.0991 sample mean; SE = 0.0025, t = -2.95, p = 0.0042, Holm p = 0.0125) and *reduces* Speaker

Shannon Entropy by -0.0029 (SE = 0.0012, t = -2.42, p = 0.0177, Holm p = 0.0177), while comment length stays flat (+0.0166, t = +1.15, p = 0.2528). Continuous Wald tests (Conflict - Negativity) confirm significant divergence on Nested Share (p = 0.0158), 90th-Percentile Depth (p = 0.0395), Mean Depth (p = 0.0121), and Evidence Links (p = 0.0500).

4.3 Paywall Moderation, Robustness Gates, and Out-of-Sample Forecasting

Testing H_4 (Model 4), paywalled outlets (M = 27) trigger a $5.38-fold jump in Paywall Mirror Link Rate over open-access outlets (M = 58): 1.089% versus 0.202% (regression slope +0.0092, SE = 0.0016, t = +5.73, p = 0.0000, \beta_{std} = +0.3106 SD; stripping archive mirrors flips the raw paywall link gap from +0.52 to -0.46 percentage points). Yet second-stage `Conflict × Paywalled` moderation is null across all outcomes (+0.0052, p = 0.5540 on reciprocity; -0.0046, p = 0.3565 on nested replies; -0.0012, p = 0.8034 on evidence; K = 3 Holm p = 1.0000; `Negativity × Paywalled` on nested replies +0.0107, Holm p = 0.0727), with positive conflict reciprocity on both open-access (+0.0117, p = 0.0252) and paywalled domains (+0.0169, p = 0.0383$).

Across all ten falsification gates, core conclusions hold: - **Per-Judge and Leave-One-Judge-Out Checks (Gate 1 & Check E):** Individual models confirm Stage-2 conflict reciprocity for Llama (+0.0117, t = +2.92), DeepSeek (+0.0144, t = +2.66), and Gemma (+0.0164, t = +3.53), while Qwen loads slightly more on negativity (t = +2.50 vs. +1.32); all four three-judge leave-one-judge-out ensembles remain significant for conflict (+0.0114 to +0.0157, p ≤ 0.0162) and null for negativity (p ≥ 0.1787). - **Unedited Newsroom URL Slugs (Gate 2):** Within lexical-slug outlets (excluding opaque path identifiers on *BBC* and *Nature*), stories with ≥ 0.60 slug-to-title overlap ($2167/3172 = 68.32% Stage 1; 1182/1797 = 65.78% Stage 2) yield +0.0140 on conflict reciprocity (t = +2.40, p = 0.0189) and +0.0001 on negativity (t = +0.03$), ruling out post-submission title edits. - **Selection, Splines, Windows, and Subsamples (Gates 3–9):** Heckman (1979) Inverse Mills Ratio adjustment (λ = +0.0228, t = +0.68; conflict +0.0137, t = +2.81), breakthrough inverse-probability weights (+0.0179, t = +3.47, p = 0.0008; combined weights +0.0202, t = +3.33), 20-vigintile volume fixed effects (+0.0135, t = +2.75), unadjusted volume (+0.0130, t = +2.70), 24-hour (+0.0129, t = +2.82) and 14-day (+0.0135, t = +2.98) horizons, no-lexicon specs (+0.0143, t = +3.28), $1%/99% winsorization (+0.0133, t = +2.88), and every top-10 outlet drop (t ∈ [+2.54, +2.90]) and single-year drop (t ∈ [+2.02, +2.84]; post-2023 interactions p ≥ 0.5219$) preserve inference.

Chronological OOS Task (2021–2023 \to 2024–2025)	Constant RMSE	Controls R^2_{OOS}	Lexical R^2_{OOS}	Shrunk 4-Judge R^2_{OOS}	DM t vs. Ctrl (p)	DM t vs. Lex (p)
Stage-1 Log Comments (N_{test} = 1422)	$1.3291	+0.8863%	+1.0654%	+1.6528%	-3.74 (0.0003)	-3.39 (0.0010$)
	$0.1076	+0.4129%	+0.4995%	+1.0461%	-2.33 (0.0217)	-2.15 (0.0342$)

Chronological OOS Task (2021–2023 \to 2024–2025)	Constant RMSE	Controls R^2_{OOS}	Lexical R^2_{OOS}	Shrunk 4-Judge R^2_{OOS}	DM t vs. Ctrl (p)	DM t vs. Lex (p)
Stage-2 Reciprocity (With Splines) (N_{test} = 759)						
Stage-2 Reciprocity (Ex-Ante t = 0) (N_{test} = 759)	$0.1076	+0.4086%	+0.4997%	+1.0542%	-2.35 (0.0206)	-2.15 (0.0335$)
Stage-2 Nested Share (Ex-Ante t = 0) (N_{test} = 759)	$0.1090	+1.6946%	+1.8861%	+2.2984%	-3.21 (0.0018)	-2.61 (0.0104$)
Stage-2 Evidence Rate (Ex-Ante t = 0) (N_{test} = 759)	$0.0678	+3.5737%	+3.5924%	+4.2182%	-2.48 (0.0149)	-2.59 (0.0111$)

Table 5 reports strict chronological out-of-sample evaluation (2021–2023 training, N = 2231 and N = 1355; 2024–2025 test, N = 1422 and N = 759; Diebold & Mariano, 2002). While unshrunk high-dimensional ordinary least squares overfits (R^2_{OOS} = -4.3383% and -5.9442%), ridge-shrunk four-judge semantic predictions beat *both* structural controls and zero-LLM dictionaries across Stage-1 log comments and strictly ex-ante (t = 0, without realized volume splines) Stage-2 reciprocity, nested share, and documentary evidence ($p \leq 0.0335$).

5. Theoretical Stress-Testing and a Revised Public-Sphere Framework

Because our pre-registered tests falsified simple binary readings of H_1, H_2, and H_0, we conducted five post-data adversarial stress tests on the 226317 raw comments to construct a revised public-sphere framework (Table 6).

5.1 Agonistic Dyadic Closure: Trench Duels Versus Habermasian Reciprocity

Standard computational schemes interpret A \to B \to A reciprocity as Habermasian dialogic engagement (Graham & Witschge, 2003; Stromer-Galley, 2007). Why, then, did Moral-Conflict Framing raise reciprocity (+0.0135) while lowering speaker entropy (-0.0029) and documentary links (-0.0073)? Decomposing Total Dyadic Reciprocity ($0.211650, or 21.16%) into an exact additive identity—**Unique-Dyad First Reciprocity Rate** (0.145407, or 14.54%) plus **Repeat-Dyad Ping-Pong Rate** (0.066243, or 6.62%, 31.30%$ of baseline reciprocity)—resolves the paradox.

Exactly **$80.44% (+0.0109, SE = 0.0035, t = +3.13, p = 0.0024) of Moral Conflict's +0.0135 reciprocity gain comes from repeat back-and-forth volleys within already-fighting pairs**, whereas **first-time unique-**

dyad reciprocity is flat (+0.0026, SE = 0.0025, t = +1.05, p = 0.2950; 19.56%)! Holding \log(V_{48}) splines fixed, Moral Conflict compresses Unique Speakers Per Comment by -0.0174 (SE = 0.0052, t = -3.34, p = 0.0013, off a 67.84% mean) and elevates Top-2 Speaker Share by +0.0074 (SE = 0.0031, t = +2.38, p = 0.0198, off 17.34%), explaining why dyads per speaker rise (+0.0130, t = +2.85, p = 0.0055) as the speaker pool contracts. Moreover, documentary erosion is not merely compositional: although A \to B \to A turns cite fewer links than non-reciprocal comments (7.77% versus 10.28%), Moral Conflict **directly depresses external citing inside non-reciprocal comments (-0.0079, SE = 0.0026, t = -3.00, p = 0.0035; within-reciprocity -0.0045, t = -0.84$)**. Both Habermasian deliberation (Habermas, 2006) and agonistic pluralism (Papacharissi, 2004; Karppinen, Moe, & Svensson, 2008; Rossini, 2020) must separate *interactive persistence (heat)* from *polycentric breadth and documentary grounding (light)*: moral-conflict headlines catalyze *agonistic dyadic closure*—protracted two-person trench duels with fewer facts.

5.2 The Two-Stage Gatekeeping Chokepoint

Estimating Model 1 across progressive comment hurdles (N = 3653) explains why Stage-1 upvotes rose while ≥ 15-comment breakthroughs did not. At **zero-to-one comment ignition** (≥ 1 comment, $43.25% base rate), both Negativity (+0.0306, t = +2.54, p = 0.0129; standalone +0.0495, t = +5.47) and Moral Conflict (+0.0300, t = +2.35, p = 0.0211; standalone +0.0526, t = +5.41) spark initial reactions. Yet effects fade by ≥ 5 comments (16.37%: conflict +0.0185, t = +1.84, p = 0.0700) and vanish across ≥ 10 (12.35%), ≥ 15 (10.90%), ≥ 30 (8.60%), ≥ 50 (6.60%), and ≥ 100 comments (3.78%, all |t| ≤ 1.62), with conditional \log(comments) \mid comments ≥ 1 (N = 1579) flat at -0.0078 (t = -0.11, reflecting newly ignited 1–4 comment threads alongside ≥ 150-comment megathreads at 2.41%, t = +1.80$). Individual psychophysiological capture (Soroka, Fournier, & Nir, 2019; Robertson et al., 2023) governs first contact, but non-personalized community curation (Wright & Street, 2007; Milli et al., 2025; Piccardi et al., 2025) acts as an institutional circuit-breaker against outrage breakthroughs.

5.3 Propositional Contestability Over Arbitrary Valence Cutoffs

Why was the pre-registered $0.50 four-cell contrast C_2 - C_1 non-significant on reciprocity (+0.0111, t = +0.96, p = 0.3389; grid check p = 0.3382) despite continuous Negativity equaling zero? Grid diagnostics across {0.35, 0.40, 0.50, 0.60} and sample medians (0.325 negativity, 0.20 conflict) reveal *sub-threshold conflict contamination*: because journalistic conflict is right-skewed (median 0.20), high cutoffs sweep moderate-blame stories into Pure Threat (41.08% of C_1 at 0.50 and 48.03% at 0.60 have C_i ≥ 0.30, versus 15.27% at 0.35). Where C_1 cleanly isolates low-blame adversity—at 0.35 (C_1 = +0.0029, p = 0.6476 vs. C_2 = +0.0302, p = 0.0001), 0.40, and the median—C_2 - C_1 is significant (+0.0273, p = 0.0023; +0.0187, p = 0.0385; +0.0199, p = 0.0437). Crucially, **Non-Negative Debate (C_3: N_i < 0.50, C_i ≥ 0.50, n = 189)**—policy or legal disagreement without tragedy valence—significantly boosts reciprocity (+0.0212, p = 0.0302; +0.0254, p = 0.0060 at 0.60), nested share (+0.0179, p = 0.0083), and 90th-percentile depth (+0.2694, p = 0.0037$). The active mechanism is **propositional contestability and responsibility attribution** (Entman, 1993; Semetko & Valkenburg, 2000; Shi et al., 2019), not negative tone.

Theoretical Clash	Key Empirical Diagnostic	Revised Public-Sphere Proposition
1. Trench Duels vs. Reciprocity	$80.44% (+0.0109, t = +3.13) of +0.0135 reciprocity is repeat dyad ping-pong; speakers/comment -0.0174 (t = -3.34); non-ABA evidence -0.0079 (t = -3.00$).	**Agonistic Dyadic Closure**: Moral conflict drives locked-in two-person duels with narrower participation and fewer documentary links.
2. Two-Stage Chokepoint	$0 \to 1 ignition positive (+0.0306, t = +2.54; +0.0300, t = +2.35); ≥ 15 breakthrough null (t = +1.52, -0.51$).	**Curation Circuit-Breaker**: Affective framing sparks individual clicks/first comments, but non-algorithmic curation blocks front-page breakthroughs.
3. Contestability vs. Valence	Continuous negativity $0.00 on five of five outcomes; C_2 - C_1 +0.0273 (p = 0.0023) at 0.35 vs. null at 0.50; Non-Negative Debate C_3 +0.0212 (p = 0.0302$).	**Propositional Contestability**: Competing actors and responsibility attribution—not sad or threatening valence—drive counter-arguing.
4. Journalistic Field Scope	Civic beats (N = 962): reciprocity +0.0165 (t = +2.43); Specialist beats (N = 1151): nested +0.0112 (t = +2.30), evidence -0.0066 (t = -2.18); interaction p = 0.4173.	**Field Precision With Broad Link Erosion**: Reciprocity precision is sharpest in civic reporting, while documentary link erosion spans beats.
5. Networked Paywall Repair	Archive mirrors jump $5.38× (+0.0092, t = +5.73), neutralizing H_4 (Holm p = 1.0000); ex-ante t = 0 OOS R^2 +1.0542% to +4.2182%$.	**Networked Epistemic Repair**: Communal mirror links equalize text access in link-permissive forums; semantic framing beats word lists OOS.

5.4 Journalistic Field Scope and Networked Paywall Repair

Across journalistic fields (Benson, 2006), reciprocity precision is strongest in National/International Wire (N = 831, M = 27: +0.0199, t = +2.76, p = 0.0105) and combined Civic/Investigative reporting (N = 962, M = 41: +0.0165, t = +2.43, p = 0.0195), whereas Technology-in-Society (N = 458, M = 11) is imprecise (-0.0031, t = -0.26, $95% CI [-0.0300, +0.0237]). Respecting interaction discipline, full-panel beat interactions are non-significant (`Conflict × Civic` +0.0066, p = 0.4173; `Conflict × Tech` -0.0079, p = 0.3599), and Specialist Science, Technology, and Culture beats (N = 1151, M = 44) still exhibit significant nested-reply expansion (+0.0112, p = 0.0267) and evidence-link erosion (-0.0066, p = 0.0320 in the full-panel interaction model). Meanwhile, the 5.38$-fold surge in communal archive mirrors on paywalled outlets demonstrates *networked epistemic repair* (Barzilai-Nahon, 2008; Carlson, 2015), equalizing full-text access for downstream commenters.

6. Limitations, Scope Conditions, and Conclusion

Four scope conditions bound our conclusions. First, despite four-way outlet, quarter, slot, and beat fixed effects and replication on N = 1182 verbatim URL-slug headlines, observational headlines reflect both underlying events and editorial choices; all findings represent conditional structural associations rather than randomized causal effects. Second, Hacker News is a high-literacy, anti-insult-moderated threaded forum without personalized outrage feeds; platforms that throttle outbound links (blocking archive-mirror repair) or algorithmically reward abuse may exhibit steeper crowding-out (Milli et al., 2025; Piccardi et al., 2025). Third, professional journalism displays substantial conflict zero-mass ($31.37% exact zero and 43.33% ≤ 0.05 in Stage 1; 23.37% and 34.30% in Stage 2), and two-part hurdle checks show reciprocity shifts strongly across the extensive onset of conflict (`Conflict > 0.05` +0.0302, t = +3.39, p = 0.0011, vs. +0.0074, t = +1.34 conditional intensity). Fourth, modest submission-time out-of-sample R^2 (+1.0542% to +4.2182%$) confirms that headline framing is a genuine but bounded influence alongside full-article substance.

Ultimately, digital engagement is not one thing. Purely negative news sparks fleeting individual clicks without altering public dialogue. Moral-conflict and blame headlines reshape public conversation—not into silent monologues or inclusive Habermasian inquiry, but into deep, concentrated, documentary-light trench duels. Evaluating the health of digital public spheres requires measuring not merely whether citizens reply, but how broadly participation circulates and whether disagreement remains anchored in shared documentary evidence.

References

- Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, 59(9), 1493–1512. https://doi.org/10.1002/asi.20857
- Benson, R. (2006). News media as a "journalistic field": What Bourdieu adds to new institutionalism, and vice versa. *Political Communication*, 23(2), 187–202. https://doi.org/10.1080/10584600600629802
- Boczkowski, P. J., & Peer, L. (2011). The choice gap: The divergent online news preferences of journalists and consumers. *Journal of Communication*, 61(5), 857–876. https://doi.org/10.1111/j.1460-2466.2011.01582.x
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & 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. https://doi.org/10.1073/pnas.1618923114
- Bright, J. (2016). The social news gap: How news reading and news sharing diverge. *Journal of Communication*, 66(3), 343–365. https://doi.org/10.1111/jcom.12232
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2), 238–249. https://doi.org/10.1198/jbes.2010.07136
- Carlson, M. (2015). Metajournalistic discourse and the meanings of journalism: Definitional control, boundary work, and legitimation. *Communication Theory*, 26(4), 349–368. https://doi.org/10.1111/comt.12088
- Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20(1), 134–144. https://doi.org/10.1198/073500102753410444

- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x
- Frieß, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3), 319–339. https://doi.org/10.1002/poi3.95
- Graham, T., & Witschge, T. (2003). In search of online deliberation: Towards a new method for examining the quality of online discussions. *Communications*, 28(2), 173–204. https://doi.org/10.1515/comm.2003.012
- Habermas, J. (2006). Political communication in media society: Does democracy still enjoy an epistemic dimension? *Communication Theory*, 16(4), 411–426. https://doi.org/10.1111/j.1468-2885.2006.00280.x
- Halpern, D., & Gibbs, J. (2012). Social media as a catalyst for online deliberation? *Computers in Human Behavior*, 29(3), 1159–1168. https://doi.org/10.1016/j.chb.2012.10.008
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161. https://doi.org/10.2307/1912352
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70. https://doi.org/10.2307/4615733
- Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2018). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, 22, 129–146. https://doi.org/10.1146/annurev-polisci-051117-073034
- Jaidka, K., Zhou, A., Lelkes, Y., Egelhofer, J., & Lecheler, S. (2021). Beyond anonymity: Network affordances, under deindividuation, improve social media discussion quality. *Journal of Computer-Mediated Communication*, 27(1), zmab019. https://doi.org/10.1093/jcmc/zmab019
- Karppinen, K., Moe, H., & Svensson, J. (2008). Habermas, Mouffe and political communication: A case for theoretical eclecticism. *Javnost - The Public*, 15(3), 5–21. https://doi.org/10.1080/13183222.2008.11008973
- Marcus, G. E., MacKuen, M., & Neuman, W. R. (2011). Parsimony and complexity: Developing and testing theories of affective intelligence. *Political Psychology*, 32(2), 323–336. https://doi.org/10.1111/j.1467-9221.2010.00806.x
- Milli, S., Carroll, M., Wang, Y., Pandey, S., Zhao, S., & Dragan, A. D. (2025). Engagement, user satisfaction, and the amplification of divisive content on social media. *PNAS Nexus*, 4(3), pgaf062. https://doi.org/10.1093/pnasnexus/pgaf062
- Muddiman, A., & Stroud, N. J. (2017). News values, cognitive biases, and partisan incivility in comment sections. *Journal of Communication*, 67(4), 586–609. https://doi.org/10.1111/jcom.12312
- Mutz, D. C. (2002). Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review*, 96(1), 111–126. https://doi.org/10.1017/s0003055402004264
- Myllylahti, M. (2013). Newspaper paywalls—the hype and the reality. *Digital Journalism*, 2(2), 179–194. https://doi.org/10.1080/21670811.2013.813214
- Neuman, W. R., Guggenheim, L., Jang, S. M., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64(2), 193–214. https://doi.org/10.1111/jcom.12088
- Noëlle-Neumann, E. (1974). The spiral of silence: A theory of public opinion. *Journal of Communication*, 24(2), 43–51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x
- Papacharissi, Z. (2004). Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2), 259–283. https://doi.org/10.1177/1461444804041444

- Piccardi, T., Saveski, M., Jia, C., Hancock, J., Tsai, J. L., & Bernstein, M. S. (2025). Reranking partisan animosity in algorithmic social media feeds alters affective polarization. *Science*, 387(6732), eadu5584. https://doi.org/10.1126/science.adu5584
- Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), e2024292118. https://doi.org/10.1073/pnas.2024292118
- Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 7(5), 812–822. https://doi.org/10.1038/s41562-023-01538-4
- Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*, 49(3), 399–425. https://doi.org/10.1177/0093650220921314
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(1), 103–122. https://doi.org/10.1111/j.1460-2466.1999.tb02784.x
- Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, 50(2), 93–109. https://doi.org/10.1111/j.1460-2466.2000.tb02843.x
- Shi, F., Teplitskiy, M., Duede, E., & Evans, J. A. (2019). The wisdom of polarized crowds. *Nature Human Behaviour*, 3(4), 329–336. https://doi.org/10.1038/s41562-019-0541-6
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. https://doi.org/10.1037/0033-2909.86.2.420
- Sjøvaag, H. (2015). Introducing the paywall: A case study of content changes in three online newspapers. *Journalism Practice*, 10(3), 304–322. https://doi.org/10.1080/17512786.2015.1017595
- Soroka, S. (2014). *Negativity in democratic politics: Causes and consequences*. Cambridge University Press. https://doi.org/10.1017/cbo9781107477971
- Soroka, S., Fournier, P., & 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. https://doi.org/10.1073/pnas.1908369116
- Soroka, S., & McAdams, S. (2015). News, politics, and negativity. *Political Communication*, 32(1), 1–22. https://doi.org/10.1080/10584609.2014.881942
- Stromer-Galley, J. (2007). Measuring deliberation's content: A coding scheme. *Journal of Deliberative Democracy*, 3(1), 12. https://doi.org/10.16997/jdd.50
- Trussler, M., & Soroka, S. (2014). Consumer demand for cynical and negative news frames. *The International Journal of Press/Politics*, 19(3), 360–379. https://doi.org/10.1177/1940161214524832
- Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Reinemann, C., Papathanassopoulos, S., Maier, M., Sheafer, T., Stanyer, J., Štětka, V., & Theocharis, Y. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, 41(1), 3–27. https://doi.org/10.1080/23808985.2017.1288551
- Wright, S., & Street, J. (2007). Democracy, deliberation and design: The case of online discussion forums. *New Media & Society*, 9(5), 849–869. https://doi.org/10.1177/1461444807081230