

Subverting democracy in the name of democracy

Alexander Bor^a and Honorata Mazepus^b

^aDemocracy Institute, Central European University, Budapest, Hungary; ^bDepartment of Political Science, University of Amsterdam, Amsterdam, Netherlands

ABSTRACT

Attempts to subvert democracy are often justified, paradoxically, as an effort to save or improve democracy. Yet, little is known about the effectiveness of these justifications. Here, we conduct two original vignette experiments, where US American and Polish respondents are introduced to a proposal of executive aggrandizement, usurping control over election results and capturing the media, respectively. We manipulate the proposal's party and the justification they provide in a 2×3 between participants experiment (total $N = 5,008$). Our experiments replicate well-known effects of partisan bias in support for the subversion of democracy: respondents in both the US and Poland are more tolerant of subversion when it is initiated by their in-party. However, respondents are surprisingly resistant to justifications, especially in the high external validity design implemented in the US. Even in Poland where we "forced" respondents to pay attention to the justifications, support for subversion is increased only among in-partisans and not by much. These results raise doubt that democratic subversion can be effectively masked as a pro-democracy intervention by pro-democratic rhetoric.


ARTICLE HISTORY Received 19 July 2025; Accepted 17 November 2025


KEYWORDS Democratic backsliding; justification; manipulation; polarization; partisan bias

What we're seeing here is ... the thrashing of the bureaucracy as we try to restore democracy

Elon Musk, Head of DOGE

Democracy remains widely regarded as the most desirable form of government, enjoying overwhelming support among citizens worldwide.¹ Paradoxically, however, democracy also appears increasingly vulnerable, marked by widespread democratic erosion even in established democracies.² This apparent contradiction – that citizens express strong, near-universal support for democracy yet simultaneously tolerate or even endorse actions that undermine it – has become central to contemporary debates in political science.

CONTACT Alexander Bor  bora@ceu.edu

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/13510347.2025.2592846>.

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

One of the strategies autocratizing incumbents use to resolve this contradiction is that they justify democratic subversion as an attempt to *save, sustain, or strengthen* democracy itself. But do these democratic justifications for subversion meaningfully shape public support for anti-democratic actions? This article addresses this question by investigating the extent to which framing subversion as necessary for preserving democracy influences citizens' attitudes, probing the potential of democratic rhetoric to legitimize autocratic practices. Specifically, we distinguish between two types of democratic justifications: majoritarian and out-party threat appeals. The former relies on the popular democratic mandate of the incumbent, waving away charges of illegitimacy with appeals to the majority who voted for the government. The latter, meanwhile, highlights that subversive actions are necessary to constrain rival political parties that pose a genuine threat to democracy.

Our inquiry aligns closely with Gerschewski and Tomini's (this issue) conceptualization of autocratization as a *tug-of-war* between the autocratizing and democratizing actors. In particular, our study focuses on the *arena of legitimacy*, where the incumbent and its opposition compete in generating narratives justifying or criticizing the incumbent's rule. While there are many battles fought in the arena of legitimacy, those pertaining to the democraticness of the incumbent and its most controversial (subversive) policies are among the most important. If incumbents can effectively spin subversive policies as benefiting democracy, they are winning in the arena of legitimacy. To the contrary, if citizens believe the narratives of the opposition, or at least disbelieve the justifications of the incumbent, autocratizing actors must look for alternative sources of legitimacy.

Our study introduces methodological innovations to reflect the complexity of the tug-of-war in the arena of legitimacy. Unlike most prior experimental studies, which use terse, direct descriptions of subversion, our article emphasizes the contested nature of legitimacy during autocratization. We employ externally valid survey experiments with large, diverse samples from the U.S. and Poland, two democracies experiencing different stages of democratic erosion. Crucially, our study incorporates dynamic discourse by embedding subversive policies within realistic vignettes that feature an incumbent justifying the policy and an opposition actor challenging it.

Our experiments replicate established findings of partisan bias: respondents in both countries are more tolerant of subversive acts initiated by their preferred party. However, we uncover significant resistance to the justifications provided. In the U.S., many respondents fail to recall the pro-democracy rhetoric presented, and even among those who remember it, the effect remains small. These results challenge the assumption that democratic rhetoric can effectively legitimize autocratic practices. We conducted a follow-up study in Poland designed to estimate the ceiling of justification effects by encouraging respondents to pay attention to the key experimental manipulations. With the revised design, we found statistically significant, but substantively small receptiveness to justifications driven by in-party respondents.

Overall, our findings offer a mixed outlook for resistance to autocratization. Encouragingly, we find little evidence that autocrats easily sway the public with democratic rhetoric, supporting theories that scepticism leads audiences to ignore untrustworthy messages. Yet, the powerful influence of partisanship remains a serious concern, as it notably biases individuals towards tolerating erosion from their own party. Our study also identifies important avenues for future research both conceptually and methodologically.

Theory

Open pursuit of authoritarian leadership is very rare in contemporary politics.³ Unlike dictators of the past, modern autocrats rely little on repression and do not strive for total control of private and public lives. How could they nonetheless prevent dissent from enacting unpopular policies? Why do citizens who live under democratic rule show no more resistance to autocratization?

The dominant answer to this question is that contemporary autocrats strategically exploit partisan identities and political polarization to secure popular support or at least passive acceptance of democratic erosion.⁴ Citizens' loyalty to their political parties often takes precedence over their commitment to democratic principles. Studies in the US and in six European countries reveal that only a small minority of voters penalize their own party's candidates for violating democratic norms, suggesting that partisan allegiance strongly outweighs concerns about democratic integrity.⁵ Worse, citizens are prone to "democratic hypocrisy" meaning they are more permissive towards democratic norm violations committed by their co-partisans, perceiving such acts as less problematic or even justified due to heightened partisan animosity.⁶ Gidengil and colleagues carefully tease apart multiple psychological mechanisms contributing to partisan bias in support for autocratization and find furthering ideological agendas to be the most prevalent in the US and Canada.⁷ We seek to replicate this effect with an original methodological approach and hence define our first hypothesis as:

- (1) H1. Democratic subversion is supported more if it is by the respondent's own-party (vs the other party).⁸

Yet, autocrats can hardly trust democratic hypocrisy and partisan motivations alone to legitimize their rule. They employ diverse arguments to explain, defend, and justify their institutional and policy changes.⁹ Examples of typical legitimation strategies involve emphasis on the economic performance delivered by the leaders, their ability to protect society from internal and external enemies, and defence of what they see as the core (often conservative) values of the societies they rule.¹⁰ It is helpful to think of these strategies as the autocrat's efforts to win the tug-of-war in the arena of legitimacy (Gerschewski and Tomini, this issue). Democratizing actors' resistance rests on two pillars. First, at a minimum, they seek to discredit the legitimizing narrative of the autocrat. They can contest the values of the autocrat, question their role in the successes, call out the pitfalls of their policies. In addition, the opposition also seeks to offer a counter-narrative highlighting which values, actors, and acts ought to benefit the nation.

Of the many ways autocrats seek to legitimize their rule, here, we focus on appeals capitalizing on citizens' near universal preference for democracy.¹¹ Modern autocrats "pretend to embrace the vogue for freedom. Of course, many abroad see through their hypocrisy. But at home – and even abroad – many others do not".¹² Remarkably, research shows that autocratic leaders use democratic rhetoric and emphasize the importance of democracy as "the only game in town" significantly more than leaders in democracies.¹³ For example, a classical study on the Greek military junta (ruling between 1967 and 1974) underscored that "even the ... most militant among the colonels spoke not of a total abandonment of democratic practices, and an alternative model of political organization for Greek society, but of a restoration of

democracy”.¹⁴ Similarly, Peru’s Alberto Fujimori justified his 1992 self-coup as “the starting point of the search for an authentic transformation that ensure a legitimate and effective democracy”. Venezuela’s Hugo Chavez famously called himself “an anti-dictator” and a “democratic revolutionary” in 2010. South Korea’s Yoon Suk Yeo declared martial law in 2024, allegedly to “protect the Republic of Korea’s freedom, democracy, and constitutional order” (see full quotes with sources and more examples in Online Appendix A).

We identify two specific types of democratic arguments that are used to justify actions that undermine liberal democratic institutions: majoritarian and out-party threat. First, majoritarian justifications emphasize incumbents’ right to introduce policies and institutional changes based on their strong popular mandate, electoral success, and firm parliamentary majority. Majoritarian justifications contradict (and often attack) liberal conceptions of democracy insisting that majority rule should be constrained by individual rights and institutional checks and balances to protect (vulnerable) minorities. A particularly salient tension between majoritarian and liberal conceptions of democracy can be witnessed in Donald Trump’s ongoing constitutional struggle to expand the Presidency’s executive authority at the expense of the judiciary and the legislature.¹⁵

In Hungary, Prime Minister Viktor Orbán and his party, Fidesz frequently invoke their parliamentary supermajority as evidence of their democratic legitimacy. They argue that this strong electoral mandate grants them the authority to enact substantial institutional reforms, dismissing any criticism of their actions as fundamentally anti-democratic or undermining the people’s will. Scheppele describes how Orbán’s government has strategically leveraged this parliamentary dominance to reshape Hungary’s constitutional framework and undermine liberal democratic safeguards, while portraying these measures as legitimate expressions of the people’s will.¹⁶

Evidence to date is mixed on whether majoritarian justifications succeed in garnering popular support for subversive policies. Psychological work demonstrated that people from five diverse societies have strong intuitions in favour of majority rule in abstract decision-making procedures.¹⁷ In an important paper, Grossman et al. demonstrate that many voters “view popularly elected leaders’ actions as inherently democratic”.¹⁸ The potency of majoritarian justifications was also demonstrated in a set of experiments with participants in the UK and Spain.¹⁹ Both British and Spanish citizens endorsed reforms subversive to liberal democracy, such as expanding executive power or curtailing judicial independence, if these are supported with a majoritarian justification.

Contrary to these findings, Kaftan and Gessler show the limits of majoritarian conceptions of democracy.²⁰ They found that German and Polish respondents object more to proposals subverting liberal aspects of democracy than to proposals subverting majoritarian aspects. These contradictory findings indicate that the persuasiveness of majoritarian arguments justifying subversion of democracy might be context/situation specific. We seek to offer a novel test of the effectiveness of majoritarian justifications and make the following hypothesis:

- (2) H2. A majoritarian justification (vs no justification) increases support for democratic subversion.

The second type of justifications focuses on *out-party threat*. These arguments portray democracy as being under threat from internal and external enemies,

necessitating defensive measures. Authoritarian and hybrid regimes frequently employ this narrative to ban or stigmatize organizations receiving funding from Western sources²¹ and to target groups identified as “social resisters” to increasing authoritarianism.²² Schedler calls this strategy “*democratic enemography*, the description or “drawing” of adversaries as democratic enemies”, and identifies it as a pernicious, yet common hallmark of today’s polarized politics.²³ Perceiving political opponents not as legitimate adversaries but as dangerous, irredeemable enemies threatening the democratic fabric significantly aids autocrats in legitimizing authoritarian actions.

Complementing this perspective, Braley and colleagues demonstrate that the so-called *subversion dilemma* may create a vicious cycle wherein fears that the rival party would subvert democracy can propel even democratically-minded citizens to support anti-democratic policies.²⁴ Thus, extreme polarization and partisan hostility can lead citizens to tolerate or even endorse democratic erosion if it aligns with their partisan interests.

We, therefore, test the following hypothesis:

- (3) H3. A justification emphasizing out-party threat (vs no justification) increases support for democratic subversion.

Can appeals to majority rule or out-party threat genuinely persuade citizens to support subversion? While it is undeniable that autocrats supply these justifications, it is unclear how much demand there is for them. Some recent findings warrant caution about citizens’ gullibility. A large body of cognitive science research suggests people possess sophisticated mechanisms of vigilance against manipulation, reducing susceptibility to political propaganda.²⁵ Recent meta-analyses and large-scale experiments consistently find small to no persuasion effects in political campaigning.²⁶ In a study about elite explanations in regular policy debates, Robison demonstrates that justifications offered by politicians are “ineffective when other speakers offer counter-explanations that focus on the official’s potential ulterior motives”.²⁷ Finally, citizens disapprove reforms that limit their freedoms and expand the power of the leaders.²⁸

Although we find it helpful to treat majoritarian and out-party threat justifications as distinct categories, in real life they are often combined, especially in populist rhetoric. Populists, who seek to “create a moralized divide between *us* and *them*”²⁹ simultaneously claim that they represent the real people constituting the silent majority, and that the corrupt elites threaten their well-being, if not existence.³⁰ Trump’s infamous speech on January 6, 2021 – which many consider pivotal in inciting the Capitol Riot – included ample examples for both of unsubstantiated claims that he won the election in a landslide (majority appeal) and that the Democrats are trying to steal the election (out-party threat).

We should acknowledge that partisan motivations (fuelling H1) could interact with the effectiveness of justifications. The perceived trustworthiness of agents matters a great deal for believing them³¹, and voters trust their own political party much more than their opposition, yielding the well-known finding of partisan cue-taking.³² Accordingly, it is plausible that justifications will be more persuasive among in-partisans than among out-partisans. However, recent methodological advances warn us that estimating these heterogeneities with sufficient precision require very large samples, beyond our financial constraints.³³ Hence, we decided to

formulate our pre-registered hypotheses as main effects and explore heterogeneities across partisan groups in post-hoc analyses.

Study 1 – Election manipulation in the United States

We conducted our first study in the United States. The US serves as a critical case for examining the paradox of democratic subversion justified in the name of democracy, as recent efforts to manipulate elections reveal how anti-democratic actions can be framed as protective measures. Election manipulation, particularly through increased partisan control over election administration, has become a pressing issue.³⁴ A consortium of non-partisan NGOs identified close to 400 legislative proposals threatening the integrity of elections in the US in 2021–22.³⁵ A particularly concerning trend involves passing laws that expand partisan influence over vote counting, certification, and oversight. To ensure external validity and policy relevance, our study models its experimental stimuli on these real-world cases, allowing us to systematically assess how justifications for election manipulation influence public perceptions of democratic legitimacy. Our study received ethical approval from the CEU Ethical Research Committee, 2024-2025/1/RD. We pre-registered materials and analyses at <https://aspredicted.org/tf29-kk2z.pdf>. Our data and code are deposited at <https://osf.io/wq6uz/>.

Data and methods

Our experiment was embedded in an omnibus survey fielded to 3,008 adults resembling the population of the United States in terms of age, gender, education, race, and geography. We report the demographic composition of our samples in Online Appendix B. We recruited our respondents through Lucid Theorem using quota sampling. The survey was fielded between Nov 1 and 5, 2024 in the days preceding the Presidential Elections.

Our survey started with an informed consent form and a screener question that filtered out inattentive respondents. Next, we asked respondents several background questions including their partisan identity, using standard questions from the American National Election Study. Respondents who identify as true independents (without leaning towards either party) were forced to choose one way or another, using a question about which party deals better with the political issues that are most important to them. We report estimates excluding true independents in Online Appendix D.

Our experimental vignette is designed to resemble a news article reporting on a newly passed state election law, allowing us to examine how different justifications for election manipulation affect public perceptions of democratic legitimacy. The vignette follows a structured format: it introduces a bill passed by the state legislature, describes its provisions expanding the Secretary of State's authority over elections, and presents reactions from supporters and opponents.

For example, in the Democratic incumbent, majoritarian justification condition, the vignette read as follows:

State Legislature Passes Controversial Bill to Prevent Fraud in Elections

The State Legislature's Democratic majority passed a bill that gives more power to the Secretary of State to oversee and intervene in the elections. The bill allows the Secretary of State to

appoint special election commissioners in any county where they suspect irregularities or fraud, and to order new elections if they find any evidence of wrongdoing.

A Democratic spokesperson stated that the new law is an “important victory for democracy”. They said that “The Democratic Party’s firm majority in the State Legislature gives us a strong mandate to legislate. The new bill reflects the will of the citizens of our state”.

Opponents of the bill claim it is undemocratic and unconstitutional. They argue that it undermines the independence of election officials. “Party leaders should not be allowed to meddle in or nullify election results” commented a Republican State Senator.

The bill also drew criticism from Democracy for the Future, which is a non-partisan NGO. They cautioned that the new bill enables the Democratic governor to “usurp control over election results”.

Our design goes beyond most prior experimental studies about support for democratic subversion, which usually rely on very brief descriptions of policies, typically without any context, justification, or commentary. By incorporating information both about the policy makers’ stated intention with the bill, and reactions from opponents, our vignette approximates much better the dynamic discourse surrounding real life instances of election manipulation.

To test our hypotheses, we manipulate two key elements in a 2×3 between participants design. First, we randomize the incumbent party introducing the bill (Democratic or Republican) to assess partisan asymmetries in responses. Second, we randomly assign one of three justifications for the law: (1) a control condition in which party representatives remain neutral and decline to comment, (2) a majoritarian justification, where the party frames the law as an expression of the electoral mandate, and (3) an out-group threat justification, in which the party claims the law is necessary to prevent election fraud orchestrated by the opposing party. We provide the full vignette in Online Appendix C.

After reading the vignette, participants were asked “If you had an opportunity to express your opinion on the new bill passed to prevent fraud in elections, would you support or oppose it?”, and indicated their answer on a five-level scale from “Definitely support the bill” to “Definitely oppose the bill”. measures follow Grossman et al.³⁶ We rescale responses to a 0–1 scale and model responses in a simple linear regression with two predictors one for partisanship – a dummy variable taking the value of 1 if the partisanship of the respondent and the subversive incumbent is aligned, and 0 otherwise – and the other for the justification – a factor variable with three levels for the three conditions with the control as a baseline.³⁷

Results

While our pre-registered hypothesis tests focus on the marginal effect of two manipulations separately, we display independent group means for each of the six experimental groups separately in [Figure 1](#). These descriptives foreshadow strong partisan differences, but negligible justification effects. They also demonstrate that out-party support is on average a bit below the scale mid-point (“neither support nor oppose”), while in-party support is a bit above it.

Do people support democratic subversion from their own-party more? Yes, we find firm support for H1. Our experiment replicates the well-known finding that people

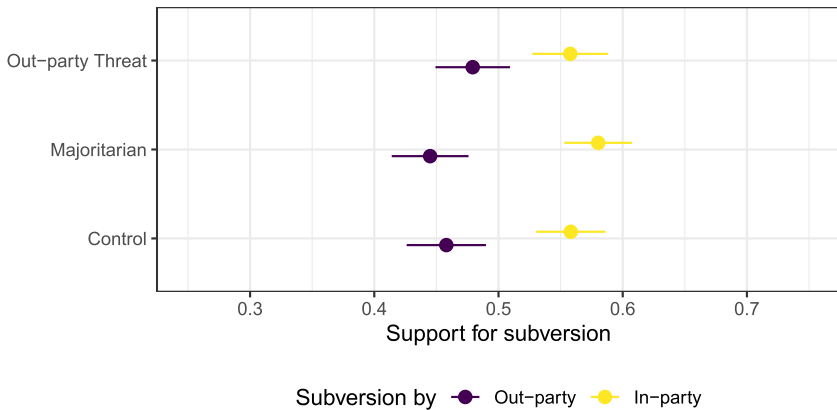


Figure 1. Support for the anti-democratic policy in the US is contingent on partisan alignment, but not on the justification provided. Points denote simple, descriptive group means. Error bars denote 95% confidence intervals.

employ a double standard when evaluating policy proposals detrimental to democracy ($\beta = 0.10$, 95% [0.08, 0.13]). This effect corresponds to 30% of a standard deviation in the outcome variable – a medium sized effects by most standards.

Does justifying subversion with a majoritarian appeal or out-party threat increase its support? No, we find no support for either H2 or H3. Respondents’ support for the policy was largely independent of the justification offered in its defence. Neither a majoritarian appeal emphasizing the popular mandate of the state legislature ($\beta = 0.004$, 95% [−0.03, 0.03]), nor a justification accusing the out-party with being a threat to democracy ($\beta = 0.01$, 95% [−0.02, 0.04]) affected support for the bill. Importantly, these are precisely estimated null effects; we can rule out with high confidence effects larger than 12% of a standard deviation in either direction. Moreover, supplementary analyses (and a glance at Figure 1) rule out the possibility that these marginal effects mask some significant interactions that cancel each other out in the aggregate (see Online Appendix D for formal tests).

What could explain the null result we obtain? One possibility is that respondents paid little attention to the manipulations. Our design approximates much better real-life discourse around subversion of democracy than most prior studies, yet this advance in external validity comes with higher demands from respondents, who must parse (a short and simple) mock newspaper article and should thus go beyond reading a simple phrase or sentence. Is it possible that in such a high external validity design, respondents simply ignore most of the details on who is subverting democracy and why? Mindful of this possibility, we embedded two factual manipulation checks after our dependent variables. First, we asked respondents to identify, which political party passed the new bill in the article [they] just read, with the option being “Republican”, “Democratic”, and “Didn’t specify”. Surprisingly, only 56% of respondents correctly identified the incumbent party in the vignette. Second, we asked respondents about the reason for passing the new bill according to the spokesperson in the vignette. Respondents again chose from three options, this time corresponding to the three experimental conditions (no justification, majoritarian, or out-party threat). Again, only 59% of respondents correctly identified the justification.

Even among those who read attentively, justifications had a limited effect. We report treatment effects among attentive respondents in Online Appendix D. These estimates are subject to post-treatment bias and should be treated with a grain of salt.³⁸ Nonetheless, it is notable that the justification effects remain small, although they turn out to be statistically significant at the 5% level (majoritarian: $\beta = 0.05$, 95% CI [0.003, 0.09]; out-party threat: $\beta = 0.06$, 95% CI [0.01, 0.11]). In contrast, the partisanship effect in this sample is substantial, increasing by 17 percentage points to $\beta = 0.27$ (95% CI [0.23, 0.31]).

Discussion

Our findings confirm that partisanship strongly conditions support for election manipulation, but they provide little evidence that the justification for subversion influences public opinion.

One possible explanation for the weak justification effects is that respondents paid limited attention to relevant bits in the experimental vignette. Our attention checks reveal that many respondents did not fully process key details of the stimulus, even though they knew they would be asked about the content afterward and despite the high quality of our sample.³⁹ While the high external validity of our design ensures that the stimulus closely resembles real-world political discourse, it may have also increased cognitive demands on respondents, reducing engagement with the manipulations. Notably, this pattern aligns with broader concerns about the public's limited attention to political news in everyday settings, where individuals lack incentives to carefully scrutinize information.⁴⁰

Beyond attention, an open question is whether these findings generalize to other democratic contexts and time periods. The study was conducted in the days leading up to the U.S. presidential election. The advantage of our timing is that the final days of a national election campaign heighten political engagement, media coverage, and partisan messaging, making election integrity a particularly salient issue. Conducting the experiment in this context ensures that respondents are highly attuned to electoral concerns, maximizing the external validity of our findings. At the same time, one might argue that this very context could have amplified partisan motivations to side with one's party, potentially limiting receptiveness to justifications.⁴¹ It remains unclear whether similar results would emerge in a different country or in a lower-salience political environment. We address these concerns in Study 2.

Study 2 – Public media capture in Poland

Study 2 replicates the experiment in Poland, incorporating design improvements such as moving factual manipulation checks before the dependent variables and requiring inattentive respondents to re-read the vignette. This allows us to assess whether increasing engagement with the stimulus affects the results and whether findings extend beyond the US context. Whereas we optimized Study 1 to maximize external validity, here, we put more emphasis on internal validity. We sacrifice some of the similarities between real world engagement with information on democratic subversion and justification to ensure that – unlike in real life – most participants engage with these materials. Study 2, thus, establishes the ceiling of the potential effect democratic justification may have.

As an example of a policy subverting democracy, Study 2 uses a media capture scenario. The struggle for independence of public television in Poland has been especially relevant since the victory of the Law and Justice (PiS) party in the 2015 parliamentary elections. The first steps PiS took to control the content of public television was to change the laws regarding the appointment of the National Broadcasting Council members, de facto leading to governmental control over the Council and the management of the public TV.⁴² The personnel changes introduced by the PiS party were quickly followed by the shift in news reporting: the main news programmes turned into a mouthpiece for the PiS government⁴³ praising the incumbents, inciting the audience against the opposition and the European Union, and imposing “a vision of the world that is often far from reality”.⁴⁴ Following the electoral victory of the Civic Coalition in 2023, the new Tusk government initiated changes to restore the role of the public media as a provider of objective information. However, given the complicated legal situation⁴⁵ and the politicization of the media, the proposals to reverse the damage caused by the PiS government remained controversial. Given this context, our experiment uses an ecologically valid scenario that sketches legal changes to the National Broadcasting Council and raises the issue of information manipulation.

Data and methods

We embedded our experiment in an omnibus survey fielded to 2,000 adult Poles recruited via Cint International. Respondents were sampled to match the Polish population in terms of age, gender, and region. The survey was fielded on December 12, 2024, and ran until February 19, 2025. Our final sample under-represents older individuals, especially respondents above the age of 65 (see demographic details in Online Appendix B).

The experimental design closely mirrors Study 1 using the same 2 (Party: In / Out) × 3 (Justification: Control / Majoritarian / Out-party Threat) between participants design. We focus our description of the design on changes to Study 1 and report full materials translated to English in Online Appendix C.

The first set of changes are to adapt the design to the Polish case. As discussed above, instead of election manipulation, we used media manipulation as the domain of autocratization. We also adapted the method for identifying in-, and out-parties in the Polish multiparty system. Specifically, we offered respondents – pre-treatment – a list of 16 major national political parties and asked them to identify their most and least liked party. Figures F.1 and F.2 in the Online Appendix display party preferences. We then pipe in the names of the least and most liked parties into the experimental vignette, manipulating whether one or the other is the incumbent. In a post-hoc decision, we exclude 76 respondents who picked the same party as their most and least liked. Results are robust to including them in the analyses.

As a second set of changes to our design, we increased attention to the stimuli. Most importantly, we moved the manipulation checks immediately after the vignette. If a respondent failed to correctly identify either the subversive party, or the justification offered, we asked them to re-read the vignette before presenting the questions with the dependent variables. For diagnostic purposes, we also included an attention check immediately before the experiment on an unrelated subject. Finally, we revised slightly the vignette adding brief headlines in bold to each paragraph. Again, these headlines sought to draw more attention to key information.

As with Study 1, all participants provided informed consent. Participants who failed either of two easy attention checks at the beginning of the survey were screened out. Our study received ethical approval from the Ethical Advisory Board of the Amsterdam Institute for Social Science Research (University of Amsterdam) issued on 2024, September 18.

Results

First, we display descriptive results, that is mean support for the subversive policy for each of the six experimental groups in Figure 2. As in Study 1, average responses across all groups hover around the scale midpoint neither supporting, nor opposing the proposition.

How attentive were respondents? Study 2 introduced stricter attention checks to ensure that respondents processed the experimental manipulations. While 98% of respondents passed a general attention check before the survey, only 43% correctly answered both factual manipulation checks on their first attempt. This number increased to 73% after prompting those who failed to reread the vignette.

Do people support democratic subversion from their own-party more? Yes, our findings strongly support H1. Consistent with Study 1, Polish respondents exhibited a significant partisan bias in their support for media manipulation. When the subversion was carried out by their in-party, respondents were more likely to endorse the policy ($\beta = 0.20$, 95% CI [0.18, 0.23]). This effect corresponds to approximately 63% of a standard deviation in the outcome variable, indicating a substantively large partisan bias. This effect is approximately twice the size of the one observed in Study 1. However, we must remember that through our adoption of the design to the Polish, multiparty system, we maximize the affective distance between the in- and out-party conditions. While in the US, many respondents may have merely lukewarm feelings towards the out-party (or the in-party), the larger menu of choice in Poland allows most to identify a party they genuinely like or and one they dislike, hence increasing the effect of partisanship.

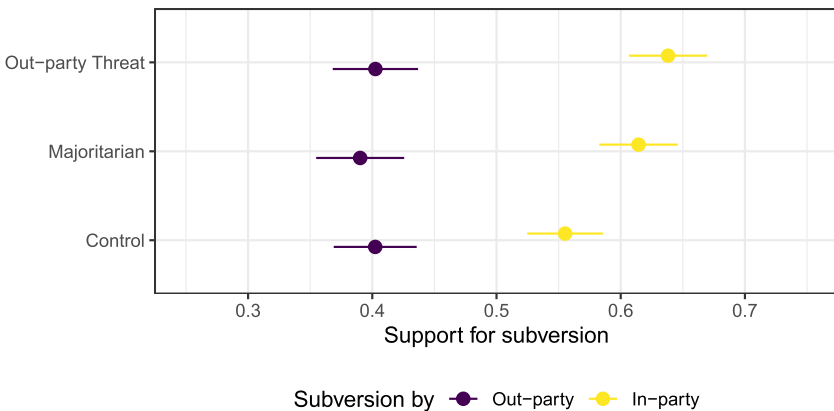


Figure 2. Justifications increase support for the anti-democratic policy in Poland, but only a little. Partisan alignment remains a very strong influence on support. Points denote group means. Error bars denote 95% confidence intervals.

Does justifying subversion with a majoritarian appeal or out-party threat increase its support? The evidence for justification effects is mixed. We find no support for H2: the majoritarian justification emphasizing the government's electoral mandate had no meaningful impact on support for subversion ($\beta = 0.02$, 95% CI $[-0.01, 0.06]$). However, the out-party threat justification had a small but statistically significant effect ($\beta = 0.04$, 95% CI $[0.01, 0.07]$). This effect remains modest in size, corresponding to only 13% of a standard deviation in the dependent variable.

In short, justification effects remained weak even after strengthening the salience of the manipulations. However, unlike in the U.S. study, here we find a significant interaction between partisanship and justification. Specifically, Polish respondents were entirely unmoved by justification attempts from the out-party but were more receptive to justifications from their in-party. Among in-party respondents, justification effects reached 19–25% of a standard deviation in support for subversion (Majoritarian: $\beta = 0.06$, 95%CI $[0.00, 0.11]$; Out-party Threat: $\beta = 0.08$, 95%CI $[0.03, 0.14]$). See formal tests of interaction effects in Online Appendix F.

Discussion

The findings from Study 2 in Poland reinforce and extend the conclusions from Study 1 (see a side-by-side presentation of main results in the the Online Appendix Figure F.3) Partisan bias in evaluating democratic subversion is robust across contexts, with citizens systematically more tolerant of subversion when it benefits their in-party. Justifications for subversion – whether framed as majoritarian legitimacy or protection from opposition threats – generally have limited persuasive power. However, unlike in the U.S., in-party justifications do shift opinions somewhat, demonstrating that partisan identity moderates the effectiveness of democratic rhetoric.

General discussion

Our findings confirm that partisanship strongly conditions support for executive aggrandizement, but they provide little evidence that the justification for subversion influences public opinion. While partisan loyalty remains a potent driver of tolerance for subversion, strategic justifications in the name of democracy appear to have only a marginal effect, and even then, only among in-party supporters. These results suggest that the public's acceptance of executive aggrandizement is primarily driven by partisan loyalty rather than the rhetorical framing used to justify it. Moreover, the precisely estimated null effects in the high external validity US sample (Study 1) indicate that even if such justifications have an impact, it is likely to be small.

Several limitations of our study should be acknowledged. First, despite efforts to enhance external validity, our experimental design remains an abstraction from real-world political decision-making, where individuals encounter justifications in more complex media environments and peer networks with much more at stake than in our experiment. Second, although we kept our mock newspaper article very simple and put a heavy finger on the scales forcing respondents to pay attention in Study 2, a substantial proportion of respondents still failed to recall key experimental details, suggesting that real-world justifications for subversion might be even less effective than observed here. Gaines and colleagues warn of the problem that survey experiments can over-estimate experimental effects, because “the survey provides a

captive audience, attending to more or less everything that the interviewer says”.⁴⁶ Third, our study focuses on two forms of democracy subversion in two national contexts, the U.S. and Poland, which differ in political polarization and democratic backsliding; further research should assess whether these findings generalize to other forms of subversion and other countries. Finally, we measure stated support for subversion rather than actual political behaviour – future work should explore whether similar partisan biases shape concrete actions, such as protesting, mobilizing, or resisting democratic backsliding.

Our findings present a nuanced picture for resistance to autocratization. On the positive side, we observe no evidence that autocrats can meaningfully shift public opinion with faux democratic rhetoric. Our data highlights that the tug-of-war in the arena of legitimation is fierce, autocrats cannot count on winning support simply by subverting democracy in the name of democracy. This aligns with theories of communication emphasizing that audiences tend to disregard messages when they doubt the source’s credibility or intentions.⁴⁷

However, our results also highlight a troubling limitation: partisanship significantly biases people’s evaluations, making them more tolerant of democratic erosion when initiated by their own party. Notably, a large co-partisan effect emerged even in our control condition, where respondents received one-sided messages from the opposition and a non-partisan NGO both criticizing the subversive party. Recent evidence shows that these messages matter: criticism can depress evaluations of politicians violating democratic norms.⁴⁸ It is therefore plausible that, absent these criticisms, the partisan gap would have been even larger, underscoring the persistent challenge partisanship poses for democratic resilience.

Our study also identifies important avenues for future research, particularly regarding whether accusations of undemocratic behaviour are similarly ineffective in shaping public opinion. Prior research has demonstrated that people readily label undesired policies as undemocratic⁴⁹, but it remains an open question if labelling a policy as undemocratic affects its desirability.

Methodologically, our findings illuminate key a trade-off in experimental studies of autocratization. While engaging designs, such as conjoint experiments, effectively identify causal mechanisms, they may poorly reflect real-world political discourse. Conversely, designs prioritizing external validity, like ours, risk encountering rational ignorance among respondents, who may ignore or forget critical justifications presented in realistic scenarios. This trade-off calls for careful consideration when designing future experiments to study democratic erosion.

Notes

1. Anderson et al., “Humanity’s Attitudes about Democracy and Political Leaders.”
2. Nord et al., “State of the World 2024”
3. Diamond, “Elections without Democracy.”
4. Graham and Svobik, “Democracy in America?”
5. Svobik et al., “In Europe, Democracy Erodes from the Right.”
6. Simonovits et al., “Democratic Hypocrisy and Out-Group Threat.”
7. Gidengil et al., “The Partisan Nature of Support for Democratic Backsliding”
8. For clarity of presentation, we relegate half of our hypotheses, concerning consistency with democracy, and corresponding results to Online Appendix E. We also renumbered our hypotheses for this manuscript but otherwise follow our pre-registrations.

9. Baturo and Tolstrup, "Strategic Communication in Dictatorships"
10. Dukalskis and Gerschewski, "What Autocracies Say (and What Citizens Hear)"; Mazepus et al., "A Comparative Study of Legitimation Strategies in Hybrid Regimes."
11. Anderson et al., "Humanity's Attitudes about Democracy and Political Leaders"
12. Guriev and Treisman, *Spin Dictators*, 17.
13. Matovski, *Popular Dictatorships*, 133.
14. Diamandouros, "Regime Change and the Prospects for Democracy in Greece" emphasis in the original.
15. Galston, *Is the Growth of Executive Power a Threat to Constitutional Democracy?*
16. Scheppele, "Autocratic Legalism"
17. Bor et al., "When Should the Majority Rule?"
18. Grossman et al., "The Majoritarian Threat to Liberal Democracy," 36.
19. Titelman et al., "When Citizens Support Democratic Backsliding in the Name of Democracy."
20. Kaftan and Gessler, "The Democracy I like"
21. for example, foreign agent laws Shyrokykh and Winzen, "International Actors and Democracy Protection"
22. Tomini et al., "Standing up against Autocratization across Political Regimes"
23. Schedler, "Rethinking Political Polarization," 355.
24. Braley et al., "Why Voters Who Value Democracy Participate in Democratic Backsliding."
25. Mercier, *Not Born Yesterday*.
26. Hewitt et al., "How Experiments Help Campaigns Persuade Voters"; Kalla and Brookman, "The Minimal Persuasive Effects of Campaign Contact in General Elections"
27. Robison, "Can Elites Escape Blame by Explaining Themselves?," 553.
28. Mazepus, "Are citizens vigilant of changing the balance of power?"
29. Dai and Kustov, "When Do Politicians Use Populist Rhetoric?"
30. Sarsfield, "Populist Rhetoric and Affective Polarization."
31. Mercier, *Not Born Yesterday*.
32. Skytte, "The Effect of Real-News Party Cues"; Zaller, *The Nature and Origins of Mass Opinion*.
33. Gelman et al., *Regression and Other Stories*.
34. Williamson, "Understanding Democratic Decline in the United States."
35. States United Democracy Center et al., "A Democracy Crisis in the Making"
36. Grossman et al., "The Majoritarian Threat to Liberal Democracy."
37. As demonstrated in OA Sections D and F, balance across experimental groups ensure that this simple strategy yields identical results to marginalizing the main effects from a regression including interaction effects as well. CF Muralidharan et al., "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments."
38. Montgomery et al., "How Conditioning on Posttreatment Variables Can Ruin Your Experiment"
39. Coppock and McClellan, "Validating the Demographic, Political, Psychological, and Experimental Results"
40. Wojcieszak et al., "Non-News Websites Expose People to More"
41. Singh and Thornton, "Elections Activate Partisanship Across Countries."
42. Weglińska et al., "Politicization as a Factor of Shaping News in the Public Service Media"
43. Chapman, "Pluralism under Attack"
44. Dopierała and Ossowski, "Polityzacja Audycji Informacyjnych w Polsce Na Przykładzie" *Wiadomości" TVP i" Faktów" TVN.*"
45. Wójcik, Restoring the Rule of Law in Poland: An Assessment of the New Government's Progress.
46. Gaines et al., "The Logic of the Survey Experiment Reexamined," 7.
47. Mercier, *Not Born Yesterday: The Science of Who We Trust and What We Believe*.
48. Hobolt and Osnabrügge, "Countering Authoritarian Behavior in Democracies."
49. Krishnarajan, "Rationalizing Democracy."

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions, grant agreement No 101130870 awarded to AB and by funding from the Challenges to Democratic Representation Research Group at the Department of Political Science (University of Amsterdam) to HM.

Notes on contributors

Alexander Bor is an ERA Fellow at the De- and Re-Democratization Group of the Democracy Institute at Central European University. His research focuses on how the human mind navigates social and political challenges like receiving and sharing political information, understanding and sustaining democracy, or selecting and evaluating leaders.

Honorata Mazepus is an associate professor of Comparative Political Behaviour at the Department of Political Science of the University of Amsterdam. Her work seeks to explain under what conditions citizens accept or contest different political entities: leaders, institutions, organizations, and decisions.

References

- Anderson, Christopher J., Damien Bol, and A. Ananda. "Humanity's Attitudes About Democracy and Political Leaders: Patterns and Trends." *Public Opinion Quarterly* 85, no. 4 (2021): 957–986.
- Baturo, Alexander, and Jakob Tolstrup. "Strategic Communication in Dictatorships: Performance, Patriotism, and Intimidation." *The Journal of Politics* 86, no. 2 (2024): 582–596.
- Bor, Alexander, Honorata Mazepus, Scott E. Bokemper, and Peter DeScioli. "When Should the Majority Rule? Experimental Evidence for Madisonian Judgments in Five Cultures." *Journal of Experimental Political Science* 8, no. 1 (2021): 41–50. doi:10.1017/XPS.2020.8.
- Braley, Alia, Gabriel S. Lenz, Dhaval Adjudah, Hossein Rahnama, and Alex Pentland. "Why Voters Who Value Democracy Participate in Democratic Backsliding." *Nature Human Behaviour* 7, no. 8 (2023): 1282–1293. doi:10.1038/s41562-023-01594-w.
- Chapman, Annabelle. "Pluralism under Attack: The Assault on Press Freedom in Poland." *Freedom House*, 2017.
- Coppock, Alexander, and Oliver A McClellan. "Validating the Demographic, Political, Psychological, and Experimental Results Obtained from a New Source of Online Survey Respondents." *Research & Politics* 6, no. 1 (2019): 2053168018822174.
- Dai, Yaoyao, and Alexander Kustov. "When Do Politicians Use Populist Rhetoric? Populism as a Campaign Gamble." *Political Communication* 39, no. 3 (2022): 383–404.
- Diamandouros, P. Nikiforos. "Regime Change and the Prospects for Democracy in Greece: 1974–1983." In *Transitions from Authoritarian Rule: Southern Europe*, edited by Guillermo O'Donnell, Philippe C. Schmitter, and Laurence Whitehead. Baltimore: The Johns Hopkins University Press, 1986.
- Diamond, Larry. "Elections Without Democracy: Thinking About Hybrid Regimes." *Journal of Democracy* 13, no. 2 (2002): 21–35.
- Dopierała, Weronika, and Szymon Ossowski. "Polityzacja Audycji Informacyjnych w Polsce Na Przykładzie" Wiadomości" TVP i" Faktów" TVN." *Com.Press* 2, no. 1 (2018): 4–25.
- Dukalskis, Alexander, and Johannes Gerschewski. "What Autocracies Say (and What Citizens Hear): Proposing Four Mechanisms of Autocratic Legitimation." *Contemporary Politics* 23, no. 3 (2017): 251–268.
- Fukuyama, Francis. *End of History and the Last Man*. New York: The Free Press, 1992.
- Gaines, Brian J., James H. Kuklinski, and Paul J. Quirk. "The Logic of the Survey Experiment Reexamined." *Political Analysis* 15, no. 1 (2007): 1–20. doi:10.1093/pan/impl008.
- Galston, William A. *Is the Growth of Executive Power a Threat to Constitutional Democracy?* < bound method Organization. get_name_with_acronym of> Organization ... , 2025.
- Gelman, Andrew, Jennifer Hill, and Aki Vehtari. *Regression and Other Stories*. Cambridge: Cambridge University Press, 2021.

- Gidengil, Elisabeth, Dietlind Stolle, and Oliver Bergeron-Boutin. "The Partisan Nature of Support for Democratic Backsliding: A Comparative Perspective." *European Journal of Political Research* 61, no. 4 (2022): 901–929.
- Graham, Matthew H., and Milan W. Svobik. "Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States." *American Political Science Review* 114, no. 2 (2020): 392–409. doi:10.1017/S0003055420000052.
- Grossman, Guy, Dorothy Kronick, Matthew Levendusky, and Marc Meredith. "The Majoritarian Threat to Liberal Democracy." *Journal of Experimental Political Science* 9, no. 1 (2022): 36–45. doi:10.1017/XPS.2020.44.
- Guriev, Sergei, and Daniel Treisman. *Spin Dictators*. Princeton: Princeton University Press, 2022. doi:10.1515/9780691224466.
- Hewitt, Luke, David Broockman, Alexander Coppock, et al. "How Experiments Help Campaigns Persuade Voters: Evidence from a Large Archive of Campaigns' Own Experiments." *American Political Science Review* 118, no. 4 (2024): 2021–2039.
- Hobolt, Sara B., and Moritz Osnabrügge. "Countering Authoritarian Behavior in Democracies." *Political Behavior* 47, no. 2 (2025): 781–800. doi:10.1007/s11109-024-09971-5.
- Kaftan, Lea, and Theresa Gessler. "The Democracy I Like: Perceptions of Democracy and Opposition to Democratic Backsliding." *Government and Opposition* 60, no. 2 (2025): 358–381.
- Kalla, Joshua L., and David E. Broockman. "The Minimal Persuasive Effects of Campaign Contact in General Elections: Evidence from 49 Field Experiments." *American Political Science Review* 112, no. 1 (2018): 148–166. doi:10.1017/S0003055417000363.
- Krishnarajan, Suthan. "Rationalizing Democracy: The Perceptual Bias and (Un)Democratic Behavior." *American Political Science Review* 117, no. 2 (2023): 2. doi:10.1017/S0003055422000806.
- Levitsky, Stephen, and Lucan A. Way. *Competitive Authoritarianism*. Cambridge University Press, 2010.
- Matovski, Aleksandar. *Popular Dictatorships: Crises, Mass Opinion, and the Rise of Electoral Authoritarianism*. Cambridge: Cambridge University Press, 2021.
- Mazepus, Honorata. "Are citizens vigilant of changing the balance of power? Testing evaluations of checks and balances reforms." *Political Research Quarterly* (2026): in print. doi:10.1177/10659129251408709
- Mazepus, Honorata, Wouter Veenendaal, and Anthea McCarthy-Jones. "and Juan Manuel Trak Vásquez. "A Comparative Study of Legitimation Strategies in Hybrid Regimes."." *Policy Studies* 37, no. 4 (2016): 350–369.
- Mercier, Hugo. *Not Born Yesterday: The Science of Who We Trust and What We Believe*. Princeton: Princeton University Press, 2020.
- Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. "How Conditioning on Posttreatment Variables Can Ruin Your Experiment and What to Do About It." *American Journal of Political Science* 62, no. 3 (2018): 760–775. doi:10.1111/ajps.12357.
- Muralidharan, Karthik, Mauricio Romero, and Kaspar Wüthrich. "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments." *The Review of Economics and Statistics* 107, no. 3 (2025): 589–604. doi:10.1162/rest_a_01317.
- Nord, Marina, Fabio Angiolillo, Ana Good God, and Staffan I. Lindberg. "State of the World 2024: 25 Years of Autocratization – Democracy Trumped?" *Democratization* 32, no. 4 (2025): 839–864. doi:10.1080/13510347.2025.2487825.
- Robison, Joshua. "Can Elites Escape Blame by Explaining Themselves? Suspicion and the Limits of Elite Explanations." *British Journal of Political Science* 52, no. 2 (2022): 553–572.
- Sarsfield, Rodolfo. "Populist Rhetoric and Affective Polarization." In *The Ideational Approach to Populism, Volume II*. London: Routledge, 2024.
- Schedler, Andreas. "Rethinking Political Polarization." *Political Science Quarterly* 138, no. 3 (2023): 335–359.
- Schedler, Andreas. "The Menu of Manipulation." *Journal of Democracy* 13, no. 2 (2002): 36–50. doi:10.1353/jod.2002.0031.
- Scheppele, Kim Lane. "Autocratic Legalism: The Hungarian Case." *University of Chicago Law Review* 85, no. 2 (2018): 545–583.
- Shyrokykh, Karina, and Thomas Winzen. "International Actors and Democracy Protection: Preventing the Spread of Illiberal Legislation in Eastern Europe and Central Asia." *Democratization* (2025): 1–23. doi:10.1080/13510347.2025.2461463.

- Simonovits, Gabor, Jennifer McCoy, and Levente Littvay. "Democratic Hypocrisy and Out-Group Threat: Explaining Citizen Support for Democratic Erosion." *Journal of Politics* 84, no. 3 (2022): 3. doi:10.1086/719009.
- Singh, Shane P., and Judd Thornton. "Elections Activate Partisanship Across Countries." *American Political Science Review* 113, no. 1 (2019): 248–253. doi:10.1017/S0003055418000800.
- Skytte, Rasmus. "The Effect of Real-News Party Cues." *American Journal of Political Science*, ahead of print, Wiley Online Library, 2025. doi:10.1111/ajps.70014.
- States United Democracy Center, Protect Democracy, and Law Forward. "A Democracy Crisis in the Making: 2024 Election Threats Emerging." *NGO Report*, 2023. <https://protectdemocracy.org/wp-content/uploads/2023/06/DCIM2023-04.pdf>.
- Svolik, Milan W., Elena Avramovska, Johanna Lutz, and Filip Milačić. "In Europe, Democracy Erodes from the Right." *Journal of Democracy* 34, no. 1 (2023): 5–20. doi:10.1353/jod.2023.0000.
- Titelman, Noam, Javier Sajuria, and Lisa Zanutti. "When Citizens Support Democratic Backsliding in the Name of Democracy." 2024. <https://osf.io/c8gzh>.
- Tomini, Luca, Sarah Gibril, and Vladislav Bochev. "Standing up Against Autocratization Across Political Regimes: A Comparative Analysis of Resistance Actors and Strategies." *Democratization* 30, no. 1 (2023): 119–138.
- Weglińska, Agnieszka, Łukasz Szurmiński, and Maria Wasicka-Sroczyńska. "Politicization as a Factor of Shaping News in the Public Service Media: A Case Study on Public Television in Poland." *Athenaeum. Polskie Studia Politologiczne* 72 (2021): 29–51.
- Williamson, Vanessa. "Understanding Democratic Decline in the United States." *Brookings Institution*, 2023. <https://www.brookings.edu/articles/understanding-democratic-decline-in-the-united-states/>.
- Wojcieszak, Magdalena, Ericka Menchen-Trevino, Bernhard Clemm von Hohenberg, Sijra de Leeuw, João Gonçalves, Sam Davidson, and Alexandre Gonçalves. "Non-News Websites Expose People to More Political Content Than News Websites: Evidence from Browsing Data in Three Countries." *Political Communication* 41, no. 1 (2024): 129–151. doi:10.1080/10584609.2023.2238641.
- Zaller, John R. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press, 1992.