Rethinking "Them":
Challenging Out-Group Stereotypes in Backsliding Democracies

Alper Sukru Gencer

Wilf Family Department of Politics, New York University

April 3, 2025

- ▶ Affective polarization (despising opponents rather than just disagreeing) is one of the gravest threats to a well-functioning democracy.
- ▶ Existing interventions (e.g., information correction and fact-checks) are often resource-intensive or prone to backfire.
- ▶ This is even more severe in backsliding democracies:
    - ▶ State-controlled media, regime propaganda, extremist platforms, grievances, and motivated reasoning challenge conventional mitigation methods.

- ▶ Research Questions:
    - ▶ Which interventions are more effective in mitigating misperceptions and affective polarization in backsliding regimes?

See **Appendix A** for additional context.

- ▶ Affective polarization (despising opponents rather than just disagreeing) is one of the gravest threats to a well-functioning democracy.
- ▶ Existing interventions (e.g., information correction and fact-checks) are often resource-intensive or prone to backfire.
- ▶ This is even more severe in backsliding democracies:
  - ▶ State-controlled media, regime propaganda, extremist platforms, grievances, and motivated reasoning challenge conventional mitigation methods.

- ▶ **Research Questions:**
  - ▶ Which interventions are more effective in mitigating misperceptions and affective polarization in backsliding regimes?

See **Appendix A** for additional context.

► **Misperceptions and Affective Polarization:**
  ► People overestimate out-group extremity, fueling social distance and hostility.
  ► These effects are amplified in backsliding democracies with state-controlled media and strong identity attachments.

► Core Hypotheses:
  ► **H1:** Correcting misperceptions reduces beliefs in out-group extremity.
  ► **H2:** Corrections reduce social distance and animosity.
► Conditional Hypotheses (Backsliding Context):
  ► **H3:** Interventions are less effective among regime supporters.
  ► **H4:** Non-confrontational interventions are more effective under motivated reasoning.
► Moderators: Prior beliefs, partisan strength, politicization.

See **Appendix B** for the full theoretical model.

- **Misperceptions and Affective Polarization:**
  - People overestimate out-group extremity, fueling social distance and hostility.
  - These effects are amplified in backsliding democracies with state-controlled media and strong identity attachments.

- **Core Hypotheses:**
  - **H1:** Correcting misperceptions reduces beliefs in out-group extremity.
  - **H2:** Corrections reduce social distance and animosity.

- Conditional Hypotheses (Backsliding Context):
  - **H3:** Interventions are less effective among regime supporters.
  - **H4:** Non-confrontational interventions are more effective under motivated reasoning.

- Moderators: Prior beliefs, partisan strength, politicization.

See **Appendix B** for the full theoretical model.

- **Misperceptions and Affective Polarization:**
  - People overestimate out-group extremity, fueling social distance and hostility.
  - These effects are amplified in backsliding democracies with state-controlled media and strong identity attachments.

- **Core Hypotheses:**
  - **H1:** Correcting misperceptions reduces beliefs in out-group extremity.
  - **H2:** Corrections reduce social distance and animosity.

- **Conditional Hypotheses (Backsliding Context):**
  - **H3:** Interventions are less effective among regime supporters.
  - **H4:** Non-confrontational interventions are more effective under motivated reasoning.

- **Moderators:** Prior beliefs, partisan strength, politicization.

See **Appendix B** for the full theoretical model.

- **Misperceptions and Affective Polarization:**
    - People overestimate out-group extremity, fueling social distance and hostility.
    - These effects are amplified in backsliding democracies with state-controlled media and strong identity attachments.

- **Core Hypotheses:**
    - **H1:** Correcting misperceptions reduces beliefs in out-group extremity.
    - **H2:** Corrections reduce social distance and animosity.
- **Conditional Hypotheses (Backsliding Context):**
    - **H3:** Interventions are less effective among regime supporters.
    - **H4:** Non-confrontational interventions are more effective under motivated reasoning.
- **Moderators:** Prior beliefs, partisan strength, politicization.

See **Appendix B** for the full theoretical model.

▶ **Challenge 1:** Little data on outgroup perceptions due to lack of transparency.
  ▶ Based on stereotypes literature, I use survey questions to find policy issues with most misperceptions.

▶ Challenge 2: High levels of motivated reasoning in backsliding democracies.
  ▶ I use socio-psychological literature to devise an effective intervention.

▶ Two Intervention Strategies:
  1. Incentivized Accuracy: Incentivized gx'uess statistics about out-group, get feedback (conventional as baseline).
  2. Conversation Exposure: Simulated social media chats showing moderate, relatable views from the out-group.
  3. Placebo: Read neutral info about trees.

More details in **Appendix C**.

▶ **Challenge 1:** Little data on outgroup perceptions due to lack of transparency.
  ▶ Based on stereotypes literature, I use survey questions to find policy issues with most misperceptions.

▶ **Challenge 2:** High levels of motivated reasoning in backsliding democracies.
  ▶ I use socio-psychological literature to devise an effective intervention.

▶ Two Intervention Strategies:
  1 Incentivized Accuracy: Incentivized gx'uess statistics about out-group, get feedback (conventional as baseline).
  2 Conversation Exposure: Simulated social media chats showing moderate, relatable views from the out-group.
  3 Placebo: Read neutral info about trees.

More details in **Appendix C**.

- ▶ **Challenge 1:** Little data on outgroup perceptions due to lack of transparency.
  - ▶ Based on stereotypes literature, I use survey questions to find policy issues with most misperceptions.

- ▶ **Challenge 2:** High levels of motivated reasoning in backsliding democracies.
  - ▶ I use socio-psychological literature to devise an effective intervention.

- ▶ **Two Intervention Strategies:**
  1. **Incentivized Accuracy:** Incentivized gx'uess statistics about out-group, get feedback (conventional as baseline).
  2. **Conversation Exposure:** Simulated social media chats showing moderate, relatable views from the out-group.
  3. **Placebo**: Read neutral info about trees.

More details in **Appendix C**.

- Meta (Facebook/Instagram) recruitment (Jan 2024).
- N = 906 Turkish citizens (18+).
- Random assignment to 3 arms.
- Balance across demographics and political alignment.

See **Appendix D** for tables and recruitment details.

Perceptions of Extremism:

▶ Regime supporters: 37–60% see opposition as willing to take undemocratic action.

▶ Opposition: 65–80% see gov't side as likely to commit undemocratic acts.
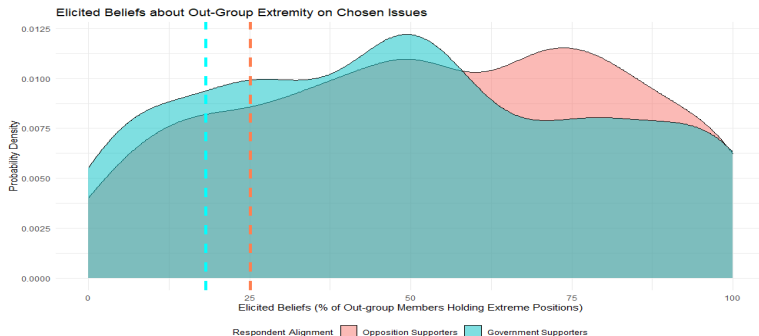
Social Distance:

▶ 50–73% of opposition uncomfortable with friendships or neighbors from the other side.

Negative Traits:

▶ Opposition sees gov't supporters as close-minded (83%), unintelligent (74%).

▶ Regime supporters attribute fewer negative traits.

See **Appendix E-1** for full descriptive tables.

Perceptions of Extremism:
- ▶ Regime supporters: 37–60% see opposition as willing to take undemocratic action.
- ▶ Opposition: 65–80% see gov't side as likely to commit undemocratic acts.

Social Distance:
- ▶ 50–73% of opposition uncomfortable with friendships or neighbors from the other side.

Negative Traits:
- ▶ Opposition sees gov't supporters as close-minded (83%), unintelligent (74%).
- ▶ Regime supporters attribute fewer negative traits.

See **Appendix E-1** for full descriptive tables.

Perceptions of Extremism:

- ▶ Regime supporters: 37–60% see opposition as willing to take undemocratic action.
- ▶ Opposition: 65–80% see gov't side as likely to commit undemocratic acts.

Social Distance:

- ▶ 50–73% of opposition uncomfortable with friendships or neighbors from the other side.

Negative Traits:

- ▶ Opposition sees gov't supporters as close-minded (83%), unintelligent (74%).
- ▶ Regime supporters attribute fewer negative traits.

See **Appendix E-1** for full descriptive tables.

Elicited Beliefs about Out-Group Extremity on Chosen Issues

▶ Suggests misperception and motivated reasoning.

See **Appendix E-1** for full descriptive tables.
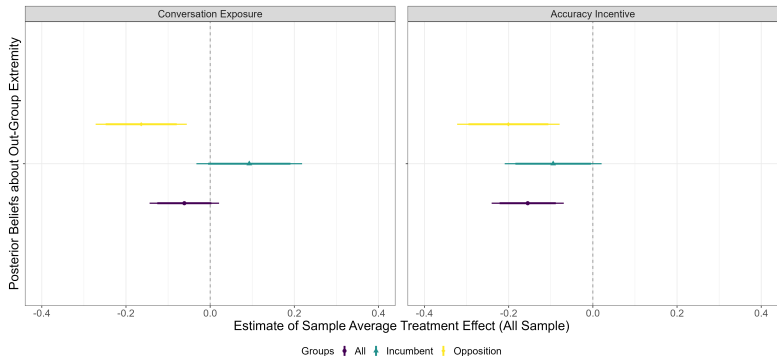
- ▶ Both interventions reduced perceived extremity.
- ▶ More effective on opposition supporters.

See **Appendix F** for full statistical models.
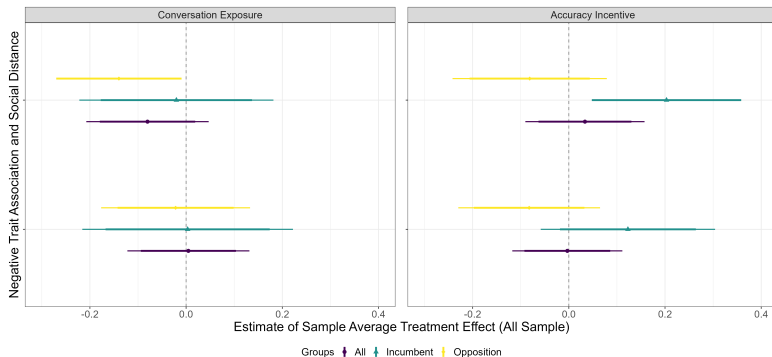
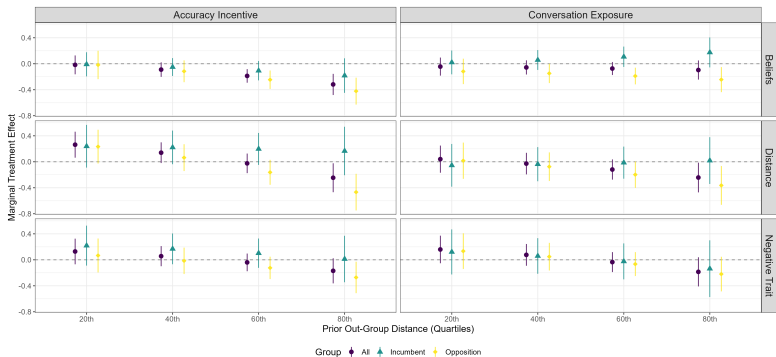- ▶ Opposition supporters reduced social distance (affective polarization).
- ▶ Regime supporters showed backlash to the conventional approach.
- ▶ No significant change in negative trait association (partisan animosity).

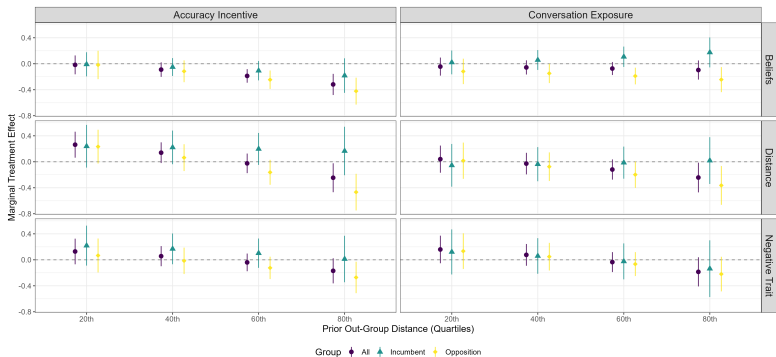See **Appendix F** for full statistical models.

▶ **Good news:** Both interventions are effective opposition supporters.
  ▶ High prior (opposition) participants are more responsive to interventions.
  ▶ Reduction in negative trait association among high prior opposition.
  ▶ Conversation is successful among highly politicized and partisan individuals.

▶ Bad news: Incentivized accuracy increases polarization
  ▶ among government supporters regardless of priors.
  ▶ Even among those without misperceptions, whereas conversation not.

See Appendix G.

► **Good news:** Both interventions are effective opposition supporters.
  ► High prior (opposition) participants are more responsive to interventions.
  ► Reduction in negative trait association among high prior opposition.
  ► Conversation is successful among highly politicized and partisan individuals.

► **Bad news:** Incentivized accuracy increases polarization
  ► among government supporters regardless of priors.
  ► Even among those without misperceptions, whereas conversation not.

See **Appendix G**.

▶ Both interventions can correct out-group misperceptions.

▶ Conversation interventions is promising for backsliding democracies.

▶ Information-based treatment risks backlash.
  ▶ This is especially true for those who follow conventional media.
  ▶ State-captured media and state propaganda worsens motivated reasoning.

▶ **Future**: Tailored and intentional intervention designs.

Thank you for listening!

For any questions and feedback:

Alper Sukru Gencer
alper.gencer@nyu.edu

Appendix A: Extended Motivation & Background

- ▶ **Problem**: Affective polarization and partisan animosity in deeply divided societies, such as backsliding democracies and post-conflict societies.
- ▶ **Research Gap**: Little work on misperceptions in these settings:
    - ▶ Higher baseline polarization, misperceptions, social distrust.
    - ▶ State-captured media primes extremist perceptions (Levendusky and Malhotra, 2016; Iyengar et al., 2019).
- ▶ **Conventional strategies fail**: prone to backlash (Taber and Lodge, 2006; Nyhan, 2021), resource intensive (Yildirim et al., 2023).

- ▶ **People overestimate extremity** (Ahler and Sood, 2018; Lees and Cikara, 2020).
- ▶ **Misperceptions cause affective polarization**: social distance, animosity (Druckman et al., 2022; Voelkel et al., 2023).
- ▶ **Role of media**: State-controlled media amplifies stereotypes (Shirikov, 2024).
- ▶ **Stereotype focus**: Attention drawn to extreme traits (Bordalo et al., 2016).
- ▶ **Implication**: Need to identify and correct most overestimated traits.

Appendix C: Intervention Details

Incentivized Accuracy Group:

▶ Participants guessed what percent of out-group holds extreme policy views.

▶ Feedback with true values provided.

▶ Raffle bonus for accurate guesses.

Conversation Exposure Group:

▶ Read simulated Facebook-style dialogues featuring out-group members.

▶ Conversations reflected moderate, humanizing political narratives.

Placebo Group:

▶ Read neutral passage on tree species in Turkey.

Appendix C: Intervention Details

Appendix D: Sample & Balance Checks

- ▶ **Recruitment:** Meta ads from Jan 24–31, 2024 targeting 18+ Turkish citizens.
- ▶ **Reached:** 127,000 users.
- ▶ **Final N:** 906 complete responses.
- ▶ **Incentives:** 200 TL raffle coupon entry.
- ▶ **Balance:** Groups were balanced across political, demographic lines.

Appendix D: Sample & Balance Checks

Table 1: Summary Table.

	Mean	Std.Dev.	Min	Max
Woman	0.476	0.500	0	1
College Graduate	0.535	0.499	0	1
18-24	0.240	0.427	0	1
25-34	0.259	0.439	0	1
35-54	0.298	0.458	0	1
+55	0.203	0.403	0	1
Turkish Ethnicity	0.860	0.347	0	1
Minority	0.352	0.478	0	1
Lower SES	0.211	0.408	0	1
Lower-Mid SES	0.291	0.455	0	1
Middle SES	0.432	0.496	0	1
Upper-Mid SES	0.061	0.239	0	1
Upper SES	0.006	0.074	0	1
Previously Voted	0.918	0.274	0	1
Voted for Ruling Coalition	0.396	0.489	0	1
Interested in Politics	0.630	0.483	0	1
Following Political News	0.539	0.499	0	1
Following News on Internet	0.668	0.471	0	1
Feeling Close to Party	0.294	0.456	0	1
Politically Knowledgeable	0.331	0.471	0	1
Full Democratic Support	0.721	0.449	0	1
Partial Democratic Support	0.912	0.284	0	1
Correction Group	0.336	0.472	0	1
Conversation Group	0.278	0.448	0	1

Appendix D: Sample & Balance Checks

Table 2: Balance Table (Mean Values).

	All Sample	Placebo	Correction	Conversation	F-test p-value
Woman	0.476	0.503	0.444	0.476	0.440
College Graduate	0.535	0.520	0.553	0.536	0.653
18-24	0.240	0.237	0.234	0.250	0.738
25-34	0.259	0.269	0.253	0.254	0.668
35-54	0.298	0.297	0.316	0.278	0.667
+55	0.203	0.197	0.197	0.218	0.547
Turkish Ethnicity	0.860	0.860	0.872	0.845	0.658
Minority	0.352	0.329	0.375	0.357	0.414
Lower SES	0.211	0.234	0.194	0.198	0.253
Lower-Mid SES	0.291	0.254	0.336	0.290	0.256
Middle SES	0.432	0.440	0.411	0.444	0.981
Upper-Mid SES	0.061	0.066	0.053	0.063	0.857
Upper SES	0.006	0.006	0.007	0.004	0.799
Previously Voted	0.918	0.914	0.938	0.901	0.652
Voted for Ruling Coalition	0.396	0.411	0.391	0.381	0.441
Interested in Politics	0.630	0.600	0.641	0.659	0.131
Following Political News	0.539	0.540	0.526	0.552	0.818
Following News on Internet	0.668	0.680	0.638	0.687	0.969
Feeling Close to Party	0.294	0.314	0.276	0.286	0.405
Politically Knowledgeable	0.331	0.311	0.345	0.341	0.410
Full Democratic Support	0.721	0.683	0.724	0.770	0.019
Partial Democratic Support	0.912	0.920	0.891	0.925	0.960
Correction Group	0.336	0.000	1.000	0.000	NA

Appendix E-1: Descriptive Statistics

Table 3: Social Distance Questions by Partisanship

I would feel uncomfortable if	Gov. Supporters (%)	Opp. Supporters (%)	P-Value
My children become friends with out-group members	25.7	52.9	0.00
Doing business with supporters of the other party	21.5	35.0	0.01
If my neighbors are out-group members	26.4	50.5	0.00
If my child married someone who supports the government	25.0	44.2	0.00

Source: Author's Original Survey Experiment (N = 350). The percentages represent participants who strongly or somewhat agree with the statements.

- ▶ **53%** of opposition supporters would feel uncomfortable if their child befriends an out-group member.
- ▶ **50.5%** of opposition supporters and **26.4%** of government supporters are uneasy with out-group neighbors.
- ▶ Opposition supporters express higher levels of discomfort in intergroup relationships than government supporters.

Appendix E-1: Descriptive Statistics

Table 4: Negative Trait Association Questions by Partisanship

Compared to in-group supporters, out-group supporters are more	Gov. Supporters (%)	Opp. Supporters (%)	P-Value
gray!6Close-minded	gray!648.6	gray!682.5	gray!60.000
Immoral	29.9	54.9	0.000
gray!6Lazy	gray!647.2	gray!662.6	gray!60.006
Dishonest	45.8	68.0	0.000
gray!6Unintelligent	gray!644.4	gray!673.8	gray!60.000
Traitorous	29.9	40.8	0.048

Source: Author's Original Survey Experiment (N = 350). The percentages represent participants who strongly or somewhat agree with the statements.

▶ Over **40%** of opposition supporters attribute negative traits to government supporters.

▶ Particularly high among opposition: close-minded (83%) and unintelligent (74%).

▶ Among government supporters, negative perceptions range from 30-50%.

Appendix F: Experimental Results

Posterior Belief Correction:

▶ Significant drop in out-group extremity beliefs in both treatments.

Social Distance & Trait Views:

▶ Some reduction in social distance, not significant for trait views.

▶ Descriptives: Social Distance M=2.1 (SD=1.0), Traits M=2.5 (SD=0.9).

By Prior Beliefs:

▶ Lower priors show stronger treatment effects.

Backlash Effects:

▶ Gov't supporters showed increased distance post-info treatment.

By Politicization & Strength:

▶ More politicized participants resisted correction more.

Ahler, D. J. and Sood, G. (2018). The parties in our heads: Misperceptions about party composition and their consequences. *The Journal of Politics*, 80(3):964–981.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.

Druckman, J. N., Klar, S., Krupnikov, Y., Levendusky, M., and Ryan, J. B. (2022). (mis) estimating affective polarization. *The Journal of Politics*, 84(2):1106–1117.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual review of political science*, 22:129–146.

Lees, J. and Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, 4(3):279–286.

Levendusky, M. and Malhotra, N. (2016). Does media coverage of partisan polarization affect political attitudes? *Political Communication*, 33(2):283–301.

Nyhan, B. (2021). Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences*, 118(15):e1912440117.

Shirikov, A. (2024). Rethinking propaganda: How state media build trust through belief affirmation.

Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American journal of political science*, 50(3):755–769.

Voelkel, J. G., Chu, J., Stagnaro, M. N., Mernyk, J. S., Redekopp, C., Pink, S. L., Druckman, J. N., Rand, D. G., and Willer, R. (2023). Interventions reducing affective polarization do not necessarily improve anti-democratic attitudes. *Nature human behaviour*, 7(1):55–64.

Yildirim, M. M., Nagler, J., Bonneau, R., and Tucker, J. A. (2023). Short of suspension: How suspension warnings can reduce hate speech on twitter. *Perspectives on Politics*, 21(2):651–663.