

Outer Limits: An Experimental Approach to Controlled Content Manipulation within the Reddit Interface

Chenchen Mao¹, Hanjing Shi¹, Haiyan Jia¹, Daniel Unguryan¹, Eric Baumer²,
and Dominic DiFranzo¹

¹ Lehigh University, Bethlehem, PA, USA {chm321, hasa23, haj616, djd219}@lehigh.edu

² University of Toronto, Toronto, ON, Canada
eps.baumer@utoronto.ca

³ Computer Science & Engineering, Lehigh University, Bethlehem, PA, USA
unguryan77@gmail.com

Abstract. Independent researchers often lack access to intervention capabilities for controlled experiments on live social media platforms. We present Outer Limits, a browser-based system for controlled content experiments within the existing Old Reddit interface, rather than in a reconstructed simulation. The system renders content locally, records study events, and contains configured voting and commenting actions so that neither constructed content nor experimental write interactions reach Reddit. In a 219-participant perceptual-fidelity study, ART ANOVAs found no significant Post Type, Participant Awareness, or interaction effects. Exploratory TOSTs met the $d = \pm 0.50$ equivalence criterion for the marginal contrasts and for Post Type within the forewarned subgroup. We also illustrate the system with a factorial study varying post frame, comment frame, and comment stance. Outer Limits combines three properties that the approaches considered here provide separately: precise control over experimental content, an existing platform interface, and containment of experimental content and interactions from the host community.

Keywords: Social media experiments · Controlled content manipulation · Browser extensions · Reddit

1 Introduction

Understanding user behavior on social media informs research on political polarization, misinformation, and public health [9, 20, 1, 8]. These problems require identifying which platform features and content change behavior, not simply documenting what users do. Controlled experiments can help identify such causal relationships [13], including how perceived group norms shape responses to online misinformation [1].

When a hypothesized mechanism depends on a platform’s social and interface context, conducting the experiment within that platform allows the effect to be

estimated in the setting in which the mechanism operates. Platform-run studies, such as Meta’s controversial “emotional contagion” experiment, illustrate this form of experimentation [18]. Such interventions, however, have required internal access or platform cooperation, which independent researchers rarely have.

Independent researchers therefore rely on observational data, which lacks randomized control [6, 22, 4, 14], or on vignettes and simulations, which provide control but do not reproduce the social embeddedness of the target platform [3, 26, 16]: participants’ actions are not visible to their real networks and carry no real reputational consequences. Findings from a reconstructed environment therefore leave open whether the same effect would emerge on the platform itself.

Experiments conducted within the platform can reduce this uncertainty. Recent client-side field experiments show that browser extensions can do this without platform cooperation, by reranking the content participants encounter in their own feeds [24, 23]. Because that content is drawn from each participant’s personalized feed, however, the specific posts and comments differ across participants. Independent researchers therefore still lack a way to present precisely matched, researcher-specified content within an existing interface while keeping that content and participants’ interactions from reaching the host community.

We present an experimental approach to controlled content manipulation within an existing social media interface, instantiated in Outer Limits. The approach links platform selection, localized content manipulation, containment, and fidelity validation. Outer Limits implements matched content variation, event logging, and containment of configured write interactions. We evaluate perceptual fidelity with 219 participants and illustrate its use in a factorial experiment.

The contribution is limited to content-level interventions in the studied Reddit setting. The validation applies only to the tested outcomes, equivalence bounds, materials, and sample.

2 Related Work

2.1 Controlled Experiments in Reconstructed Social Media Environments

Independent researchers frequently use reconstructed environments to conduct controlled social media experiments without relying on platform cooperation. Static vignette studies present participants with hypothetical scenarios, such as text descriptions or screenshots of social media posts [3, 26]. These approaches allow researchers to manipulate content while holding other stimulus features constant, but static stimuli offer limited interaction with the surrounding platform environment. This may constrain how participants engage with the material compared with a live platform.

More interactive alternatives include the (Mis)information Game [7], the Mock Social Media Website Tool [15], and the Truman Platform [11]. These systems combine experimental control with an interactive interface that participants can browse, react to, and engage with.

The main limitation is not necessarily perceived artificiality, because participants may find these environments engaging. A reconstructed environment, however, does not reproduce the social embeddedness of the target platform. Participants’ actions are not visible to their actual networks and carry none of the reputational consequences of a real like, comment, or share. When anticipated audiences and platform conventions shape how participants interpret content and decide whether to engage, findings from a reconstructed environment leave open whether the same effect would emerge on the target platform.

2.2 Independent Experiments on Existing Platforms

An alternative is to intervene within an existing platform interface rather than reconstructing it. Historically this has depended on platform-controlled infrastructure: Meta’s “emotional contagion” study altered the news feeds of nearly 700,000 users [18], while a 61-million-person Facebook experiment examined how messages containing social cues affected voter turnout [5]. These studies demonstrate the scale and directness of platform-based experimentation but generally required platform access or cooperation.

Independent researchers have often relied instead on observational data, even as platforms have increasingly restricted API access [6, 10, 28]. User-donation approaches, including Mozilla’s RegretsReporter, Digital Footprints, and MovieLens, partially address these access barriers [22, 4, 14]. These approaches provide access to naturally occurring behavior but do not by themselves provide randomized manipulation of content or interface features. Causal identification therefore requires additional assumptions or research designs.

Client-side reranking experiments provide a more direct route to independent intervention within existing platforms. Piccardi et al. describe a browser-extension method that intercepts and reranks existing content in participants’ social media feeds without platform cooperation and apply this method in a field experiment [23, 24]. Because that content comes from each participant’s own feed, however, the specific posts and comments differ across participants, and their wording remains outside the researcher’s control.

Intervening within a live platform also carries a cost that reconstructed environments do not: experimental actions can affect users or communities that did not consent to participate. The r/ChangeMyView moderation team reported an unauthorized experiment that introduced undisclosed AI-generated comments without prior community consultation [25]. These concerns motivate a central design requirement of Outer Limits: constructed content is not published to Reddit, and designated participant interactions do not generate corresponding platform write requests. This containment addresses spillover into the host platform and community, though it does not by itself resolve questions of deception, consent, or debriefing.

Approach	Interface	What is controlled	Social context	Platform effects
Static vignette	Static image or text	Displayed content	None	None
Standalone simulation	Reconstructed	Content and features	Known to be constructed	None
Client-side reranking	Existing feed	Exposure and ordering	Participant's own network	Real
Outer Limits	Existing Reddit	Matched content variants	Existing community	Contained

Table 1. Design space for social media experiments. The table compares design characteristics, not validated outcomes, and does not imply a hierarchy of methods.

2.3 Positioning Outer Limits

These approaches differ in the interface participants use, what the researcher controls, the social context participants perceive, and what reaches the platform (Table 1). Vignettes and simulations give the researcher full control over content, but the environment is built for the study and participants' actions reach no real audience. Client-side reranking keeps participants on the existing platform and controls which posts appear, but not their wording. Outer Limits controls what the posts say, presents them within the existing Reddit community, and prevents constructed content and participant interactions from reaching the platform.

Outer Limits is designed for studies requiring this combination. Standalone simulations or client-side reranking experiments may remain preferable when a study prioritizes, respectively, complete control over the full experimental environment or naturalistic exposure to existing personalized feeds and ordinary downstream platform behavior.

3 Design Principles

Outer Limits begins with an existing platform interface rather than a reconstructed simulation. This design is useful when interpretation of the treatment depends on platform context, while the causal comparison requires matched, researcher-specified content. Starting from an existing interface does not imply naturalistic recruitment, unrestricted platform use, or a field experiment.

The approach rests on four principles.

Platform Selection: Researchers select a platform whose interface matters to the research question and can be modified in the participant's browser. This approach is most useful when posts, comments, or other presentation cues are part of the experimental context.

- Localized Content Manipulation:** Researchers change only the posts, comments, or other elements required by the experimental design, while keeping the surrounding interface and nonmanipulated content unchanged. This allows precise and consistent manipulation without access to the platform’s servers.
- Containment:** Constructed content and participant actions such as votes and comments should remain within the study and should not reach or affect the host platform or its community.
- Fidelity Validation:** Manipulated content should appear naturally integrated into the host interface. Building on an existing platform preserves its established mechanics but limits how much researchers can customize [12]. Inconsistencies may reveal the intervention and change how participants interpret the content. Covert manipulations therefore require evidence that altered content does not create substantial differences on the selected fidelity measures; disclosed interventions need not meet the same standard. Perceptual fidelity contributes to experimental realism, but does not establish it in full. Participants need only find the situation credible enough to take it seriously and engage with the content as intended [2].

4 Outer Limits: Design and Implementation

Outer Limits implements this approach through a Chrome extension and backend for controlled content experiments within the existing Reddit interface. We selected Reddit because its structure suits social media experiments. Its division into thousands of topic-specific communities (“subreddits”) offers substantial diversity for research questions [27], and the platform accommodates content types from simple text to rich media. Reddit also has an established culture of “throw-away accounts” for discussing sensitive topics [21], making accounts without long-standing identities familiar on the platform.

4.1 System Architecture and Data Flow

Outer Limits consists of two components (Figure 1).

The Chrome extension runs in the participant’s browser. It fetches the participant’s assigned experimental condition from the backend, replaces the designated content in the Reddit page, logs the participant’s interactions, and prevents those interactions from reaching Reddit.

The backend is an Express API with a MongoDB database. It stores the experimental conditions that researchers configure, assigns participants to conditions, and receives the interaction records the extension returns.

A session begins when the participant enters a randomly generated Prolific identifier, which serves as a pseudonym in the study backend. The backend either retrieves an existing assignment for that identifier or assigns the participant to a condition, balancing enrollment across conditions. The identifier subsequently links the participant’s condition assignment to their interaction records. Uninstalling the extension ends its ability to modify pages or record interactions.

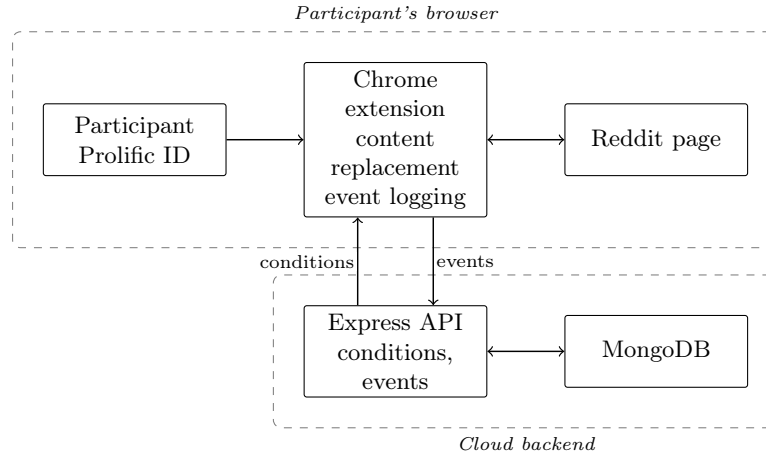


Fig. 1. Outer Limits architecture. The extension fetches the assigned condition from the backend, replaces content in the Reddit page, and logs interactions. Constructed content and participants’ votes and comments stay within the participant’s browser and the study backend; they are not sent to Reddit.

4.2 Researcher Configuration and Deployment

Researchers specify experiments through two CSV files. A post file defines each target URL and the content to be displayed there: title, body text, image, displayed author, timestamp, vote count, and the condition to which the post belongs. A comment file links comments to posts and specifies their text, displayed author, timestamp, and vote count. An upload script writes these records to the study database.

Because the extension reads all experimental content from the database at runtime, researchers can revise stimuli, add conditions, or change a factorial design without modifying or rebuilding the extension. We hosted the backend on Render and used MongoDB Atlas for storage; both are implementation choices rather than requirements of the approach.

For participant deployment, we published the extension through the Chrome Web Store so that participants could install it directly rather than manually load an unpacked extension. This distribution route was intended to reduce participant setup steps and the installation support required during study enrollment. The listing address is omitted for anonymous review.

4.3 Runtime Manipulation and Containment

Content replacement. When a participant opens a target page, the extension requests the assigned condition from the backend and uses the returned fields to replace the corresponding elements in the page DOM: the post’s title, body text, image, displayed author, timestamp, and vote count, together with the

comment thread. The surrounding interface (the subreddit header, sidebar, navigation, and any nonmanipulated content) continues to load from Reddit, so the intervention remains localized to the experimental content unit. Constructed content is rendered only in the participant’s browser and is never published to Reddit.

Interaction logging. The extension logs designated study interactions: post views, upvotes, downvotes, comment submissions, and subsequent comment edits. Each record carries the participant identifier, a timestamp, the action type, and the relevant post or comment metadata, and is sent to the backend.

Containment of write interactions. Logging alone does not prevent an interaction from reaching Reddit; containment is implemented separately for each write action. For votes on manipulated content, the extension intercepts the participant’s click before Reddit’s own event handler executes, prevents the event from propagating to that handler, updates the displayed vote state locally, and submits an interaction record to the backend. For comment submissions, the extension captures the submitted text and stores it as a study record without generating the corresponding write request to Reddit. From the participant’s perspective, the interface responds to voting and commenting, but no corresponding action reaches Reddit’s servers.

Software availability. Outer Limits is open-source. Its repository includes the Chrome extension, the backend, an upload script, example CSV files, and a step-by-step deployment guide. The address is omitted for anonymous review.

5 Validation: Assessing Perceptual Fidelity

The validation tested whether participants perceived manipulated posts as realistic as authentic posts and whether forewarning them about possible alterations changed those perceptions. Forewarning served as a boundary condition: if ratings remained similar when participants expected alterations, casual browsing would be less likely to explain the result.

RQ1. Do participants perceive manipulated posts as realistic as authentic posts within the existing Reddit interface?

RQ2. Does explicitly forewarning participants that content may have been altered change these perceptions?

5.0.1 Method We conducted a validation study on Prolific using a 2×2 between-subjects factorial design. The factors were **Post Type** (authentic vs. manipulated) and **Participant Awareness** (uninformed vs. informed).

For study materials, we selected one image-based and one text-only post from Reddit’s **r/funny** community. In the experimental condition, Outer Limits replaced their title, text, image, and comments with content from the same subreddit. The URL remained unchanged (Figures 2 and 3).

(a) Authentic image-based post



(b) Manipulated image-based post



Fig. 2. Image-based validation stimuli shown at a readable scale. Outer Limits retained the target URL and surrounding Reddit interface while replacing the title, post text, image, and comments in panel (b).

After a pre-survey, participants installed the extension and entered their Prolific identifier. The backend assigned them to one of four conditions. Participants in the informed condition were told that the posts might have been altered. All participants viewed the two target links and completed a post-survey containing content-recall checks and perceptual-fidelity measures, followed by a debriefing. The study received IRB approval at [Anonymous] University (protocol #2171134-1).

5.0.2 Variables and Measurements Participants rated four retrospective statements on a 7-point scale (1 = Strongly Disagree, 7 = Strongly Agree): the way interactive elements (e.g., upvote/downvote buttons and comment placement) functioned seemed different; the overall look and layout of the posts or page appeared different from usual; the content of the posts, including text, images, and comments, seemed altered; and “I did not notice anything unusual.” They also rated confidence from “not confident at all” to “very confident” and described anything unusual in an open-ended response.

5.0.3 Participants Between November and December 2024, we recruited 225 participants via Prolific. Six participants were removed for failing condition-specific content-recall checks, which required correctly identifying the content shown in their assigned condition; 219 participants remained. Compensation was \$6.50. An a priori power analysis ($f = 0.25$, power = .95) required $N = 210$. Mean age was 33.9 years ($SD = 9.9$); 63% identified as male and 33% as female. Demographics did not differ by condition (all $p > .18$).

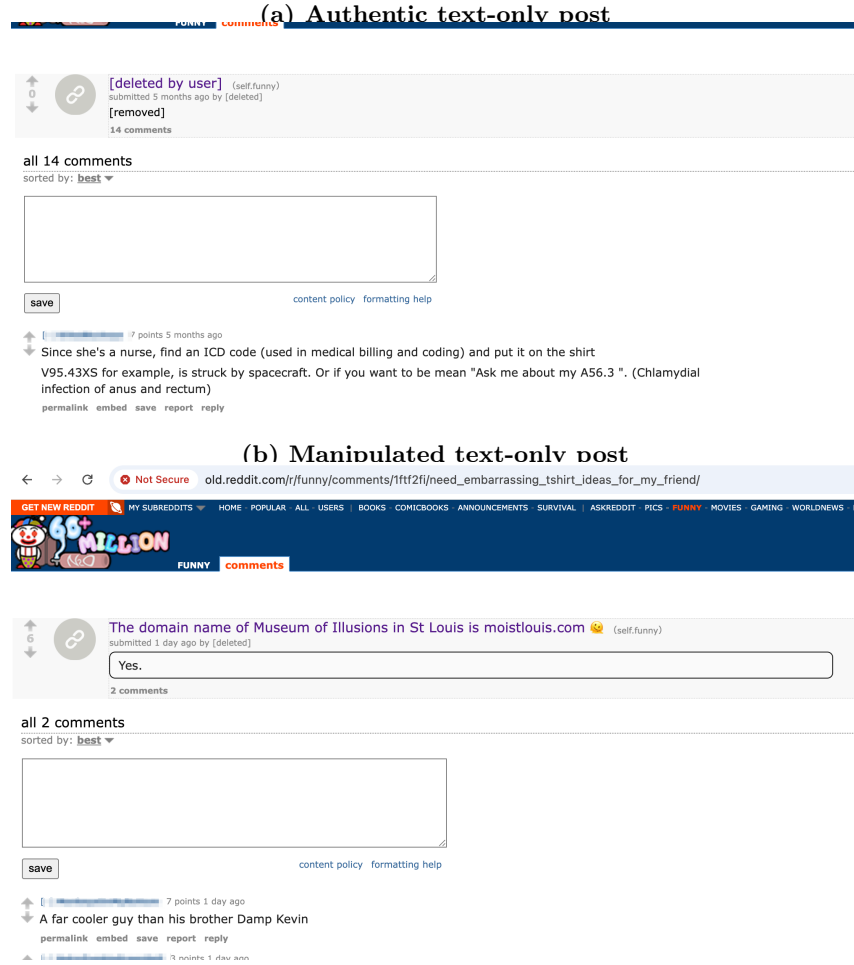


Fig. 3. Text-only validation stimuli shown at a readable scale. The authentic post’s title and body had been deleted by its owner. Outer Limits retained the target URL and surrounding Reddit interface while replacing the title, post text, and comments in panel (b).

5.1 Validation Results

5.1.1 Quantitative Analysis We used ART ANOVAs to test Post Type, Participant Awareness, and their interaction. Because nonsignificant difference tests do not establish equivalence, we also conducted exploratory TOSTs using $d = \pm 0.50$ bounds [19]. The TOSTs examined the two marginal contrasts and Post Type within the forewarned subgroup; the bounds were not specified in advance.

No ART effect was significant (Table 2). All reported TOST contrasts met the $d = \pm 0.50$ equivalence criterion, supporting statistical equivalence within

Response	ART ANOVA: $F(p)$			TOST ($d = \pm 0.50$)	
	Post Type	Awareness	Interaction	Marginal Post/Aware.	Post within forewarned
Interaction	1.53 (.217)	0.05 (.830)	0.39 (.531)	Eq./Eq.	Eq.
Layout	1.34 (.249)	0.78 (.378)	0.65 (.422)	Eq./Eq.	Eq.
Content	1.98 (.161)	0.01 (.904)	0.37 (.545)	Eq./Eq.	Eq.
Nothing					
unusual	2.07 (.152)	0.04 (.841)	0.08 (.774)	Eq./Eq.	Eq.
Confidence	0.08 (.771)	0.49 (.483)	0.07 (.790)	Eq./Eq.	Eq.

Table 2. ART ANOVA and exploratory TOST results. ART cells report $F(p)$; all effects were nonsignificant. Eq. denotes that both one-sided tests rejected effects at or beyond $d = \pm 0.50$ at $\alpha = .05$ for the marginal Post Type/Awareness contrasts and Post Type within the forewarned subgroup ($n = 58$ vs. 49).

these bounds for the tested outcomes and setting. This does not establish zero difference or experimental realism in full.

5.1.2 Qualitative Analysis Two authors calibrated the coding scheme on an initial 10% of responses and independently coded a second 10%, reaching Krippendorff’s $\alpha > 0.8$ across seven themes: No Changes Noticed, Layout Issues, Content Suspicion, Clear Alterations, Uncertain, Technical Issues, and Other. They then independently coded the full dataset and resolved remaining disagreements through discussion. We fitted a Poisson regression to theme-by-condition counts with theme category, Post Type, Participant Awareness, and their interaction as predictors. Neither Post Type ($\beta = -0.03, p = .910$), Participant Awareness ($\beta = -0.11, p = .647$), nor their interaction ($\beta = 0.35, p = .259$) predicted theme frequencies. The qualitative results were consistent with the quantitative results.

6 Illustrative Case Study: A Factorial Reddit Study

We illustrate Outer Limits with a factorial study of how post frame, comment frame, and comment stance shape health beliefs and behavioral intentions [1, 17]. Direct deployment would require publishing vaccine-related experimental content, while a standalone simulation would remove the Reddit context.

The $2 \times 2 \times 2$ design varied post frame, comment frame, and comment stance while holding usernames, timestamps, metrics, and other page elements constant. Participants completed a pre-survey, installed the extension, viewed their condition, and completed a post-survey followed by debriefing. The study received IRB approval (protocol #2171134-2). We report the design rather than substantive outcomes.

7 Discussion

7.1 Contribution

Outer Limits combines matched content manipulation within the existing Reddit interface with containment of constructed content and configured write interactions. By separating the interface participants experience from the platform consequences their actions would otherwise produce, it supports studies in which platform context matters but unrestricted downstream behavior is unnecessary. For computational social science, this configuration makes matched content comparisons possible when platform APIs do not provide intervention access. It also allows researchers to study sensitive content without publishing it to host communities.

Methodologically, this approach complements observational platform studies and reconstructed experiments rather than replacing either. It allows theory-derived content mechanisms to be tested under randomized, matched conditions while retaining the interface cues through which those mechanisms may operate. This combination can help researchers examine whether effects attributed to content depend on the interface context in which participants encounter it. Evidence from such experiments could also inform assumptions used in computational models, although the present study does not test that use.

7.2 Relation to Existing Approaches

Compared with reconstructed environments, Outer Limits retains the host interface instead of requiring researchers to reproduce selected platform conventions, but it offers less control over the surrounding environment and remains vulnerable to platform changes [16]. Piccardi et al. likewise use a browser extension to intervene without platform cooperation [23, 24], but they rerank the content participants encounter in their own feeds rather than specifying it. Outer Limits instead assigns matched, researcher-specified posts and comments, records designated events, and contains configured write interactions. This combination is useful when exact content contrasts, existing platform cues, and bounded platform consequences all matter, but not when naturalistic feed use is required.

7.3 Ethical Considerations

Containment is both a technical and an ethical boundary. It prevents constructed content and configured interactions from reaching the host community, but it does not eliminate the ethical questions that covert manipulation raises. Participants act under the belief that their votes and comments carry ordinary platform consequences; in fact these interactions are contained. The same belief may contribute to behavioral credibility but also constitutes deception. We addressed it through IRB review and full debriefing, and researchers deploying this approach should do the same, in addition to documenting and testing which write interactions are intercepted and protecting backend data.

7.4 Limitations

The current implementation targets desktop Old Reddit and does not support newer Reddit interfaces, mobile clients, or video content. Because content replacement depends on the rendered page structure, changes to Reddit’s DOM can require updates and renewed fidelity assessment. The validation itself covers two posts from one subreddit, evaluated by a Prolific sample; the equivalence criterion was met using bounds selected after data collection. Fidelity is a property of particular materials and settings, not an automatic consequence of technical integration, so each new deployment requires its own assessment.

7.5 Future Work

The most direct test of whether the existing platform interface affects an estimated treatment effect would compare the same manipulation in Outer Limits and a reconstructed simulation. The approach could also support matched studies of message framing, moderation cues, social norms, and misinformation responses in settings where community context is part of the treatment. Adaptive computational social science experiments could vary content in real time according to prespecified participant actions. Such extensions would require auditable adaptation rules, controls against unintended treatment variation, and renewed review of consent, deception, and containment. Applications to other interfaces or platforms would require renewed technical, fidelity, and ethical review in each setting.

8 Conclusion

This paper presents an experimental approach to controlled content manipulation within an existing social media interface and implements it in Outer Limits for Old Reddit. A 219-participant study supports perceptual fidelity within the tested equivalence bounds and setting. An illustrative factorial case applies Outer Limits to matched post and comment variations. Relative to the reconstructed simulations and feed-reranking methods examined here, Outer Limits combines researcher-specified content, the host interface, and bounded downstream consequences in one experimental design.

References

1. Aghajari, Z., Baumer, E.P.S., DiFranzo, D.: What’s the norm around here? Individuals’ responses can mitigate the effects of misinformation prevalence in shaping perceptions of a community. In: Proceedings of the 2023 CHI conference on human factors in computing systems. Chi ’23, Association for Computing Machinery, Hamburg, Germany (2023). <https://doi.org/10.1145/3544548.3580946>

2. Aghajari, Z., Baumer, E.P.S., Hohenstein, J., Jung, M.F., DiFranzo, D.: Methodological Middle Spaces: Addressing the Need for Methodological Innovation to Achieve Simultaneous Realism, Control, and Scalability in Experimental Studies of AI-Mediated Communication. *Proc. ACM Hum.-Comput. Interact.* **7**(CSCW1), 73:1–73:28 (Apr 2023). <https://doi.org/10.1145/3579506>
3. Atzmüller, C., Steiner, P.M.: Experimental vignette studies in survey research. *Methodology* **6**(3), 128–138 (2010). <https://doi.org/10.1027/1614-2241/a000014>
4. Bechmann, A., Vahlstrup, P.B.: Studying facebook and instagram data: The digital footprints software. *First Monday* **20**(12) (Dec 2015). <https://doi.org/10.5210/fm.v20i12.5968>
5. Bond, R.M., Fariss, C.J., Jones, J.J., Kramer, A.D.I., Marlow, C., Settle, J.E., Fowler, J.H.: A 61-million-person experiment in social influence and political mobilization. *Nature* **489**(7415), 295–298 (Sep 2012). <https://doi.org/10.1038/nature11421>
6. Bruns, A.: After the ‘APIcalypse’: social media platforms and their fight against critical scholarly research. *Information, Communication & Society* **22**(11), 1544–1566 (2019). <https://doi.org/10.1080/1369118X.2019.1637447>
7. Butler, L.H., Lamont, P., Wan, D.L.Y., Prike, T., Nasim, M., Walker, B., Fay, N., Ecker, U.K.: The (mis) information game: A social media simulator. *Behavior Research Methods* **56**(3), 2376–2397 (2024)
8. Charles-Smith, L.E., Reynolds, T.L., Cameron, M.A., Conway, M., Lau, E.H., Olsen, J.M., Pavlin, J.A., Shigematsu, M., Streichert, L.C., Suda, K.J., others: Using social media for actionable disease surveillance and outbreak management: a systematic literature review. *PloS one* **10**(10), e0139701 (2015)
9. Conover, M., Ratkiewicz, J., Francisco, M., Goncalves, B., Menczer, F., Flammini, A.: Political polarization on twitter. *Proceedings of the International AAAI Conference on Web and Social Media* **5**(1), 89–96 (Aug 2021). <https://doi.org/10.1609/icwsm.v5i1.14126>
10. Davidson, B.I., Wischerath, D., Racek, D., Parry, D.A., Godwin, E., Hinds, J., Van Der Linden, D., Roscoe, J.F., Ayravainen, L., Cork, A.G.: Platform-controlled social media APIs threaten open science. *Nature Human Behaviour* **7**(12), 2054–2057 (2023)
11. DiFranzo, D., Taylor, S.H., Kazerooni, F., Wherry, O.D., Bazarova, N.N.: Upstanding by design: Bystander intervention in cyberbullying. In: *Proceedings of the 2018 CHI conference on human factors in computing systems*. pp. 1–12. Chi ’18, Association for Computing Machinery, Montreal QC, Canada (2018). <https://doi.org/10.1145/3173574.3173785>
12. Epstein, D.A., Liu, F., Monroy-Hernández, A., Wang, D.: Revisiting piggyback prototyping: Examining benefits and tradeoffs in extending existing social computing systems. *Proc. ACM Hum.-Comput. Interact.* **6**(CSCW2) (Nov 2022). <https://doi.org/10.1145/3555557>
13. Gilbert, D.T., Fiske, S.T., Lindzey, G.: *The Handbook of Social Psychology*. McGraw-Hill (1998)
14. Harper, F.M., Konstan, J.A.: The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems* **5**(4) (Dec 2015). <https://doi.org/10.1145/2827872>
15. Jagayat, A., Boparai, G., Pun, C., Choma, B.: Mock social media website tool (1.0). Computer software.[accessed 2024 Sep 10]. <https://docs.studysocial.media> (2021)

16. Jagayat, A., Choma, B.L.: A primer on open-source, experimental social media simulation software: Opportunities for misinformation research and beyond. *Current Opinion in Psychology* **55**, 101726 (2024). <https://doi.org/10.1016/j.copsyc.2023.101726>
17. Janz, N.K., Becker, M.H.: The Health Belief Model: a decade later. *Health Education Quarterly* **11**(1), 1–47 (1984). <https://doi.org/10.1177/109019818401100101>
18. Kramer, A.D.I., Guillory, J.E., Hancock, J.T.: Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* **111**(24), 8788–8790 (2014). <https://doi.org/10.1073/pnas.1320040111>
19. Lakens, D.: Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science* **8**(4), 355–362 (2017). <https://doi.org/10.1177/1948550617697177>
20. Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J., Zittrain, J.L.: The science of fake news. *Science* **359**(6380), 1094–1096 (2018). <https://doi.org/10.1126/science.aao2998>
21. Leavitt, A.: "This is a throwaway account": Temporary technical identities and perceptions of anonymity in a massive online community. In: *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. pp. 317–327. Cscw '15, Association for Computing Machinery, Vancouver, BC, Canada (2015). <https://doi.org/10.1145/2675133.2675175>
22. Mozilla: Mozilla investigation: YouTube algorithm recommends videos that violate the platform's very own policies (Jul 2021), <https://foundation.mozilla.org/en/blog/mozilla-investigation-youtube-algorithm-recommends-videos-that-violate-the-platforms-very-own-policies/>
23. Piccardi, T., Saveski, M., Jia, C., Hancock, J., Tsai, J.L., Bernstein, M.S.: Reranking partisan animosity in algorithmic social media feeds alters affective polarization. *Science* **390**(6776), eadu5584 (Nov 2025). <https://doi.org/10.1126/science.adu5584>
24. Piccardi, T., Saveski, M., Jia, C., Hancock, J.T., Tsai, J., Bernstein, M.: Reranking Social Media Feeds: A Practical Guide for Field Experiments. *ACM Transactions on Social Computing* **9**(1), 2:1–2:17 (Mar 2026). <https://doi.org/10.1145/3800557>
25. r/ChangeMyView Moderation Team: Unauthorized experiment on CMV involving AI-generated comments (Apr 2025), https://www.reddit.com/r/changemyview/comments/1k8b2hj/meta_unauthorized_experiment_on_cmv_involving/, reddit community announcement, accessed 13 July 2026
26. Vraga, E.K., Bode, L.: Addressing COVID-19 Misinformation on Social Media Preemptively and Responsively. *Emerging Infectious Diseases* **27**(2), 396–403 (Feb 2021). <https://doi.org/10.3201/eid2702.203139>
27. Weninger, T., Zhu, X.A., Han, J.: An exploration of discussion threads in social news sites: a case study of the Reddit community. In: *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining*. pp. 579–583. Asonam '13, Association for Computing Machinery, Niagara, Ontario, Canada (2013). <https://doi.org/10.1145/2492517.2492646>
28. Ying, N.: The evolving landscape of web scraping on social media platforms | d-lab (Mar 2025), <https://dlab.berkeley.edu/news/evolving-landscape-web-scraping-social-media-platforms>