

# MANIPULATION TECHNIQUES

A Guide to Common Manipulation Techniques in Online Spaces



OVERVIEW

# Executive Summary

For many in the field of violence prevention, stopping the spread of hateful narratives is of utmost importance. As many as [83% of students encounter hate online](#) and [52% of Americans have experienced harassment](#) on social media. While tracking these narratives is essential, building resilience to them is an outcome that must be pursued at scale. PERIL's ongoing research on prebunking, which finds that identifying the manipulation techniques used in harmful narratives is an effective way of preventing their spread, is one way of achieving this goal. As such, this guide is designed to help prevention researchers, advocates, and organizations identify and understand manipulation techniques, enabling them to create their own evidence-based prebunking messages.

The rapid spread of hateful narratives across social media presents a unique challenge – traditional, largely reactive measures like content moderation, fact-checking, and debunking struggle to keep pace and yield results. To address this challenge, the [Polarization & Extremism Research & Innovation Lab \(PERIL\)](#) at American University uses “prebunking” to teach audiences to recognize common manipulation techniques associated with hateful narratives before they encounter them. This document provides a brief primer on some of the key manipulation techniques that PERIL commonly finds in its ongoing research on hateful narratives and manipulation tactics – or “codebooking.” As documented in PERIL’s work on [secondary harms in the wake of crisis events](#), predictable or common manipulation techniques that underpin hateful narratives tend to fall into four distinct (but often co-present) buckets:

| Emotional Manipulation                                                                                                                                                                                           | Identity Grievance Manipulation                                                                                                         | Information Manipulation                                                                                                          | “Just a Joke”/Trolling Manipulation                                                                                            |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| Used to elicit emotional responses from intended audiences – most often by evoking negative emotional reactions to circumvent critical thinking (e.g., fear, anger, sadness, frustration, shame, disgust, etc.). | Pits groups of people against each other to create further polarization, hate, and violence, especially against vulnerable populations. | Used to present misinformation, disinformation, malinformation, or other faulty forms of information and reasoning as legitimate. | Stirs engagement or lowers defenses by framing otherwise inflammatory content or hateful narratives as a “joke” or “trolling.” |

## RECOGNIZING MANIPULATION AND BUILDING RESILIENCE

Recognizing these types of manipulation techniques is a vital step in building resilience to hateful narratives by informing prebunking messages. PERIL uses the above “buckets” of manipulation techniques as the basis of its four-part integrated prevention pipeline:

| Monitor                                                                                                     | Understand                                                                                                                                | Prebunk & Test                                                                                                                      | Scale & Sustain                                                                                                                           |
|-------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| Track hateful narratives across mainstream and fringe platforms and identify new narratives as they spread. | Document hateful narratives, tropes, and the manipulation tactics behind them on an ongoing basis and produce timely briefs for partners. | Develop short prebunking videos and test them in randomized trials to evaluate their effectiveness and screen for backfire effects. | Expand the reach of and disseminate tested prebunking videos through trusted creators, AI tools, physical exhibits, and partner training. |

Recognizing how common manipulation techniques work alongside emerging and persistent hateful narratives is essential for addressing the spread of hate – and therefore the violence associated with it – long-term.

# 01 Emotional Manipulation

## SUMMARY

Emotional manipulation is used to elicit emotional responses from intended audiences – most often by evoking negative emotional reactions (e.g., fear, anger, sadness, frustration, shame, disgust, etc.). Often, the aim of this is to override critical thinking through evocative language, imagery, and tone. When this technique is used, the primary focus is on the exploitation of feelings as opposed to fact-based arguments.

## COMMON TYPES OF EMOTIONAL MANIPULATION

|                                                                                                                                                                                                                                                                                                                                            |                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p><b>Dead Cat Strategy</b></p> <p>Introducing a shocking or absurd claim to divert attention from a more damaging story.</p> <p><b>Example:</b> <i>A political figure makes a deliberately outrageous post the same day that a damaging investigative report is released.</i></p>                                                         | <p><b>Emotional Hijacking</b></p> <p>Content engineered to trigger strong emotional responses (e.g., outrage, fear, disgust) that override critical thinking; emotionally aroused audiences are more likely to share without verification.</p> <p><b>Example:</b> <i>Fabricated atrocity images are posted with inflammatory captions to provoke moral outrage and rapid sharing.</i></p>                                           |
| <p><b>FOMO Exploitation</b></p> <p>Creating artificial urgency or fear of being "left behind" to pressure targets into rapid, unreflective action.</p> <p><b>Example:</b> <i>Extremist recruiters telling potential recruits or supporters that "the revolution is happening now – don't be on the wrong side of history."</i></p>         | <p><b>Moral Targeting</b></p> <p>Identifying and exploiting the specific moral foundations (e.g., care, fairness, loyalty, authority, sanctity) that resonate with target audiences, then framing propaganda accordingly.</p> <p><b>Example:</b> <i>Propaganda targeting loyalty-oriented audiences frames dissent as betrayal; propaganda for care-oriented audiences frames the same issue through victimhood narratives.</i></p> |
| <p><b>Moral Outrage</b></p> <p>This tactic appeals to righteous indignation and frames issues as fights between good and evil. This is often an extreme version of more generalized emotional manipulation or hijacking.</p> <p><b>Example:</b> <i>Suggesting that a group of people are inherently evil/Satanic or anti-American.</i></p> | <p><b>Outrage Farming</b></p> <p>Deliberate creation of polarizing content to generate maximum engagement through anger, exploiting algorithmic amplification of high-engagement content.</p> <p><b>Example:</b> <i>A page posts a fabricated inflammatory story about a cultural flashpoint, knowing the ensuing outrage will drive massive engagement.</i></p>                                                                    |

## RESPONSE RECOMMENDATIONS

Countering emotional manipulation techniques requires messages that call out specific manipulative emotions (e.g., fear, anger, sadness, etc.), promote checking for sources and evidence, prioritize critical thinking over emotional responses, and use preventative education where possible. For a specific example of how emotional manipulation was used in the wake of a mass violence crisis, see PERIL's brief on [manipulation techniques in the aftermath of the 2026 Tumbler Ridge shooting](#).

## 02 Identity Grievance Manipulation

### SUMMARY

Identity grievance manipulation techniques pit groups of people against each other to create further polarization, hate, and violence, especially against vulnerable populations. This type of hate mostly focuses on villainizing certain races, genders, religions, and sexualities.

### COMMON TYPES OF IDENTITY GRIEVANCE MANIPULATION

#### Audience Segmentation and Micro-Targeting

Using data analytics to divide a population into psychographic segments, then delivering tailored propaganda to exploit specific fears, values, or grievances.

**Example:** Cambridge Analytica used Facebook data to [build psychographic profiles](#) and serve customized political ads calibrated to personality traits.

#### Grievance Exploitation

Identifying real or perceived injustices and framing them to direct anger toward designated enemies, positioning the extremist group as the only willing solution.

**Example:** Far-right recruiters frame economic decline as caused by immigration rather than structural change, offering their movement as the answer.

#### Identity Appropriation

Creating fake personas claiming membership in a specific demographic group to manipulate discourse by appearing to speak from within that community while saying things that other communities will find either reprehensible or artificially agreeable.

**Example:** The Russian Internet Research Agency operated accounts [posing as Black American activists](#) to sow division and amplify racial tensions.

#### Scapegoating

The systematic attribution of complex social problems to a specific outgroup, simplifying complicated causality into a villain narrative that focuses grievance on that outgroup and justifies hostility against them.

**Example:** Economic hardship caused by global financial systems is consistently misattributed to the Jewish community, providing a concrete target for diffuse anxiety.

#### Us vs. Them Framing (In-group/Out-group Polarization)

A systematic construction of rigid in-group/out-group boundaries through propaganda emphasizing existential threat and dehumanization.

**Example:** Extremist content frames opponents not as fellow citizens but as existential threats, using dehumanizing language that portrays compromise as surrender.

#### Wedge Driving

Identifying and exploiting existing social, political, or cultural fault lines, amplifying both sides to deepen polarization and weaken social cohesion.

**Example:** Russian operations during the 2016 U.S. presidential election simultaneously [operated right and left-wing Facebook pages](#) to maximize polarization.

### RESPONSE RECOMMENDATIONS

Since many of these manipulation techniques rely on the unjust targeting of certain groups, responses should first and foremost denounce and reject such targeted hate and appropriately center those groups in responses. Messages should also focus on positives, such as building alliances with other communities to enhance solidarity. Focus should be shifted from divisive language toward language that emphasizes a shared sense of community and common goals – language that is deliberate and inclusive. Messages should also focus on empathy and trust, rather than stereotyping. Education on social cohesion and prevention of hate should be promoted as well, such as PERIL's [Bystander Intervention Playbook](#).

# 03 Information Manipulation

## SUMMARY

Information manipulation techniques are used to present misinformation, disinformation, malinformation, or other faulty forms of information and reasoning as legitimate. For information manipulation, teaching recognition of the strategies themselves rather than trying to determine the intent of those spreading the information is often prudent.

## COMMON TYPES OF INFORMATION MANIPULATION

### Conspiracy Theory

Claims that events or situations are secretly manipulated by powerful, hidden groups. The technique uses the idea of being “in on a secret” to create false closeness and a sense of power and knowledge. Conspiratorial logic often relies on narrative to replace actual evidence.

**Example:** “[Pizzagate](#)” conspiracy that suggested there was an elite sex-trafficking ring being operated by the Clinton family out of a pizzeria in Washington D.C.

### Deepfake Deployment

Use of AI-generated synthetic media (e.g., video, audio, images) to fabricate events or statements attributed to real individuals.

**Example:** A fabricated video of a candidate making inflammatory statements is released days before an election, timed so debunking cannot reach all audiences.

### Echo Chambers

Information environments where users are predominantly exposed to belief-reinforcing content, produced by the interaction of algorithms, platform design, and self-selection.

**Example:** A user watching videos skeptical of mainstream media is recommended increasingly extreme content, constructing a purely conspiratorial media diet.

### False Dilemmas

Presents two options as if they are the only two possibilities, while ignoring other options. This works to narrow the perception of options.

**Example:** Often used to drive partisan polarization, with messages like either you support a specific candidate or you support the downfall of America.

### Faulty Assertions

Logical fallacies, false equivalencies, pseudoscience, and/or other faulty forms of reasoning are used to artificially prop up a narrative.

**Example:** Misrepresenting or fabricating scientific data to reject climate or medical science.

## RESPONSE RECOMMENDATIONS

Given that information manipulation is based on faulty reasoning and evidence, one of the strongest lines of defense is fostering critical thinking skills in an audience. For example, messaging that counters faulty assertions typically identify the fallacies/faulty logic at play, provide counter evidence, focus on arguments, not actors, and educate about how to engage in more sound logic and arguments. Conspiracy theories can be reduced by bolstering critical thinking while also engaging in empathetic communication. False dilemmas often require providing alternative options and introducing nuance that moves beyond either-or dichotomies. All forms of information manipulation can also be addressed with digital literacy and critical thinking educational programs, as well as targeted psychological resilience measures (e.g., PERIL’s prebunking videos and digital literacy programs, like [DUCC](#)).

## 04 “Just a Joke”/Trolling Manipulation

### SUMMARY

This type of manipulation stirs engagement and creates emotional responses through framing otherwise inflammatory content or hateful narratives as a “joke” or “trolling.” Often, the intent of the users is unclear, as posts are framed as innocent questions, such as in the Just Asking Questions trope, or purposely misleading, such as in the Rage-bait trope – the goal being increased engagement to further spread content or narratives.

### COMMON TYPES OF “JUST A JOKE”/TROLLING MANIPULATION

#### Fight-Picking

This tactic provokes or challenges audiences into conflict and is often used to increase engagement or emotional intensity of a message or situation and encourage reactive responses instead of thoughtful conversation.

**Example:** Starting a post with “Hey, idiot liberals/conservatives” or the like.

#### Hashtag Hijacking

Co-opting a trending or popular hashtag by flooding it with unrelated or adversarial content, disrupting the original conversation or associating it with a different message.

**Example:** [#DoctorsSpeakUp](#) was hijacked by anti-scientific counter publics to spread anti-vaccination messaging.

#### Just Asking Questions (JAQing Off / Sealioning)

Framing disinformation as innocent questions to avoid accountability while introducing misleading premises. Related to sealioning: persistent, polite demands for evidence designed to exhaust opponents.

**Example:** An account repeatedly posts “has anyone actually verified that [documented event] happened the way they say?” across forums.

#### Rage-Bait

Content specifically designed to provoke anger through misleading framing, selective editing, or fabrication to drive engagement, sharing, and polarization.

**Example:** A manipulated video clip removes context to make a public figure appear offensive, generating millions of angry shares before context emerges.

#### Trolling (Strategic)

Deliberate posting of inflammatory content to disrupt discussion, provoke emotional reactions, or waste opponents' energy; distinguished from casual trolling by coordination and strategic purpose.

**Example:** Coordinated teams target policy discussion forums with inflammatory comments to derail constructive conversation.

### RESPONSE RECOMMENDATIONS

Because online actors rely on engagement to spread their messaging, a lack of response is effective in countering such engagement farming. However, when a response is necessary, de-escalatory strategies are often effective since these tactics are meant to provoke audiences into conflict. This can include focusing on the arguments and messages, not the actors themselves, as well as pointing out the emotional manipulation inherent in such messages. Preemptive education against this technique is also useful, where possible. For example, PERIL has developed further resources on how to approach this type of hate through its [“Not Just a Joke” guide](#) and more general media and digital literacy programs for youth audiences through its [DUCC program](#).

# Further Resources/References

- American Psychological Association.** 2023. "Using Psychological Science to Fight Misinformation: A Guide for Journalists." <https://www.apa.org/topics/journalism-facts/psychology-misinformation-guide-journalists>
- Anti-Defamation League (ADL).** 2023. *Online Hate and Harassment: The American Experience 2023*. [https://www.adl.org/sites/default/files/pdfs/2023-12/Online-Hate-and-Harassmen-2023\\_0\\_0.pdf](https://www.adl.org/sites/default/files/pdfs/2023-12/Online-Hate-and-Harassmen-2023_0_0.pdf)
- BBC News.** 2019. "Russian Trolls' Chief Target Was 'Black US Voters' in 2016." <https://www.bbc.com/news/technology-49987657>
- Benkler, Yochai, Robert Faris, and Hal Roberts.** 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.
- Bradshaw, Samantha, and Philip N. Howard.** 2019. "The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation." Working Paper 2019.3. Oxford, UK: Project on Computational Propaganda, Oxford Internet Institute.
- Christie, Christina, Anne Gilliland, Christine Ong, Wendy Slusser, Mark Hansen, Arif Amlani, Tyrone Howard, Mark Berkman, Sarah Krongard, Seul Lee, and Zhaopeng Ding.** 2024. "The Rise of Social Media Hate." UCLA Initiative to Study Hate. <https://studyofhate.ucla.edu/smash-social-media-hate/>
- European Commission.** "6 tips to spot and stop information manipulation." <https://ec.europa.eu/stories/6-tips/>.
- Frenkel, Sheera, and Kevin Roose.** 2018. "Zuckerberg, Facing Facebook's Worst Crisis Yet, Pledges Better Privacy." *New York Times*. <https://www.nytimes.com/2018/03/21/technology/facebook-zuckerberg-data-privacy.html>
- Holiday, Ryan.** 2012. *Trust Me, I'm Lying: Confessions of a Media Manipulator*. New York: Portfolio/Penguin.
- McCauley, Clark, and Sophia Moskalenko.** 2017. "Understanding Political Radicalization: The Two-Pyramids Model." *American Psychologist* 72(3):205–16.
- Moghaddam, Fathali M.** 2005. "The Staircase to Terrorism: A Psychological Exploration." *American Psychologist* 60(2):161–69.
- Paul, Christopher, and Miriam Matthews.** 2016. *The Russian "Firehose of Falsehood" Propaganda Model: Why It Might Work and Options to Counter It*. Santa Monica, CA: RAND Corporation.
- Phillips, Whitney.** 2018. *The Oxygen of Amplification: Better Practices for Reporting on Extremists, Antagonists, and Manipulators Online*. New York: Data & Society Research Institute.
- Polarization & Extremism Research & Innovation Lab (PERIL).** *Be the Sixth Player: A Bystander Intervention Playbook for Fans, Families & Community*. <https://perilresearch.com/wp-content/uploads/2026/07/Be-the-Sixth-Person-Complete-Playbook.pdf>
- Polarization & Extremism Research & Innovation Lab (PERIL).** *DUCC*. <https://perilresearch.com/projects/ducc/>
- Polarization & Extremism Research & Innovation Lab (PERIL).** *Mitigating Harms From Crisis Events: Mass Violence: Manipulation and Harmful Narratives in the Aftermath of the 2026 Tumbler Ridge Shooting*. <https://perilresearch.com/wp-content/uploads/2026/06/Mass-Violence-Crisis-Tumbler-Ridge-One-Pager.pdf>
- Polarization & Extremism Research & Innovation Lab (PERIL).** *Resources*. <https://perilresearch.com/resources/>
- Polarization & Extremism Research & Innovation Lab (PERIL) and the Southern Poverty Law Center (SPLC).** *Not Just a Joke: Understanding & Preventing Gender- & Sexuality-Based Bigotry*. <https://perilresearch.com/resource/not-just-a-joke-understanding-preventing-gender-sexuality-based-bigotry/>
- Quandt, Thorsten.** 2018. "Dark Participation." *Media and Communication* 6(4):36–48.
- Samuelson, Kate.** 2016. "What to Know About Pizzagate, the Fake News Story With Real Consequences." *Time*. <https://time.com/4590255/pizzagate-fake-news-what-to-know/>
- Search for Common Ground, Multi-Faith Neighbors Network, and Polarization & Extremism Research & Innovation Lab (PERIL).** 2023. *The Peacemaker's Toolkit: A Reference Guide for Reconciliation in Your Community*. <https://perilresearch.com/wp-content/uploads/2023/12/Peacemakers-Toolkit.pdf>
- Shane, Scott, and Sheera Frenkel.** 2018. "Russian 2016 Influence Operation Targeted African-Americans on Social Media." *New York Times*. <https://www.nytimes.com/2018/12/17/us/politics/russia-2016-influence-campaign.html>
- Starbird, Kate, Ahmer Arif, and Tom Wilson.** 2019. "Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations." *Proceedings of the ACM on Human-Computer Interaction* 3(CSCW): 1–26.
- Temple University Libraries.** 2026. "'Fake News,' Misinformation & Disinformation: How to Identify (and Avoid) False and Misleading Information." <https://guides.temple.edu/fakenews/recognize>
- Wardle, Claire, and Hossein Derakhshan.** 2017. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Strasbourg: Council of Europe.
- Woolley, Samuel C., and Philip N. Howard,** eds. 2018. *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. New York: Oxford University Press.
- Zhao, Bei, Wujiong Ren, Yuan He, and Hongzhong Zhang.** 2025. "Demystifying Hashtag Hijacking in the Public Opinion Game: Attention, Narratives, and Social Bots." *Information, Communication, & Society* 29(2): 634–656.