# Interventions to Counter Polarization: Lessons from the Global North and Applications to the Global South

Rachel Berwald, Joshua Goetz, Dan Harker, Connor Warshauer, Sabrina Habchi, AP Turek, and Nikita Savin

December 7th, 2023

## TABLE OF CONTENTS

## INTRODUCTION

Researchers, journalists, and policymakers frequently cite polarization—or the divergence of political attitudes in society toward ideological extremes—as a key source of political dysfunction. A growing body of scholarship argues that polarization undermines democratic norms and institutions (Graham and Svolik, 2020; Hetherington and Rudolph, 2015; Kingzette et al., 2021; McCoy and Somer, 2019; Levitsky and Ziblatt, 2018), while some work indicates that polarization can even lead to political violence (Piazza, 2023). Moreover, evidence suggests that the rise of social media in recent years has fueled political polarization by entrenching users in homogenous online communities (Bail et al., 2018; Garimella and Weber, 2017; Quattrociocchi et al., 2016). To address this issue, scholars and practitioners have sought to develop strategies for effectively mitigating polarization and its most pernicious effects.

This report reviews the literature on depolarization interventions and identifies patterns across intervention types and geographic regions. We focus on 8 loose categories of interventions: informational; contact and exposure-based; educational; mass media; identity-based; empathy and perspective based; and echo chamber breaking. The bulk of depolarization research addresses contexts in the Global North. To compensate for this lopsidedness, we supplement the review with scholarship on polarization-adjacent trends in the Global South, such as prejudice and intergroup hostility.

Our findings indicate that contact-based interventions, such as those that encourage participation in intergroup sports and nonpolitical conversations, show the most promise in mitigating polarized attitudes. Interventions that focus on education and mental exercises, however, yield mixed results. One major obstacle to the success of depolarization interventions is that subjects with deeply entrenched prejudicial attitudes, arguably the target demographic of these efforts, are often the most resistant to their effects. Moreover, many depolarization studies are constrained by small, under-representative samples. Future research should investigate how to counter prejudicial attitudes among the most polarized while drawing from as broad samples as possible.

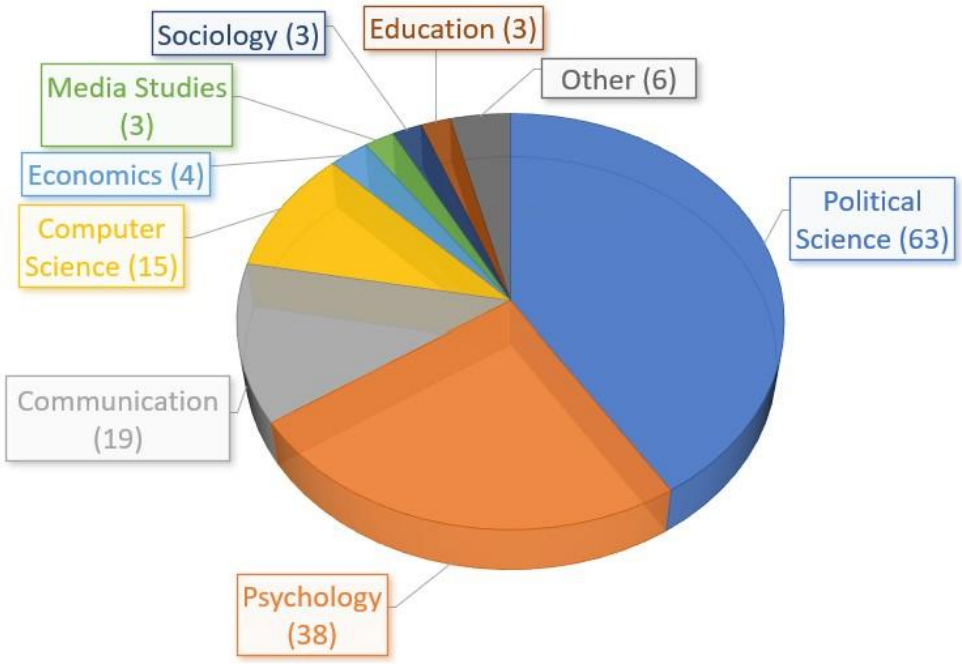
This report proceeds as follows. First, we provide a visual summary of our findings and an explanation of the search terms we used to gather relevant scholarship. We then give an overview of the findings for each intervention type, with a special focus on the distinctions between studies in the Global North and South. We then briefly summarize some of the working papers from leading scholars in the field. We conclude with a discussion and insights for further research.

**Visual Summary of the Literature:**

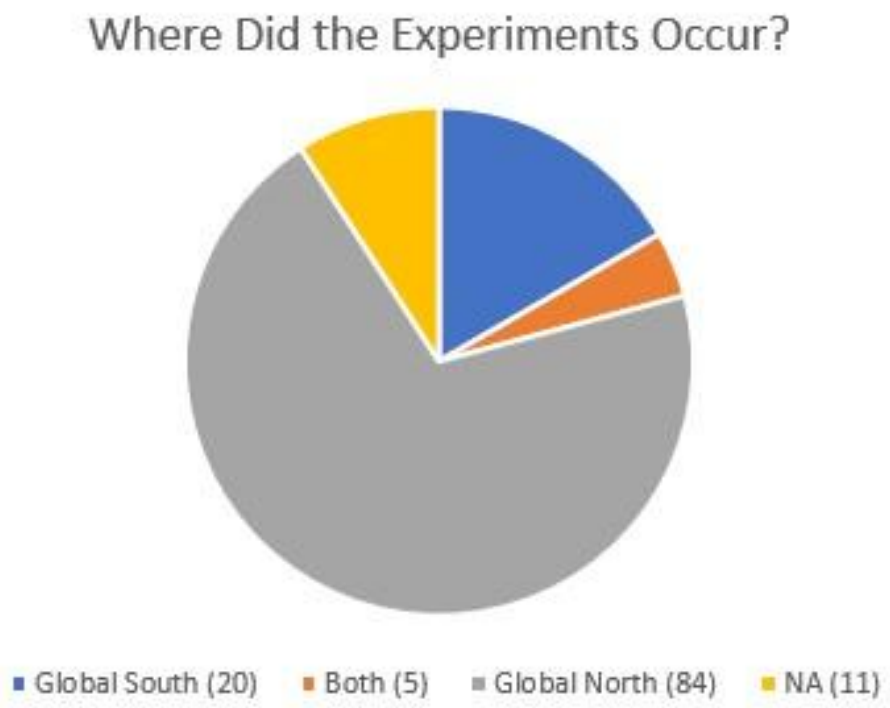
Word Cloud - The word cloud shows which keywords appeared most frequently in the surveyed literature.[1]


Disciplines covered in literature review - The relevant literature for this project primarily was published in political science or psychology journals. However, as the topic of polarization is naturally interdisciplinary, articles were drawn from journals in 11 other fields as well.
Note: "Other" includes business; computational social science; law; neuroscience; and women's, gender, and sexuality studies
Note: The total count exceeds the total number of articles because some articles were multidisciplinary

[1] This was generated on freewordcloudgenerator.com using the titles of articles in the master literature repository

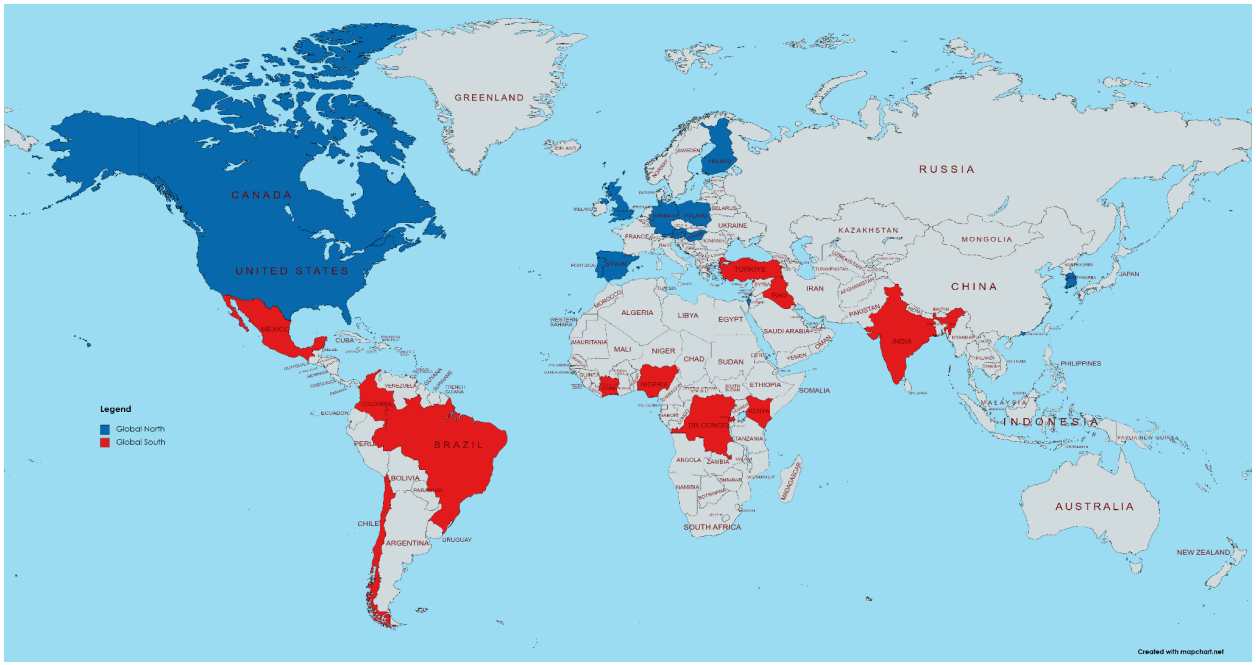



Disproportionate focus on the Global North in the literature - The experimental literature has a heavy focus on the Global North. While Europe was overrepresented compared to the Global South, the main driver of the disparity was the United States, which was the location of more than half of the studies conducted in the Global North. Note that we were specifically looking for interventions in the Global South, so the literature is probably even more skewed than this graphic suggests.




Countries examined in literature review - This map highlights countries that were the location of experiments included in the literature review. Countries in the Global South are colored in red while the Global North is in blue. While the countries in which the experiments took place appear reasonably well-distributed around the globe, this figure also highlights the dearth of studies taking place in the majority of the world's countries.[2]

[2] Image credit: This map was generated using mapchart.net




Overall, our review covers the 71 unique studies we found. Of these, 32 were successful at reducing polarization between groups, 21 had mixed results, and 17 had either null or negative results. We also subdivided the studies into eight intervention types—information-based, contact and exposure, educational, mental exercise, mass media, empathy/perspective-based, echo chamber breaking. In the following section, we discuss the search protocol we used for building our evidence base and literature review. The remainder of the paper will explain each intervention in further detail. We conclude with discussing key takeaways and potential paths forward for future research.

---

## SEARCH PROTOCOL

### *Introduction*

This search protocol outlines the procedures for creating an evidence base and literature review for the effectiveness of experimental interventions in mitigating polarization. The procedures include:

1. Search Terms
   1.1. Identify and define key terms corresponding to the topics
2. Search Process
   2.1. Databases and sources
   2.2. Use and application of search terms and boolean search operators
   2.3. Identify key (“bullseye”) articles from which to draw new search terms

2.4. Strategies and best practices
2.5. Example of information notation and a “bullseye” article

While this search protocol was developed to support a specific evidence base initiative, it is designed to be readily adapted or modified for future efforts to assemble literature and evidence for EGAP’s programming on other topics. A quick guide to do so is provided at the end. If any topics relevant to EGAP’s project are not represented in these search terms, please let us know and we will update our search protocol accordingly.

1. *Search Terms*

**1.1. Identify and define key terms corresponding to the topics:**

The topic can be broken down into four parts:

1. the issue being addressed (e.g., depolarization, prejudice, etc.)
2. the study methodology (e.g., experimental, quasi-experimental, observational, etc.)
3. the type of intervention (e.g., inoculation (pre-bunking), refutation strategy, educational intervention, etc.)
4. The measurement of the issue being addressed (e.g., animosity, trust, support for democracy, etc.)

A more comprehensive list of subtopics is shown in Table 1 (similar terms are written in parentheses). Definitions of these key terms will follow.

Table 1: List of subtopics

| **Issue** | **Methodology** | **Intervention** | **Measurement** |
|---|---|---|---|
| Depolarization<br>Bias Reduction (Prejudice Reduction)<br>Counter-extremism<br>Reconciliation<br>Conspiracy theories (hyper-partisan content)<br>Echo chambers (homophily) | Experiment (lab, lab-in-the-field, survey, online field experiment)<br>Quasi-experiment<br>RCT (random, random assignment)<br>Vignette (hypothetical, imagined<br>Observational study | Information manipulation<br>Framing<br>Priming<br>Social Identity Salience<br>Online intervention<br>Network-based<br>Nudge<br>Refutation Strategy<br>Boost<br>Educational interventions<br>Accuracy prompt<br>Debunking (correcting)<br>Friction<br>Inoculation (pre-bunking) | Animosity (Partisan animosity)<br>Happiness<br>TRI framework<br>Affective response (Affect, emotional response)<br>Trust (Trust game, Trust ratings, Trust scores)<br>Political efficacy<br>Tolerance<br>Civility<br>Prejudice (Bias, Hate)<br>Attitudes<br>Romantic desire |

| | | | |
|---|---|---|---|
| | | Lateral reading<br>Media-literacy tips<br>Rebuttals of science denialism<br>Self-reflection tools<br>Social norms<br>Warning labels (fact-checking labels, context labels, context tags, credibility labels, credibility tags)<br>Platform alteration<br>Politician messaging<br>Retractions<br>Perspective-taking<br>Gameplay | Belief in… (Belief in info post-treatment / Rumor acceptance)<br>Reliance on…<br>Support for… (...democracy, …violence)<br>Competence (understanding potentially of contemporary political issues)<br>Threat perception<br>Sharing (likelihood to share)<br>Manipulativeness (susceptibility, persuasion)<br>Reasoning skills (critical thinking, digital literacy, political consciousness)<br>Confidence (certainty)<br>Cooperation (Agreement)<br>Dictator Game<br>BIAT (IAT)<br>Judgments<br>Feeling thermometers<br>Trait ratings<br>Selection tasks<br>Perceptions (misperceptions)<br>Assessment (evaluation)<br>Likert Scale<br>True/False<br>Resistance to treatment<br>Identification of (...misinformation, …propaganda)<br>Backfire effects |

These are the definitions/explanations of key terms:

Issues:

- **Polarization –** Broadly defined as a divergence in attitudes. Polarization can include party-based, affective polarization, partisan, etc.
- **Bias Reduction (Prejudice Reduction) -** Methods of reducing the prejudice, bias, or negative feelings that one segment of society feels towards another.
- **Counter-extremism -** Methods of combating extremist ideology and preventing its spread
- **Reconciliation -** Building trust, peace, and a sense of normalcy between groups that have formerly been in conflict
- **Conspiracy theories (hyper-partisan content)** – An improbable explanation of events, usually for political ends
- **Echo chambers (homophily)**-- an environment (usually online) where similar ideas are reinforced and opposing ideas are not introduced or discussed

Methodology:

- **Experiment (lab, lab-in-the-field, survey, online field experiment) -** A scientific research design relying on random assignment and treatment of certain units.
- **Quasi-experiment -** A research design that meets many of the criteria of an experiment, but in which the researcher does not have full control over the assignment or treatment
- **RCT (random, random assignment) -** A randomized controlled trial is a way to test an intervention or treatment, by randomly dividing the units being studied into treatment and control groups, and subjecting the treatment group to an intervention
- **Vignette (hypothetical, imagined) -** A short story about hypothetical characters in prescribed circumstances to which interviewees are invited to respond.
- **Observational study -** Involves systematically observing and recording behaviors, events, or phenomena as they occur naturally, without direct intervention or manipulation by the researcher.

Intervention Type:

- **Information manipulation (framing?)** ***an intervention which implies selecting some aspects of a perceived reality and making them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation and/or treatment recommendation for the item described (Entman 1993:52)*****)**
- **Priming** - Exposing a subject to a stimulus that influences her subsequent thoughts, feelings, or behaviors.
- **Social Identity** – The attachment an individual has to a group they belong to. This can be gender, religion, race, ethnicity, etc.
- **Salience** – The relative importance of one aspect of an individual's identity over another
- **Online intervention** - An intervention that takes place via an online interface (e.g., via a survey or on social media) rather than in person.
- **Network-based -** A study that uses or changes the characteristics of social networks to influence outcomes or behaviors.
- **Nudge**

- **Refutation Strategy -** A strategy designed to discredit false or misleading information.
- **Boost** - A deliberate manipulation or intervention introduced during an experiment to amplify or enhance a particular effect, treatment, or condition being studied.
- **Educational interventions**– These can include media literacy workshops, health literacy, or any other intervention that aims to inform people about a given topic or teach a certain skill
- **Accuracy prompt** - An intervention designed to shift social media users' attention toward the accuracy of the online content they are consuming.
- **Friction -** An intervention designed to compel social media users to pause and consider the veracity of news headlines before sharing them.
- **Lateral reading -** A practice in which a reader searches for information about a source's credibility and accuracy while she is reading it.
- **Media-literacy tips -** Designed to enhance subjects' ability to consume media in a thoughtful and critical way.l
- **Rebuttals of science denialism**
- **Self-reflection tools -** Mechanisms designed to encourage research subjects to consider how they engage with online content and how that content affects them.
- **Social norms -** Widely shared attitudes and beliefs about acceptable standards of behavior in online contexts.
- **\Warning labels (fact-checking labels, context labels, context tags, credibility labels, credibility tags)**
- **Platform alteration**
- **Politician messaging -** Online content generated directly by political elites.
- **Retractions**
- **Perspective-taking** - Understanding a situation or viewing a concept through the point of view of someone else
- **Gameplay**

Measurement

- **Animosity (Partisan animosity)** - The degree to which subjects experience negative feelings, hostility, or antagonism toward individuals or groups based on their differing political affiliations, beliefs, or ideologies.
- **Happiness** - The degree to which subjects experience an emotional state characterized by positive feelings, life satisfaction, and an overall sense of well-being.
- **TRI framework -** Developed in Hartman, et al., (2022), TRI is a broad framework for measuring multiple outcomes. It stands for thoughts (related to correcting misperceptions or finding commonalities), relationships (encouraging dialogue or contact), and institutions (related to public discourse and political structures).
- **Affective response (Affect, emotional response)** - A subject's emotional response to an intervention.
- **Trust (Trust game, Trust ratings, Trust scores) -** A subject's willingness to become vulnerable to another party under the presumption that the party will act in the subject's best interest.
- **Tolerance -** A subject's willingness or acceptance of diverse beliefs, opinions, cultures, or lifestyles.

- **Civility** - A subject's willingness to adhere to norms of politeness, consideration, and cooperation.
- **Prejudice (Bias, Hate) -** Negative opinions or assumptions about someone based on preconceived notions or group membership
- **Attitudes** - A subject's thoughts or feelings about some entity. This can include support for governmental policies or systems, the extent to which these thoughts or feelings can be manipulated through interventions and the level of confidence (how strong or weak) that exists.
    - **Belief in… (Belief in info post-treatment / Rumor acceptance)**
    - **Reliance on…**
    - **Support for… (...democracy, …violence)**
    - **Confidence (certainty)**
    - **Perceptions (misperceptions)**
    - **Assessment (evaluation)**
    - **Threat perception**
- **Perceptions** – The interpretation of real or imagined stimuli, which can include stimuli perceived to be harmful (threatening).
- **Romantic desire** - A subject's emotional and cognitive inclination or longing for a deep, affectionate, and intimate relationship with another person.
- **Competence (understanding) -** The extent to which a subject demonstrates knowledge, proficiency, or a skill in a specific domain.
- **Sharing (likelihood to share)** - A subject's propensity to share content.
- **Manipulativeness (susceptibility, persuasion)** - A subject's receptivity to an intervention or new content.
- **Reasoning skills (critical thinking, digital literacy)** - A subject's ability to analyze, evaluate, and synthesize information from various sources, make logical judgments, and effectively navigate and comprehend digital content.
- **Cooperation (Agreement)** - A subject's willingness to comply with others in working toward a common goal.
- **Dictator Game** - A classic experimental paradigm where one participant, the "dictator," is given the power to unilaterally distribute a sum of resources between themselves and another participant reflecting altruism and fairness tendencies.
- **BIAT (IAT)** - A psychological tool used to measure the strength of automatic and unconscious associations between concepts in a person's mind, often revealing implicit biases or attitudes that may not be readily expressed through explicit self-report measures.
- **Judgments** - Cognitive assessments or evaluations subjects make often involving the application of personal attitudes and beliefs.
- **Feeling thermometers** - A visual scale that allows respondents to rank their attitudes on a given subject from "cold" (disapproving, negative) to "hot" (approving, positive).
- **Trait ratings -** The systematic evaluation or assessment of an individual's enduring personality characteristics, behaviors, or attributes based on standardized scales or measures.
- **Selection task** - A cognitive psychology task used to study reasoning and decision-making, involving a logical rule where participants must choose cards to confirm or refute a conditional statement.

- **Likert Scale** - A Likert scale is a commonly used measurement tool in social science research that presents a series of statements or items for participants to rate their level of agreement or disagreement, typically using a range of response options (e.g., "strongly agree," "agree," "neutral," "disagree," "strongly disagree").
- **True/False -** A measurement in which respondents indicate whether a statement is true or false.
- **Resistance to treatment** - The phenomenon where subjects do not respond as expected to an intervention, potentially hindering the desired outcome or change.
- **Backfire effects** - Instances in which interventions have the opposite effects intended or expected by the researchers.

### 1.2. Fruitful search terms

We have done some preliminary searches with some of the terms above. Below, we report the most successful terms, along with estimates of how many relevant articles each of these terms turn up.

- "Polarization" (Google Scholar):
- "Depolarization" AND "Politics" (Google Scholar):
- What kinds of interventions help mitigate polarization? (Elicit)

Based on this preliminary search, it appears that there will be no shortage of articles on the topic, and thus we will probably not need to cast a very wide net in order to find relevant literature.

## 2. *Search Process*

### 2.1. Database and Sources

In conducting our search, we use the following list of databases and search engines:

- Proquest
- WorldCat
- Usa Gov
- Microsoft Academic
- Bielefeld Academic Search Engine (BASE)
- CORE
- Semantic Scholar
- Homeland Security Digital Library
- The Library of Congress
- SAGE Knowledge
- Google Scholar
- Women Also Know Stuff (WAKS)
- World Bank reports
- Elicit
- Litmaps
- ResearchRabbit

- JSTOR
- Pre-registered studies from: EGAP, AEA, JPAL

### 2.2. Use and application of search terms and Boolean search operators

The search terms we used were taken from our foundational understanding of the literature and brainstorming conducted as a team. As our search progressed, we developed our own search terms taken from relevant articles and the lists of alternative terms (Section 1.2). Table 1 (in Section 1.2) displays the most effective search terms we used and their alternatives, divided by issue, methodology, and intervention.

The search engine used affected the number of relevant results we received, with certain terms working more effectively on one engine than on others. This was discovered through trial and error. When a string of search terms was unproductive in one search engine, we tried the same phrase in a different search engine and, at times, received more valuable results. Furthermore, some search engines (e.g. Google Scholar) display results from a myriad of sources and search engines while others (e.g. Oxford Handbooks) only show results from a single publisher. The former would inherently gather more results than the latter.

We added Boolean operators and modifiers to our search terms to achieve more focused hits during our search. These Boolean operators and modifiers can be used in combination with each other to increase the specificity of results.

a. **AND**: when used in a search, results must include both terms
   i.e. depolarization AND experiment
b. **NOT**: when used in a search, results must exclude a specific word or phrase
   i.e. depolarization NOT experiment
c. **OR**: when used in a search, results must include either term
   i.e. depolarization OR experiment
d. **“__”**: when used in a search, results must include the exact phrase in the quotations
   i.e. “depolarization,” “experiment”
   i. Note: This modifier was particularly helpful during our search
e. **(__)**: groups statements together
   i.e. (depolarization OR misinformation) AND experiment

We conducted our search using these multiple strategies. Since most search engines automatically order results by relevance to the search terms, we only looked at the results from the first two to three pages. of search results. Of these results, we took note of the articles that were at least partially relevant to the project themes. For each relevant article, we recorded the following information: author(s), title, year of publication, journal/publisher, discipline, topic, method, intervention type, independent and dependent variables, location of study, the success of an intervention, and the relevance to the project, (which we denoted as a “bullseye.”). We also noted the exact search terms and search engines used for each relevant article found.

### 2.3. Identify key relevant (“bullseye”) articles to draw new search terms

“Bullseye” pieces are articles that focus on Polarization and employ an experimental intervention. These key parts are key pieces that are directly relevant to our search and can be used to generate additional search terms and phrases. We used the following list to identify our “bullseye” pieces, arranged according to their priority:

- **Issue being addressed:** focuses on polarization
- **Methodology:** experimental interventions are our highest priority
- **Region:** we have a preference for global south studies, but are also very interested in any northern studies that have generalizable findings. As a priority, region is subordinate to methodology and issue

If an article is a “bullseye” piece:

- The article’s keywords are used to generate alternative search terms for Table 1.
- The article’s bibliography is systematically reviewed for other relevant pieces.

If an article is relevant but not a “bullseye” piece:

- It is included in the repository, but it does not alter the search process.

**2.4. Strategies and best practices**

We conducted a straightforward and replicable search protocol, utilizing a combination of the different databases. After selecting a database, we used varying levels of specification, and Boolean operators and modifiers to find relevant articles. The following are the levels of specification we used:

**1.2.4.1. No Specification**: To begin, we cast a wide net to gather a large number of relevant articles by conducting a broad search of the main themes targeted in this project (e.g. depolarization and experiments).

Example search: depolarization, experiments

**2.4.2. Single Specification:** Next, we began narrowing our search by specifying either a type of intervention or a specific region in combination with one of the main themes as a whole.

Example search: depolarization, nudge
Example search: polarization, Sub-Saharan Africa

While we knew this strategy would not produce articles as relevant as those produced by dual specification, discussed next in point 2.4.2., this single specification was done to view all the possible effects of a specific type of cross-party program or all the possible approaches toward collective action outcomes.

**3.2.4.2. Dual Specification:** Finally, we specified our search by narrowing our search terms down to a particular type of intervention in combination with a specific region or measurement.

Example search: depolarization, inoculation, Latin America

When we did not get many relevant hits, we would either add more or adjust the search terms, adjust the Boolean operators and modifiers, or conduct an identical search using a different search engine. This combination of strategies allowed us to continue finding relevant articles.

### 2.5. Example of Information notation and a "Bullseye"

As an example, we will look at the article *Exposure to opposing views on social media can increase political polarization* by Chrisopher Bail, et al. which was found using Elicit. This article was published in PNAS in 2018. It takes place in the United States and utilizes a field experiment as its primary method. As the authors find, exposing conservatives to liberal twitter bots (an experimental intervention) increases personal polarization (one of our key issues). Since the authors set out to test interventions to mitigate the polarizing effects of 'echo chambers'--which in this case proved less polarizing than extra-echo chamber exposure–we marked their intervention as unsuccessful. Finally, we determined how relevant the article was to the search criteria. In this case, the article was relevant to the topic as it evaluated the effect of an experimental intervention on polarization, and therefore was what we call a "bullseye." While the study did not take place in the Global South, we maintain
that studies like these hold important implications for a variety of contexts.

When we did not get many relevant hits, we would either add more or adjust the search terms, adjust the Boolean operators and modifiers, or conduct searches on a different search engine. This combination of strategies allowed us to continue finding relevant articles.

In addition to providing an example of a bullseye, we'd also like to provide some examples of what we believe to be "outer ring" articles (ones that aren't bullseyes but still merit inclusion in the literature review) and articles that are not relevant enough to warrant inclusion. Outer ring articles may be one of two categories: 1) focused on the topic of depolarization interventions but not an experiment, and 2) focused on experimental interventions regarding a related topic like racism, sexism, or bias.

Outer ring example, Type 1: The paper *Toolbox of Interventions Against Online Misinformation and Manipulation* by Kozyreva, et al. (2022) is an example of an outer ring article. This piece succinctly summarizes the different types of interventions against misinformation and manipulation available to researchers, providing explanations of each type of intervention as well as helpful examples. The article itself, however, is not an experimental study and contains no formal test. Thus, it is not a bullseye. But the topic is extremely relevant, so the paper merits inclusion.

Outer ring example, Type 2: The article *Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq* by Salma Mousa (published in *Science*, 2020) would be considered an outer ring article. Mousa employs a randomized experiment in Iraq to create youth soccer teams that were either all-Christian or mixed Christian/Muslim. She then observes the behaviors of the Christian players in the treated and control group, to see whether the intervention (being placed on a team with Muslims) increases social cohesion. Social cohesion and reconciliation are closely related to depolarization, but the topics do not perfectly overlap, so

this is not a bullseye piece. Methodologically, this piece is extremely relevant to our literature review, as it is an RCT, and thus it merits inclusion in our literature repository.

Irrelevant piece, example: The article *When do the Advantaged See the Disadvantages of Others? A Quasi-Experimental Study of National Service* by Cecilia Hyunjung Mo & Katharine Conn (published in *APSR*, 2018) is not quite relevant enough to be included in our literature repository. The authors employ a quasi-experimental research design to test whether participation in Teach for America causes advantaged Americans to adopt the beliefs of disadvantaged Americans. Methodologically, this study does fall within our research team's field of interest, because we are interested in quasi-experiments as well as true ones. However, the topic of this piece is too far removed from the topic of depolarization. In order to be relevant, this piece would have to focus on prejudice, bias, or some type of negative feelings between the two groups that could be reduced through an intervention. This piece does not deal with any sort of animosity between groups - rather, it looks at differing viewpoints between two segments of society. Thus, it falls outside of the scope of this literature review.

Irrelevant piece example: Sheri Berman's "Causes of Populism in the West," https://www.annualreviews.org/doi/abs/10.1146/annurev-polisci-041719-102503. While populism does tend to go hand-in-hand with polarization, sources that focus more on the party implications of populism would be a better fit.

*Conclusion*

For this task, we were asked to create a search protocol on the effects of experimental interventions on polarization. While we do not yet have an answer for how the interventions affect collective action, we do have some initial conclusions based on our search.

First, and perhaps most importantly, after just an initial search we encountered a large and current body of literature relevant to the topic. The breadth of polarization scholarship seems to indicate a pronounced academic and policy-based interest in countering polarization.

Second, we found that interventions across the literature offered competing conclusions about the efficacy of giving people new information to correct polarized or misinformed attitudes. On the one hand, only a handful of studies suggest that people are surprisingly willing to update their attitudes when provided with new information. On the other, several papers offer evidence that people resist new information when it contradicts previously held beliefs, especially when those attitudes are closely linked to personal identity. Further work is needed to elucidate the nuances in these apparently competing claims and identify any existing consensuses in the literature.

Third, the majority of studies focus on the Global North, in particular the United States and Western Europe. While we anticipated that this would be the case prior to conducting our search, it further emphasizes the need for studies on randomized interventions in the Global South.

The process we used to conduct our search was methodical yet flexible, and it allowed us to gain a comprehensive understanding of the existing literature. Our search protocol can easily be adapted and applied to nearly any topic.

Often, topics are broken down into an independent variable and an outcome or dependent variable. Most will ask, what is the effect of the independent variable on the dependent variable? As is generally the case, most topics have two broad components: 1) an independent variable and 2) a dependent variable.

After a topic or question has been established, we recommend the following steps:

1. **Break down the topic**: Create a list with synonyms and examples for independent and dependent variables. Come up with as many as possible since this will serve as a list of search terms that will be used in the subsequent steps. This step may lead to there being multiple potential outcomes. For instance, an intervention related to voting may impact multiple aspects of elections such as trust, voter turnout, and voter education.

2. **Order preferences:** What would an ideal article look like? What methods would the authors use? When and where would it be published? Questions like these help determine how relevant articles are to a specific topic. Examples of these bullseye criteria can include: themes under the independent variable(s), themes under the outcome variable(s), methods, location, population studied, type of publication, date of publication, language, author, article impact, etc. It is vital to include or exclude and order these criteria based on the topic. For instance, the publication date may be immensely important when researching a more recent topic. Alternatively, the location of the study may become more important when narrowing one's search to focus on a specific region or country.

3. **Note search engines**: Some search engines are better than others for given fields of research. One should focus their time on using these search engines. If one does not know where to begin, Google Scholar is a good place to start as it pulls from a wide variety of search engines. When using Google Scholar, one should then note the URL of each relevant result to identify additional search engines to use next.

4. **Begin a broad search:** Begin by typing the keywords from the topic into a specified search engine. This allows for a very broad understanding of existing literature that may potentially be identical to one's own topic.

5. **Narrow focus:** Using the list of synonyms and examples created in Step 1, begin by changing the original independent variable term to a synonym or example, keeping the dependent variable term the same. Alternatively, use the initial independent variable term and change the dependent variable term to a synonym or example. It is up to the discretion of the researcher whether the independent or dependent variable is changed first. To narrow the focus even more, one should change both variable terms to synonyms or examples. Boolean operators and modifiers are of particular importance in this step.

6. **Adjust:** Which terms have been garnering the most relevant results? How can these terms be used to adjust the search? Perhaps changing the terms of the independent variable leads to more relevant results than using synonyms or examples from the dependent variable. In this case, continue going through the list of synonyms and examples for the independent variable created in Step 1. Another useful suggestion is to use quotation marks around phrases to ensure that, at a minimum, results contain the exact phrase needed. During this stage, it is important to be flexible about the search terms and punctuation one is using and to cast as diverse and wide a net as possible to gather the largest number of relevant results. It is also useful to review, and perhaps adjust, the list of ordered preferences written in Step 2 to ensure that one is finding articles that are most relevant.

7. **Shift focus:** Repeat Steps 5 and 6, narrowing the focus of the search and adjusting as necessary based on the results one is gathering.

8. **Zoom out:** Deciding when to complete a search is one of the most challenging steps. While it is usually not harmful to continue searching, one should consider moving past the search stage when they feel comfortable with their understanding of themes, narrow and broad, in the literature. If one feels uncomfortable with their grasp on the existing literature or has the time and resources, it may be a good idea to repeat Steps 5-7 to gather additional relevant articles.

Overall, we suggest that starting broad and then narrowing focus is the most effective manner to conduct a search. It is also vital to adjust one's search terms based on the absence or presence of relevant results. Flexibility and remaining grounded in the initial search terms and ordered preferences for relevance are essential to a successful search.

---

INFORMATION BASED INTERVENTIONS

| **Summary of Informational Interventions** |
| --- |
| ● 12 interventions<br>○ Global South: 4<br>○ Global North: 8<br>● Successful: 6<br>● Unsuccessful (backfire): 6<br>● In person: 7<br>● Online surveys/virtual: 5 |

Informational interventions in the field of polarization studies are a common, though highly varied and wide-ranging set of interventions. In this survey of recent publications, meaningful variations exist in terms of the type of information, the form of intervention, and the metric for

polarization. More conventional studies, generally from the Global North, tend to make use of high levels of political literacy among audiences. Most scholarship revolves around the direct manipulation of information related to contemporary politics, as in studies by Houston (2021), Pereira et al. (2022), Baysan (2022), Chong et al. (2014), and Cubillos Ramdoh (2022). Others utilize mock candidates to isolate specific mechanisms around polarization, as in publications by Bassan-Nygate and Weiss (2022), or Rogowski and Sutherland (2015). Another approach is that used by Lees and Cikara (2019) and Ruggeri et al. (2021), and Strandberg et al. (2021). These studies focus on the relationship between voters and candidates, ingroups and outgroups, as related to polarization. The majority of publications around informational polarization use in-person interventions, as with Baysan (2022), Strandberg et al. (2021), and Bassan-Nygate and Weiss (2022).

Measuring polarization tends to be divided along affective and ideological lines; for instance, Cubillos Ramdoh (2022) uses a battery of indicators in his survey of Chilean voters, including a ten point positivity scale, favorability of pro-choice legislation, and socioeconomic policies. This multifaceted approach represents the standard in the literature, particularly given the interaction of certain aspects of identity on polarization. Most studies vary measurement of key variables depending on the type of intervention. Baysan (2022) and Gerber, Huber, and Washington (2010) use voting patterns as an outcome variable after some pre-election treatment, while most experimental manipulations are measured immediately after treatment, as with Pereira et al. (2022).

The literature on informational interventions features strong representation from Global South perspectives. Particular attention is given to the role of information—media sources, misinformation drives, authoritarian threats—in developing democracies, which in turn informs the shape and direction of studies from the Global South. Baysan (2022) in particular focuses on the rising authoritarianism in Turkey when constructing her intervention, while Pereira et al. (2022) are attuned to pervasive political rumors spread in the buildup to Brazilian elections in 2018. These reflect not only the capacity to apply methods derived in Western settings—Ruggeri et al. (2021) study polarization in 26 different nations—but to adapt analyses to local conditions, as with Chong et al. (2014), capturing global nuances in polarization.

The most salient variation between publications is clearly efficacy. In just this brief overview of informational interventions, the literature is split down the middle when it comes to the depolarizing power of such interventions. Work in the Global North, in addition to studies focused on social psychological mechanisms, tends towards a positive view of informational manipulation. Strandberg et al. (2021) and Houston (2021), for instance, identify meaningful depolarizing devices among American respondents, though these findings tend to reflect abstract polarization, as with studies on meta-perception. Both Lees and Cikara (2019) and Ruggeri et al. (2021) find beliefs on outgroup hostility highly inflated, demonstrating certain interventions—showing respondents what the outgroup *actually* thinks of them—can reduce

polarization. These are in line with Rogowski and Sutherland (2015) and Bassan-Nygate and Weiss (2022), who show reduced polarization *in fictional settings with imagined candidates*. While of undisputed value to the field, these aforementioned studies mirror manipulations of abstract polarization, and in Global North settings, which might explain success in depolarizing interventions.

Both publications focused on actual contemporary political issues and work done in the Global South strongly dispute the efficacy of informational interventions. Levendusky (2018) demonstrates the limits on depolarizing strategies. His work strongly suggests that common approaches to more abstract polarization reduction, including self-affirmation techniques and increasing partisan ambivalence, may actually increase polarization. His work demonstrates the powerful role heterogenous effects play in polarization, in that strong partisan identities tend to become *more* polarized after treatment. This closely parallels the findings of Cubillos Ramdoh (2022) in a completely different setting; despite links, affective and ideological polarization rarely interact directly, excepting when affective polarization *increases* ideological polarization. Baysan (2022) comes to a similar conclusion in Turkey, that information may strengthen polarization via stronger group identities. Work by Chong et al. (2014) and Pereira et al. (2022) somewhat contests this claim; informational interventions were shown to be either totally ineffective, or to depress participation in aggregate.

It is difficult to come to a clear conclusion regarding informational interventions based on this abridged examination of current works in the field. While the increasing representation of Global Southern perspectives is certainly welcome, as is the growing complexity of polarizing mechanisms examined, there exists no clear consensus in terms of the form and effect of informational interventions. This brief overview suggests to the author a dissonance between generalized abstract polarization (i.e. Lees and Cikara, Ruggeri et al., Bassan-Nygate and Weiss) and more grounded analysis of context-specific informational interventions in polarization. Experiments in depolarization appear most successful when more abstract, dealing with fictitious candidates or policy outcomes (Rogowski and Sutherland, for instance). These tended also to be highly generalizable, as Ruggeri et al. (2022) ably demonstrate. Yet when more specific and politically conversant methods were applied—as with Baysan (2022)—the results contradict the more optimistic findings elsewhere in the literature. Levendusky (2018) aptly put the issue: "on the methodological front such [failed interventions] show the importance of avoiding the file-drawer problem by publishing null results, *especially for treatments that have been shown to work in other contexts*" (pg. 590, italics added). Thus, it seems appropriate for the literature to refocus around a set of successful and generalizable techniques, avoiding the apparent contradictions between abstract polarization and contextual informational interventions. Moreover, given the potential to heighten polarization as shown by Baysan (2022) and Cubillos Ramdoh (2022), particularly in developing democracies, future informational interventions must be cognizant of local conditions and prepared to report unsuccessful depolarization strategies.

| **Highlights from information-based interventions** |
| --- |
| ● Intense variation in type of intervention, measurement of polarization, and informational content<br>● Most studies in person, and from Global North, though strong showing from other perspectives<br>● Split down the middle in terms of efficacy of informational interventions on depolarization<br>● Apparent divergence in literature between abstract manipulations and use of real political contexts<br>● Most conventional interventions, when adapted to local conditions, appear to have no effect or exacerbate existing polarization<br>● Potential for future publications to focus on uniform strategies, closer agreement on polarization measures |

---

CONTACT & EXPOSURE INTERVENTIONS

| **Summary of Contact and Exposure Interventions** |
| --- |
| ● 12 interventions<br>  ○ Global South: 2<br>  ○ Global North: 10<br>● Successful: 7<br>● Mixed results: 4<br>● Unsuccessful: 1<br>● In person: 4<br>● Online surveys/virtual: 8 |

*Discussion*

One of the simplest and most well-known methods for reducing prejudice is through contact. The contact hypothesis, attributed to Allport (1954), posits that prejudice between members of different groups if they take part in interactions which meet certain criteria, such as having a common goal and equal status. Because of the potential of contact to reduce animosity between groups, it is hypothesized to be effective at reducing polarization as well as prejudice.

The contact hypothesis is very well studied, and this literature review does not come close to encompassing the vast majority of the papers on this topic. However, we do believe that this selection provides a representative sample of the literature and analyzes it in some depth. Mixed in with the pure contact hypothesis interventions are similar interventions which seek to expose people to the views and arguments of outgroup members. This exposure is often similar to contact (e.g. Balietti et al (2021)), so the two methods are analyzed together.

A cursory glance at the literature leads to the premature conclusion that these interventions are clearly effective, given that all but one study report successful or mixed results. However, looking under the hood reveals less reason for optimism, and much more caveats.

For starters, it is not clear whether the effects of these successful interventions last beyond the intervention period. If effects fade after only a few weeks, then interventions need to be sustained in order to be truly effective.  . However, most studies do not even assess long term effects. Of the 12 studies examined, only four of them measure outcome variables more than 3 weeks after the intervention (highlighted in green above). Of these, one study (McDermott et al. (2018)) was limited to studying university students and relying on self-reported measures, and another (Bail et al. (2017)) found that the intervention had backfired. Of the two successful studies, Santoro and Broockman (2022) only found an effect among one treatment group, and Mousa (2020) found a significant effect only in behavioral measures and not attitudinal ones.

Taken together, this literature review yields almost no hard evidence that polarization can be durably reduced through any of the interventions tested. Many of the studies point to promising methods for reducing polarization, but without rigorous testing of an intervention's long-term effects, it is difficult to say whether it is truly effective.

Some comfort may be found in a meta-analysis of the prejudice reduction literature by Paluck et al. (2021). The authors found that contact interventions, as well as many other types of interventions to reduce prejudice (e.g. perspective-taking) on average do seem to reduce prejudice, but only to a modest extent. Paluck et al. highlight concerning evidence of publication bias, which indicates that the effectiveness of these interventions are probably overstated in the literature. They also note concerning pitfalls that the majority of the literature falls victim to, including failure to pre-register experiments, small sample size, overreliance on self-reported measures, and (as mentioned earlier) a short or nonexistent delay between intervention and outcome variable measurement. Taken together, this means that the vast majority of studies require more contextualization before they can serve as guidelines for policymakers. At the same time, they point to several "landmark studies" which avoid the methodological pitfalls and have very relevant findings. Three of these studies were related to the contact hypothesis: Mousa (2020), Lowe (2020), and Scacco & Warren (2018), the last of which is discussed elsewhere in this report.

Interestingly, the two "landmark" studies also happen to be the only two studies in this section which were conducted in the Global South. Mousa's experiment took place in Iraq while Lowe experimented in India. Both researchers took advantage of pre-existing sports leagues, which was a convenient choice because team sports generally meet the criteria for Allport's contact hypothesis. The main finding from Mousa's paper was that contact can durably reduce discriminatory behaviors towards the outgroup, but does not seem to affect attitudes towards the outgroup. Lowe's main finding was that cooperative contact can significantly reduce discriminatory behavior, but competitive contact has mixed effects. The main limitation of both studies is that the participants of both studies were exclusively male.

While Mousa and Lowe mainly studied observable behaviors, most researchers examining exposure or online contact looked at self-reported attitudes and policy positions. A number of these studies stood out for their clever research designs and interesting findings. The existence of heterogeneous effects across studies suggests that interventions can only successfully reduce polarization under certain conditions. The common thread between these appears to be that cross-partisan exposure and contact are more effective at reducing polarization when partisan labels are downplayed or when non-political factors are highlighted - or perhaps more broadly, when similarities are heightened and differences are downplayed. For example, Messing & Westwood (2014) find that when participants only see the partisan affiliation of a news article, they are more likely to read one that shares their affiliation. However, when this is supplemented with non-political information—in this case the social endorsement the article receives—participants are more likely to read across the aisle.

In a similar vein, Bail et al. (2018) find that political polarization increases when social media users are exposed to a bot that retweets output from high-profile outgroup members compared to a control condition where social media users see their normal feed. Conversely, Balietti et al. (2021) found that exposure to opposing viewpoints decreases polarization, but their sample is limited to pairs of people who are similar along demographic and non-political dimensions. A possible interpretation of these seemingly opposing findings is as follows: When people are exposed to an opposing argument, and the only accompanying information about the person making the argument is that they are different to the subject (communicated through a partisan label), they will react negatively to the argument and the person making it. But when people are exposed to an opposing argument, and the only accompanying information about the person making the argument is that they are similar to the subject (communicated through demographic information, common interests, or common humanity), then they will react more positively to the argument and the person making it.

The results of Santoro & Broockman (2022) support this interpretation. In their study, cross-partisan discussions on a non-political topic convey information about the similarities of

the participants (e.g. common interests or values), and these conversations drastically reduce affective polarization. Meanwhile, cross-partisan discussions of political topics convey information about differences (e.g. they naturally expose the partisan divide between the participants), and these conversations have no effect. While this research remains suggestive rather than conclusive, these papers do suggest that highlighting commonalities vs. differences is important for depolarization.

Overall, the literature provides a number of promising avenues for policymakers and researchers to pursue, but does not provide any definitive recommendations regarding how to reliably reduce polarization. These papers suggest that the contact hypothesis is correct, and that political attitudes can be changed upon exposure to opposing viewpoints under certain conditions, but there is no clear-cut avenue from these findings to policy implementation.

Perhaps unsurprisingly, the main conclusion from this section of the literature is that we need more research. However, we do not just need more experiments, we need more high quality experiments. The field is inundated with experimental interventions seeking to decrease social ills such as depolarization and prejudice. Examining prejudice reduction alone, the Paluck et al. meta-analysis identified over 300 published manuscripts. Future research on depolarization, in addition to following methodological best practices, should either focus on developing new theories and interventions (e.g. researchers could follow Balietti's advice that future work should attempt more realistic interventions), apply existing interventions to understudied contexts (e.g. researchers could study women in the Global South more extensively), or seek to replicate previous studies. Funding organizations could focus more on funding research with high-quality methods, like those of Mousa and Lowe.

| **Main Takeaways from Contact & Exposure Interventions** |
| --- |
| **Takeaway 1:** The contact hypothesis does seem to be effective at reducing polarization, prejudice, and discrimination, but the magnitude and durability of these reductions may be smaller than most researchers report.<br><br>**Takeaway 2:** Applying the contact hypothesis to virtual environments does not eliminate its effectiveness; parasocial contact and exposure to online content can also result in reduced prejudice and polarization.<br><br>**Takeaway 3:** There is evidence that attitudinal change and behavioral change occur through different mechanisms. When choosing among interventions, policymakers should first decide which they want to prioritize.<br><br>**Takeaway 4:** There is evidence that cooperative contact can reduce discriminatory outgroup behavior in the Global South, but may not be effective at changing attitudes |

**Takeaway 5:** Countries in the Global South—and particularly women living in countries in the Global South—remain very understudied.

**Takeaway 6:** There is promising suggestive evidence that social media can be used as a tool to combat depolarization rather than exacerbate it.

**Takeaway 7:** There is suggestive evidence that exposure interventions work better when commonalities (e.g. common identities or interests) are highlighted and/or when differences (e.g. political labels) are downplayed.

---

EDUCATIONAL INTERVENTIONS

**Summary of Educational Interventions**

- 8 interventions
  - Global South: 3
  - Global North: 4
  - Crossnational: 1
- Successful: 4
- Mixed results: 3
- Unsuccessful: 1
- In person: 7
- Online surveys/virtual: 1

Educational interventions are used to teach individuals a specific skill or about a given policy or group of people. There were a total of eight studies whose interventions dealt with teaching participants. Only three of these studies took place in the Global South—namely Nigeria, Côte d'Ivoire, and India—and one was cross-national. The remainder took place in the Global North. Four of these studies were successful in their efforts to depolarize participants' beliefs, three had mixed results, and one had null findings. Finally, all but one of the studies took place in person.

The studies varied in what they were teaching and how they relayed the information. Some studies, like Sousa, et al. (2005) and Neto, et al. (2016), involved adjusting the curriculum of school-aged children. Similarly, Scacco & Warren (2018)'s study took place in a classroom setting, but it emphasized the contact between adult students rather than the content learned. To look at propensity to believe and share misinformation, Badrinathan (2021) conducted an hour-long media literacy lesson and Gottlieb, et al. (2022) had participants watch four-minute long videos, some of which took social identity into consideration. Guidi, et al. (2021-2022)

analyzed the effects of a short lesson conducted in a survey, and Chang, et al. (2019) looked at the impacts of existing diversity trainings.

Regardless of the content relayed in the intervention, the success rates varied based on whether they were conducted in the Global North or the Global South. The findings divided up by region are below.

*Global North*

| **Main Takeaways from Educational Interventions in the Global North** |
| --- |
| **Takeaway 1:** Use of contact hypothesis (usually across racial or ethnic lines) was often successful. However, the three studies that utilized the contact hypothesis only used a sample of school-aged students.<br><br>**Takeaway 2:** Educational interventions are successful at reducing polarizing beliefs among school-age children.<br><br>**Takeaway 3:** While one study looked at the efficacy of educational interventions on adult populations, more evidence is needed to fully understand the probability of success. |

Overall, all of the studies reported successes at reducing polarization. While the cross-national study (Chang, et al., 2019) had mixed results, the majority of the successful individual cases within the larger analysis were in the United States. Successful interventions tended to be long, with durations ranging from around two hours to six months. These studies, however, were conducted with students making longer-term studies more feasible. Finally, in person interventions were more common than online ones.

Sousa, et al. (2005) and Neto, et al. (2016) conducted similar studies with Portuguese school children to look at the effects of participation in a cross-cultural music program, four or six months long respectively, on attitudes towards light- versus dark-skinned individuals. Sousa, et al. (2005) found that among those in the cross-cultural music program, negative stereotyping of dark-skinned individuals was significantly reduced. These effects were seen among nine and ten year olds but not seven and eight year olds. Neto, et al. (2016) found similar results, adding that the effects of the intervention persisted after three months and, for those who were able to be surveyed again, after two years.

Janzen, et al. (2023) also conducted their intervention with a sample of middle-school aged children that was grounded in the contact hypothesis. The students toured a mosque and were subsequently asked about their views towards Islam. After eleven days, they reported more positive feelings. Unlike Neto, et al. (2016), these gains did not persist and returned to baseline levels after four months.

As with any school-based intervention, buy-in from both teachers, students, parents, and the school district can impact the likelihood of success. These studies may also struggle to generalize to the broader population as what may be successful with younger students may not translate to the adult population. Furthermore, the school setting is inherently unique. Given how students are required to attend certain classes and activities, it may be easier to conduct longer-term, and ultimately more successful, interventions.

Outside of school and in the workplace, Chang, et al. (2019) analyzed the effects of diversity training programs in 63 countries on attitudes and behaviors towards women. While this study was cross-national, the successful results were primarily in the United States. On average, the diversity training programs had a significant effect on reducing prejudicial attitudes among men, especially those who were more discriminatory to begin with, but had no significant effect on attitudes. Female employees reported that, after participating in diversity training, men were more likely to nominate them for awards or hire more women. The results, however, were sometimes only significant among participants in the United States.

Finally, Guidi, et al. (2021-2022) conducted the only virtual survey experiment of the Global North educational interventions. Individuals who learned that vaccine passports are standard practice for other types of diseases and infections (besides COVID) generally became more supportive of a COVID vaccine passport, show a willingness to get vaccinated if passports are introduced, and depolarize their beliefs towards passports. No long-term effects were measured.

In conclusion, educational interventions tend to be successful in the Global North. In particular, interventions that take place in schools and last for an extended period of time are widely effective at reducing polarizing attitudes. While there are some indications that educational intervention can be successful among adults, more research is needed to establish this conclusively.

*Global South*

| **Main Takeaways from Educational Interventions in the Global South** |
| --- |
| **Takeaway 1:** Pre-existing beliefs are likely the most important predictor of whether or not an intervention will be successful.<br><br>**Takeaway 2:** Interventions that are successful in the Global North may not be successful in the Global South. This is true regardless of the duration's length and the topic taught.<br><br>**Takeaway 3:** Studies tended to show mixed results in the short term, while one study showed null findings two weeks after the intervention. Effects after two weeks were not analyzed. |

**Takeaway 4:** While educational interventions show some promise, more research is needed, especially that which takes social identity into account, to fully understand both their short and long term effects.

There were a total of only four educational interventions that took place in the Global South. In particular, Nigeria, India, and Côte d'Ivoire. Among these, three of them had mixed results and one had null findings. Two of the studies dealt with misinformation and polarizing beliefs. Additionally, the lengths of studies varied from four minutes to sixteen weeks. The studies also varied with regards to when the dependent variable was measured, ranging from immediately to one month after the intervention.

Sacco & Warren (2018) combined education with the contact hypothesis where they had young men in the city of Kaduna, Nigeria—a city known for religious-based violence—participate in a computer skills class composed of a religiously homogeneous or heterogeneous group of students. Individuals in the control did not take a class. Each course lasted for sixteen weeks (four hours each week). Outcomes were measured one month after the course was completed. The authors found that participation in the class was associated with a change of behaviors but not attitudes. They further found that, contrary to their predictions based on the contact hypothesis, students in the control and heterogeneous groups behaved roughly the same while those in the homogeneous group behaved more discriminatory.

Gottlieb, et al. (2022) and Badrinathan (2021) both crafted intervention to correct misinformation related to polarizing topics and attitudes. In Côte d'Ivoire, Gottlieb, et al. (2022) showed participants one of four videos explaining the importance of digital literacy. Videos were four minutes long, and outcomes were measured immediately. Interventions were only successful at improving an individual's ability to identify fact versus misinformation when they took social and political identity into account.

Badrinathan (2021) had individuals in India participate in an hour-long media literacy workshop where they were taught how to reverse image search and navigate a fact checking website. Outcomes were measured after two weeks. The workshop had no significant effect on an individual's ability to identify misinformation, as any effects of treatment are likely dependent on pre-existing beliefs.

In summary, the success rates of interventions in the Global South are mixed at best. No education-based intervention measured outcomes after a month, so it is unclear how long any effects last. However, looking at the relationship between social (namely ethnic, religious, and political) identity may help craft more successful interventions. Regardless, more research is needed on long-term effects and the role social identity may play in the efficacy of interventions in the Global South.

MENTAL EXERCISE INTERVENTIONS

| **Summary of Mental Exercise Interventions** |
| --- |
| ● 4 interventions<br>○ Global South: 0<br>○ Global North: 4<br>● Successful: 1<br>● Mixed results: 2<br>● Unsuccessful: 1<br>● In person: 1<br>● Online surveys/virtual: 3 |

Mental exercise interventions are those that encourage people to think more critically about a given subject or behavior. For instance, these interventions can include deliberation or inoculation against misinformation. Participants are encouraged to think about the content they are exposed to in the intervention and, ideally, about that same process to the world outside of the study.

*Global North*

| **Main Takeaways from Mental Exercise Interventions in the Global North** |
| --- |
| **Takeaway 1:** Mental exercise interventions often produce mixed results at best or struggle to generalize beyond the scope of the study.<br><br>**Takeaway 2:** Pre-existing beliefs may influence an individual's responses to the intervention, rendering it at least partially ineffective.<br><br>**Takeaway 3:** Further evidence is needed to see if mental exercise interventions could be effective in the Global South. |

There were a total of four studies that implemented mental exercise interventions. All of which occurred in the Global North. In total, one was successful, two had mixed results, and one was not successful. Only one study was in person while the remaining three took place online, and outcomes were measured either during or immediately after the study. Each used a different methodology. Munger (2017) looked at the effects of social sanctioning on Twitter. Fishkin, et al (2019) had small group deliberation between Democrats and Republicans. Schmid-Petri & Bürger (2021) used inoculation to correct misinformation in a survey. Finally, Kvam, et al. (2022) created a simulation for participants.

Pre-existing beliefs and stereotypes impacted individuals' responses to the treatment. For instance, Munger (2017) used Twitter bots that varied race (Black or white) and status (high or low follower count) to reprimand individuals who used racial slurs. Decreases in the use of racial slurs were only seen when the bot that did the sanctioning was a white male with a high follower count. Similarly, Schmid-Petri & Bürger (2021) used an inoculation-based treatment to combat climate change misinformation. Their treatment was ineffective at reducing belief in climate change misinformation as pre-existing beliefs are too strong.

Fishkin, et al (2019) conducted small group deliberations between Democrats and Republicans on extremely polarizing topics that were associated with a greater consensus among the two groups when compared to the control. Deliberation also reduced levels of affective polarization between the groups but did not alter feeling thermometer scores within the in-group. These effects were strongest among more extreme partisans. Unlike Munger (2017) and Schmid-Petri & Bürger (2021), pre-existing stereotypes and feelings were able to be overcome, at least to an extent and in the short-term, by small group deliberations.

The fourth and final mental exercise intervention, by Kvam, et al. (2022), was a simulation. Individuals were asked to make a series of decisions on recommendations for an alien civilization. When they were forced to choose between two options (rather than many), they were less likely to select the less polarizing option. While this study was the only successful intervention, it is unlikely to generalize to the real world. Individuals more often than not make decisions with a myriad of potential choices, and what individuals choose within the simulation may not translate to what they would decide in the real world.

In summary, there is some indication that addressing pre-existing beliefs may be essential for mental exercise interventions to be successful. Even so, the evidence is limited as to if this type of intervention is effective for depolarization. Further research is needed to see if they may work in the Global South.

MASS MEDIA INTERVENTIONS

| **Summary of Mass Media Interventions** |
| --- |
| ● 12 interventions<br>○ Global South: 3<br>○ Global North: 9<br>● Successful: 5<br>● Mixed results: 3<br>● Unsuccessful: 4 |

- In person: 10
- Online surveys/virtual: 2

*Introduction*

Mass media interventions offer a potential channel for promoting social cohesion and post-conflict reconciliation. However, a nuanced examination of these interventions reveals varying degrees of success, significant contextual dependencies, and in many cases limited generalizability.

*Mass Media Interventions in the Global North*

In the Global North, evidence suggests that musical interventions may be moderately effective at countering prejudice and polarization. In their 2014 paper, Greitemeyer and Schwab conducted experiments among German and Austrian university students and found that music with pro-integration lyrics effectively decreased discrimination against Turkish migrants. Similarly, Bodner and Bergman (2017) exposed Jewish and Muslim Arab university students in Israel to different iterations of Jewish national songs, yielding less prejudicial intergroup attitudes among respondents as well as a heightened ability to read the emotions of members of cultural outgroups. Music-based educational interventions have also been applied among schoolchildren in the Global North. Sousa, Neto, and Mullet (2005) found that cross-cultural music education could produce significant reduction in stereotypes among treated groups. In testing the durability of their 2005 intervention, Neto, Conceicao, and Mullet (2016) found a persistent decrease in preferences for light-skinned people even after two years. Nonetheless, Vuoskoski, Clarke, and DeNora cautioned that media interventions may not significantly affect implicit associations. Their 2016 study found that while music can influence empathy, it tends not to significantly impact implicit associations, a potentially profound limitation that could impede deeper attitudinal shifts.

By contrast, LaMarre, Knobloch-Westerwick, and Hoplamazian (2012) test the capacity of music-based interventions to *intensify* racial attitudes and find evidence that music interventions may inflame prejudicial attitudes as well as mitigate them. The authors exposed White American participants to different music genres and find that that “radical white power rock” listeners demonstrate higher levels of prejudice, allocating more funding to white student organizations and less to Black and Arab student organizations.

Their results notwithstanding, several of these studies are constrained by their reliance on small, unrepresentative samples. Most of the musical interventions used with adults recruited study participants from college campuses, and disproportionate numbers of those selected were female. This limitation undermines the generalizability of these studies, especially since it is uncertain

whether the treated subjects demonstrated high levels of outgroup prejudice in the first place. According to Gubler et al. (2022), pro-integration messages tend to be most effective among those with pre-existing positive attitudes, while those with strong anti-outgroup attitudes show minimal changes in empathy and policy attitudes. Unpleasant affect from cognitive dissonance may explain the intransigence of entrenched attitudes among these individuals.

*Global North*

| **Main Takeaways from Mass Media Interventions in the Global North** |
| --- |
| **Takeaway 1:** There is suggestive evidence that exposure to music promoting reconciliatory and integrative attitudes toward outgroups can be moderately successful in reducing prejudice. However, most of these studies tend to draw from small samples of university students, limiting their generalizability.<br><br>**Takeaway 2:** Studies that use music to foster tolerance and inclusion among young school children have found that culturally-educational music programs can durably diminish children's negative attitudes toward ethnic minorities.<br><br>**Takeaway 3:** Media interventions tend to promote increased empathy for outgroups only among people who already have minimally prejudicial attitudes. These interventions appear to be much less effective at reversing people's entrenched attitudes and beliefs. |

*Mass Media Interventions in the Global South*

Research conducted in the Global South suggests that radio drama interventions can be potent tools for reducing prejudice. In her seminal 2009 paper, Paluck found that people in post-genocide Rwanda who listened to a radio soap opera featuring pro-reconciliation narratives were more likely to engage in prosocial behaviors with outgroups including intermarriage, open dissent, and cooperation. However, listeners did not necessarily demonstrate significant shifts in attitudes or beliefs. Bilali and Vollhardt (2013) corroborated Paluck's findings by examining how simply priming the radio drama affected respondents' polarized attitudes, arguing that the primes alone had a positive impact on respondents' ability to adopt the historical perspective of members of cultural outgroups. The authors' 2015 paper, however, yielded more mixed results. After deploying a similar priming technique in the Democratic Republic of the Congo, they found that while respondents exhibited higher levels of victim consciousness, they also demonstrated more lock-step allegiance to ingroup leaders and greater skepticism about the prospect of peace with outgroups. Bilali and Vollhardt suggest that the inconsistent results in Rwanda and the DRC may be due to the statuses of each country's conflicts. Respondents in the DRC, an active conflict zone, may be less receptive to pro-reconciliation media interventions than their counterparts in post-conflict Rwanda.

Global South studies also address the relationship between multimedia content on social media and polarization. In their working paper, Ventura et al. conducted an experiment in which Brazilian subjects deactivated their WhatsApp multimedia capabilities prior to the 2022 Brazilian general election. They found that while these respondents were systematically less exposed to misinformation than the control group, the intervention did not have a significant impact on their levels of polarization. Lee and Hahn similarly subvert the assumption that social media use is necessarily linked to greater personal polarization. In their 2017 paper, they found that South Korean legislators were in fact less polarized after actively using social media for an extended period. The authors' results, however, are largely correlational and have yet to be rigorously tested in a Global Southern context.

*Global South*

| **Main Takeaways from Mass Media Interventions in the Global South** |
| --- |
| **Takeaway 1:** Media interventions like radio dramas that promote reconciliation and peacebuilding can have a potent positive effect on intergroup behaviors. This is especially the case when the shows successfully transport listeners from their reality and immerse them in the perspectives of the characters.<br><br>**Takeaway 2:** However, priming interventions like radio dramas can have a less significant positive effect or even promote backlash when deployed in active conflict zones, rather than in post-conflict contexts.<br><br>**Takeaway 3:** Limiting people's consumption of multimedia on platforms like WhatsApp can reduce their exposure to misinformation but has no effect on affective polarization. |

*Conclusion*

Pro-integration musical interventions have shown promise as tools for mitigating prejudice in the Global North. This is particularly the case for school age children who continued to demonstrate reduced levels of prejudice years after initial interventions. However, several of the studies focusing on musical interventions drew their samples from unrepresentative college populations that were disproportionately female. Moreover, evidence suggests that these types of interventions may not be particularly effective at reducing bias among people who have firmly entrenched prejudicial attitudes. Further research in the Global North should investigate whether representative samples of adults subscribing to a variety of prejudicial orientations experience significant attitudinal shifts as a result of media interventions.

In the Global South, research indicates that radio interventions can be potent in reducing intergroup hostility. However, these findings seem to be limited to post-conflict scenarios and may produce backlash effects if deployed to respondents actively experiencing conflict. Other studies in the Global South push back against the paradigm that social media use is directly

linked to polarization. Further research is needed to investigate why multimedia content seems to drive misinformation but not polarization, as well as why radio interventions appear to be so sensitive to case-specific conditions.

---

IDENTITY-BASED INTERVENTIONS

| **Summary of Identity-based Interventions** |
| --- |
| ● 4 interventions<br>  ○ Global South: 3<br>  ○ Global North: 1<br>● Successful: 3<br>● Mixed results: 1<br>● Unsuccessful: 0<br>● In person: 3<br>● Online surveys/virtual: 1 |

Identity based interventions imply the usage of stimuli which activates particular aspects of people's identity. In total, eleven studies addressed this issue. Three of them were based on evidence from Global South (Africa and Latin America). The remaining studies were conducted predominantly in the United States. One study was based on data from South Korea.

Identity-based interventions generally did not produce significant effects. They were successful only in four studies. One of them was conducted in the Global South. While one study was a literature review, six remaining studies had either mixed or negative results. Scholars' attention was focused predominantly on national identity or domestic partisan identities. However, some studies referred also to religious identity (Smith and Boas, 2020), belonging to humanity (Masullo, 2023), fan identity (Ronconi, 2022), and racial discrimination (Iyengar & Westwood, 2014). Only three studies were published in political science journals. Others were conducted within either psychology or communication science. Studies varied in length of intervention but there is no connection between whether intervention was successful and how it was measured.

| **Main Takeaways from Identity-based Interventions in the Global South** |
| --- |
| **Takeaway 1:** Highlighting belonging to outgroups though does not cause the growth in disliking but activates stereotypical thinking among respondents.<br><br>**Takeaway 2:** Social rivalries which are accompanied by positive role-models (e.g. football players who follow rules) lead to the growth of social cohesion after the end of the rivalry.<br><br>**Takeaway 3:** Elites' debates on new political issues (e.g. sexual minorities rights) entail new |

| |
|---|
| political cleavages and realignment. |

Two of three studies did not reflect to what extent results of studies conducted in the countries of the Global North may be valid for the countries of the Global South. This issue was extensively discussed in Smith and Boas (2020). Scholars adopted the concept of issue polarization which was developed in the US and used it in cross-country study in Latin America (Brazil, Chile, and Peru). They stated that while the two-party system in the US binded together attitudes on sexuality politics and religion, Latin America experienced low levels of mass partisanship and multiparty systems. However, their results showed that even weak party systems elite priming could trigger electoral realignments. Thus, the results of studies conducted in the countries of the Global North can be used in braving polarization in the Global South.

| **Main Takeaways from Identity-based Interventions in the Global North** |
|---|
| **Takeaway 1:** Elite pacts can produce only temporary depolarizing effects<br><br>**Takeaway 2:** Priming broad-based identities (e.g. national or human) relieve inner rivalries<br><br>**Takeaway 3:** Following political leaders in social media decrease polarization among Twitter users |

Clark (2022) stated that popular identity remedies tend to be ineffectual or only temporarily successful. They analyzed secondary literature and focused their attention on three cases: post-civil war US, Spain after Franco's death, and contemporary Turkish politics. Notably, they also extensively referred to Venezuela's Puntofijo Pact (1958) and concluded that formal and informal pacts between political elites, which were aimed to overlook divisive issues and pursue cooperation above all, were characterized by the same problems and resulted in only temporary peace as major cleavages remained under the surface.

Levendusky (2018) stated that national identity ameliorated attitudes toward outgroups. This is in line with Masullo (2023) who found out that a sense of common humanity caused an increase in feeling competent to form relationships with people with different political views. Both used short-term interventions and measured their dependent variables immediately after exposure to treatment. Both of them used a feeling-thermometer in order to measure affective polarization. However, Levendusky (2018) also used social distance measures (whether the respondent is comfortable having close friends from the other party), and whether the other party's ideas are so extreme they are dangerous for the health of the nation. Along with affective polarization Masullo (2023) also measured relationship skills toward an outgroup by asking participants to rate their agreement with the following: "I am confident that I have the skills to develop positive relationships with those who disagree with me politically".

Two studies show how social media can be used as a tool for depolarization. Lee and Hahn (2018) showed that increased social media use over time was associated with lower levels of aggregate polarization among Twitter users in South Korea. However, they mentioned that Twitter was not a dominant political social media outlet in South Korea and the political significance of depolarization might be quite limited. Masullo (2023) showed that a sense of common humanity could be primed using meme-like posts on Facebook, and, as a result, lead people to have more positive attitudes toward their political outgroup.

Although the majority of interventions were not successful these results also provide valuable insights for depolarization strategies. Myrick (2021) suggested that stressing a security threat from China could overcome affective polarization. However, their survey experiment had limited ability to explain either polarization in US foreign policy or affective polarization among the American public. Instead, responses to external threats reflected the domestic political environment.

---

EMPATHY/PERSPECTIVE-BASED INTERVENTIONS

| **Summary of Empathy/Perspective-based Interventions** |
| --- |
| ● 4 interventions<br>○ Global South: 0<br>○ Global North: 4<br>● Successful: 3<br>● Mixed results: 1<br>● Unsuccessful: 0<br>● In person: 3<br>● Online surveys/virtual: 1 |

Perspective-based interventions have been shown to increase positive feelings toward out-group members. Experimental research has verified these findings in contexts that include non-judgmental conversation, narrative writing, virtual reality and social media usage.

| **Main Takeaways from Empathy / Perspective-based Interventions in the Global North** |
| --- |
| **Takeaway 1:** Perspective-based interventions help increase empathy and decrease animosity towards a political, racial, social, or other out-group.<br><br>**Takeaway 2:** Strategies such as structured narrative writing, fostering dialogue, or taking the perspective of a member of an out-group (like in a video game) can decrease negative sentiment towards an out-group. |

**Takeaway 3:** More research is needed to establish if similar interventions can work in the Global South.

**Takeaway 4:** Perspective-based interventions may struggle to be scaled up because they are often labor intensive or require increased guidance on the part of the enumerator.

Narrative writing has been shown to decrease political polarization and can be employed in two separately effective ways: (1) first-person perspective taking and (2) cooperative contact. Narrative first-person perspective taking is an intervention that involves individuals writing a story about an out-group member that has demonstrated animosity toward a part of their identity in a way that is relevant to themselves. For instance, in the context of a depolarization experiment, participants were shown information about a person who was a hostile member of the opposing political party and told to write a story in which that person was giving a lecture about an issue the participant personally cared about (Warner et al 2020). In cooperative contact, individuals recategorize people based on what they have in common. Participants in the same depolarization experiment were asked to imagine the hostile outgroup member was their partner in a class project for their (shared) favorite class and were told that they both would need to do well to get an A in the class (Warner et al 2020). Both these writing strategies increased positive feelings and a sense of similarity with a political outgroup member.

Depolarization in social media requires more than just exposure to alternative political views. Although exposure has been shown to increase engagement with the opposing content, participants only increased their empathy of the outgroup member (understanding of why some people might identify with the stated views) when they are asked to recall a time they disagreed with a friend prior to browsing the social media account of a political outgroup member (Saveski et al 2022).

The non-judgmental exchange of narratives is an intervention that has decreased prejudice toward outgroups that occurs by fostering a dialogue with an individual in which they share an experience they had that is parallel to that of an outgroup member. It has been used in an experimental setting for groups such as transgender folks and undocumented immigrants. For instance, when canvassers asked people to “tell [them] about a time someone showed [them] compassion when [they] really needed it,” people reported increased positive feelings toward undocumented immigrants than canvassers who did not engage in a non-judgmental exchange of narratives (Broockman and Kalla 2016).

This type of perspective taking can also be applied to both virtual and written settings. Individuals who were assigned to embody a Black character in a video game showed reduced

implicit racial bias as compared to those assigned to embody a White character (Banakou et al 2016).

Overall, the existing evidence suggests that empathy and perspective-based interventions may be an effective strategy to reduce polarization. However, issues of scalability and generalizability across contexts remain.

---

ECHO CHAMBER BREAKING INTERVENTIONS

| **Summary of Echo Chamber Breaking Interventions** |
| --- |
| ● 15 interventions<br>○ Global South:<br>○ Global North:<br>● Successful:<br>● Mixed results:<br>● Unsuccessful:<br>● In person:<br>● Online surveys/virtual: |

An influential body of scholarship suggests that one major contributor to accelerating polarization has been the creation of media echo chambers (Sunstein 2001). The echo chamber hypothesis suggests that individuals who only receive news and political information from one side of the political spectrum may adopt increasingly extreme views. There are potential reasons individuals might find themselves in echo chambers: people may unintentionally or intentionally form politically homophilic social networks, they might selectively consume politically agreeable news content to avoid cognitive dissonance, and algorithmic news feeds may curate one-sided news feeds to increase engagement. Echo chambers formed from any of these causes might contribute to polarization. As a result, interventions designed to break echo chambers constitute a potentially promising strategy for reducing polarization. In this section, we review the existing experimental literature evaluating the efficacy of such strategies.

| **Main Takeaways from Identity-based Interventions in the Global North** |
| --- |
| **Takeaway 1:** Results for if breaking echo chambers can reduce polarization are mixed at best.<br><br>**Takeaway 2:** Social interventions (those that encourage individuals to be more receptive to counterattitudinal information) may be the most effective strategy to break echo chambers. |

These interventions tend to be more promising than friend recommendations, making people aware of echo chambers, content recommendation, and highlighting homogeneity in news flows.

**Takeaway 3:** Studies that successfully break echo chambers struggle with their real world applicability due to selective exposure.

*Does Breaking Echo Chambers Reduce Polarization?*

One set of experimental studies attempts to measure *in principle* whether or not interventions that expose individuals to a more diverse media diet can reduce polarization. The results of such studies are mixed, but generally positive.

Warner (2010) randomly assigned students to read either liberal, conservative, moderate, or mixed editorial content on Iran. Attitudes toward Iran were measured both before and after the stimulus. Warner found that students exposed to conservative content became more militant toward Iran, whereas those exposed to moderate content became less militant. Students exposed to liberal and mixed media did not significantly change their attitudes. These results suggest that for those in a conservative echo chamber, introducing liberal media may mitigate polarization. Interestingly, the results also suggest that liberal echo chambers may not be create polarization in the first place.

Karlsen et al (2017) randomly assign survey respondents to read an argument that either suggests gender equality has gone too far, not far enough, or a mix of both. Using pre-survey data, they classify each respondent as having received confirmatory, contradictory, or mixed stimulus. They then measure attitudes after the intervention, assessing the percentage in each treatment group whose attitudes polarize, moderate, or remain stable. Across the board, they find polarization to be rare, but even more so in the group receiving mixed stimulus. Respondents exposed to mixed arguments were also more likely to moderate their opinion. These results suggest that breaking echo chambers by introducing mixed perspective media content may reduce polarization.

Rhodes (2021) exposed survey respondents to a series of ten news clips, randomly assigning respondents to read either all true news, all fake news, or half and half. He also randomly assigned survey respondents to receive read entirely politically agreeable news clips or a mix of agreeable and not agreeable clips. After reading the clips, respondents were asked to assess the believability of the news. Rhodes found that Democratic respondents who read mixed news were more likely to rate Democratic-friendly fake news as false compared to Democrats who only read Democratic-friendly news. He found no effect for Republicans. These results suggest that breaking echo chambers may lead to a reduction in belief in fake news, which might also reduce polarization, but only for Democrats.

Balietti et al (2021) asked a group of respondents to write an essay on wealth redistribution and asked another group of respondents to read those essays. They measure Wave 2 respondents' attitudes toward redistribution both before and after reading the essay. Respondents updated their stance in the direction of the essay that they read, although respondents who read pro-redistribution essays updated substantially more than respondents who read anti-redistribution essays. These results may suggest that breaking echo chambers by exposing individuals to contrary information may reduce polarization substantially for conservatives, and less so for liberals.

Giese et al (2019) asked three waves of respondents to craft a message on the flu vaccine for the next wave of respondents. Each respondent in the second and third wave was randomly assigned to receive a "chain history"–positive or negative after the first wave, and positive-positive, positive-negative, negative-positive, or negative-negative in the second wave. In addition to crafting their own message, respondents were asked to rate the message they received. Giese et al find that almost no evidence of attitude change in any direction, regardless of chain history. These findings suggest that breaking echo chambers might not be effective at reducing polarization when attitudes are highly resilient.

Overall, the research evaluating the effectiveness of breaking echo chambers for reducing polarization is promising. Most studies conclude that breaking echo chambers by introducing individuals to politically mixed information reduces polarization (Warner 2010; Karlsen et al 2017; Rhodes 2021; Balietti et al 2021; but see Giese et al 2019). Researchers find contradictory results in assessing how party moderates these effects; Warner (2010) and Balietti et al (2021) suggest that breaking echo chambers may only reduce polarization for Republicans, while Rhodes (2021) suggests that breaking echo chambers only works for Democrats. More research should be conducted on this issue.

Research on this question can only be taken so far, however. Even if breaking echo chambers in principle reduces polarization, policy-makers require specific strategies for how to break echo chambers. Randomly assigning respondents to mixed media diets is an intervention available only to researchers. In the next section, we move on to assess the research on specific strategies for breaking echo chambers.

*How Can Echo Chambers Be Broken?*

**Content Recommendation**

One intervention strategy for breaking echo chambers consists of directly recommending diverse content to individuals, especially on social media. Such intervention would most likely take the form of directly altering social media or news aggregation algorithms to promote contrarian content. There is no experimental, or even empirical, evidence assessing the efficacy of this type of intervention, but numerous scholars have developed theoretical arguments on the question.

Grossetti et al (2020) analyze a large Twitter dataset and use computational models to show that Recommender Systems create a news feed that is less diverse than individuals' social network. They construct an algorithmic solution that could be implemented on top of existing Recommender Systems that would avoid algorithmically-created echo chambers. Similarly, Chitra and Musco (2020) develop an opinion dynamics model that incorporates a network administrator. They show that if the network administrator increases the strength of certain connections in the network in order to increase user engagement (i.e. by disproportionately highlighting posts from certain friends), polarization can greatly increase. They show that adding in an algorithmic constraint that encourages the administrator to avoid polarization mitigates these harms without significantly harming engagement.

Fabbri et al (2022) assess what-to-watch-next algorithms for video streaming, relying on a classification of certain videos as radicalizing and others as non-radicalizing. They develop an algorithm that effectively avoids recommending a substantial number of consecutive radicalizing videos while still recommending high-engagement content. Madsen et al (2020) use rational choice theory to show that rational Bayesians will still develop echo chambers, especially in large networks. They assess the effect of periodic educational broadcasts, showing that these broadcasts only slightly mitigate echo chamber formation.

Three of these studies assess the effectiveness of content recommendation to reduce algorithmically created echo chambers, and all three show that interventions are effective (Grossetti et al 2020; Chitra and Musco 2020; Fabbri et al 2022). The one study that assesses the effectiveness of content recommendation to reduce naturally occurring echo chambers concludes that it is largely ineffective (Madsen et al 2020). Policymakers should therefore primarily consider content recommendation as a solution to algorithmic echo chambers. No experimental or observational evidence exists on these questions, however, and policymakers should refrain from coming to any firm conclusions on the effectiveness of content recommendation until researchers assess the question.

**Friend Recommendation**

Friend recommendation interventions seek to ameliorate echo chambers by providing potential connections to social media users that would diversify the political alignment of their network.

These recommendations might, but do not necessarily have to, arise from social media company's people-to-friend suggestions. As with content recommendation, the bulk of the existing literature on these questions uses theoretical and computational methods rather than empirical or experimental methods.

Gillani et al (2018) is a rare exception. They randomly assign certain Twitter users to receive a set of recommendations of accounts to follow in order to diversify their network, in addition to visual depiction of homophily in their Twitter network. They find that treated users have significantly more diverse connections one week after receiving recommendations, but that the effect decays to zero after two and three weeks. These small effects may also be exaggerated, because recruitment messages informed potential participants that the study would show them their Twitter echo chambers, meaning that the sample likely overrepresents individuals motivated to break out of echo chambers.

In a very similar study, Matias et al (2017) randomly assigned some Twitter users to see the gender breakdown of the accounts they follow, in addition to a recommendation of women to follow. They found mixed results; in the first wave of their study, they found that those who received recommendations significantly increased the number of women they followed three weeks later, whereas in the second study they found no effects. Moreover, even in the first study, the average increase in women followed was below one percentage point.

Two theoretical studies in network sciences attempt to establish that altering the structure of social networks is a viable means to reduce polarization. Chen et al (2018) use opinion dynamics models to show that the risk of conflict in a social network can be minimized by altering only a few connections in the broader network. Interian et al (2021) develop an algorithm that finds the minimum number of connections that must be added between individuals in a network to achieve acceptably low levels of polarization. They conclude that only a small number of connections are needed in most cases.

Whereas Chen et al (2018) and Interian et al (2021) focus only on which connections would most reduce polarization, Garimella et al (2018) incorporate the possibility that friend recommendations may be rejected. Specifically, they model which recommendations are most likely to both be accepted as new connections and to reduce the network's polarization. They develop an algorithm that identifies the most efficient friend recommendations in this context.

Morales and Cointet (2021) study the polarization effects of a variety of friend recommender systems. They use opinion dynamics models to study the effects of a variety of well-known recommender systems on real world Twitter data, finding that while certain recommender systems increase polarization, random friend recommendations reduce it.

Donkers and Ziegler (2021) challenge much of this work by introducing the concept of ideological echo chambers, in which members are not only isolated from opposing views but actively biased against such views. In this way, Donkers and Ziegler's model is arguably more consistent with much of the psychological literature than other author's more simplistic models. They find that in such cases, simply altering friend recommendation algorithms is insufficient to reduce attitude polarization.

The evidence on friend recommendation as a tool for breaking echo chambers is therefore not very encouraging. Although some theoretical studies show that friend recommendation may be effective, particularly in reducing algorithmically driven echo chambers, these models rely on psychological assumptions they may not hold. The only two empirical studies find small or null effects of friend recommendation after a few weeks. Although the evidence is not sufficient to warrant fully abandoning friend recommendation as an echo chamber breaking tool, policymakers would be wise to look elsewhere.

**Echo Chamber Awareness**

Echo Chamber Awareness interventions seek to encourage individuals to break out of their echo chambers simply by making them aware of how one-sided their news diet is. Individuals may be motivated to avoid echo chambers, but not realize the lack of diversity in their information intake. In such cases, simply informing individuals about their news consumption patterns may be an effective depolarization tool.

Two studies in this vein were previously reviewed in the *Friend Recommendation* section. In addition to recommending new social connections, the treatments studied in both Gillani et al (2018) and Matias et al (2017) included informing respondents about their echo chamber. Gillani et al (2018) showed respondents a visual aid depicting the homophily in their network, whereas Matias et al (2017) informed respondents of how many women they follow on Twitter. As reviewed earlier, neither of these interventions was found to be substantially effective.

Munson, Lee, and Resnick (2013) randomly assign some respondents to have access to a browser extension that gives users constant feedback concerning the ideological balance of their news consumption. They find that users accessing the widget have significantly more balanced scores than users not receiving feedback. The substantive effect, however, appears to have been quite small; even in the treatment condition, users changed their behavior very little. Moreover, recruitment did not rely on compensation and instead advertised the tool as a means for users to break out of an echo chamber, ensuring that the sample was exclusively individuals already motivated to break out of their echo chambers. The effects in the study should therefore be seen as a best-case scenario.

Jeon et al (2021) randomly assign respondents to either read about echo chambers, or play a game in which they try to create an echo chamber, designed to show them the negative consequences of echo chambers. Before and after the experiment, they ask participants about their willingness to seek out information from alternative perspectives that they may disagree with. They found that game-players were significantly more likely to report a change in willingness to seek out new information, although the effect size was small. Jeon et al also do not directly measure behavioral change, but only query self-reported willingness to break out of echo chambers which may be subject to social desirability bias or inaccurate self-perceptions. Moreover, other studies have shown that even initially promising diversity-seeking behavior may decay over time (Gillani et al 2018).

The research on echo chamber awareness interventions is not particularly promising. Although a few studies have found significant effects, the magnitude of the change induced by interventions tends to be very small (Matias et al 2017; Munson, Lee, and Resnick 2013; Jeon et al 2021). These studies also either sample only from populations most likely to be responsive to the interventions (Matias et al 2017; Munson, Lee, and Resnick 2013) or do not assess real behavior change as a dependent variable (Jeon et al 2021). Policymakers should therefore not expect making people aware of their echo chambers to have substantial effects on polarization.

**Social Interventions**

Social interventions attempt to use social cues to encourage individuals to be more receptive to politically contrary information. If individuals see out-partisan content as coming from friends or in general being socially popular, they may be more likely to engage with it.

Gao et al (2017) pilot a Reddit-alternative that allowed users to see how members of each party reacted to recommended news stories. They found that users engaged with more articles in their software tool, and more often reported reading an article due to curiosity about other users' reactions. The pilot only included eight respondents, lacked randomization, and did not measure any dependent variable directly related to echo chamber or polarization reduction.

Messing and Westwood (2014) instruct respondents to choose between four headlines to read more about. Respondents were randomly selected to be able to see the articles' publisher, other users' reactions to the article, or both. They find that when users can only see the source of articles, they are much more likely to choose articles from politically agreeable sources (i.e. Republicans choose more Fox News articles). When they can only see social cues, individuals select articles with more likes. When respondents can see both, they select only articles with high social scores, and the publisher no longer significantly predicts selection. These results suggest that social cues can overwhelm partisan ones, encouraging individuals to read out-partisan information they would have otherwise avoided, thus breaking echo chambers.

Balietti et al (2019) instructed Phase 1 respondents to fill out a large survey on non-political questions, before asking them to write an essay on redistribution. Phase 2 respondents filled out the same survey, and read an essay from a Phase 1 respondent. Phase 2 respondents were randomly assigned to receive an essay from a close or distant non-political match, and were asked to self-assess their non-political similarity to their Phase 1 match. They find that although all respondents update their beliefs on redistribution in the direction of the essay they read, respondents update much more when they believe themselves to be similar to the author of the essay. These results suggest that social ties can make out-partisan information more appealing and more likely to be internalized into attitude change.

The limited research on social interventions is therefore fairly encouraging. Both Messing and Westwood (2014) and Balieti et al (2019) find that social cues can reduce opposition to out-partisan information. More research is needed to discover ways to implement social interventions at scale, especially as offline social networks become increasingly socially homophilic, thus reducing the opportunity for cross-partisan social interactions. Still, policymakers should proceed with cautious optimism regarding social interventions.

**Highlighting Homogeneity in Heterogeneous News Flows**

Munson and Resnick (2010) attempt a novel intervention in which respondents were assigned to rate their satisfaction with a curated news feed. Respondents were randomly assigned to have either a homogenous politically-friendly news feed, or a mixed feed. Respondents with mixed feeds were also randomly assigned to have politically friendly articles either highlighted or put at the top of the feed, with the hope that highlighting politically friendly articles would increase respondents' satisfaction with a feed that actually did contain diverse information. They found that 75% of respondents prefer homogenous news feeds, and that among these respondents, highlighting the politically friendly articles did not increase satisfaction with heterogeneous feeds.

*Summary*

Are interventions designed to break people out of their echo chambers effective at reducing polarization? With some exceptions, the literature reviewed here does not suggest much optimism. The research generally suggests that echo chambers do contribute to polarization, and that breaking people of their echo chambers may be an effective way to reduce polarized attitudes. Breaking echo chambers appears to be easier said than done, however.

Of all the intervention types for breaking echo chambers reviewed here, only social interventions have empirical evidence suggesting their effectiveness. Some strategies, like content

recommendation, have not been sufficiently tested empirically, although some theoretical research suggests their viability. Other strategies, like friend recommendation and echo chamber awareness do appear very effective in experimental tests.

Although these results allow us to tentatively recommend in favor of social interventions and against friend recommendation and echo chamber awareness, substantially more research is needed to develop confidence in these recommendations. Even the most highly researched interventions have only been assessed in two or three studies.

Moreover, researchers and policymakers must think creatively to expand the universe of possible interventions to break echo chambers. Almost all of the existing research focuses on interventions that are exclusively focused on social media, despite the fact that traditional media consumption and offline social networks may be the larger contributors to echo chambers (Gentzkow and Shapiro 2011). Relatedly, many of the interventions researchers have considered, such as friend and content recommendation can only be implemented by social media companies themselves, who control the algorithms manipulated by researchers in these studies. Interventions that target individuals more broadly across society may be more feasible, scalable, and effective, and researchers and policymakers should turn their attention to these possibilities.

---

WORKING PAPERS

Ongoing research tests the proposition that intergroup contact and empathy building can facilitate depolarization. In their working paper, Mousa, Scacao, and Naumann explore the role of contact and empathy-based educational programs in mitigating hostilities between migrants and their host communities. Broockman and Kalla investigate whether exposure to cross-cutting media can moderate people's attitudes, even in the face of motivated reasoning that may make them wary of opposing viewpoints. Conversely, Scacao, Siegel, and Weiss suggest in their working paper that partisans avoid outgroup contact even when it is readily available and that this avoidance impedes depolarization. Aruguete et al. use games to argue that uncivil partisan discourse can motivate people to behave in an untrustworthy manner, exacerbating polarization. Other avenues of research explore the potential for elite vouchers (Mousa and Mironova) and participation in institutions such as financial markets (Jha, Shayon, and Weiss) to encourage depolarization.

Additional research related to the contact hypothesis focuses on testing contact interventions in novel situations, as well as exploring and validating the underlying mechanisms and assumptions at play. Ghosh et. al's paper explores the effectiveness of intergroup contact in youth camps in India, building upon a budding experimental literature on contact in the Global South. Blair et al. examine how in-party, rather than cross-party contact can facilitate depolarization without requiring participants to leave the comfort of their own party. Lowe and Jo examine an entirely

different, easy, and low-cost method to reduce political polarization at the highest levels of government - seating cross-partisans next to each other in Parliament. All three interventions yield somewhat successful results. Greene et al.'s paper takes a step back by testing a well-known but untested aspect of the contact hypothesis - the idea that equal status matters. By varying study participants' relative social status in an experimental setting, they are able to verify that equal status is indeed critical for promoting tolerance. Cai's paper addresses the mechanisms driving polarization and perceived polarization in the first place, and finds that selective memory - people's tendency to remember extreme statements more than moderate statements - is a contributing factor both in the lab and in the field.

Current depolarization research also explores other intervention types. Ramón Enríquez, et al. (forthcoming) find that a depolarization nudge on Facebook encourages citizens to hold the government accountable because they have more accurate information. Ramón Enríquez, Larreguy, & Lujambio (forthcoming) also look at social media, finding that exposure to counterattitudinal content decreases levels of polarization. Alternatively, Grady, et al. (2023) use an intervention rooted in the contact hypothesis that found that individuals living in areas that received the intervention viewed the out-group more positively.

---

## DISCUSSION

This review offers a glance at the current state of depolarization research, with special attention paid to variation in outcomes across intervention types and between the Global North and South. While a clear consensus has yet to emerge on which approaches to depolarization are consistently most effective, this report offers several important takeaways for researchers and policymakers dedicated to mitigating global polarization.

First, evidence suggests that sustained contact and relationship building with out-partisans and members of out-groups, as well as empathy and perspective-building exercises, may hold the most promise for countering polarization. Successful interventions in this vein include programs in which participants competed on the same soccer team as members of religious out-groups or engaged in nonpartisan conversations with out-partisans. However, many of these studies did not examine their effects' durability and would benefit from rigorous testing of their core findings.

Second, informational and cognitive exercise-based interventions often struggle to produce significant effects, especially among individuals with deeply entrenched prejudicial attitudes. The techniques employed in these studies include exposing participants to the political views of out-partisans on social media and informing voters about candidates and policies prior to elections. Not only do these types of interventions frequently fail to reduce polarized attitudes, but in many cases they can backlash and cause committed partisans to double down. This is likely due to the cognitive dissonance the interventions inspire by confronting individuals with

information that contradicts closely held beliefs. Young children or people with less established attitudes, by contrast, tend to be more receptive to informational interventions.

Third, the bulk of depolarization interventions focus on the Global North, and their successful application in the Global South appears to be highly context dependent. For example, mass media interventions that successfully compelled participants in Rwanda to take the historical perspectives of ethnic out-groups failed to produce the same results in the Democratic Republic of the Congo (DRC), potentially because Rwanda had emerged from active conflict while the DRC was still gripped by national violence. These kinds of country specific conditions are often critical in determining a Global Southern intervention's efficacy. Further research should seek to rigorously challenge the findings of Global North studies while carefully considering local conditions when applying to interventions to the Global South.

---

## PRIORITIES AND CHALLENGES FOR FUTURE RESEARCH

The literature on depolarization interventions suggests a number of challenges and areas for future research. One difficulty in the literature is the interdisciplinary determinants of polarization. While we see these experiments across different disciplines, it is less common to see interdisciplinary approaches in the design of the interventions themselves. One promising area is in better bridging the work on psychological approaches with those working on informational or learning approaches. Many interventions have elements of both, necessitating more consideration of not only different potential mechanisms that interventions can operate through, but also potential synergies between them.

In terms of research design, another area for improvement is in a greater attention to long term vs. short term effects, as well as the development of better measures of polarization. The effects of depolarization interventions are often only measured in the short term, often even at the point of the intervention. Relatively few studies explore long-term effects of these interventions. It may be possible that even successful interventions need to be sustained or reinforced for effects to persist.

Relatedly, the development of more consistent measures of polarization is also important. In general, the standards for measuring electoral polarization have far outpaced the development of consistent attitudinal measures of polarization. One challenge in this area is that attitudinal measures are more context specific. In addition to the design of the measures, there may also be culture or context-specific barriers to measurement tools that are used in other countries: which questions and concepts are sensitive or politically controversial, or even the relative effectiveness of different tools for eliciting answers without social desirability bias.

One point of caution for polarization interventions is the potential for backlash and heterogenous effects. While heterogenous effects are fairly standard across a wide range of interventions (especially information interventions), the potential for backlash that exacerbates attitudes such as prejudice and hostility necessitate more caution in the design of interventions specifically intended to address polarization. The types of heterogenous effects are also more pronounced in polarization interventions. For example, while it wouldn't be unusual to see interventions that have stronger effects among some groups than others, or subject to attenuating effects by context, polarization interventions can have completely divergent effects among different groups, such as increasing hostile attitudes in some groups even while reducing them in others.

Polarized political settings also provide a challenge for other types of interventions, even if not designed to address polarization directly. For example, recent work on Covid-19 mitigation efforts have shown that attitudes are linked to political polarization and mistrust in government (Druckman et. al. 2020). Even if not explicitly political in nature, messaging about Covid-19 from the government can have unintended results: either only increasing compliance among supporters (Bokemper et. al. 2021) or even decreasing vaccine intention overall (Urrunaga-Pastor et. al. 2021).

As a result, risk-mitigation strategies are essential to depolarization interventions, even those that are not necessarily fostering interactions between groups or with members of out-groups. Even when there is no risk in conflict between groups, these types of interventions should take care that they're not reinforcing stereotypes rather than learning across groups, or increasing animosity or resentment rather than empathy. It also suggests that these types of interventions should be facilitated, moderated, and monitored to a greater extent than other types of interventions, and perhaps should not be thought of as one-time inducements without careful measurement and monitoring. It also suggests that measures include assessing potential adverse effects, in addition to measuring the outcomes of interest in the study.

Last, but certainly not least, both researchers and implementing partners need better frameworks for understanding positionality. Not all researchers and implementing partners are equally well-positioned to conduct these interventions. Both researchers and implementing partners need to be considered neutral or non-partisan actors, for example. This is for both ethical reasons but also for programmatic ones: if credibility and trust are unobservable factors that affect the success of interventions, understanding the dimensions of trust and credibility in the implementation process will provide results that can be interpreted more clearly, making it easier to see how similar programs might work in different contexts.

---

APPENDIX A: SUMMARIES

**Information Interventions**

*Studies from Global North*

**Strandberg et al. (2020):** Respondents provided self-reported survey ratings of Hillary Clinton and Donald Trump, which was manipulated by researchers to reflect more moderate views. Nearly all participants failed to correct these moderated views, suggesting susceptibility to false feedback and polarizing cognitive biases.

**Houston (2021):** When respondents in the United States are provided with information on education spending, partisan attitudes regarding education tend to converge, though political endorsements by moderates are either ineffectual, or increase polarization.

**Lees & Cikara (2019):** group meta-perceptions (beliefs concerning outgroup opinions of ingroup) consistently overestimate animosity, suggesting intergroup interactions to be more hostile than interpersonal interactions. This animosity was experimentally reduced via an informational intervention, though results mixed.

**Ruggeri et al. (2021):** Expanding on Lees and Cikara, this study demonstrates the generalizability of exacerbated outgroup animosities across 26 countries. Animosity experimentally reduced by demonstrating true outgroup opinion, in line with Strandberg et al., though again effects highly heterogenous.

**Gerber, Huber, & Washington (2010):** Exogenous treatments—such as encouraging voter registration—can lead to polarization through expression of latent beliefs, while also increasing participation more generally. This outcome implies the potency of group identification in changing how individuals perceive the world.

**Rogowski & Sutherland (2015):** By experimentally reducing ideological divergence of fictional candidates, respondents can be led to express less polarized views. These results generally according to ideological polarization with highly polarized individuals being more receptive to affective depolarization.

**Bassan-Nygate & Weiss (2022):** This study suggests that affective polarization is tied to salience of electoral competition. By priming audiences in Israel about cooperation between political elites, affective polarization can be meaningfully reduced regardless of ideological polarization.

**Levendusky (2018):** Common depolarization strategies fail to reduce affective polarization in aggregate, as evidenced by a cross-national study in the United States. Respondents with more moderate ideological views were receptive to depolarization, while the same treatment elicited greater affective polarization among more ideological individuals.

*Studies from Global South*

**Pereira et al. (2022):** Third-party fact-checking ineffectual when confronting misinformation, specifically politicized rumors, as this experimental survey from Brazil demonstrates. Fact-checking was shown to be ineffective regardless of political engagement, socioeconomic factors, and ideological polarization.

**Baysan (2022):** Exposing potential voters to information on executive performance shown to increase affective polarization in Turkey. Potential for sorting mechanism, whereby voters use new information to clarify relevant political identity, thus stronger group identification increases net polarization and participation.

**Chong et al. (2014):** An experimental survey in Mexico argues informing respondents on government corruption decreases turnout and identification with both incumbents and the opposition but does not change prior beliefs. Information on corruption was associated with disengagement from politics generally.

**Cubillos Ramdoh (2022):** Affective and ideological polarization are moderately associated according to surveys in Chile. However, experimental manipulation of one does not alter the other, though information from a co-partisan shown to increase ideological polarization of an affectively polarized individual.

**Contact and Exposure Interventions**

*Brief Descriptions of Articles*

Bruneau et al (2021): Researchers paired non-Muslim American students with Muslim non-American students in virtual discussion settings. They measured the level of dehumanization and meta-dehumanization felt among American students with respect to the Muslim students. The researchers supplemented these quasi-experimental findings with cross-sectional and longitudinal data from 5 countries (US, Spain, Hungary, Israel, and Greece). The main finding was that high-quality contact decreased dehumanization and meta-dehumanization, whereas high quantity of contact was only weakly correlated with reductions in these variables.

- Intervention duration: 1 semester (2 hours a week)
- Delay between intervention and measurement: Less than one week
- Outcome variable: Trait ratings, “Ascent of Human” scores, perceived level of dehumanization
- All outcomes self-reported
- Results: Successful

McDermott et al. (2018): Placed volunteers in a university psychology course into parasocial contact with transgender people, and measured their transnegativity before and after the intervention. Found that levels of transnegativity decreased significantly immediately after the intervention, and did not rebound in the long term.

- Intervention duration: 4 hours (split into two sessions across two weeks)
- Delay between intervention and measurement: 7 weeks
- Outcome variable: Likert scale / transphobia scale
- Outcome was self-reported
- Results: Successful

Harwood et al. (2016): Surveyed American students about their attitudes towards Arabs. Then showed them a video discussing either a) musical, or b) technological collaboration between c) two Arab students, or d) an Arab student and a White student. The participants took the baseline survey again after viewing the video, and also answered questions regarding the relationship between the people in the video and their personal desire to interact with Arabs. The results were mixed. The cross-cultural video tended to have more prejudice-reducing effects than the video with two Arabs. The video showing technological collaboration resulted in greater prejudice reduction than the musical collaboration video.

- Intervention duration: 4 minutes
- Delay between intervention and measurement: Instantaneous
- Outcome variable: Trait ratings, Likert scale, survey questions
- All outcomes self-reported
- Results: Mixed

Diamond & Lobitz (1973): During the Stanford student riots in spring 1970, conducted a contact-based intervention with 164 students (95 control) and 37 police officers. Three different forms of contact were tested, and resulted in both groups (students and police officers) experiencing significant attitudinal depolarization towards the other group.

- All outcomes self-reported
- Results: Successful

Bail et al. (2018): Researchers conducted a field experiment that offered a large group of Democrats and Republicans financial compensation to follow bots on social media that retweeted messages by elected officials and opinion leaders with opposing political views. The intervention - intended to induce depolarization - backfired. Republican participants expressed substantially more conservative views after following a liberal Twitter bot, whereas Democrats' attitudes became slightly more liberal after following a conservative Twitter bot—although this effect was not statistically significant.

- Intervention duration: 1.5 months

- Delay between intervention and outcome measurement: 1.5 months
- Outcome variables: “Ideological consistency scale” (questions about social policies)
- Outcome was self-reported
- Results: Unsuccessful

Levendusky (2013): The author hypothesizes that exposure to like-minded partisan news media causes consumers to decrease affect for the other party, lower trust in the other party, and lower support for bipartisanship. Furthermore, the author hypothesizes that exposure to partisan media from the opposite side will have no impact on the consumer’s evaluation of their own party. The author tests this by exposing participants to co-partisan or cross-partisan media and surveying them before and after exposure. The author finds support for all hypotheses. Exposure to like-minded media heightens variables associated with polarization, while cross-partisan media is mostly tuned out.

- Intervention duration: 8-9 minutes
- Delay between intervention and outcome measurement: Instantaneous
- Outcome variables: Feeling thermometers, trait evaluations, trust, support for bipartisanship
- All outcome variables self-reported
- Results: Successful (although it was successful at increasing polarization)

Santoro & Broockman (2022): Paired outpartisans together to discuss randomly assigned topics over video call. When the participants discussed a topic unrelated to their area of disagreement (e.g. their perfect day), affective polarization was dramatically reduced. However, the effect decayed in the long term. When the participants discussed topics over which they disagreed (e.g. why they support their party), there was no effect. Neither conversation topic resulted in changes in attitudes regarding democratic accountability.

- Intervention duration: 10 minutes
- Delay between intervention and outcome measurement: 3 months
- Outcome variables: Outparty warmth, meta-perceptions of outparty members, attitudes towards democratic accountability
- All outcome variables self-reported
- Results: Mixed

Velez & Lavine (2017): Randomly assigned university students to dormitories that had varying degrees of racial diversity. Found that for students assigned to racially heterogeneous dormitories, the students’ political attitudes became polarized on the dimension of authoritarianism. Authoritarian students expressed more support for nationalism and military intervention, but less support for refugees. Meanwhile, non-authoritarian students expressed more tolerance of refugees and less support for nationalism and military intervention. The authors supplemented these findings with a nationally representative survey that found a

significant interaction effect of authoritarianism and local diversity on racial resentment, immigration beliefs, and political intolerance.

- Intervention duration: 1 semester
- Outcome variables: Scales measuring racial resentment and political intolerance, questions measuring beliefs about immigration
- All outcome variables self-reported
- Results: Successful (in that the outcome variables were successfully changed by the intervention)

Messing & Westwood (2014): Exposed people on social media to news headlines that had two randomized qualities: partisan label and social endorsement. Without taking into account social endorsement, people strongly chose to read the news articles that had the label of their party. However, when the social endorsement variable was included (operationalized as a label saying "[N] people recommend"), participants read the articles that were recommended by peers, while the partisan label variable had little effect.

- Intervention duration: N/A
- Delay between intervention and outcome measurement: N/A (Instantaneous)
- Outcome variable: Choice of article to read
- Outcome was observed rather than self-reported
- Results: Successful

Balietti et al (2021): Researchers paired participants with partners who had differing opinions on a divisive political topic: wealth distribution. Partners were matched based on similarity along non-political dimensions. Participants then read essays written by their partners, in which the partners explained their position on wealth distribution. The researchers then surveyed respondents on their polarization and updated positions on the issue. The intervention succeeded in reducing polarization, and it caused people on both sides of the issue to adopt more pro-redistribution positions.

- Intervention duration: Duration of survey
- Delay between intervention and outcome measurement: Instantaneous
- Outcome variables: Political stance on redistribution, feeling of closeness towards partner
- All outcomes self-reported
- Results: Successful

Mousa (2020): Added Muslim players to some teams in a Christian soccer league in Iraq, and measured whether cooperative contact with the Muslim players over a two-month season reduced prejudice by Christian players. Measured prejudice largely through behavioral measures, with a long time delay between intervention and measurement (usually months). Found that the Christian players on mixed teams significantly reduced their prejudice towards Muslims in on-the-field behavior, but their off-the-field prejudices remained. Despite some behavioral

changes, the Christian players did not significantly change their attitudes towards Muslims in general.

- Intervention duration: 2 months
- Delay between intervention and outcome measurement: Up to 6 months
- Outcome variables: A variety of behavioral measures related to outgroup engagement, plus attitudinal measures
- Most outcomes observed, some self-reported
- Results: Mixed

Lowe (2020): Created a cricket league in which some teams had players from multiple castes and some players were homogeneous with respect to caste. Measured the effect of cooperating with vs. competing against members of another caste, vs. a control group who did not play at all. Found that, across multiple behavioral measures, cooperative contact significantly increased cross-caste behavior, while competitive contact had mixed effects, but on balance decreased cross-caste behavior and was marginally significant.

- Intervention duration: 1 month
- Delay between intervention and outcome measurement: 1-3 weeks
- Outcome variables: Real-world and gameplay behavior towards outgroup
- Outcomes were observed rather than self-reported
- Results: Mixed

---

REFERENCES

Aruguete, Natalia & Calvo, Ernesto & Scartascini, Carlos & Ventura, Tiago. (2020). Trustful Voters, Trustworthy Politicians: A Survey Experiment on the Influence of Social Media in Politics *. 10.13140/RG.2.2.26137.65123.

Badrinathan, S. (2021). Educative Interventions to Combat Misinformation: Evidence from a Field Experiment in India. American Political Science Review, 115(4), 1325–1341. https://doi.org/10.1017/S0003055421000459

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. Proceedings of the National Academy of Sciences, 115(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Balietti, S., Getoor, L., Goldstein, D. G., & Watts, D. J. (2021). Reducing opinion polarization: Effects of exposure to similar people with differing political views. Proceedings of the National Academy of Sciences, 118(52), e2112552118. https://doi.org/10.1073/pnas.2112552118

Balietti, S., Getoor, L., Goldstein, D. G., & Watts, D. J. (2021). Reducing opinion polarization: Effects of exposure to similar people with differing political views. Proceedings of the National Academy of Sciences, 118(52), e2112552118. https://doi.org/10.1073/pnas.2112552118.

Banakou, D., Hanumanthu, P. D., & Slater, M. (2016). Virtual Embodiment of White People in a Black Virtual Body Leads to a Sustained Reduction in Their Implicit Racial Bias. Frontiers in Human Neuroscience, 10, Article 601. https://doi.org/10.3389/fnhum.2016.00601.

Bassan-Nygate, L., & Weiss, C. M. (2022). Party competition and cooperation shape affective polarization: evidence from natural and survey experiments in Israel. Comparative Political Studies, 55(2), 287-318.

Batista Pereira, F., Bueno, N. S., Nunes, F., & Pavão, N. (2022). Fake news, fact checking, and partisanship: the resilience of rumors in the 2018 Brazilian elections. The Journal of Politics, 84(4), 2188-2201.

Baysan, C. (2022). Persistent polarizing effects of persuasion: Experimental evidence from turkey. American Economic Review, 112(11), 3528-3546.

Bilali, R., & Vollhardt, J. R. (2013). Priming effects of a reconciliation radio drama on historical perspective-taking in the aftermath of mass violence in Rwanda. Journal of Experimental Social Psychology, 49(1), 144-151. https://doi.org/10.1016/j.jesp.2012.08.011.

Bilali, R., & Vollhardt, J. R. (2015). Do mass media interventions effectively promote peace in contexts of ongoing violence? Evidence from Eastern Democratic Republic of Congo. Peace and Conflict: Journal of Peace Psychology, 21(4), 604–620. doi:10.1037/pac0000124.

Blair, R., Gottlieb, J., Schenk, M., & Woods, C. (2023, June 12). Depolarizing Within the Comfort of Your Party: Experimental Evidence from Online Workshops. https://doi.org/10.31219/osf.io/eha8r

Bodner, E., & Bergman, Y. S. (2017). The power of national music in reducing prejudice and enhancing theory of mind among Jews and Arabs in Israel. Psychology of Music, 45(1), 36-48. https://doi.org/10.1177/0305735616640599.

Broockman, D., & Kalla, J. (2022, April 1). Consuming cross-cutting media causes learning and moderates attitudes: A field experiment with Fox News viewers. https://doi.org/10.31219/osf.io/jrw26

Bruneau, E., Hameiri, B., Moore-Berg, S. L., & Kteily, N. (2021). Intergroup Contact Reduces Dehumanization and Meta-Dehumanization: Cross-Sectional, Longitudinal, and Quasi-Experimental Evidence From 16 Samples in Five Countries. Personality and Social Psychology Bulletin, 47(6), 906–920. https://doi.org/10.1177/0146167220949004

Cai, P. (2023, November 8). Selective Memory and Perceived Political Polarization. Retrieved from [URL]: https://ssrn.com/abstract=4598389 or http://dx.doi.org/10.2139/ssrn.4598389. Unpublished working paper.

Chang, E. H., Milkman, K. L., Gromet, D. M., Rebele, R. W., Massey, C., Duckworth, A. L., & Grant, A. M. (2019). The mixed effects of online diversity training. Proceedings of the National Academy of Sciences, 116(16), 7778–7783. https://doi.org/10.1073/pnas.1816076116

Chen, X., Lijffijt, J., & De Bie, T. (2018). Quantifying and Minimizing Risk of Conflict in Social Networks. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1197–1205. https://doi.org/10.1145/3219819.3220074.

Chitra, U., & Musco, C. (2020). Analyzing the Impact of Filter Bubbles on Social Network Polarization. Proceedings of the 13th International Conference on Web Search and Data Mining, 115–123. https://doi.org/10.1145/3336191.3371825.

Chong, A., De La O, A. L., Karlan, D., & Wantchekon, L. (2015). Does corruption information inspire the fight or quash the hope? A field experiment in Mexico on voter turnout, choice, and party identification. The Journal of Politics, 77(1), 55-71.

Clark, F. (2022). A comparative analysis of depolarization strategies. [Bachelor's thesis, University of Vermont]. ScholarWorks. https://scholarworks.uvm.edu/cgi/viewcontent.cgi?article=1532&context=hcoltheses

Cubillos Ramdohr, P. P. (2022). Affective vs. ideological polarization in a Latin American country: evidence from two survey experiments.

Diamond, M. J., & Lobitz, W. C. (1973a). When Familiarity Breeds Respect: The Effects of an Experimental Depolarization Program on Police and Student Attitudes toward Each Other. Journal of Social Issues, 29(4), 95–109. https://doi.org/10.1111/j.1540-4560.1973.tb00105.x.

Donkers, T., & Ziegler, J. (2021). The Dual Echo Chamber: Modeling Social Media Polarization for Interventional Recommending. Fifteenth ACM Conference on Recommender Systems, 12–22. https://doi.org/10.1145/3460231.3474261.

Enríquez, J. R., Larreguy, H., & Lujambio, O. (2023, September 29). How social media reinforces or ameliorates political polarization. Unpublished working paper.

Enríquez, J. R., Larreguy, H., Marshall, J., & Simpson, A. Accountability under Polarization. Unpublished working paper.

Fabbri, F., Wang, Y., Bonchi, F., Castillo, C., & Mathioudakis, M. (2022). Rewiring What-to-Watch-Next Recommendations to Reduce Radicalization Pathways. Proceedings of the ACM Web Conference 2022, 2719–2728. https://doi.org/10.1145/3485447.3512143.

Ferroni, M. F., Fisman, R., & Golden, M. (2023, August 31). How Tolerant are Legislators and Citizens of Corruption? Descriptive and Experimental Evidence from Three Countries (Version 1.1). Unpublished working paper.

Fishkin, J., Siu, A., Diamond, L., & Bradburn, N. (2021). Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on "America in One Room." American Political Science Review, 115(4), 1464–1481. https://doi.org/10.1017/S0003055421000642

Gao, M., Do, H. J., & Fu, W.-T. (2017). An Intelligent Interface for Organizing Online Opinions on Controversial Topics. Proceedings of the 22nd International Conference on Intelligent User Interfaces, 119–123. https://doi.org/10.1145/3025171.3025230.

Garimella, K., Morales, G. D. F., Gionis, A., & Mathioudakis, M. (2016). Reducing Controversy by Connecting Opposing Views. https://doi.org/10.48550/ARXIV.1611.00172.

Garimella, K., & Weber, I. (2017). A long-term analysis of polarization on Twitter (ICWSM 2017). In Proceedings of the eleventh international AAAI conference on web and social media. Retrieved from https://ingmarweber.de/wp-content/uploads/2017/05/A-Long-Term-Analysis-of-Polarization-on-Twitter.pdf

Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline *. The Quarterly Journal of Economics, 126(4), 1799–1839. https://doi.org/10.1093/qje/qjr044.

Gerber, A. S., Huber, G. A., & Washington, E. (2010). Party affiliation, partisanship, and political beliefs: A field experiment. American Political Science Review, 104(4), 720-744.

Ghosh, Arkadev & Kundu, Prerna & Lowe, Matt & Nellis, Gareth. (2023). Creating Cohesive Communities: A Youth Camp Experiment in India. Unpublished working paper.

Giese, H., Neth, H., Moussaïd, M., Betsch, C., & Gaissmaier, W. (2020). The echo in flu-vaccination echo chambers: Selective attention trumps social influence. Vaccine, 38(8), 2070–2076. https://doi.org/10.1016/j.vaccine.2019.11.038.

Gillani, N., Yuan, A., Saveski, M., Vosoughi, S., & Roy, D. (2018). Me, My Echo Chamber, and I: Introspection on Social Media Polarization. https://doi.org/10.48550/ARXIV.1803.01731.

Gottlieb, J., Adida, C. L., & Moussa, R. (2022). Reducing Misinformation in a Polarized Context: Experimental Evidence from Côte d'Ivoire. https://doi.org/10.26085/C3Q30T

Grady, C., Wolfe, R., Dawop, D., & Inks, L.. How contact can promote societal change amid conflict: An intergroup contact field experiment in Nigeria. Proceedings of the National Academy of Sciences (PNAS), Unpublished Working Paper.

Graham, M. H., & Svolik, M. W. (2020). Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States. American Political Science Review, 114(2), 392–409.

Greene, K., Rossiter, E., Seira, E., & Simpser, A. (2023, April 2). Interacting as Equals: How Contact Can Promote Tolerance Among Opposing Partisans. Retrieved from [URL]: https://ssrn.com/abstract=4456223 or http://dx.doi.org/10.2139/ssrn.4456223

Greitemeyer, T., & Schwab, A.K. (2014). Employing music exposure to reduce prejudice and discrimination. Aggressive behavior, 40 6, 542-51.

Grossetti, Q., Du Mouza, C., & Travers, N. (2019). Community-Based Recommendations on Twitter: Avoiding the Filter Bubble. In R. Cheng, N. Mamoulis, Y. Sun, & X. Huang (Eds.), Web Information Systems Engineering – WISE 2019 (Vol. 11881, pp. 212–227). Springer International Publishing. https://doi.org/10.1007/978-3-030-34223-4_14.

Gubler, J. R., Karpowitz, C. F., Monson, J. Q., Romney, D. A., & South, M. (2022). Changing Hearts and Minds? Why Media Messages Designed to Foster Empathy Often Fail. The Journal of Politics, 84(4), 2156-2171.

Guidi, S., Romano, A., & Sotis, C. (2021). Depolarizing the COVID Vaccine Passport Forum Collection: Vaccines and the Law. Yale Law Journal Forum, 131, 1010–1046.
Harwood, J., Qadar, F., & Chen, C.-Y. (2016). Harmonious Contact: Stories About Intergroup Musical Collaboration Improve Intergroup Attitudes. Journal of Communication, 66(6), 937–959. https://doi.org/10.1111/jcom.12261.

Hetherington, M., & Rudolph, T. (2015). Why Washington Won't Work: Polarization, Political Trust, and the Governing Crisis. Chicago: Univ. Chicago Press.

Houston, D. M. (2021). Polarization and the politics of education: What moves partisan opinion?. Educational Policy, 35(4), 566-589.

Interian, R., Moreno, J. R., & Ribeiro, C. C. (2021). Polarization reduction by minimum‐cardinality edge additions: Complexity and integer programming approaches. International Transactions in Operational Research, 28(3), 1242–1264. https://doi.org/10.1111/itor.12854.

Iyengar, S., & Westwood, S. (2015). Fear and loathing across party lines: New evidence on group polarization. American Journal of Political Science, 59(3), 690-707. https://doi.org/10.1111/ajps.12152

Janzen, O., Diekmann, I., Tsolak, D., & Salentin, K. (2023). Do guided mosque tours alleviate the prejudice of non-Muslims against Islam and Muslims? Evidence from a quasi-experimental panel study from Germany. Zeitschrift Für Religion, Gesellschaft Und Politik. https://doi.org/10.1007/s41682-023-00161-4.

Jeon, Y., Kim, B., Xiong, A., Lee, D., & Han, K. (2021). ChamberBreaker: Mitigating the Echo Chamber Effect and Supporting Information Hygiene through a Gamified Inoculation System. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 1–26. https://doi.org/10.1145/3479859.

Jha, S., Shayo, M., & Weiss, C. M. Financial Market Exposure Increases Generalized Trust, Particularly Among the Politically Polarized. Unpublished working paper.

KALLA, J., & BROOCKMAN, D. (2020). Reducing Exclusionary Attitudes through Interpersonal Conversation: Evidence from Three Field Experiments. American Political Science Review, 114(2), 410-425. doi:10.1017/S0003055419000923.

Karlsen, R., Steen-Johnsen, K., Wollebæk, D., & Enjolras, B. (2017). Echo chamber and trench warfare dynamics in online debates. European Journal of Communication, 32(3), 257–273. https://doi.org/10.1177/0267323117695734.

Kingzette, J., Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., & Ryan, J. B. (2021). How Affective Polarization Undermines Support for Democratic Norms. Public Opinion Quarterly, 85(2), 663–77.

Kvam, P. D., Alaukik, A., Mims, C. E., Martemyanova, A., & Baldwin, M. (2022). Rational inference strategies and the genesis of polarization and extremism. Scientific Reports, 12(1), Article 1. https://doi.org/10.1038/s41598-022-11389-0.

LaMarre, H., Knobloch-Westerwick, S. & Hoplamazian, G. (2012). Does the Music Matter? Examining Differential Effects of Music Genre on Support for Ethnic Group. Journal of Broadcasting & Electronic Media, 56:1, 150-167, DOI: 10.1080/08838151.2011.648683.

Lee, H. & Hahn, K. (2018). Partisan selective following on twitter over time: polarization or depolarization? Asian Journal of Communication, 28(3), 227-246, DOI: 10.1080/01292986.2017.1384845.

Lee, H., & Hahn, K. (2018). Partisan selective following on twitter over time: Polarization or depolarization? Asian Journal of Communication, 28(3), 227-246. https://doi.org/10.1080/01292986.2017.1384845

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. Nature human behaviour, 4(3), 279-286.

Levendusky, M. (2013). Partisan Media Exposure and Attitudes Toward the Opposition. Political Communication, 30(4), 565–581. https://doi.org/10.1080/10584609.2012.737435.

Levendusky, M. (2018). Americans, not partisans: Can priming American national identity reduce affective polarization? Journal of Politics, 80(1), 59-70. https://doi.org/10.1086/693987

Levendusky, M. S. (2018). Americans, not partisans: Can priming American national identity reduce affective polarization?. The Journal of Politics, 80(1), 59-70.

Levitsky, S., & Ziblatt, D. (2018). How Democracies Die. New York: Crown.

Lowe, M. (2021). Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration. American Economic Review, 111(6), 1807–1844. https://doi.org/10.1257/aer.20191780.

Lowe, M., & Jo, D. (2023, February 21). Legislature Integration and Bipartisanship: A Natural Experiment in Iceland. Unpublished working paper.

Madsen, J. K., Bailey, R. M., & Pilditch, T. D. (2018). Large networks of rational agents form persistent echo chambers. Scientific Reports, 8(1), 12391. https://doi.org/10.1038/s41598-018-25558-7.

Masullo, G. (2023). A new solution to political divisiveness: Priming a sense of common humanity through Facebook meme-like posts. New Media & Society. Advance online publication. https://doi.org/10.1177/14614448231184633.

Matias, J. N., Szalavitz, S., & Zuckerman, E. (2017). FollowBias: Supporting Behavior Change toward Gender Equality by Networked Gatekeepers on Social Media. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 1082–1095. https://doi.org/10.1145/2998181.2998287.

McCoy, J., & Somer, M. (2019). Toward a Theory of Pernicious Polarization and How It Harms Democracies: Comparative Evidence and Possible Remedies. The Annals of the American Academy of Political and Social Science, 681(1), 234–71.

McDermott, D. T., Brooks, A. S., Rohleder, P., Blair, K., Hoskin, R. A., & McDonagh, L. K. (2018). Ameliorating transnegativity: Assessing the immediate and extended efficacy of a pedagogic prejudice reduction intervention. Psychology & Sexuality, 9(1), 69–85. https://doi.org/10.1080/19419899.2018.1429487.

Messing, S., & Westwood, S. J. (2012). Selective Exposure in the Age of Social Media. Communication Research. https://doi.org/10.1177/0093650212466406.

Messing, S., & Westwood, S. J. (2014). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. Communication Research, 41(8), 1042–1063. https://doi.org/10.1177/0093650212466406.

Mousa, S. (2020). Building social cohesion between Christians and Muslims through soccer in post-ISIS Iraq. Science, 369(6505), 866–870. https://doi.org/10.1126/science.abb3153.

Mousa, S. (2021, October 26). Can “Vouching” for Stigmatized Individuals Overcome the Trust Deficit in Post-ISIS Iraq? Evidence from a Survey Experiment in Mosul. https://doi.org/10.17605/OSF.IO/KMZ5W

Mousa, S., Scacco, A., & Naumann, L. (2023, March 22). Intergroup Contact, Empathy Education, and Refugee-Native Integration in Lebanon. https://doi.org/10.17605/OSF.IO/FDBP3.

Munger, K. (2017). Tweetment Effects on the Tweeted: Experimentally Reducing Racist Harassment. Political Behavior, 39(3), 629–649. https://doi.org/10.1007/s11109-016-9373-5.

Munson, S. A., & Resnick, P. (2010). Presenting diverse political opinions: How and how much. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1457–1466. https://doi.org/10.1145/1753326.1753543.

Munson, S., Lee, S., & Resnick, P. (2021). Encouraging Reading of Diverse Political Viewpoints with a Browser Widget. Proceedings of the International AAAI Conference on Web and Social Media, 7(1), 419–428. https://doi.org/10.1609/icwsm.v7i1.14429.

Myrick, R. (2021). Do external threats unite or divide? Security crises, rivalries, and polarization in American foreign policy. International Organization, 75(4), 921-958. https://doi.org/10.1017/S0020818321000175

Neto, F., da Conceiçao Pinto, M., & Mullet, E. (2016). Can music reduce anti-dark-skin prejudice? A test of a cross-cultural musical education programme. Psychology of Music, 44(3), 388-398. https://doi.org/10.1177/0305735614568882.

Neto, F., da Conceiçao Pinto, M., & Mullet, E. (2016). Can music reduce anti-dark-skin prejudice? A test of a cross-cultural musical education programme. Psychology of Music, 44(3), 388–398. https://doi.org/10.1177/0305735614568882.

Paluck, E. L. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. Journal of Personality and Social Psychology, 96(3), 574-587. https://doi.org/10.1037/a0011989. PMID: 19254104.

Piazza, J. A. (2023). Political Polarization and Political Violence. Security Studies, 32(3), 476-504. DOI: 10.1080/09636412.2023.2225780

Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echo chambers on Facebook. SSRN Electronic Journal. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2795110

Ramaciotti Morales, P., & Cointet, J.-P. (2021). Auditing the Effect of Social Network Recommendations on Polarization in Geometrical Ideological Spaces. Fifteenth ACM Conference on Recommender Systems, 627–632. https://doi.org/10.1145/3460231.3478851.

Rhodes, S. C. (2022). Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals to Be Less Critical of Political Misinformation. Political Communication, 39(1), 1–22. https://doi.org/10.1080/10584609.2021.1910887.

Rogowski, J. C., & Sutherland, J. L. (2016). How ideology fuels affective polarization. Political behavior, 38, 485-508.

Ronconi, J. (2022, June). Divided for good: Football rivalries and social cohesion in Latin America. (Graduate Student Bravo Working Paper No. 2022-003). https://economics.brown.edu/sites/default/files/papers/Graduate%20Student%20Bravo%20Working%20Paper_2022-003.pdf.

Ruggeri, K., Većkalov, B., Bojanić, L., Andersen, T. L., Ashcroft-Jones, S., Ayacaxli, N., ... & Folke, T. (2021). The general fault in our fault lines. Nature Human Behaviour, 5(10), 1369-1380.

Santoro, E., & Broockman, D. E. (2022). The promise and pitfalls of cross-partisan conversations for reducing affective polarization: Evidence from randomized experiments. Science Advances, 8(25), eabn5515. https://doi.org/10.1126/sciadv.abn5515.

Saveski, Gillani, N., Yuan, A., Vijayaraghavan, P., & Roy, D. (2022). Perspective-Taking to Reduce Affective Polarization on Social Media. Proceedings of the International AAAI Conference on Web and Social Media, 16, 885–895. https://doi.org/10.1609/icwsm.v16i1.19343.
Scacco, A., & Warren, S. S. (2018). Can Social Contact Reduce Prejudice and Discrimination? Evidence from a Field Experiment in Nigeria. American Political Science Review, 112(3), 654–677. https://doi.org/10.1017/S0003055418000151.

Scacco, A., Siegel, A., & Weiss, C. M. (2023, July 26). Outgroup Avoidance. Unpublished working paper.

Schmid-Petri, H., & Bürger, M. (2022). The effect of misinformation and inoculation: Replication of an experiment on the effect of false experts in the context of climate change

communication. Public Understanding of Science, 31(2), 152–167. https://doi.org/10.1177/09636625211024550.

Smith, A., & Boas, T. (2020, August). Religion, sexuality politics, and the transformation of Latin American electorates [Paper presentation]. The Annual Meeting of the American Political Science Association 2020. https://www.ucis.pitt.edu/clas/sites/default/files/Smith%20Boas%202020%20Charlemos.pdf.

Sousa, M. D. R., Neto, F., & Mullet, E. (2005). Can music change ethnic attitudes among children? Psychology of Music, 33(3), 304-316. https://doi.org/10.1177/0305735605053735.

Sousa, M. D. R., Neto, F., & Mullet, E. (2005). Can music change ethnic attitudes among children? Psychology of Music, 33(3), 304–316. https://doi.org/10.1177/0305735605053735.

Strandberg, T., Olson, J. A., Hall, L., Woods, A., & Johansson, P. (2020). Depolarizing American voters: Democrats and Republicans are equally susceptible to false attitude feedback. Plos one, 15(2), e0226799.

Sunstein, C. R. (2018). Republic: Divided democracy in the age of social media (Third printing, and first paperback printing). Princeton University Press.

Velez, Y. R., & Lavine, H. (2017). Racial Diversity and the Dynamics of Authoritarianism. The Journal of Politics, 79(2), 519–533. https://doi.org/10.1086/688078.

Ventura, T., et al. (2023). WhatsApp Increases Exposure to False Rumors but has Limited Effects on Beliefs and Polarization: Evidence from a Multimedia-Constrained Deactivation. Available at SSRN: https://ssrn.com/abstract=4457400 or http://dx.doi.org/10.2139/ssrn.4457400.

Vuoskoski, J. K., Clarke, E. F., & DeNora, T. (2017). Music listening evokes implicit affiliation.Psychology of Music, 45(4), 584-599. https://doi.org/10.1177/0305735616680289.

Warner, B. R. (2010). Segmenting the Electorate: The Effects of Exposure to Political Extremism Online. Communication Studies, 61(4), 430–444. https://doi.org/10.1080/10510974.2010.497069.

Warner, B. R., Horstman, H. K., & Kearney, C. C. (2020). Reducing political polarization through narrative writing. Journal of Applied Communication Research, 48(4), 459-477. https://doi.org/10.1080/00909882.2020.1789195.

Zillmann, D., Aust, C. F., Hoffman, K. D., Love, C. C., Ordman, V. L., Pope, J. T., Seigler, P. D., & Gibson, R. J. (1995). Radical Rap: Does It Further Ethnic Division? Basic and Applied Social Psychology, 16(1-2), 1-25. doi:10.1080/01973533.1995.9646098.