Harvard Kennedy School Misinformation Review[1]
November 2026, Volume 7, Issue 7
Creative Commons Attribution 4.0 International (CC BY 4.0)
Reprints and permissions: misinforeview@hks.harvard.edu
DOI: https://doi.org/10.37016/mr-2020-216
Website: https://misinforeview.hks.harvard.edu

Commentary

Feed algorithms shape the misinformation supply, not only what users believe

Debates about platform effects keep asking whether ranked feeds change an individual's beliefs. That question matters, but it misses a second effect. Ranking systems reward some messages, teach the people who post what travels, and gradually change the supply of information available to everyone. Treating engagement ranking as a selection environment clarifies why mixed findings about short-term persuasion can coexist with serious ecosystem risks. It also points toward better research, audits, and platform design.

Author: David E. Clementson
Affiliations: Grady College of Journalism and Mass Communication, University of Georgia, USA
How to cite: Clementson, D. E. (2026). Feed algorithms shape the misinformation supply, not only what users believe. *Harvard Kennedy School (HKS) Misinformation Review*, *7*(7).
Received: July 27th, 2026. Accepted: October 23rd, 2026. Published: November 19th, 2026.

Introduction

Research on feed algorithms can appear contradictory. Experimentally changing Facebook and Instagram feeds during the 2020 U.S. election altered what people saw but did not measurably change several political attitudes over three months (Guess et al., 2023). A later experiment on X found that its algorithmic feed increased engagement and shifted some political opinions relative to a chronological feed (Gauthier et al., 2026). These findings show that persuasion effects vary across platforms, populations, and moments.

Both studies ask a necessary question: does ranked exposure, a feed ordered by predicted engagement rather than by time posted, change the person viewing it? Misinformation research should also ask what repeated ranking does to the messages available for anyone to view. Feed algorithms allocate scarce attention, reward successful message variants, and give producers immediate feedback about what to make next. (Producers here are the accounts that create and post content, from individual creators to coordinated influence operations, and not the platforms.) A system with those three properties is a selection environment, a setting in which many messages compete for reach, a few are amplified, and the features that won amplification are copied into the next round. It can reshape the supply of public information even when no measurable attitude changes occur during a particular experiment. Platforms and regulators should audit the content ecosystems that ranking creates, with that

[1] *A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.*

obligation scoped to the largest engagement-ranked systems, where selection pressure is strongest. Interventions should change the rewards attached to accuracy as well as remove individual falsehoods.

Ranking is a gate, not a neutral pipe

Gatekeeping research describes how the vast universe of possible messages becomes the much smaller set that reaches an audience (Shoemaker & Vos, 2009). A newspaper editor, for example, does this openly, surveying everything happening in the world on a given day, judging a few items newsworthy, assigning those news items to reporters, and printing the select reports. Social platforms briefly seemed to abolish the gatekeeper, since anyone could publish without an editor's permission. But no one can attend to everything, so ranking systems now make the same kinds of decisions about which posts appear, in what order, and for whom. These systems encode their criteria as predictions about what a viewer will do next, such as watch, share, comment, or return. As DeVito (2017) showed in an early analysis of Facebook, algorithmic selection embeds values even when those values are expressed as technical signals. The point is not that editorial gatekeeping was neutral, but that every system of distribution applies criteria.

This makes a feed more than a pipe carrying messages to willing consumers. It is an institution that allocates visibility at a scale and speed no editorial desk could match. A chronological feed and an engagement-ranked feed may contain content from the same eligible pool, but they create different rewards for producers. Those rewards matter because the content pool is not fixed. It responds to the distribution system.

Repeated ranking reshapes the supply of misinformation

To create a selection environment, three things need to happen: 1) producers generate message variants, 2) the platform's ranking system selects some variants for greater reach, and 3) producers retain what worked when creating the next round. This cycle does not require a platform to favor misinformation intentionally. It requires only that some feature of a post reliably produces the behavior that ranking systems are built to predict. Those systems do not assess accuracy. They estimate how likely a viewer is to click, share, or keep watching, and distribute accordingly. If a feature that raises those estimates, such as how a post is worded or how surprising it sounds, appears more often in false claims than accurate ones, falsehood wins reach without anyone designing it that way.

Several measured features of online diffusion make that connection consequential. False news spread farther, faster, deeper, and more broadly than true news in a large study of Twitter cascades, partly because falsehood carried greater novelty (Vosoughi et al., 2018). More recent work found that misinformation sources evoke more outrage than trustworthy sources and that outrage makes misinformation more likely to be shared, sometimes without users opening the linked article (McLoughlin et al., 2024). Platform feedback also changes later expression. Users who received more positive feedback for moral outrage became more likely to express outrage again, which is evidence of social learning rather than merely stable preference (Brady et al., 2021).

Together, these findings support a supply-side mechanism. If outrage, novelty, or identity affirmation reliably earns distribution, producers who use those features gain visibility and learn to repeat them. Other producers imitate successful formats. Claims are simplified, recombined, or sharpened to compete in the next cycle. Over time, the population of circulating messages can become better adapted to engagement ranking.

A misleading claim can gain an advantage through correlated features such as novelty or anger, then lose that advantage when the same claim becomes familiar. Producers can respond with a new allegation, a stronger emotional cue, or a format that is harder to moderate. Selection is therefore dynamic,

concerning the reproduction and modification of messages rather than the average reach of a fixed set of true and false posts.

This mechanism explains why removing a single false post is necessary but insufficient to change what gets produced next. Moderation acts on an item after production, while selection pressure acts on the incentives that generate the next item. Repeated ranking outcomes become production rules. The relevant harm is, therefore, not limited to whether one user believes one post; it also includes a changing information supply in which misleading messages may become more numerous, more polished, and easier to encounter.

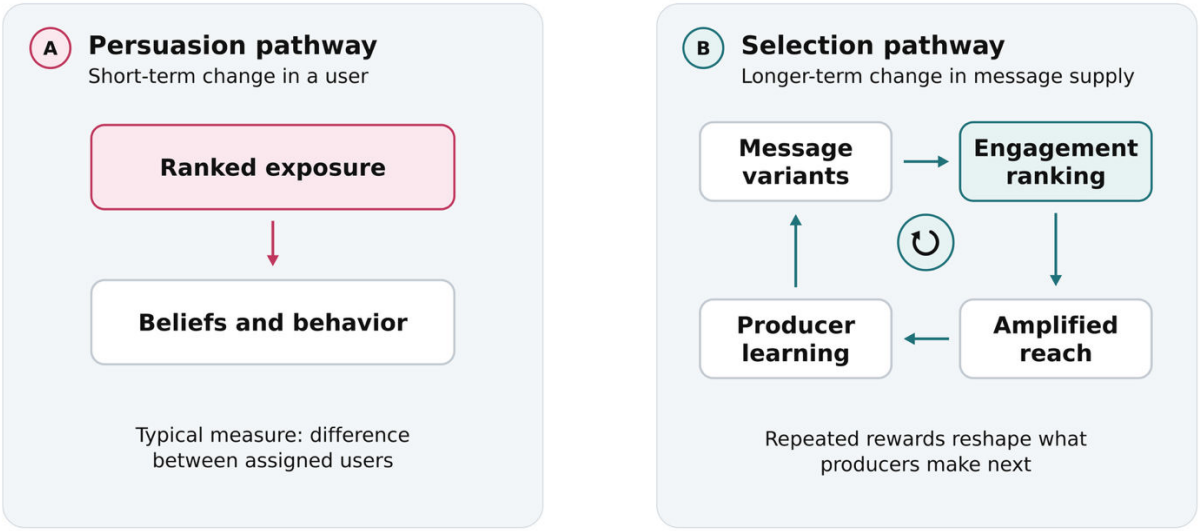


Figure 1. Two effects of feed ranking. *Persuasion studies test whether exposure changes a user. A selection-environment approach also examines how repeated ranking and feedback change the population of messages and producers over time.*

Persuasion studies tend to answer only one of two questions

The distinction between *persuasion* and *selection* changes how evidence should be interpreted. The field experiments that dominate this debate all measure persuasion. In the 2020 Facebook and Instagram experiment (a randomized experiment that switched consenting U.S. users to a reverse-chronological feed), chronological feeds changed exposure and reduced time spent on the platforms, yet produced no significant changes in several measures of political knowledge, polarization, or attitudes over three months (Guess et al., 2023). On X in 2023, assignment to the algorithmic feed increased engagement, promoted more conservative content, and shifted some issue positions in a conservative direction, although it did not change partisanship or affective polarization (Gauthier et al., 2026). Another field experiment found that down-ranking posts containing anti-democratic attitudes and partisan animosity reduced out-party animosity, while up-ranking them increased it (Piccardi et al., 2025).

These studies rule out a simple claim that ranking either always persuades or never persuades. Effects depend on the platform, ranking system, available content, outcome, and period studied. That is valuable knowledge, but not a test of how ranking changes the message population over months or years. Even a chronological-feed group still draws from a supply of posts created by producers who have learned inside the platform's existing reward structure.

Selection effects also operate at a different unit of analysis. Persuasion studies usually compare individuals, while a selection study tracks changes among messages and producers, as the outrage-feedback result of Brady et al. (2021) begins to do. It asks whether certain claims reproduce faster, whether successful framings mutate into new variants, and whether producers change their output after

receiving distribution. It also separates a post's immediate treatment from the history that produced the pool it was drawn from. Randomizing feed order today does not erase yesterday's rewards from the messages being ranked.

Time scale matters as well. Short experiments with adults cannot, by themselves, answer questions about cumulative exposure, producer adaptation, or developmental effects. Those possibilities should be treated as open empirical questions, not inferred from either null or positive short-term results. This is not an argument to discount field experiments. It is an argument to align the design and unit of analysis with the causal process being claimed.

Measure and redesign the selection environment

Audits should compare identical content pools under engagement, chronological, and accuracy-sensitive ranking. Outcomes should include the prevalence and reach of low-credibility claims, the concentration of attention among repeat misinformation producers, and changes in a producer's next posts after unusually high or low distribution. Researchers could estimate a message reproduction rate by tracking how often a claim or framing reappears after amplified exposure. Experiments can vary the feedback shown to producers as well as the feed shown to consumers.

Platforms can support this work by providing qualified independent researchers with privacy-protecting access to ranking exposures, content histories, and producer-level feedback. They should also report selection audits as system-level risk assessments.

Why require selection audits, and of whom

Selection effects are invisible in the data platforms now publish. Transparency reports count how much violating content users saw, not which attributes the ranking system rewarded or whether those rewards changed what producers made next. Outside researchers cannot reconstruct it, because the comparison requires running one content pool under alternative ranking rules, which only the operator can do. When an effect can be measured only by the party with the least incentive to measure it, a disclosure rule is what makes the evidence exist, and reporting a comparison is not the same as certifying which claims are true.

The obligation should attach to scale and ranking design rather than to speech. Selection pressure requires prediction-based ranking, attention scarce enough that ranking determines reach, and feedback fast enough for producers to learn from it. Services that display posts chronologically, or that are too small for ranking to determine who sees what, meet none of these conditions. A self-hosted forum or volunteer-run server runs no ranking model to compare against alternatives, so an audit would return no finding while its fixed costs in legal review and reporting fell on an operator with no compliance staff. The EU Digital Services Act draws this line, reserving risk assessment, independent audit, and researcher data access for platforms designated as very large, above 45 million average monthly users in the Union, and exempting micro and small enterprises, those with fewer than 50 employees and under 10 million euros in annual turnover (European Parliament & Council of the European Union, 2022).

Scope matters because compliance costs have second-order effects. In the weeks after the General Data Protection Regulation took effect, for instance, websites cut third-party technology vendor use by about 15 percent while concentration in that vendor market rose about 17 percent, as large providers absorbed compliance costs more easily than small ones (Johnson et al., 2023). The opposite failure is equally instructive. The same regulation's researcher access provision took three years and a delegated act to implement, and early applicants report slow review and frequent rejection (Pierri et al., 2026), burden without the evidence that justified it.

Interventions and safeguards

Interventions should address rewards as well as removals. One option is to incorporate credibility or accuracy signals into ranking and test the results against engagement-only and chronological baselines. Another is to increase the reliable supply available in feeds. In a field experiment in France and Germany, asking people to follow balanced news accounts improved political knowledge and the ability to distinguish true from false claims (Altay et al., 2025). The reranking experiment by Piccardi et al. (2025), likewise, shows that small changes to distribution can alter exposure and attitudes without deleting lawful speech.

These interventions require safeguards. Accuracy signals can reproduce institutional bias, and a credible audit must examine errors across languages, communities, and political groups. Researchers should preregister outcome measures and publish null results. Platforms should test for displacement, since producers who lose reach in one format often adopt another. The aim is not to create a single authorized account of reality. It is to discover which selection rules improve an information environment's accuracy, diversity, and resilience without treating attention as the only value.

Conclusion

Feed algorithms have at least two distinct effects. They can change what a person sees and, in some settings, what that person believes or does. They also allocate rewards among competing messages and producers. The first is a persuasion effect, while the second is a selection effect that can alter the information supply over time.

Evidence from Facebook, Instagram, and X shows why this distinction is needed. Short-term persuasion effects are neither universal nor absent. They vary with design and context. Meanwhile, research on false-news diffusion, outrage, and social feedback identifies features that can help misleading content win repeated competitions for attention. Those findings do not prove that every engagement-ranked system necessarily increases misinformation. They establish a mechanism that can be measured and governed.

Misinformation research should therefore add ecosystem outcomes to its standard individual outcomes. Platforms should expose enough data to reconstruct what their systems reward. Regulators should require counterfactual selection audits, scoped to the systems that can generate selection effects. Designers should test ranking objectives that value credible information rather than relying only on post-by-post moderation. The guiding questions are complementary: who is persuaded now, and what information environment is being selected for later? Answering both would move debate beyond the false choice between algorithmic omnipotence and algorithmic innocence.

Bibliography

Altay, S., Hoes, E., & Wojcieszak, M. (2025). Following news on social media boosts knowledge, belief accuracy and trust. *Nature Human Behaviour, 9*(9), 1833–1842. https://doi.org/10.1038/s41562-025-02205-6

Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances, 7*(33), Article eabe5641. https://doi.org/10.1126/sciadv.abe5641

DeVito, M. A. (2017). From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism, 5*(6), 753–773. https://doi.org/10.1080/21670811.2016.1178592

European Parliament & Council of the European Union. (2022). Regulation (EU) 2022/2065 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act). *Official Journal of the European Union, L 277*, 1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

Gauthier, G., Hodler, R., Widmer, P., & Zhuravskaya, E. (2026). The political effects of X's feed algorithm. *Nature, 652*(8109), 416–423. https://doi.org/10.1038/s41586-026-10098-2

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., … Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science, 381*(6656), 398–404. https://doi.org/10.1126/science.abp9364

Johnson, G. A., Shriver, S. K., & Goldberg, S. G. (2023). Privacy and market concentration: Intended and unintended consequences of the GDPR. *Management Science, 69*(10), 5695–5721. https://doi.org/10.1287/mnsc.2023.4709

McLoughlin, K. L., Brady, W. J., Goolsbee, A., Kaiser, B., Klonick, K., & Crockett, M. J. (2024). Misinformation exploits outrage to spread online. *Science, 386*(6725), 991–996. https://doi.org/10.1126/science.adl2829

Piccardi, T., Saveski, M., Jia, C., Hancock, J. T., Tsai, J. L., & Bernstein, M. S. (2025). Reranking partisan animosity in algorithmic social media feeds alters affective polarization. *Science, 390*(6776), Article eadu5584. https://doi.org/10.1126/science.adu5584

Pierri, F., Araujo, T., Kruikemeier, S., Lorenz-Spreen, P., Vanden Abeele, M. M. P., Vandenbosch, L., Gonçalves-Sá, J., & Grabowicz, P. A. (2026). Research opportunities and challenges of the EU's Digital Services Act. *Communications of the ACM*, *69*(9), 40–43. https://doi.org/10.1145/3779412

Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Funding
No funding has been received to conduct this research.

Competing interests
The author declares no competing interests.

Copyright
This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.