



Restoring institutional confidence in backsliding democracies: Evidence from Mexico

Brett Bessen^{a,1}, Susan Stokes^{b,1,2} , and Andres Uribe^{c,1}

Affiliations are included on p. 9.

Contributed by Susan Stokes; received June 24, 2025; accepted October 27, 2025; reviewed by Alejandro Monsivais Carrillo and Kurt Weyland

Declining confidence in public institutions afflicts many democracies, a trend apparently exacerbated by backsliding leaders. These are leaders who gradually undermine the institutions that sustain democratic competition and accountability. Does the rhetoric of backsliders undermine the public's confidence in the institutions under attack and can rebuttals of presidential diatribes restore this confidence? We explore the impact of backsliding leaders' anti-institutional rhetoric in the context of Mexico. With text-as-data analyses, we demonstrate the harshness of President Andrés Manuel López Obrador's (2018–2024) anti-institutional diatribes against the agency that oversees national elections. With survey experiments, we demonstrate that these diatribes can indeed undermine public confidence. Yet our research also uncovers the potential for rebuttals to restore confidence. Counternarratives offered by organizations viewed as above the fray of Mexican politics restored public confidence—surprisingly, even among the president's supporters. Our findings suggest strategies for breaking out of the cage of intense partisanship and countering democracy-degrading rhetoric. Though presidential haranguing of democratic institutions can have a powerful effect, there remains room for public confidence to be restored by more positive accounts.

democratic backsliding | democratic erosion | confidence in elections | Mexico | misinformation

A loss of confidence in public institutions afflicts many democracies. Brader and Kent document a 50-y decline in Americans' confidence in a long list of political and societal institutions, including Congress, the executive branch, the presidency, and the Supreme Court (1). Similar trends have been documented for a number of other democracies (2).

Complex reasons lie behind this decline. But among them are politicians' statements and narrative frames, which sometimes appear designed to increase public skepticism. Democratic backsliders, in particular, have good reasons to want the public to view coequal branches and agencies of government as ineffective, incompetent, and corrupt. Presidents and prime ministers who seek to aggrandize their power and undermine accountability find it helpful to pursue this strategy of *trash-talking democracy* (3, 4). Voters in effect say to themselves, *I may not like the president's attacks on the press or his stacking of the courts, but he tells me that these institutions are hollow and corrupt; we need him in office to make democracy operate as it is supposed to*. A skeptical population is less likely to object when leaders act to undermine these institutions (3–6).

Sometimes the narratives that these politicians craft about institutions are simply false, as when Donald Trump claimed that election regulators in several states had deprived him of victory in the 2020 presidential election (7). Yet backsliding leaders also offer dark visions that are not in all instances false but that render highly misleading pictures. Examples are Trump's vilification of the federal bureaucracy as a “deep state” and of federal agencies as “incompetent and really corrupt.”*

Former Mexican President Andrés Manuel López Obrador's depictions of federal agencies and coequal branches as deeply corrupt fall into this latter category: not entirely groundless but highly misleading. This is true of his statements about Mexico's election administration body, the focus of this study. Internationally, and indeed among many Mexicans, the body is regarded as highly professional and independent, and as having played a central role in Mexico's democratization in the 1990s (8, 9). Measures of electoral integrity derived from Varieties of Democracy surveys confirm these positive assessments (see ref. 10, 488). The image conjured by President López Obrador's statements diverged sharply from this view.

Significance

The rise of democratic backsliders, leaders and ministers who win elections fairly but then undermine democratic institutions, is a major challenge to 21st-century democracies. Backsliding leaders criticize coequal branches of government and public agencies, perhaps reducing public confidence in democratic institutions. Is their rhetoric effective? And can it be rebutted? We explore these questions in Mexico. A Mexican president's diatribes reduced institutional confidence. But rebuttals attributed to actors outside the partisan fray restored it. Supporters of the backsliding leader, whom we might have expected to be immune to rebuttals, were moved by these rebuttals to view the institution as more effective and honest. Our research demonstrates that well-attributed rebuttals can counter backsliders' democracy-degrading rhetoric.

Author contributions: B.B., S.S., and A.U. designed research; performed research; analyzed data; and wrote the paper.

Reviewers: A.M.C., El Colegio de la Frontera del Norte; and K.W., University of Texas, Austin.

The authors declare no competing interest.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹B.B., S.S., and A.U. contributed equally to this work.

²To whom correspondence may be addressed. Email: sstokes@uchicago.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2516520122/-DCSupplemental>.

Published December 3, 2025.

*Donald Trump, February 11, 2025, Oval Office news conference. Trump was referring to the U.S. Agency for International Development.

This study asks: Is backsliders' anti-institutional rhetoric effective? Does it indeed sully the image of vital institutions, in voters' eyes? And, critically, are counterarguments capable of restoring public confidence? We offer answers to these questions in the context of Mexico during the administration of President López Obrador (2018–2024).

Much recent research into backsliding presidents and electoral institutions has focused on the United States and Europe. We broaden the lens by studying Mexico, a large and important democracy in which attacks on democratic institutions have been intense and, in part, successful.[†] For instance, Mexico has recently moved to popular election as the method for selecting all federal judges, a move widely seen as bringing the judiciary more firmly under the control of the ruling party (15). In some ways Mexico is a "hard" case in which to demonstrate the effectiveness of democratic trash-talk; confidence in democratic institutions is lower at baseline in this middle-income, new democracy than in a wealthy and longstanding one like the United States (16, 17, 454). Experimental treatments might be less likely to reduce this confidence to a point below an already-low floor.

We also broaden the lens with regard to the nature of the institution under attack. Recent comparative studies of democratic erosion focus on presidents undermining legislatures and courts and find that populist rhetoric and democratic trash-talk boost public support for executive overreach (3, 5). Extending this research to elections and electoral administration bodies can shed light on a key institution that has come under attack in major democracies such as the United States, Brazil, and Mexico. The public's loss of faith in the integrity of elections, furthermore, can severely debilitate democracies. Hence, our focus on López Obrador and the National Electoral Institute (Instituto Nacional Electoral, INE). Throughout his term, López Obrador waged a rhetorical campaign against the INE, accusing it of being corrupt, partisan, and ineffective. The INE was highly salient in the López Obrador years, with the government undertaking several efforts to weaken it and large numbers of citizens taking to the streets to rally in its support.

A further contribution of our study is our simultaneous exploration of democratic trash-talk and rebuttals of it. *That is, we ask both what impact does presidential trash-talk have on a key institution and can more positive statements about these same institutions reverse the effects of democratic trash-talk?* Prior studies have tended to look at these effects separately. Thus, since Donald Trump's first presidential run, scholars have explored whether his diatribes against institutions influence public perceptions. (They have concluded that they do; see refs. 18–21.) Other scholars have explored the effectiveness of rebuttals of Trump's attacks. (Here the findings are mixed; see refs. 18 and 22.) By contrast, we combine these two steps of the process, exposing the same sets of survey respondents first to democratic trash-talk and then to rebuttals of it.

We go to some lengths to design our tests so that they parallel real-world dynamics. Yet there are limits to these parallels. In our study, people read the president's harsh words about an institution; a subset is then exposed to a more positive (and indeed more accurate) account. The rebuttals are in important ways effective. Yet in reality, the president has communicative advantages, which our study does not capture. In our design, the rebutters in effect get the last word. While prebunking studies—where rebuttals precede misleading information—suggest that this ordering could be reversed (23), the incumbent has structural

advantages in the battle over public opinion. A president can respond and repeat his attacks and do so through a loud bull-horn. Future research should consider avenues by which pro-democracy groups can compete for attention to dispute misleading claims.

In what follows, we first examine the nature of then-president López Obrador's rhetorical attacks. Using computational text analysis methods, we show that the president launched a remarkable barrage of rhetorical ammunition against the INE. We then turn to the consequences of this rhetoric. We deploy two preregistered survey experiments to demonstrate that these verbal attacks do, in fact, influence the public. In general, López Obrador's tongue-lashings of the INE reduced our respondents' confidence in it. But how respondents reacted depended on their partisanship: Exposure to the president's diatribes reduced confidence in the INE among López Obrador's base but had no detectable effect among opposition-party supporters.

Crucially, these rhetorical attacks could be effectively rebutted. We identify two factors that jointly influence people's responses to the rebuttals: the *source* to which the rebuttal is attributed and the *partisan preference* of the respondent. Rebuttals attributed to international experts restored confidence in the INE; those attributed to opposition parties or to domestic experts were less effective. Turning to voter partisanship, though supporters of the president and of the opposition held different views at baseline, the president's supporters did respond to rebuttals. As long as rebuttals did not emanate from suspect sources, his supporters were swayed to seeing the INE as a more credible institution. Indeed, rebuttals undid the much of the impact of the president's institutional criticisms.

López Obrador's Institutional Discourse

Andrés Manuel López Obrador was elected President of Mexico in 2018 on a leftist, populist agenda, proclaiming his intention to enact sweeping wealth redistribution and anticorruption policies. López Obrador had run for president twice before, in 2006 and 2012. In 2006, he narrowly lost a bitter contest to conservative Felipe Calderón. López Obrador claimed fraud and demanded a national recount of the vote, but the Federal Electoral Court found little evidence of fraud or miscounts and denied his request. López Obrador ran for the presidency again in 2012, this time losing by a larger margin; again he claimed fraud and protested. In 2018, he constructed a new political coalition, under the banner of the National Regeneration Movement (*Movimiento de Regeneración Nacional*, Morena), and swept to victory.

In office, López Obrador implemented parts of the progressive agenda that he had promised, emphasizing poverty reduction efforts and infrastructure investment. He tripled the minimum wage, created a new system of cash transfers for the elderly, and oversaw a reduction in the national poverty rate of over 25%. But he also pursued elements of an austerity program, curtailing access to healthcare and defunding higher education and science (24, 25).

And he attempted to undermine the constraints and accountability mechanisms that bind the presidency. López Obrador clashed with the courts, expanded control over universities and civil society, and demonized the press and opposition parties. Given his attempts to undermine the independence of coequal institutions and agencies, scholars identify him as among the list of democratic backsliders (4, 11, 14, 26).

Perhaps the most high-profile of López Obrador's institutional fights was with the INE, the autonomous institution charged with administering national elections. The INE is a crucial institution

[†] For important studies of democratic backsliding in Mexico, see refs. 10–13 and 14.

in Mexican democracy. In the late 1990s, its predecessor, the Federal Election Institute (IFE), played a key role in Mexico’s democratic transition (14). In the 21st century, international observers have consistently characterized the INE as competent, independent, and nonpartisan.

López Obrador disagreed. He never stopped blaming the IFE for overseeing and ratifying his 2006 electoral defeat. In office, he clashed with its successor, the INE, over a range of issues, from campaign finance rules to restrictions on his use of the office of the presidency to campaign for Morena candidates. In 2022, López Obrador advanced a constitutional reform to dismantle the INE. When this reform floundered in the legislature, he pivoted to a legislative package he referred to as “Plan B,” which sought to reduce the funding and autonomy of the institution. In 2023, Plan B was approved by the Morena-controlled Congress but was later struck down by the Supreme Court amid public protests in support of the institution. López Obrador continued to clash with the INE in the run-up to the June 2024 presidential elections, which his chosen successor, Claudia Sheinbaum, won easily.

Analyzing López Obrador’s Discourse. López Obrador’s anti-INE rhetoric was indeed extraordinarily harsh—as harsh, we will show, as the rhetoric that he directed against the partisan opposition. We analyze that rhetoric here. In addition to demonstrating the nature and harshness of his rhetoric, this analysis shapes our choices of what kinds of statements to include in the survey experiments described later, boosting the external validity of the treatments.

To give a taste of the president’s discourse, consider his depictions of the INE:

- It was created because the “authoritarian, antidemocratic, corrupt regime needed it to commit fraud, so they could stay in power, so they could keep looting” from the people (April 29, 2021).[‡]
- Its leadership is “conservative, corrupt,” and “violates the Constitution by earning more than the president” (May 9, 2023).
- Its defenders want “to maintain an instrument that permits them to use electoral fraud to conserve the privileges of the corrupt, conservative oligarchy that has dominated Mexico for a long time” (November 17, 2022).

Quantitative textual analysis reveals that these sentiments were constant themes of López Obrador’s rhetoric toward the INE. We constructed a large corpus of every statement he made during his morning press conferences or *mañaneras*, held almost every weekday during his term. In these *mañaneras*, the president fielded questions from reporters, made policy pronouncements, and opined about news and politics. The press conferences often lasted multiple hours, shaping that day’s news cycle and driving discussions in both legacy and social media. The office of the presidency transcribed each *mañanera* and uploaded the transcription to its website. We scraped, compiled, and cleaned these transcripts to construct our corpus.

Our sample runs from December 7, 2018, to July 16, 2024, shortly after the presidential elections that selected López Obrador’s successor, Claudia Sheinbaum. Given the size of this corpus, we analyze López Obrador’s speech using computational methods. After removing utterances by other speakers—

[‡]The president was referring to prior elected governments after the fall of the single-party regime in 2000.

Table 1. Frequency of López Obrador’s rhetorical attacks on the INE

	Corrupt	Incompetent	Expensive	Conservative
Attacks	206	183	93	28

journalists, invited speakers, and other government officials—and short fragments, we segmented the full sample of López Obrador’s speech into sentences. This leaves us with nearly 420,000 sentences.

We first use simple keyword searches to identify sentences that mention the INE or other political institutions. To situate López Obrador’s discussions of the INE in the broader context of his political rhetoric, we track references to three other government institutions: the judiciary, the Congress, and the military. We also compare his language about the INE to his discussions of the major political parties: López Obrador’s Morena and the two major opposition parties, the Institutional Revolutionary Party (PRI) and the National Action Party (PAN). (We define our keyword searches to capture references to the institutions and major individuals associated with them. See *SI Appendix, Table S1* for details.)

Having identified sentences in which the president mentions political institutions, we then employ a large language model—OpenAI’s GPT-4o model—to measure the overall valence of each sentence—whether it describes the institution in a positive, negative, or neutral tone (27, 28). We then select four rhetorical frames or dimensions that López Obrador frequently employed in discussing institutions and their personnel: honesty, political ideology, competence, and cost-effectiveness. We thus assess whether López Obrador referenced institutions as honest or corrupt; conservative or liberal; competent or incompetent; and efficient or spendthrift. The prompts we use for each query and the results of human validation are presented in *SI Appendix, Tables S2–S4*.

López Obrador’s tone toward the INE was harsh. Table 1 shows the frequency of the president’s distinct frames of attack in discussing the INE—i.e., how many sentences invoke the “negative” side of each of the four dimensions of honesty, political ideology, competence, and cost-effectiveness. López Obrador most often criticized the INE as corrupt or dishonest, a critique he invoked more than 200 times. But he also sometimes focused on what he viewed as other faults in the INE: its incompetence (183 mentions), its wastefulness and inflated salaries (93 mentions), or its bias in favor of conservative interests (28 mentions).

Fig. 1 allows us to gauge the president’s anti-INE diatribes against his tone when discussing other institutions. It plots the average score for each rhetorical target along each dimension. We scale each sentence from −1 (bad, incompetent, conservative, corrupt, and expensive) to 0 (neutral/neither) to 1 (good, competent, liberal, honest, and inexpensive) and then average across all references to each institution. We also record the number of references to each target.

The INE was not López Obrador’s most common target; he directed more invective toward the courts. (Among his complaints about the courts, however, was their repeatedly getting in the way of his proposals for restructuring the INE.) But the president portrayed the INE in a more negative light than any other government institution, a degree of negativity only rivaled by his characterization of his political opposition. (Observers of Mexican politics in the López Obrador years will not be surprised by the frequency with which the president lavished praise on the Mexican military, his favorite institution.)

The INE was an outlier in the extent to which López Obrador characterized it as incompetent, corrupt, and, particularly, as wasteful of resources—a charge he almost never levied against other targets. Across these dimensions, he leveraged a language of criticism largely reserved for his electoral opponents, proclaiming the INE as just bad as—and, often, beholden to—his political enemies.

Elite Discourse and Confidence in Democratic Institutions: Theory and Expectations

We have seen that the president directed an acerbic discourse at a key institution whose independence he hoped to reduce. But did his rhetoric shake public confidence in Mexico's principal electoral institution?

There are reasons to expect that López Obrador's attacks would in general undermine confidence in the INE. Democracy-degrading discourse by backsliders tends to work, reducing confidence in the target institution (3, 19). In theory, such discourse functions like other forms of persuasion: by introducing new ideas about targeted institutions or by rebalancing the importance of existing beliefs (29). If Mexicans accept the message that the INE is corrupt, inefficient, and costly, their confidence in it would decline. Looking toward our experiment, *we expect that exposure to López Obrador's discursive attacks would, on average, reduce confidence in the National Electoral Institute.*

There are several reasons to seek experimental evidence in this setting. Monsiváis Carrillo (30) offers observational evidence of the persuasive power of López Obrador's anti-INE discourse (30). And in other settings, such as in the United States during President Trump's attack on the 2020 presidential election, rhetorical attacks depress confidence in elections (18, 20, 21). Yet the causal arrow could go in the other direction: It might be that voters with low trust in the political system would be more prone to support an institutional critic like López Obrador. Our experimental evidence from Mexico can help to establish whether attacks reduce confidence, whatever selection effects might also be at play.

Our survey experiments can also help disentangle López Obrador's election losses in 2006 and 2012 from his *disparagement* of the institution as contributing factors in his followers' distrust in the INE. Much research shows that voters whose candidate loses tend to have less confidence in elections than those whose candidate wins (*inter alia*, 31–33). Indeed, both election losses and candidate claims of electoral fraud can jointly reduce voter confidence in elections (34). López Obrador's supporters' confidence fell after the 2006 loss and did not rebound after their candidate again lost in 2012 (35). If only sour grapes were at play, we would expect his supporters' confidence in the INE to have rebounded after his 2018 victory. But their leader's continued disparagement of the INE could be expected to pull in the opposite direction (30). Our survey experiments focus on the leader's discourse and hold constant any election results, thus allowing us to isolate the effect of the rhetoric, *per se*.

We also expect the impact of López Obrador's discourse to depend on whether a voter supports the president's party or an opposition party (or whether she has no partisan inclination). Scholars of political communication teach us that the acceptance of political messaging depends on the receiver's prior attitudes toward the source of the information (36). Individuals tend to accept information presented by in-party messengers and reject information presented by out-partisans (37, 38), even more so in the kinds of polarized settings that nurture democratic erosion.

The same is true when the public is faced with democracy-disparaging rhetoric. People who support an incumbent who voices the attack are more receptive to the attacker's message than are those who oppose this leader or their party (5, 19). Opposition party supporters are relatively immune to such attacks. Indeed, in some cases, attacks may backfire among opposition supporters, increasing support for the targeted institution (39).

Hence we expect López Obrador's discourse to dampen confidence more among his party's (Morena) supporters. Conversely, we expect little change in institutional confidence levels among opposition supporters, who are likely to reject claims that the institution is undemocratic or corrupt.

Prospects for Rebuttal. Whether disparaging rhetoric about democratic institutions can be effectively countered is a pressing question. Studies of misinformation offer grounds for believing that rebuttals, including simple informational corrections, can be effective (40–42). A common strategy explored by researchers is known as *debunking*. It involves presenting people with factual corrections after their exposure to misleading information. Debunking has been effective on topics ranging from vaccination myths to policy rumors (43, 44). Most of this research focuses on the United States and Europe. But several studies suggest that debunking interventions are effective in the Global South (45–47).

Still, political misinformation, and electoral misinformation in particular, can prove difficult to correct. Though some efforts have been successful (22), the overall results are mixed. Political rumors in Brazil proved resistant to fact-checking (48), as did unfounded fraud claims in the United States (18). In Mexico, corrections attributed to the INE's then-president, Lorenzo Córdova, did little to dispel rumors that the INE committed fraud in 2006 (49). Difficulties in correcting electoral misinformation may be rooted in motivated reasoning, in which individuals dismiss corrections to claims that are tied to their political identities (50, 51). The effectiveness of corrections to misleading claims about electoral authorities is an open question, especially in polarized societies.

We investigate whether simple informational rebuttals can reverse the effects of López Obrador's anti-INE polemics and restore voter confidence. We designed informational treatments in which all treatment groups read highly critical presidential statements about the INE. We culled these statements from the presidents' morning news conferences—which, as we saw in the previous section, he repeated frequently and were often picked up by the press. A subset of respondents was then randomly assigned to read corrections that rebutted López Obrador's claims and highlighted the INE's contribution to the Mexico's democratization. With the rebuttal language, we have mimicked real-world discussions of the INE, using frames deployed by the institution's defenders in the controversy surrounding López Obrador's "Plan B." *We expect higher levels of confidence in the INE among individuals who read these corrections than among those who read only López Obrador's critical portrayal of the institution.*

That said, not all rebuttals will be equally effective. Rebuttals that come from credible sources tend to be more persuasive (44, 52–54), though this effect may hinge on citizens' ability to discern source credibility (55). Partisanship also matters. Citizens are more likely to be persuaded by criticisms coming from its copartisans (56). In highly polarized democracies, people may become accustomed to suspecting that even ostensibly independent organizations in fact have hidden partisan motives (57, 58). Members of the public who hold such suspicions might

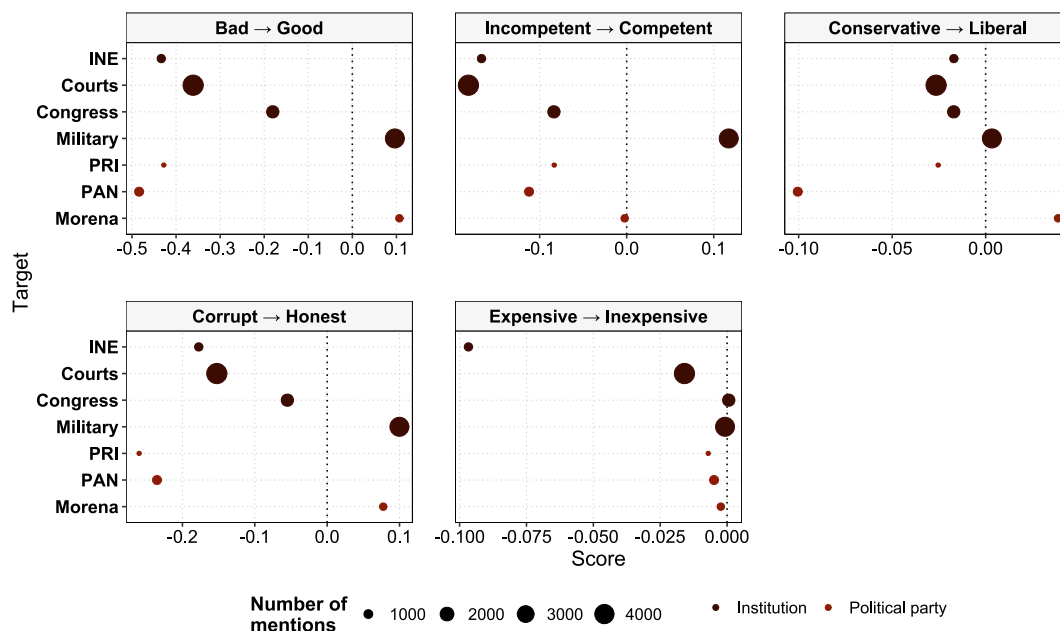


Fig. 1. How López Obrador discussed the INE and other institutions. Figure shows how López Obrador framed references to the National Electoral Institute and other government institutions, as well as major national political parties, during his daily press conferences. References were classified with a large language model (GPT-4o).

discount rebuttals attributed to these organizations. We test the effectiveness of rebuttals attributed to partisan, domestic nonpartisan, and international nonpartisan sources, with the expectation that those farther from domestic partisan politics will be more effective.

Survey Experiment

Research Design. We assess the impact of presidential discourse, and the effectiveness of rebuttals, with two survey experiments. In both, the design involves exposing some respondents only to the president's institutional criticisms, others to these criticisms followed by a rebuttal, and then to compare levels of confidence in the institution. (Preregistration materials are [posted online](#).) The respondents in each study were drawn from an online panel of Mexicans who were over the age of 18, with soft quotas for socioeconomic status, age, gender, and party support.

We conducted Study 1 ($n = 1,001$) from November 29 and December 9, 2023, toward the end of López Obrador's six-year presidential term, and Study 2 ($n = 3,001$) between April 17 and May 27, 2024, in the lead up to Mexico's June 2, 2024, presidential elections. We oversampled opposition supporters and non- or minor party-supporters to test for subgroup effects. See [SI Appendix, section 2.1](#) for additional information on the samples. [SI Appendix, section 2.5](#) presents substantively similar results among respondents with lower socioeconomic status.

Respondents were assigned to one of four (Study 1) or five (Study 2) conditions by simple random assignment. Table 2 presents the text of the vignettes from Study 2, translated into English. The experimental vignettes from Study 1 and the original Spanish text from both studies are available in the [SI Appendix, section 2.2](#). The treatment conditions are as follows:

Control: A neutral description of the INE.

Attack: A newspaper-style vignette reporting President López Obrador's disparagement of the INE.

International Expert: The same attack followed by a rebuttal from an international, nonpartisan source (the European Union).

Partisan: The attack plus a rebuttal attributed to a domestic partisan actor.

Domestic Expert (Study 2 only): The attack followed by the rebuttal from a domestic, nonpartisan source.

Recall that respondents in the rebuttal conditions first read the attack, followed by a rebuttal. If the rebuttals, in fact, reduce the effect of the initial attack, we will observe smaller negative effects on confidence in the rebuttal conditions. In addition, since the substantive content across the rebuttal conditions is identical, we can attribute any differences in outcomes between them to the information source.

Following treatments, respondents were immediately asked a series of outcome questions. We crafted these questions to probe respondents' attitudes toward the INE. Our main dependent variable is confidence in the INE, measured with four Likert items:

1. *The INE ensures impartiality in elections,*
2. *The INE performs its function adequately,*
3. *I trust the INE to carry out fair elections, and*
4. *The INE will count the votes fairly in the upcoming federal elections.*

We combine these items to create an index, with values ranging from zero to four; higher values indicated greater confidence in the INE ($\alpha = 0.92$). Across treatment groups, the mean value is 2.43 (s.d. = 1.17) in Study 1 and 2.46 (s.d. = 1.06) in Study 2, indicating general ambivalence about the electoral body.

At the end of both studies, participants read a debriefing statement that corrects the misleading statements in the attack condition. This statement is basically the same as the international expert rebuttal (which is also accurate), but in the rebuttal it is attributed to generic international experts. In Study 2, we

Table 2. Experimental vignettes from Study 2 (English)

Control	The National Electoral Institute (INE) is the independent body responsible for organizing elections in Mexico.
Attack	<p>The National Electoral Institute (INE) is the independent body responsible for organizing elections in Mexico. But President López Obrador claims that the National Electoral Institute acts against the people. President López Obrador asserts that the National Electoral Institute (INE) lacks impartiality and is “subject to interest groups created by the old regime.” He maintains that before his term, “there was no democracy; it was the government of the minority and the rich.”</p> <p>He also claims that “the INE was involved in fraud in 2006 that prevented us from winning the elections.” Referring to protests in favor of the INE, he states that when they say “Don’t touch the INE, the INE is untouchable,” it really means not touching corruption, not touching privileges, not touching the narco-state.”</p>
Intl. expert	<p>[Respondent reads text from attack condition, followed by:]</p> <p>However, international experts from the European Union affirm that the National Electoral Institute (INE) conducts “fair and clean” elections.</p> <p>According to international experts, the INE ensures clean elections in Mexico. This allowed the opposition to win the presidency in 2000 after decades of one-party rule.</p> <p>The elections organized by the INE have allowed the alternation of political parties in charge of the federal government.</p> <p>European Union election observers state that the elections organized by the INE since 2000 have been “fair, open, and well-organized, carried out in an environment where freedom of expression, assembly, and association were respected.”</p> <p>According to these experts, “the INE has maintained its independence and significant levels of public trust, and it organizes elections effectively and transparently.”</p>
Partisan	<p>[Respondent reads text from attack condition, followed by:]</p> <p>However, the National Action Party (PAN) affirms that the National Electoral Institute (INE) conducts “fair and clean” elections.</p> <p>According to the PAN and other opposition parties ...[Same text as International Expert.]</p> <p>The opposition parties state that ...</p> <p>According to these parties ...</p>
Domestic expert	<p>[Respondent reads text from attack condition, followed by:]</p> <p>However, nonpartisan experts from a Mexican center for the study of public administration affirm that the National Electoral Institute (INE) conducts “fair and clean” elections.</p> <p>According to these Mexican experts ...[Same text as International Expert.]</p> <p>The experts from this institute, which is not affiliated with any political party, state that ...</p> <p>According to these experts ...</p>

Participants in the expert, partisan, and domestic rebuttal conditions first read the text of the attack then the corresponding rebuttal.

reassess respondents’ views of the INE after the debriefing. The procedure confirms the effectiveness of the expert rebuttal and offers evidence of its generalizability.

Results

In the experiment, López Obrador’s rhetorical attacks indeed reduce confidence in the INE. This result can be seen in the *Upper* panel of Fig. 2, which shows the average effects of the treatments compared to the control condition. In Study 1, the president’s attacks reduce confidence in the INE by 0.22 points on the five-point response scale compared to the control ($P < 0.01$). His attacks reduce confidence in the INE in Study 2, as well, though the effect is smaller (reducing confidence in the INE by 0.10 points relative to the control, $P < 0.05$). (See *SI Appendix, section 2.3* for model specification and tabular results. As preregistered, we report one-tailed P -values for directional hypotheses.)

The center panel shows the effect of the rebuttals. The *international expert* rebuttal restored confidence in the INE; neither the *domestic* nor the *partisan* rebuttal made a significant difference. The international rebuttal increases confidence in the INE by 0.24 points in Study 1 ($P < 0.01$) and by 0.13 points

in Study 2 ($P < 0.01$ —see the *Center* panel of Fig. 2). The *Lower* panel shows that the international expert rebuttal boosts confidence to a larger degree than does the domestic expert rebuttal.

Since we study how much people’s confidence is buffeted both by anti-INE diatribes and pro-INE tributes, we are able to compare the two effects. In fact, in both studies, *the positive effect of the international expert rebuttal matches or exceeds the negative effect of the initial attack, reversing its effects*. Not only is the rebuttal effective, then, it entirely neutralizes the president’s attacks.

Study 1 leaves unanswered the question: Was the effectiveness of the rebuttal reliant on its being attributed to an expert, or to an international organization? To answer this question, in Study 2 we add a *domestic* expert rebuttal while retaining the international one. The text of the rebuttals remains unchanged; the only change is that one is attributed to a nonpartisan Mexican think-tank, the other to an international organization.

In fact, it is the international element of the treatment that moves people. The middle panel of Fig. 2 shows that only the international expert treatment shifts confidence significantly. Those exposed to the president’s attack followed by the domestic think-tank rebuttal do not experience a revival of their institutional confidence. Confidence in the INE is modestly—

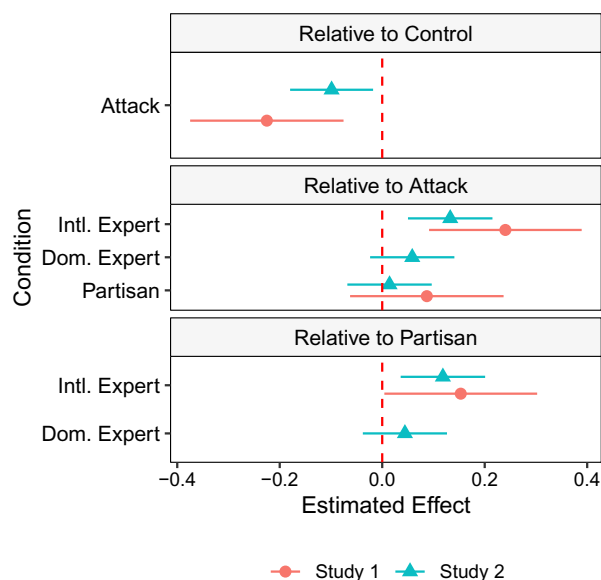


Fig. 2. Effect on confidence in the National Electoral Institute. Figure shows the effect of moving from the control (*Top* panel), the rhetorical attack (*Center* panel), or partisan rebuttal (*Lower* panel) to the treatment condition on the y-axis. 90% CIs are shown.

but not significantly—greater in the domestic rebuttal condition than the attack condition. In turn, the *Bottom* panel shows that average levels of confidence are not statistically different between people exposed to the domestic partisan rebuttal and the domestic nonpartisan (think-tank) one.

To summarize the findings thus far, confidence in the INE declines significantly among those who read López Obrador's attacks but it is restored among those who read the international nonpartisan rebuttal. The effects of the domestic expert and partisan rebuttals are positive but not measurably different from zero.

Conditional Effects by Party Support. In the polarized setting of Mexican politics and with a president who placed the INE at center stage, we expect Mexicans' views to be powerfully shaped by their prior partisan commitments.

And indeed, partisans of various stripes came into our experiment with sharply diverging levels of confidence in the INE. This can be seen by comparing average levels of support across different partisan groups among subjects assigned to our control condition, and who therefore are not exposed either to the diatribes or to the rebuttals (Fig. 3). Supporters of the partisan opposition—the PAN, PRI, PRD coalition—are more confident in the INE than nonpartisans or Morena supporters in the baseline condition. Mean support appears slightly lower among nonpartisans and minor-party supporters than Morena supporters but is not significantly so (*SI Appendix, Table S13*). This pattern is atypical—government supporters are usually more confident in electoral institutions—but is consistent with the idea that López Obrador's rhetoric affected confidence in the INE among his supporters (30).

Are López Obrador's supporters, in fact, more responsive to his verbal attacks on the INE? We find that they are indeed. The presidents' diatribe significantly diminishes confidence in the INE among *Morena* supporters but has no significant effect on non-Morena respondents.

These party-moderated effects come through in the *Top-Left* panel of Fig. 4. The president's diatribe reduces confidence in the

INE among his supporters but has no statistically significant effect on opposition-party supporters or nonpartisans. Fig. 4 reveals a trend toward a backlash against the president's harsh words among opposition party supporters—they are a bit more positive about the INE than are their fellow opposition voters in the control group. And the figure hints at a decline in confidence among minor-party supporters and nonsupporters. But neither of these two impacts is statistically different from zero.

Do responses to the rebuttals also depend on party support? Given their leader's almost nonstop disparagement of the INE, one might expect the president's supporters to ignore positive statements about it. But, surprisingly, the expert rebuttal ameliorates the effect of López Obrador's anti-INE diatribes even among his supporters. The upper-right panel in Fig. 4 shows that, compared to the attack condition, the international expert rebuttal increases Morena voters' average confidence in the INE.[§]

Averaging across the two studies, exposure to the expert correction nearly halves the effect of the attack, reducing it by 43% (*SI Appendix, Table S17*). This is a hopeful result for those who desire to restore confidence in Mexico's democratic institutions. The governing party's supporters—people whom we might expect to have hardened anti-INE attitudes—are responsive to more positive, and more accurate, accounts of the country's election administration body.

Turning to opposition voters, the expert rebuttal does not boost their confidence in the INE. Nor are these opposition voters swayed against the INE by the president's attacks, as we saw earlier. They express greater confidence in the INE in the control condition and are relatively unmoved either by the attack or the rebuttal.

The international expert rebuttal also boosts confidence in the INE among non- and minor-party supporters. Notably, the expert rebuttal boosted confidence more than the president's diatribes reduced it. That is, non- or minor party supporters who

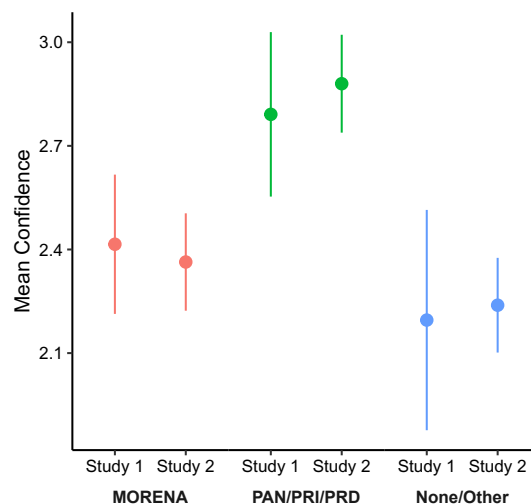


Fig. 3. Mean confidence in the National Electoral Institute in the control condition. The confidence measure ranges from 0 (low confidence) to 4 (high confidence). Party support is based on respondent vote intention, measured pretreatment. 95% CIs are shown.

[§]As in Fig. 3, party support is measured as vote intention in a snap election. The international expert rebuttal boosted confidence in the INE by 0.36 points in Study 1 ($P < 0.01$), by 0.07 points in Study 2 ($P = 0.22$), and by 0.15 points on average ($P < 0.05$) among Morena supporters. We suspect that the effect of the expert rebuttal is smaller—and nonsignificant—in Study 2 because the effect of the attack is also less pronounced, leaving the rebuttal with less to counteract.

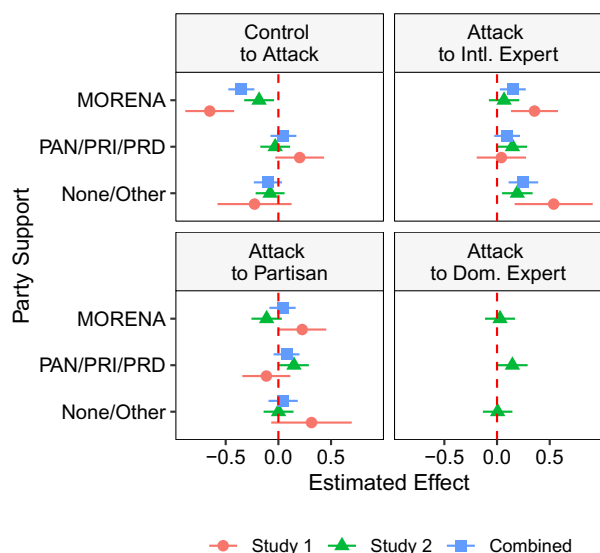


Fig. 4. Figure shows the average effect of moving from control to the attack (Top-Left), the attack to the international-expert rebuttal (Top-Right), the attack to the partisan rebuttal (Bottom-Left), and the attack to the domestic-expert rebuttal (Bottom-Right). 90% CIs.

read both the attack and the rebuttal ended up marginally *more* confident in the INE than were those assigned to the control condition. These individuals, then, are the most persuadable group; their confidence in the INE decreases slightly (though not significantly) with the attack, then increases more substantially with the expert rebuttal.

In sum, a rebuttal attributed to international experts is strikingly effective in restoring confidence among supporters of the president and those who do not support a major party. In contrast, the rebuttal attributed to partisan opposition is largely ineffective, as is the rebuttal attributed to domestic experts. The Lower Right-hand panel of Fig. 4 shows that a rebuttal attributed to nonpartisan domestic experts slightly boosts the confidence of opposition party supporters (by 0.15 points, $P < 0.05$), but not that of any other group. That the domestic reassurance fails to land is striking, given that the message is identical to that in the international expert treatment. We suspect that the domestic rebuttal fails because domestic experts are *perceived as partisan*—and likely as anti-Morena—despite explicit claims to the contrary in the vignette.

International Rebuttals: Using the Debriefing to Compare Generic and E.U. Messages. A plausible interpretation of these results is that something about the specific international experts we invoke, ones from the European Union, resonated with our respondents. Our debriefing at the end of the surveys allows us to test for this possibility. In both surveys, we debriefed respondents whom we had exposed to the president's diatribes. The debriefing repeats the corrective information included in the international rebuttal treatment. But instead of attributing the language to the E.U., we attribute it to generic "international experts." After the debriefing, we again assess confidence in the INE, which allows us to examine changes caused by the debriefing.

We focus on individuals whom we expose to López Obrador's attack and who did not subsequently receive a rebuttal. (For those who did read the rebuttal, the rebuttal in effect served as a debriefing.) Among this group, we find that the debriefing increases confidence in the INE by an average of 0.18 points ($P < 0.001$). See *SI Appendix, Table S18* for details. And the debriefing is effective across partisan groups, as shown in

Fig. 5. Hence, generic international experts' rebuttals are at least as effective as those attributed to E.U. observers. The result suggests that international source credibility, in the eyes of our samples, arose from the source appearing to be above the fray of Mexican politics, rather than from some attribute specific to the E.U.

Discussion and Conclusion

The denigration of political institutions—trash-talking democracy—is not the only strategy that backsliding leaders employ to maintain public support. They also polarize the public, claim majoritarian legitimacy (59), and filter out damaging information by censoring and manipulating the press. But institutional trash-talk is an important tool at their disposal. Our study asks whether this strategy can indeed undermine public confidence in core democratic institutions and whether these messages can be effectively rebutted.

The answer to both questions is, Yes. A Mexican president lashed out frequently against the country's election administration body, the guarantor of electoral integrity. The president's messages underlined the institution's presumed ineffectiveness, conservatism, profligacy, and corruption; our text analysis reveals his disdain for the election body to have been as great as his disdain for his partisan rivals.

We crafted survey experiments to replicate the dimensions and intensity of these real-world diatribes. And indeed, exposure to them undermined people's institutional confidence. The misleading diatribes worked but were more effective among some voters (his core base) than others (opposition and minor-party voters). Such attacks polarize public confidence, leaving government supporters less confident in electoral institutions than other citizens.

But the rebuttals also worked. Here, too, some segments of the electorate were more responsive than others. Though one could imagine the president's core supporters willfully ignoring contrary information, this was not the case. Indeed, one of our most important findings is that the right rebuttal, one that respondents find credibly removed from polarized domestic

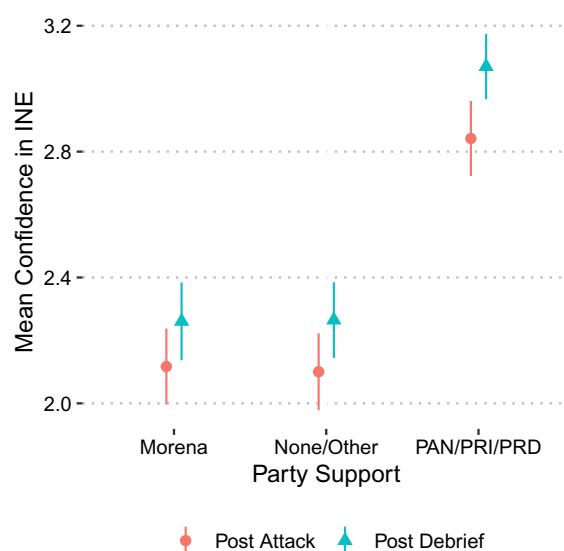


Fig. 5. Effect of the debriefing on confidence in the INE. Figure shows confidence in the INE after López Obrador's attack (circles) and the debriefing (triangles) by party support, with 90% CIs.

politics, not only works but *works on the backsliding leader's own followers*. Hence, confidence in this key electoral institution seems to be malleable, perhaps to a surprising degree given the high levels of affective polarization among López Obrador's supporters (13).

If official-party supporters' views were malleable, opposition-party supporters' views were less so. One could imagine that sustained and harsh criticisms might undermine their confidence in the targeted institution. Perhaps the high visibility of the López Obrador–INE conflict left opposition voters' views fairly immovable; confidence in a less visible institution might have been more malleable. Recall that at the time of our survey, the INE had been the subject of high-profile conflict, including marches led by opponents of the president, among others, to protect it from intervention.

The failure of opposition voters to respond to the president's criticisms by rallying to the INE's defense might seem puzzling. We uncovered a hint of such a backlash effect, but not a sizable enough one that it could be confidently distinguished from no effect at all. In a study of López Obrador's diatribes against the Mexican judiciary, Cella et al. suggest that rhetorical denigration of democratic institutions is a strategy that risks less of a backlash, from the backslider's point of view, than does a strategy of deepening partisan polarization (3), a point that Stokes extends to backsliding leaders in general. Our study offers additional evidence along these lines (4). The strategic implication of our study for a backsliding leader like López Obrador is that he could hack away at the image of institutions like the INE and the courts and thus undermine them among his supporters without stirring a backlash among opposition supporters. (Still, we show that he had to worry about rebuttals to his claims.)

Stepping back, democratic erosion has taken place in a surprisingly large and diverse set of countries. Even when a leader who has undermined democratic institutions exits office, as the Mexican president did in late 2024, adverse effects of backsliding are not immediately undone. In the United States, persistent loss of confidence in election integrity among large segments of the electorate is a phenomenon that the country is likely to struggle with for years to come.

Our study says that it is not a waste of time to offer voters alternative viewpoints and corrective information. In line with recent work on democratic backsliding, we find that presidential discourse can indeed reduce public confidence in critical democratic institutions, in ways that facilitate the expansion of executive power (3, 5, 19). But we also find that public confidence may be rebuilt, even among those who are most under the sway of presidential rhetoric.

In practice, the kinds of rebuttals that succeed in counteracting presidential rhetoric may be hard to disseminate at scale. International, nonpartisan experts have a much smaller megaphone than

the president; simply gaining the attention of voters is no easy task. Here, domestic experts with larger or more regular audiences could help. But, as our results show, this may undermine the rebuttal if they are perceived as partisan or biased. Reaching voters with such rebuttals requires sustained attention and thoughtful investment.

Our study offers succor both to backsliding leaders and to defenders of democracy. When President López Obrador left office in 2024, he left behind a presidency that was much strengthened. Autonomous government oversight agencies were brought under the control of the executive. The alliance between the military and the presidency was stronger than one would have expected in the aftermath of a leftist executive. Mexico was poised to embark on an experiment of electing all of its federal judges. The president's favored successor won the 2024 election decisively enough that criticisms of the INE were moot.

And yet the López Obrador presidency was filled with delays and frustrations for a president set on turning his country into a more autocratic democracy. The courts, journalists, opposition party leaders in congress, civil society organizations—all at one time or another threw sand into the gears of the president's project of executive aggrandizement. As Gamboa shows in the context of Colombia under Álvaro Uribe, slowing the process of autocratization can buy critical time for democracies' defenders to regroup, strategize, and fight back (60). Indeed, in some instances, populist attacks on democracy generate their "own antibodies," as autocratic threats spur mobilization by as pro-democracy groups (61, 71). Our study shows that these efforts at slowing down the autocratization process can meaningfully extend to struggles over public opinion. It is far from quixotic to try to disabuse voters of dire pictures of their institutions.

Materials and Methods

The *SI Appendix, section 1*, presents the keywords used to classify references to the institutions in Fig. 1, as well as human validation of the text classifier. *SI Appendix, section 2*, details the experimental samples, treatments, and regression estimates of the treatment effects.

Data, Materials, and Software Availability. All data and replication code are available on the Harvard Dataverse (62).

ACKNOWLEDGMENTS. We thank Agustina Giraudy, Jake Grumbach, Eli Rau, Laron Williams, and Kevin Kromash for their thoughtful comments, and Laura Sandino for invaluable research assistance.

Author affiliations: ^aDepartment of Political Science and International Relations, Tecnológico de Monterrey, Monterrey 64700, Nuevo Leon, Mexico; ^bDepartment of Political Science, University of Chicago, Chicago, IL 60637; and ^cDepartment of Political Science, University of Wisconsin-Madison, Madison, WI 53706

- H. E. Brady, T. B. Kent, Fifty years of declining confidence & increasing polarization in trust in American institutions. *Daedalus* **151**, 43–66 (2022).
- R. J. Dalton, *Citizen Politics: Public Opinion and Political Parties in Advanced Industrial Democracies* (CQ Press, 2018).
- L. Cella, I. Çınar, S. Stokes, A. Uribe, Building tolerance for backsliding by trash-talking democracy: Theory and evidence from Mexico. *Comp. Pol. Stud.* (2025).
- S. Stokes, *The Backsliders: Why Leaders Undermine Their Own Democracies* (Princeton University Press, Princeton, NJ, 2025).
- B. R. Bessen, Populist discourse and public support for executive aggrandizement in Latin America. *Comp. Polit. Stud.* **57**, e00104140231223738 (2024).
- L. Bartels, *Democracy Erodes from The Top* (Princeton University Press, 2023).
- A. C. Eggers, H. Garro, J. Grimmer, No evidence for systematic voter fraud: A guide to statistical claims about the 2020 election. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2103619118 (2021).
- Organization of American States, Informe Final de la Misión de Observación Electoral de la Organización de los Estados Americanos sobre el Proceso Electoral Federal 2011–2012 del 1 de julio de 2012 celebrado en los Estados Unidos Mexicanos (Organization of American States, Washington, DC, 2013). https://www.oas.org/es/sap/deco/MOE_informe/Mexico2012.pdf.
- European Union Election Observation Mission, Final Report: Presidential and Parliamentary Elections, Mexico, 2 July 2006 (European Union Election Observation Mission, Mexico City/Brussels, 2006). <https://repositoriodocumental.ine.mx/xmlui/bitstream/handle/123456789/151095/VE-informe-20230331-205.pdf>.
- A. Monsiváis-Carrillo, Deceitful autocratization: Subverting democracy through electoral reform in Mexico. *Latin Am. Policy* **15**, 486–508 (2024).
- J. A. Aguilar-Rivera, Dinámicas de la autocratización: México 2021. *Rev. Cienc. Polit.* **42**, 355–382 (2022).
- A. A. Aguilar Aguilar, R. C. Cornejo, A. Monsiváis-Carrillo, Is Mexico at the gates of authoritarianism? *J. Democr.* **36**, 50–64 (2025).

13. R. Castro Cornejo, The AMLO voter: Affective polarization and the rise of the left in Mexico. *J. Polit. Latin Am.* **15**, 96–112 (2023).
14. M. Sánchez-Talanquer, K. F. Greene, Is Mexico falling into the authoritarian trap? *J. Democr.* **32**, 56–71 (2021).
15. J. Ríos-Figueroa, *Mexico 2024: The Battle for the Constitution* (ConstitutionNet, 2024).
16. C. Welzel, H. D. Klingemann, Evidencing and explaining democratic congruence: The perspective of 'substantive' democracy. *World Values Res.* **1**, 2390577 (2008). <http://dx.doi.org/10.2139/ssrn.2390577>.
17. N. Kerr, B. A. King, M. Wahman, The global crisis of trust in elections. *Public Opin. Q.* **88** (suppl. 1), 451–471 (2024).
18. N. Berinski *et al.*, The effects of unsubstantiated claims of voter fraud on confidence in elections. *J. Exp. Polit. Sci.* **10**, 34–49 (2023).
19. K. Clayton *et al.*, Elite rhetoric can undermine democratic norms. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2024125118 (2021).
20. S. Bowler, T. Donovan, Confidence in US elections after the Big Lie. *Polit. Res. Q.* **77**, 20–34 (2024).
21. B. Sinclair, S. S. Smith, P. D. Tucker, "It's largely a rigged system": Voter confidence and the winner effect in 2016. *Polit. Res. Q.* **71**, 854–868 (2018).
22. J. M. Carey *et al.*, The narrow reach of targeted corrections: No impact on broader beliefs about election integrity. *Polit. Behav.* **47**, 1–14 (2024).
23. J. M. Carey, B. Fogarty, M. Gehrke, B. Nyhan, J. Reifler, Prebunking and credible source corrections increase election credibility: Evidence from the US and Brazil. *Sci. Adv.* **11**, eadv3758 (2025).
24. A. Maldonado-Maldonado, *The Fight Over Science in Mexico* (International Higher Education, 2023).
25. O. López, Historic: How Mexico's welfare policies helped 13.4 million people out of poverty. *Guardian* (2025).
26. S. Haggard, R. Kaufman, *Backsliding: Democratic Regress in the Contemporary World* (Cambridge University Press, 2021).
27. M. Heseltine, B. Clemm von Hohenberg, Large language models as a substitute for human experts in annotating political text. *Res. Polit.* **11**, 20531680241236239 (2024).
28. P. Törnberg, Large language models outperform expert coders and supervised classifiers at annotating political social media messages. *Soc. Sci. Comput. Rev.* **43**, e08944393241286471 (2024).
29. D. Chong, J. N. Druckman, Framing theory. *Annu. Rev. Polit. Sci.* **10**, 103–126 (2007).
30. A. Monsiváis-Carrillo, Happy winners, sore partisans? Political trust, partisanship, and the populist assault on electoral integrity in Mexico. *J. Polit. Latin Am.* **15**, 72–95 (2023).
31. C. J. Anderson, A. Blais, S. Bowler, T. Donovan, O. Lijthaug, *Losers' Consent: Elections and Democratic Legitimacy* (OUP Oxford, 2005).
32. G. Kernell, K. J. Mullinix, Winners, losers, and perceptions of vote (mis)counting. *Int. J. Public Opin. Res.* **31**, 1–24 (2019).
33. A. M. Daniller, D. C. Mutz, The dynamics of electoral integrity: A three-election panel study. *Public Opin. Q.* **83**, 46–67 (2019).
34. V. Hernández-Huerta, F. Cantú, Public distrust in disputed elections: Evidence from Latin America. *Br. J. Polit. Sci.* **52**, 1923–1930 (2022).
35. F. Cantú, O. García-Ponce, Partisan losers' effects: Perceptions of electoral integrity in Mexico. *Elect. Stud.* **39**, 1–14 (2015).
36. J. R. Zaller, *The Nature and Origins of Mass Opinion* (Cambridge University Press, 1992).
37. K. J. Mullinix, Partisanship and preference formation: Competing motivations, elite polarization, and issue importance. *Polit. Behav.* **38**, 383–411 (2016).
38. J. N. Druckman, E. Peterson, R. Slothuus, How elite partisan polarization affects public opinion formation. *Am. Polit. Sci. Rev.* **107**, 57–79 (2013).
39. S. Bowler, M. Carreras, J. L. Merolla, Trump tweets and democratic attitudes: Evidence from a survey experiment. *Polit. Res. Q.* **76**, e10659129221137348 (2022).
40. N. Walter, R. Tukachinsky, A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Commun. Res.* **47**, 155–177 (2020).
41. M. S. Chan, C. R. Jones, K. Hall Jamieson, D. Albarracín, Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol. Sci.* **28**, 1531–1546 (2017).
42. T. Wood, E. Porter, The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Polit. Behav.* **41**, 135–163 (2019).
43. B. Nyhan, J. Reifler, When corrections fail: The persistence of political misperceptions. *Polit. Behav.* **32**, 303–330 (2010).
44. A. J. Berinsky, Rumors and health care reform: Experiments in political misinformation. *Br. J. Polit. Sci.* **47**, 241–262 (2017).
45. E. Porter, Y. Velez, T. J. Wood, Correcting COVID-19 vaccine misinformation in 10 countries. *R. Soc. Open Sci.* **10**, 221097 (2023).
46. M. Winters *et al.*, Debunking highly prevalent health misinformation using audio dramas delivered by WhatsApp: Evidence from a randomised controlled trial in Sierra Leone. *BMJ Glob. Health* **6**, e006954 (2021).
47. S. Badrinathan, S. Chauchard, "I Don't Think That's True, Bro!" Social corrections of misinformation in India. *Int. J. Press Polit.* **29**, 394–416 (2024).
48. F. Batista Pereira, N. S. Bueno, F. Nunes, N. Pavao, Fake news, fact checking, and partisanship: The resilience of rumors in the 2018 Brazilian elections. *J. Polit.* **84**, 2188–2201 (2022).
49. R. Castro Cornejo, Who believes in fraud in the 2006 Mexican presidential election? Election denialism, partisan motivated reasoning, and affective polarization. *Lat. Am. Res. Rev.* **59**, 588–609 (2024).
50. C. S. Taber, M. Lodge, Motivated skepticism in the evaluation of political beliefs. *Am. J. Polit. Sci.* **50**, 755–769 (2006).
51. B. M. Tappin, G. Pennycook, D. G. Rand, Bayesian or biased? Analytic thinking and political belief updating. *Cognition* **204**, 104375 (2020).
52. T. G. Van der Meer, Y. Jin, Seeking formula for misinformation treatment in public health crises: The effects of corrective information type and source. *Health Commun.* **35**, 560–575 (2020).
53. J. J. Guillory, L. Geraci, Correcting erroneous inferences in memory: The role of source credibility. *J. Appl. Res. Mem. Cogn.* **2**, 201–209 (2013).
54. S. Botero, R. C. Cornejo, L. Gamboa, N. Pavao, D. W. Nickerson, Says who? An experiment on allegations of corruption and credibility of sources. *Polit. Res. Q.* **68**, 493–504 (2015).
55. R. Weitz-Shapiro, M. S. Winters, Can citizens discern? Information credibility, political sophistication, and the punishment of corruption in Brazil. *J. Polit.* **79**, 60–74 (2017).
56. S. Botero, R. Castro Cornejo, L. Gamboa, N. Pavao, D. W. Nickerson, Under friendly fire: An experiment on partisan press, fragmented opposition and voting behavior. *Elect. Stud.* **60**, 102044 (2019).
57. J. N. Druckman, S. Klar, Y. Krupnikov, M. Levendusky, J. B. Ryan, *Partisan Hostility and American Democracy: Explaining Political Divisions and When They Matter* (University of Chicago Press, 2024).
58. S. Lewandowsky, U. K. H. Ecker, J. Cook, Beyond misinformation: Understanding and coping with the post-truth era. *J. Appl. Res. Mem. Cogn.* **6**, 353–369 (2017).
59. R. Castro Cornejo, A. Monsiváis-Carrillo, A. A. Aguilar Aguilar, "Presidential approval as majoritarian legitimacy: How populists undermine democracy with the people's support" in *Presented at the Annual Meeting of the American Political Science Association* (2025).
60. L. Gamboa, *Resisting Backsliding: Opposition Strategies against the Erosion of Democracy* (Cambridge University Press, 2022).
61. K. Weyland, *Democracy's Resilience to Populism's Threat: Countering Global Alarmism* (Cambridge University Press, 2024).
62. B. Bessen, S. Stokes, A. Uribe, Replication data for: Restoring institutional confidence in backsliding democracies: Evidence from Mexico. Harvard Dataverse. <https://doi.org/10.7910/DVN/HQJD6F>.