

ARTICLE

Citizens as a democratic safeguard? The sequence of sanctioning elite attacks on democracy

Marc S. Jacob^{1,2} 

¹Center for Comparative and International Studies, ETH Zurich, Zurich, Switzerland

²Department of Political Science, Stanford University, Stanford, California, USA

Correspondence

Marc S. Jacob, Department of Political Science, Stanford University, Encina Hall West, Stanford, CA 94305, USA.
Email: msjacob@stanford.edu

Abstract

In many elections worldwide, citizens support politicians who have undermined democracy while in office. Why? For citizens to safeguard democratic institutions, they must not only disapprove of a politician's undemocratic conduct but also be willing to retract support from her at the next election. This article examines under which conditions citizen evaluations of undemocratic elite conduct become consequential for behavioral actions and whether specific segments of the electorate, such as politically educated, liberal, antimajoritarian, and moderate partisans, react more forcefully to such elite violations. Evidence from a survey experiment in Poland, closely following the sequence of presidential elections, reveals that citizens firmly dislike attacks on core electoral institutions, irrespective of whether they are committed by incumbent or oppositional copartisans. However, neither the electorate's nor any segment's dissent translates into revised vote choices. The study has implications for why undemocratic elite behavior often remains unpunished and citizens rarely avert democratic backsliding.

[The] shifting sector of the electorate must play a basic role in the workings of a democratic system, for the fear of loss of popular support powerfully disciplines the actions of governments.

– Key (1966, 10)

In the fiercely fought run-off of the Polish 2020 presidential elections, Poland's citizens were faced with a consequential choice: either confirm incumbent President Andrzej Duda, who supported the copartisan Law and Justice (*Prawo i Sprawiedliwość*, PiS) party in its quest to reform democratic institutions to its advantage or vote for Warsaw Mayor Rafał Trzaskowski, candidate of the oppositional alliance Civic

Coalition (*Koalicja Obywatelska*, KO). Among others, Trzaskowski promised to reestablish democracy and veto legislation that would induce any further democratic decay (BBC, 2020). And yet, despite having the viable option of casting a vote in favor of the opposition candidate, 51% of voters supported President Duda, allowing him to stay in office for a second term. After 5 years of gradual democratic backsliding (Chiopris et al., 2021; Sadurski, 2019), the Polish electorate missed a singular opportunity to sanction a powerful politician for subverting checks and balances.

That voters confirm political leaders who undermined democratic institutions is by no means unique to the Polish case. To name but a few, President Erdoğan of Türkiye, Prime Minister Orbán of Hungary, and President Chávez of Venezuela have been repeatedly reelected, despite their well-documented attacks on core democratic principles, including judicial independence and the integrity of elections. Why do voters often fail to protect democracy at the ballot box? And

Verification Materials: The data and materials required to verify the computational reproducibility of the results, procedures and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: <https://doi.org/10.7910/DVN/XERHRE>.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *American Journal of Political Science* published by Wiley Periodicals LLC on behalf of Midwest Political Science Association.

under which conditions are they willing to safeguard democracy by retracing support from politicians who attacked democratic institutions?

While a growing body of literature has begun to seek answers to these questions, we still have a limited understanding of how voters process, evaluate, and eventually respond to high-stakes voting situations such as the Polish 2020 presidential runoff election.

If citizens are to protect democratic institutions from a politician who attacked democracy, they need to look back at their initially supported politician's tenure, conclude that the politician's behavior violated democratic principles, and disapprove of her conduct. But merely expressing disapproval of elite attacks is not sufficient; it must also be consequential. In other words, discontent with a leader's undemocratic conduct must translate into withdrawn support at the ballot box. Therefore, whether citizens take issue when a copartisan leader violates democratic principles and, if so, to what extent this assessment has behavioral consequences, is at the core of studying electorates' ability to contain undemocratic elite behavior.

Whereas previous studies have mainly focused on either citizens' perceptual (e.g., Krishnarajan, 2023; Simonovits et al., 2022) or behavioral (e.g., Graham & Svulik, 2020) reactions to undemocratic elite conduct, this article examines a representative sample of Polish citizens as it both evaluates *and* responds to a politician who attacked core democratic principles in the sequence of presidential contests, just as they did when President Duda sought reelection in 2020. Specifically, I placed survey respondents in a situation where each first chooses between two presidential candidates running on two different party tickets (PiS and KO); she learns whether her endorsed candidate won or lost the election and undermined or sustained core electoral institutions, and she is finally asked to evaluate the politician's behavior and report whether she would confirm or reject her in another, subsequent election.

The experiment reveals that Polish citizens consider elite attacks undemocratic and disapprove of them. Nevertheless, this evaluation is inconsequential, as the survey electorate is mainly unwilling to withdraw self-reported electoral support from its initially supported presidential candidate.

If the electorate at large does not retract support, are segments of it readier to act in line with negative assessments of elite attacks? In close elections, shifts among only a tiny portion of voters can be sufficient to tip outcomes, rendering small segments pivotal (Ashworth & Fowler, 2020).

To examine whether at least certain types of voters would be willing to punish past undemocratic elite conduct in their political behavior, I identify characteristics of citizens that are deemed to strengthen their

commitment to democratic governance: the politically educated (Karp et al., 2003), citizens who embrace liberal and reject majoritarian forms of democratic rule (Grossman et al., 2022; Wunsch et al., 2022), and moderate partisans (Aarslew, 2023; Graham & Svulik, 2020). Reassuringly, many of these segments disapprove of undemocratic elite behavior more than the average voter does. But still, despite the more forceful dissent with such conduct, their evaluation remains without behavioral consequences: none of these segments withdraws more electoral support from their copartisans who violated democratic principles than the average voter.

These findings contribute to our understanding of not only why the Polish electorate failed to safeguard democracy in the 2020 presidential election but why voters—even those who are believed to be firmly committed to democratic rule—fall short of halting democratic backsliding induced by their democratically elected leaders more generally.

Crucially, Poland's democratic trajectory and societal context, particularly a strongly polarized society (Tworzecki, 2019), has much in common with other prominent cases of democratic decay such as Hungary and Türkiye (Laebens & Öztürk, 2021; McCoy et al., 2018; McCoy & Somer, 2019). Often seen as an exemplary case of democratization after the Eastern European revolutions (Bunce, 2003, 184–85), for a long time, Polish democracy has been characterized by elites committed to democratic conduct and peaceful electoral turnovers. This promising democratic path, however, was abruptly halted as PiS took over executive powers in 2015. Among others, the PiS government undermined the independence of the Polish judiciary and altered established legislative procedures (Sadurski, 2019, 3–4). PiS's "illiberal playbook" (Pirro & Stanley, 2022) resembles several other cases of backsliding, such as Türkiye and Hungary, enabling us to learn from Polish citizens' reactions to elite attacks on democracy about similar phenomena in other countries.

This article advances the study of citizens' potential to contain undemocratic elite behavior in three critical ways. First, it highlights the sequential nature of how citizens can intervene if elite attacks on democracy occur. While there is an emerging game theoretic literature addressing the temporal dynamics of backsliding (Chiopris et al., 2021; Grillo & Prato, 2023; Gratton & Lee, 2023; Helmke et al., 2022), empirical studies have refrained from confronting citizens with the sequence of elite attacks and documenting their evaluation and (self-reported) behavioral reactions. In contrast to survey experimental work featuring multiple candidate choices (Carey et al., 2022; Frederiksen, 2022; Graham & Svulik, 2020; Gidengil et al., 2022; Svulik, 2023; Wunsch et al., 2022), this study allows for (1) assessing citizens' willingness to revise vote choices in the

sequence of two elections, closely aligning with how citizens can safeguard democracy in the real world, and (2) to what extent evaluations of elite behavior have behavioral consequences.

Second, while the bulk of research has focused on incumbent misconduct (Aarslew, 2023; Simonovits et al., 2022) or neglected government status entirely, this article jointly studies citizens' reactions to winning and losing (or incumbent and opposition) politicians. To assess how voters react to attacks on core democratic institutions, I select two typical cases of attempts to undermine the integrity of elections that have occurred around the world: an *incumbent* seizing control over an electoral commission and an *opposition* candidate not conceding defeat. While incumbents are at the core of recent trends of democratic backsliding (Bermeo, 2016; Svulik, 2020), politicians in opposition may also distort or contribute to democratic stability and deserve more attention in theoretical and empirical work.

Third, and closely related, this article offers an explanation for why also opposition politicians' attacks on democracy often remain unpunished in contexts where one political force is mainly responsible for inducing democratic decay. In such settings, withdrawing support from a copartisan candidate as a current opposition supporter would bolster the undemocratic incumbent's chances of getting reelected. Hence, in cases where incumbent supporters fail to hold their copartisans accountable for subverting democracy, oppositional partisans will similarly be less willing to sanction their copartisan leaders for attacking democratic institutions. As a result of this dynamic, oppositional partisans, while being opposed to undemocratic elite behavior in principle, have no electoral alternative available to keep their own copartisan politicians in check.

CITIZENS, REPEATED ELECTIONS, AND ELITE ATTACKS ON DEMOCRACY

Citizens as a democratic safeguard

Several democracies have experienced sequences of autocratization in recent decades. However, these transformations differ from previously known regime changes, such as self-coups (*autogolpes*). Today, most incumbents gradually undermine democratic institutions while keeping parliamentary business and elections largely intact (Bermeo, 2016; Svulik, 2020). This development raises the question of whether the public can prevent political elites from subverting democracy. Citizens and their behavior in elections may serve as a powerful check on undemocratic elite conduct, namely when democratically elected elites

fear being punished by the electorate for undermining democratic institutions (e.g., Svulik, 2020; Weingast, 1997).

Previous research has built on this mechanism and sought to test citizens' readiness to withdraw electoral support from politicians who have shown undemocratic conduct. Most works have leveraged candidate-choice experiments (Carey et al., 2022; Frederiksen, 2022; Graham & Svulik, 2020), in which, along with party affiliation and policy platform, political candidates propose undemocratic reforms or have supported them in the past. Based on several choices between such candidate profiles, these studies infer the extent to which voters are willing to abandon undemocratic politicians. Although to different degrees, these works conclude that although many citizens do not withdraw support from undemocratic politicians, there is almost always a segment in the electorate that is willing to do so. For example, Graham and Svulik (2020) point to moderate voters and Wunsch et al. (2022) to mainly liberal-oriented citizens as potential guardians of democracy.

While this scholarship has advanced our understanding of voters' willingness to prioritize democratic institutions over other concerns such as partisanship, I argue that this literature has mainly overlooked the *sequence* of how voters can safeguard democratic institutions.¹ Namely, citizens vote, elected politicians then sustain or subvert democracy, and citizens evaluate and respond to this behavior in another subsequent election. This sequence is analogous to the timing in the literature of retrospective voting in electoral democracies, which theorizes that voters first observe an incumbent's performance in public office, evaluate her record, and finally decide to reelect her or instead support a challenger (Barro, 1973; Fearon, 1999; Fer-john, 1986). Similar to the setup of these models, I argue that this timing of events is a key component of the way in which citizens can contain elite attacks on democracy. This sequence, in turn, should inform empirical work when assessing citizens' evaluation of and response to undemocratic elite conduct.

Repeated elections and sanctioning elite attacks on democracy

Accountability mechanisms in representative democracies adhere to a distinct temporal logic (Manin et al., 1999). Voters express their support for a political candidate or party in an election; the electoral outcome determines which candidates will be in government and opposition; voters evaluate their performance and finally have the opportunity to reconsider their initial

¹ But see Graham and Svulik (2020) and Svulik (2023) for two observational studies on short-term vote shifts between two elections with an elite violation of democratic principles in between.

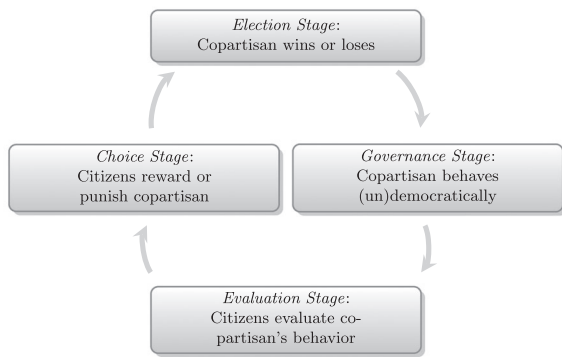


FIGURE 1 Sequence of the “Citizens as a Democratic Safeguard” mechanism. *Note:* The cycle illustrates the sequence of election outcomes (*election stage*), elite conduct to democratic institutions (*governance stage*), and citizen evaluation and behavior (*evaluation and choice stages*).

choice by either remaining with their initially preferred candidate or selecting another one.

This electoral cycle has strongly informed the vast literature on voting in democracies with repeated elections. In this line of reasoning, voters look back and judge the incumbent politician based on her performance in office. If the incumbent’s record meets the voter’s expectations, the latter will *reward* and reelect the former in office. This is because the voter expects a similar payoff from having that incumbent in office for another term. However, if the incumbent falls short of providing sufficient benefits to the electorate, the voter will *punish* her and choose the challenger (Barro, 1973; Ferejohn, 1986; Healy & Malhotra, 2013).

This distinct sequence of events can be similarly applied to situations where politicians violate democratic norms. As illustrated in Figure 1, the sequence of elite attacks on democracy and the stages in which citizens may safeguard democracy in elections are as follows:

1. After an election, citizens learn whether their copartisan candidate has won or lost (*election stage*).
2. Over the course of a copartisan politician’s tenure, citizens learn whether she attacked democratic institutions or refrained from doing so (*governance stage*).
3. Citizens evaluate whether this behavior violated democratic principles and they (dis)approve of it (*evaluation stage*).
4. Citizens decide to abandon or remain loyal to their initially preferred candidate at the next election (*choice stage*).

In this sequence, an individual citizen may not directly affect the outcome of an election, but she does learn about who won and lost the contest (*election*

stage). Similarly, it is for politicians to decide whether to follow or violate democratic principles during their tenure (*governance stage*). Violations may occur at any point between one election and the next, including refusing to accept ballot outcomes immediately after Election Day and attacks on democratic institutions that occur later during a politician’s tenure. By contrast, citizens only have the option to intervene in the next election if their copartisan has chosen to behave undemocratically.

Two deliberate actions are necessary for citizens to do so. First, at the *evaluation stage*, citizens need to identify elite attacks on democratic institutions as undemocratic and inappropriate. Second, at the *choice stage*, if voters find their copartisan politician’s behavior to run against democratic norms, they need to consider this violation severe enough to punish her by withdrawing electoral support.

In scenarios where political parties wield substantial power within a political system or when politicians who have displayed undemocratic behavior opt not to seek reelection or retire, this cycle can extend to political parties as well. Within this sequence, citizens reassess their voting preferences and reevaluate their support for the initially elected party whose leaders had engaged in actions undermining democratic principles during their previous term.

The next two sections focus on the specific stages where citizens play an active role in the safeguard cycle (*evaluation and choice stages*), while also examining the expectations surrounding their potential as democratic safeguards during these pivotal moments.

Evaluation stage: Citizens assess elite attacks on democracy

In order for citizens to act as a safeguard against undemocratic behavior by political elites, they must first be able to observe and evaluate such conduct. However, the extent to which the public is aware of such behavior depends on various factors, including the severity of the elites’ undemocratic conduct, the level of citizen engagement in political affairs, and the quality of media reporting. Severe violations of electoral principles, such as refusing to accept election outcomes or hijacking electoral institutions, represent the most extreme forms of attacks on democratic institutions and are likely to attract the greatest amount of public attention.

Although citizens’ awareness of undemocratic behavior by political elites is a prerequisite for a forceful response to such conduct, they must also recognize and acknowledge such behavior as undemocratic. Whereas Albertus and Grossman (2021) show that citizens across the Americas detect attacks on democratic norms, Krishnarajan (2023) provides evidence

that citizens evaluate undemocratic elite behavior committed by politicians who closely align with their policy preferences as less severe than those who advance disliked policies.

This study deliberately focuses on attacks on *electoral* institutions to minimize ambiguity regarding whether such elite conduct violates democratic principles.² While attempts to undermine institutions that constrain the executive branch (such as the judiciary) might be seen as necessary to give the elected ruling party more leeway in implementing policies, attacks on the process through which political officials are selected concern the representation of citizens in politics. Undemocratic actions such as refusing to concede and transgressing presidential norms to undermine the balanced composition of electoral commissions represent a violation of the democratic selection process and should thus be easily identified as undemocratic by citizens.

At the same time, it is less clear whether citizens disapprove of undemocratic behavior exhibited by politicians from their own party. As demonstrated by Simonovits et al. (2022), when respondents realize that their fellow party member holds power rather than being in opposition, they are more inclined to support policy reforms that undermine democratic principles. This could be because a politician's support base may also gain advantages from having an electoral commission composed of members from their own party or by dismissing the uncomfortable truth of losing an election.

Choice stage: Citizens punish or reward politicians' attacks on democracy³

Assuming that voters consider their initially preferred candidate's behavior undemocratic and disapprove of it, are they willing to withdraw support from this politician or even switch to an out-partisan candidate? Most existing work, while neglecting the sequential dimension of policing undemocratic elite behavior, has found that citizens tend to withdraw support from undemocratic politicians, yet to different degrees (e.g., Carey et al., 2022; Frederiksen, 2022; Graham & Svulik, 2020). Precisely this behavior would need to occur for citizens to safeguard democracy:

Hypothesis 1. Citizens should be more likely to withdraw support from an undemocratic than democratic copartisan.

The extent to which citizens are willing to punish, however, may depend on electoral outcomes: the intensity of citizens' punishing behavior may be contingent on whether the copartisan is in government or opposition. Mazepus and Toshkov (2022) found that citizens who voted for a losing party in the previous elections are more supportive of checks and balances than winning parties. Similarly, evidence from Brazil suggests that electoral losers maintained their support for democratic governance, even in the face of efforts by incumbent forces to induce democratic backsliding (Cohen et al., 2023). It is thus reasonable to expect that electoral losers would demonstrate greater resistance to political elites' endeavors to subvert democratic governance, while also displaying a stronger inclination to hold their own party accountable in subsequent elections for such conduct, in contrast to those who emerged as winners in the previous election. This is because democratic institutions provide the defeated party with a chance to participate and potentially secure victory in the upcoming election.

Hypothesis 2. Citizens are more likely to withdraw support from an undemocratic copartisan who lost an election than from an undemocratic copartisan who won.

But even if the electorate at large is unwilling to retract support from incumbent or opposition politicians, there might be segments that are readier to sanction elite attacks than others. In cases where election outcomes are close, it may suffice when democratically inclined voters revise their choices and tip the outcome against the incumbent. First, citizens have varying levels of knowledge about the function and purpose of certain democratic institutions (cf. Karp et al., 2003). Importantly, for citizens to punish undemocratic elite behavior, they need to have an adequate understanding of why the democratic process—including accepting electoral outcomes and impartial electoral commissions—constitutes a core principle of democratic governance.

Hypothesis 3. When citizens are uninformed about the democratic functions of institutions that elites might attack, they should be less likely to withdraw support from copartisans who behave undemocratically than their informed counterparts.

Citizens' diverging conceptions of what democracy constitutes (Davis et al., 2021; Osterberg-Kaufmann et al., 2020; Wunsch et al., 2022) may furthermore affect citizens' readiness to punish undemocratic elite conduct. While the bulk of research has focused on the extent to which citizens subscribe to the protection of minority rights and checks and balances, more

² See section entitled "Incumbent and Opposition Attacks on Electoral Institutions" for an elaboration.

³ Hypotheses 1 through 4b have been preregistered (<https://osf.io/hkt4x>). Hypothesis 5 was not preregistered but added due to its theoretical significance (Aarslew, 2023; Graham & Svulik, 2020).

recent studies assert that a considerable share of voters agrees that it is legitimate for the majority to decide over the minority without institutional constraints (Grossman et al., 2022). Holding such majoritarian understandings of democracy, in turn, may influence whether voters hold undemocratic politicians accountable for their attacks on democratic institutions. In the eyes of majoritarian citizens, institutions ensuring that minorities and opposition are considered in the political process are not democratically legitimate. Political actors intending to extend their power may leverage this weak commitment to minority rights and co-opt institutions that restrict their political power.

Hypothesis 4a. Holding a liberal understanding of democracy increases citizens' willingness to punish undemocratic copartisan behavior.

Hypothesis 4b. A majoritarian understanding reduces citizens' readiness to abandon copartisan politicians who have shown undemocratic conduct.

Besides knowledge and divergent understandings of democracy, previous work has pointed to intense polarization within society that prevents citizens from holding undemocratic elites accountable in elections (McCoy et al., 2018). According to these explanations, moderate voters exhibit a higher likelihood of distancing themselves from undemocratic politicians compared to citizens who possess a stronger partisan identification (Graham & Svobik, 2020; Svobik, 2020). This occurs because this segment of the electorate tends to exhibit greater indifference toward the partisan and policy orientation of political candidates, with concerns over democratic values playing a more dominant role. In contrast, stronger partisans are more likely to prioritize ideological considerations over democratic concerns.

Hypothesis 5. Stronger partisans are less likely to vote against undemocratic copartisan politicians than moderate partisans.

INCUMBENT AND OPPOSITION ATTACKS ON ELECTORAL INSTITUTIONS

Recent trends of democratic backsliding have been mainly driven by incumbent elites who take advantage of their powers (Bermeo, 2016; Svobik, 2020). However, the behavior of opposition forces toward democratic institutions can also have consequences for democratic stability. In this study, I thus examine undemocratic behavior displayed by both types of actors, with a focus on attacks on electoral institutions—a defining feature of democracies worldwide. If citi-

TABLE 1 Four forms of (un)democratic behavior by incumbent and opposition politicians.

| Conduct | Win | Lose |
|--------------|---|--------------------|
| Undemocratic | Seize control over electoral commission | Not concede defeat |
| Democratic | Maintain autonomy of electoral commission | Concede defeat |

Note: The matrix distinguishes between political candidates' electoral success (win vs. lose) and their postelection conduct (undemocratic vs. democratic).

zens are willing to penalize undemocratic conduct, we can expect such punitive behavior to be particularly pronounced when it comes to elite attacks on core electoral institutions. Hence, examining citizens' willingness to withdraw support from politicians who engage in such conduct presents an easy case to study public responses to elite attacks, as undermining the integrity of elections should be a clear violation of core democratic principles in the eyes of citizens.

This article focuses on two common violations of electoral principles: an incumbent seizing control over a multipartisan electoral commission and a defeated candidate refusing to accept electoral outcomes (see Table 1). In the first scenario, incumbents can use their constitutional powers to undermine democratic institutions throughout their tenure, often with severe consequences. I present survey respondents with an attack on the integrity of election management bodies (EMBs), which are responsible for tasks such as preparing for Election Day, administering the election, verifying returns, and settling disputes (Pastor, 1999, 8–9). I select EMBs because attempts to seize control over these bodies are a clear violation of core democratic principles, and yet, comparative evidence suggests that this type of incumbent attack is a common phenomenon in contemporary democracies. As shown in Figure 2a, between 2% and 16% of EMBs in democracies have not been autonomous in their work over the last two decades.⁴

While incumbents have the power to induce democratic backsliding, opposition politicians' undemocratic behavior may have less severe effects on democratic stability due to their exclusion from power. Nevertheless, opposition politicians who uphold their support for democracy, such as by accepting electoral outcomes, signal to citizens that current power arrangements are legitimate and have been assigned due process, even when the out-partisan party prevailed in the electoral contest. Moreover, when opposition politicians violate democratic principles, citizens can reasonably expect them to transgress demo-

⁴ I categorize cases in which the EMB is fully or almost independent of the government as "autonomous," while all other cases were classified as "not autonomous." See details on coding scheme in the online supporting information (pp. 33–34).

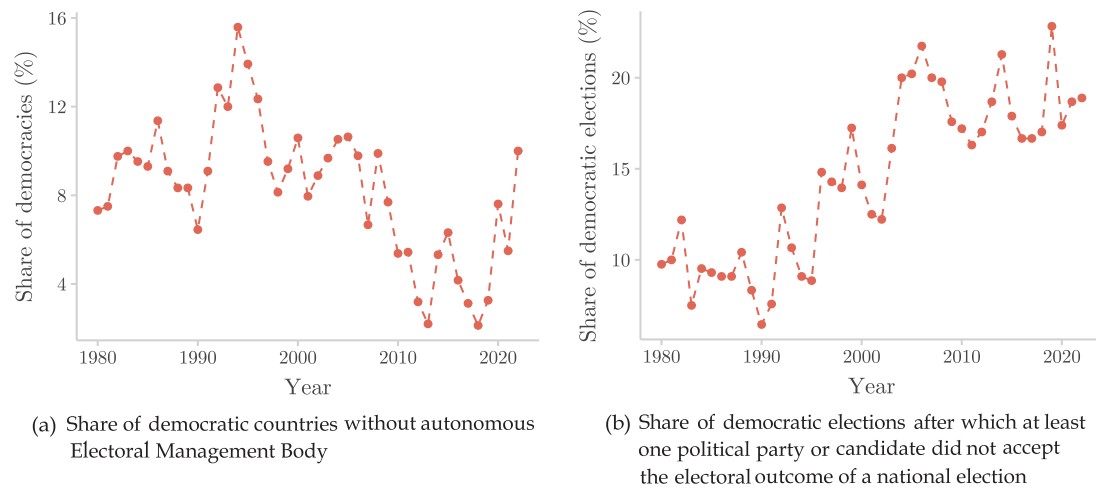


FIGURE 2 Autonomous electoral commissions and refused concessions in electoral and liberal democracies, 1980–2022. *Note:* See details on coding scheme in Appendix I in the online supporting information (pp. 33–34). *Source:* Varieties of Democracy (Coppedge et al., 2023).

cratic norms when in government, barring them from assuming executive power by retracting votes.

One of the most consequential forms of undemocratic opposition behavior occurs immediately after elections. That is, defeated elites do not accept the outcome and refuse to acknowledge the opponent's victory, a behavior that has occurred in around 9%–23% of democratic elections annually since 1995 (Panel [b] in Figure 2). Political forces who just lost an election and value stability more than violence will choose to admit their defeat and wait for the next opportunity to prevail in the electoral contest (Przeworski, 1991, 31). If they value stability less than violence, the defeated political camp may resort to the force of arms. This violent form of opposition behavior represents the most extreme form of not conceding defeat and has occurred in various places, including incumbent presidents refusing to leave the presidential palace after being voted out of public office (e.g., in the Gambia in 2017, see Kora & Darboe, 2017). In more consolidated democracies, not accepting defeat takes a less radical fashion but may still have far-reaching ramifications for democratic stability. In several instances, defeated political parties and candidates challenged the election outcome even after independent inquiries did not find evidence of voter fraud.⁵

Overall, elite attacks on electoral institutions have been a common phenomenon not only in recent cases of democratic backsliding but throughout the last decades. This article examines under which conditions citizens are willing to sanction political elites who have attacked electoral institutions while in government or opposition.

DEMOCRATIC BACKSLIDING IN POLAND

Poland's transformation toward democracy after the collapse of socialist rule in 1989 has often been considered a prime example of successful democratization after authoritarianism. Besides implementing institutional reforms that distributed power among branches of government, including the judiciary, Poland has been characterized by a democratically minded civil society, contributing to the consolidation of democratic rule (Rychard, 1998). In 2015, however, when Law and Justice (PiS) gained the majority in parliament and formed a government while winning the presidency earlier that year, democratic institutions have increasingly come under attack.

Since assuming power, the PiS government has consistently targeted the judiciary, media, and public administration to consolidate authority within the executive branch. Following his victory in the 2015 presidential election as a PiS candidate, President Duda declined to take oaths of office from several judges elected by parliament. Instead, he appointed a new President of the Constitutional Tribunal, resulting in the selection of three additional judges nominated by PiS and the disruption of the Tribunal's balanced composition (Pech et al., 2021, 6). Furthermore, the government expanded state-controlled media by acquiring additional media companies and replacing key figures with dedicated supporters (Zgut, 2022, 302–304), while systematically lowering employment standards for civil servants and showing favoritism toward party loyalists in hiring (Mazur, 2021, 106–107). The PiS administration also introduced measures that restricted the activities of non-governmental organizations, reduced the retirement age of judges in ordinary courts, and made alterations

⁵ Besides the 2020 U.S. presidential election (Arceneaux and Truex, 2022), defeated candidates running for President of Ghana have repeatedly refused to concede, even before finally prevailing over the rival candidate in a runoff election (Parku, 2014).

to established parliamentary procedures (Sadurski, 2019, 3–4).

The Polish president has occasionally played an ambiguous role in PiS's quest to centralize power in the executive branch of government. Due to the semipresidential design of Poland's political system (Sedelius & Mashtaler, 2013), where the president's main duty lies in the maintenance of the state apparatus (Słomka, 2019, 179), their authority is comparatively limited in contrast to the extensive powers of other presidents, such as those in the United States, Türkiye, and Brazil. Nonetheless, in addition to appointing members of constitutional bodies such as the Constitutional Tribunal and the National Electoral Commission, the Polish president may veto legislation that passed parliament. Whereas President Duda appointed many copartisan members to constitutional bodies and usually did not veto institutional reforms that passed the legislature, he sometimes refused to sign controversial legislation put forward by the copartisan PiS government into law. Among others, Duda vetoed a PiS bill that would have prevented businesses outside the European Economic Area from holding controlling stakes in Polish media (Reuters, 2021).

Despite the president's predominantly administrative responsibilities in Polish politics, he possesses the powers to reduce democratic backsliding by complying with democratic norms and vetoing legislation attacking democratic institutions. Citizens can withdraw support from an incumbent president every 5 years if she chooses to undermine democracy during her tenure. Moreover, as the government and not the president is mainly in charge of policymaking, citizens may prioritize sustaining democracy in presidential contests while following policy concerns in parliamentary elections. Poland's institutional background and recent experience with democratic backsliding thus provide a suitable environment in which to study citizens' willingness to hold incumbent and opposition elites accountable for behaving undemocratically in repeated elections.

In line with the comparative evidence on undemocratic elite behavior toward electoral institutions, electoral procedures, and the membership of Poland's Electoral Management Body (EMB), the National Electoral Commission (*Państwowa Komisja Wyborcza*, or PKW), have been subject to controversial debates in Polish politics, mainly because of a PiS bill that passed in 2018. Under the new law, seven out of nine members of the Commission are no longer nominated by the heads of Poland's highest courts but by parliament. Specifically, the seven members, who still need to be judges at a Polish court, are now selected by parliamentary parties and then elected in the legislature. The other two members continue to be nominated

by the heads of the Supreme Administrative Court (*Naczelny Sąd Administracyjny*) and Constitutional Tribunal (*Trybunał Konstytucyjny*), respectively.⁶

Notwithstanding the politicization of the Commission's membership, President Duda appointed judges nominated by both the incumbent party (PiS) and the main opposition alliances (Civic Coalition, or *Koalicja Obywatelska*; Polish Coalition, or *Koalicja Polska*; and The Left, or *Lewica*).⁷ It was the president's responsibility to appoint all judges to the Commission before and after the reform took effect, rendering him a critical actor in ensuring the functioning of democratic institutions. Even though the Commission's membership remains balanced, the recent reform of the appointment system suggests that the PiS government was willing to restructure established institutions in charge of administering elections. The case of the Commission hence provides an opportunity to examine how citizens may react to elite attacks that would undermine the balanced, multipartisan membership of a core democratic institution ensuring the integrity of elections.

EXPERIMENTAL DESIGN: SIMULATING THE SEQUENCE OF VOTING

In the remainder of this article, I report the results of an experiment implemented in Poland that closely follows the sequence of sanctioning undemocratic politicians.⁸ Drawing on Poland's institutional setup and the president's responsibility to appoint members of the Electoral Commission, I fielded a vignette experiment with a representative sample⁹ of Polish citizens ($N = 2910$) in which voters are confronted with a copartisan candidate running for president who, shortly after polling day, either attacks the integrity of electoral institutions or refrains from doing so.¹⁰

First, respondents were asked to imagine an upcoming presidential election in which two candidates compete, one running on a Law and Justice (PiS) and the other on a Civic Coalition (KO) ticket. No further information about the candidates, other than

⁶ However, the head of the Supreme Court (*Sąd Najwyższy*) was excluded from nominating a member after the reform took effect.

⁷ The main opposition forces in Poland are alliances between parties.

⁸ This survey experiment was preregistered at <https://osf.io/hkt4x>. Appendix A in the online supporting information (pp. 1–4) contains details on the experiment in Polish and English; Appendix D (pp. 11–16) evaluates the preregistered hypotheses and regression models, as well as additional statistical tests. For a power calculation on the realized sample size, see Appendix F (p. 31).

⁹ See descriptive statistics and composition of the sample in Appendix B in the online supporting information (pp. 5–8). The survey was administered online with YouGov in July/August 2021. Appendix H (pp. 31–32) shows that treatment groups are balanced on selected covariates.

¹⁰ Survey experiments featuring hypothetical scenarios have become a common tool in experimental research on democratic backsliding, see Simonovits et al. (2022).

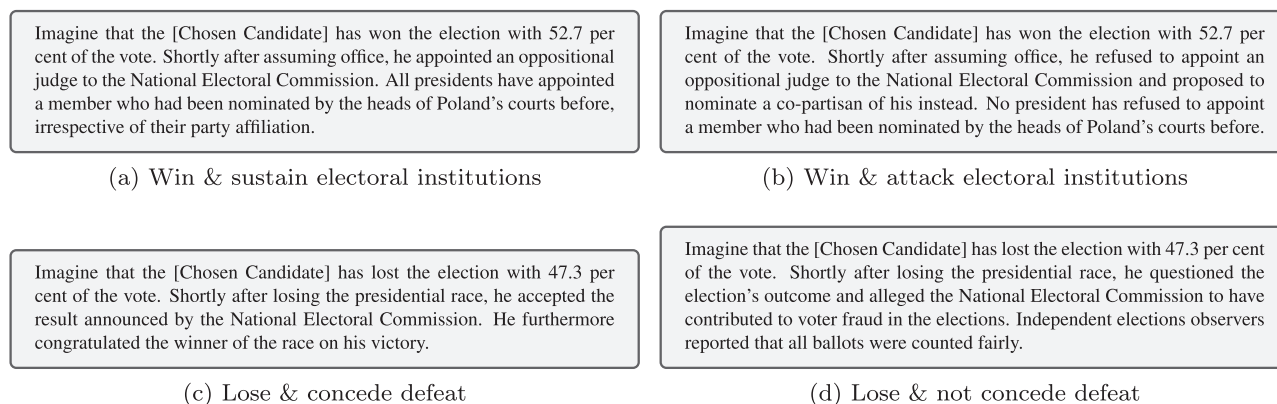


FIGURE 3 Hypothetical postelection scenarios presented to respondents. *Note:* Respondents were exposed to the behavior of the preferred candidate (referred to as “Chosen Candidate,” that is, either the Law and Justice (PiS) candidate or Civic Coalition (KO) candidate) selected before viewing the scenario vignettes.

that both candidates are male and around 50 years old, was provided to ensure that factors other than partisanship did not influence respondents' decisions in the experiment. Respondents were then asked to report (1) their likelihood of voting for each of the two candidates (from “extremely unlikely” to “extremely likely”) and (2) which of these two candidates they would vote for or whether they would not go to the polls.

After making their choice, all participants were randomly assigned to one out of four postelection scenarios providing information about (1) whether the preferred candidate won or lost the (hypothetical) election and (2) how the candidate behaved toward the integrity of elections.¹¹ Figure 3 displays the exact wording of the scenarios. In the *Win&Sustain* condition, the preferred candidate wins the presidential election and continues to appoint out-partisan judges to the Electoral Commission. In the *Lose&Concede* condition, the candidate was defeated in the election but sustained democracy by congratulating the winner.

The other two treatment conditions describe postelection situations where candidates attack democracy. In the *Win&Attack* condition, the candidate wins the election but refrains from appointing out-partisan nominees to the Electoral Commission.¹² Lastly, in the *Lose&NotConcede* condition, the defeated candidate challenges the election outcome despite no evidence of voter fraud.

It is important to acknowledge that the outcome of winning or losing an election has significant implications for the strategies through which elites

can undermine democracy. This prevents us from directly studying whether respondents respond differently when both incumbent and opposition politicians commit the same attack on democratic institutions. However, by comparing how respondents react to two distinct scenarios that vary in both incumbency status and undemocratic behavior, we can gain valuable insights into whether citizens exhibit stronger punitive responses in one scenario compared to the other.

After seeing one of the four scenarios, respondents were asked whether they considered the politician's behavior democratic or not and whether they approve of her behavior (*evaluation stage*). Finally, they were again asked to report how likely they were to vote for each candidate in another hypothetical presidential election (*choice stage*).

The key quantity of interest in this experiment is the shift in self-reported voting before and after learning about the preferred presidential candidate's postelection behavior. The voting likelihood scale is measured on a 7-point scale, ranging from *extremely unlikely* to *extremely likely* to vote for the candidate. Consequently, the main outcome variable used in the analysis ranges from -6 to 6 , where negative values indicate a decrease in support for the candidate, zero represents no change in voting likelihood, and positive values indicate an increase in voting likelihood after exposure to the treatment vignette.

This preregistered measure aligns with the response we would need to see for citizens to act as a democratic safeguard in repeated elections. To assess citizens' withdrawal of support from undemocratic versus democratic copartisans, I present simple difference-in-means estimates for the treatment conditions. Additionally, I provide similar estimates for respondents' evaluation of undemocratic elite conduct, including perceptions of copartisans' democratic

¹¹ Respondents who selected “would not go to the polls” were not assigned to any treatment group and immediately completed the experiment ($N = 631$).

¹² The vignette text also stated that previous presidents have appointed out-partisan members to highlight that their preferred candidate deviates from previously practiced norms.

behavior and approval ratings. Further details, including preregistered OLS models and evidence from additional statistical tests, are discussed in Appendix D in the online supporting information (pp. 11–16).

As I have argued, while the sequence of (1) voting, (2) observing electoral outcomes and candidate behavior, and (3) reconsidering vote choice closely resembles sanctioning mechanisms in electoral democracies, this design differs from existing experimental studies due to its sequential setup. The bulk of research in this area uses candidate-choice experiments to investigate citizens' willingness to defect from undemocratic politicians (Carey et al., 2022; Frederiksen, 2022; Graham & Svolik, 2020; Svolik, 2020, 2023).

However, when attacks on core democratic institutions such as elections are concerned, voters are not asked to make several choices between political candidates but only have a single decision to make: either remain loyal to or abandon a copartisan politician. In this sense, my experimental design presents a higher cost for respondents to withdraw support but more closely reflects real-world settings, such as presidential voting in Poland, Türkiye, and the United States.

Furthermore, my design combines quasi-behavioral outcome measures with attitudinal items. Previous research has mainly focused on either voting intentions (e.g., Frederiksen, 2022; Graham & Svolik, 2020; Wunsch et al., 2022) or attitudes toward elite behavior and political institutions (Simonovits et al., 2022). An exception is Mazepus and Toshkov (2022), who implement a between-subject design and include voting likelihood and evaluation items after providing information about the government's interference in appointing judges. Comparing attitudinal and quasi-behavioral outcome measures allows for studying whether the mere disapproval of elite attacks on electoral institutions also translates into (self-reported) political behavior.

Lastly, the adopted experimental design allows for evaluating *within*-subject behavior. Existing experimental work has either used a between-subject or a repeated candidate choice design to evaluate citizen responses to manipulated elite behavior. Both designs, however, do not allow for drawing conclusions about individual shifts in voting behavior. By contrast, pre- and postmeasuring of voting preferences provides insights into the magnitude of shifting behavior after learning about copartisan candidates' conduct toward electoral institutions (cf. Svolik, 2023).

Despite these strengths, the experimental setup comes with limitations. First, as the vignettes do not vary any other features of copartisan politicians, the design only allows for testing how citizens react to undemocratic elite behavior in a simplistic information environment. While this enables me to examine the effect of undemocratic elite conduct on citizen behavior, it precludes studying the role of addi-

tional candidate characteristics on voter evaluation and behavior. Furthermore, whereas respondents are confronted only for a few moments with two electoral contests and the observed candidate behavior in between, elections are held only every few years in the real world, and the timing of elite attacks varies. As with any survey experiment examining citizen responses to undemocratic elite behavior, it remains uncertain whether the effects of punishment behavior in the real world are stronger or weaker than those indicated by the experimental treatment effects.

However, it is important to note that the experimental design enables a comparison of how citizens perceive and respond to undemocratic elite actions, which aligns with a central objective of this study.

Second, it is essential to highlight that survey respondents approach the experiment with different viewpoints and political experiences. Most importantly, as described above, the PiS party has attacked democratic institutions, while the Civic Coalition has been in opposition at the national level throughout the recent period of democratic backsliding.

Although the experimental vignettes provide explicit information about the electoral outcome, supporters of the oppositional Civic Coalition may be more aware of the detrimental consequences of undermining democracy, as their political camp has not benefited from PiS's reforms. While this experience could lead to a more negative evaluation of undemocratic elite behavior, withdrawing support from their copartisan would essentially help the PiS candidate win the election. Hence, as the electoral alternative (i.e., PiS) has already subverted Polish democratic institutions in the real world, many KO voters will likely remain loyal to their copartisan candidate, even if confronted with a copartisan's hypothetical undemocratic conduct. By contrast, PiS voters support a party that has already attacked democratic institutions in the past and could hence be less sensitive when assessing and responding to undemocratic conduct, leading to a more benign reaction to copartisans who violate democratic principles.

EXPERIMENTAL RESULTS

Scenario 1: Winning copartisan attacks electoral commission

Do voters approve of winning copartisans' attempts to gain control over the Electoral Commission and consider such behavior undemocratic? Figure 4b reports respondents' mean evaluation of whether the candidate's conduct complied with democratic principles by treatment condition. Table 2 reports the corresponding difference-in-means estimates together with bootstrap confidence intervals and *p*-values. The

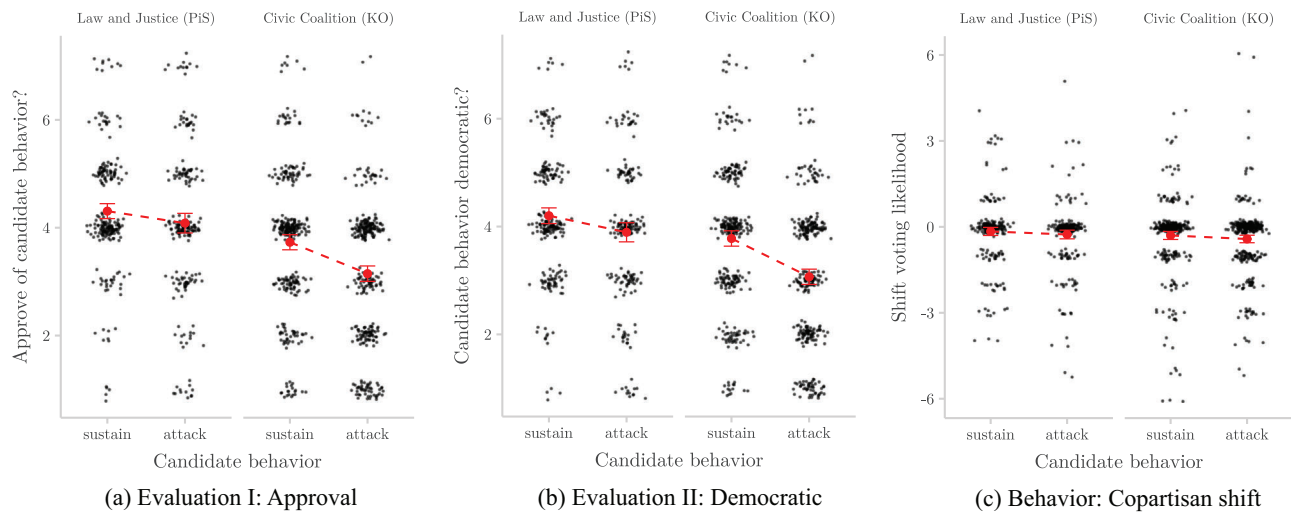


FIGURE 4 Copartisan wins: Mean responses for evaluation and behavioral outcomes by treatment condition and partisan camp. *Note:* Copartisan shifts are calculated by subtracting pretreatment from posttreatment support for candidate chosen before treatment assignment. Black dots plot original data points. Bars represent 95% confidence intervals.

TABLE 2 Copartisan wins: Difference-in-means between democratic and undemocratic treatment conditions.

| | Pooled | Law and Justice (PiS) | Civic Coalition (KO) |
|------------------|-------------------------------------|-------------------------------------|---------------------------|
| Approval | -.453** (-.610, -.296) | -.223 [†] (-.450, .006) | -.582** (-.782, -.388) |
| Democratic | -.560** (-.715, -.406) | -.307** (-.535, -.081) | -.711** (-.911, -.508) |
| Copartisan shift | -.120 [†] (-.262, .020) | -.107 (-.306, .096) | -.123 (-.312, .072) |
| <i>N</i> | 1137 | 469 | 668 |

Note: Pooled sample includes Law and Justice (PiS) and Civic Coalition (KO) voters. Copartisan shifts are calculated by subtracting pretreatment from posttreatment support for candidate chosen before treatment assignment. Ninety-five percent bootstrap confidence intervals in parenthesis. [†] $p < .1$; * $p < .05$; ** $p < .01$.

pooled sample and both voter camps evaluate attacks on the Electoral Commission as undemocratic. In a similar vein, as Figure 4a depicts, both partisan groups tend to disapprove of their copartisan attacking electoral institutions,¹³ where the difference-in-means estimate between democratic and undemocratic candidate behavior is more substantial for KO than for PiS supporters.¹⁴

Moving on to the choice stage, Figure 4c displays the mean shifts in self-reported voting likelihoods for the initially supported candidate by treatment condition and partisan group, while Table 2 presents the corresponding difference-in-means estimates. The more negative the estimate, the more subjects were willing to withdraw support from the preferred

candidate who attacked the Electoral Commission, compared to the one who did not. In contrast to the strong effect observed at the evaluation stage, the results for the quasi-behavioral outcome suggest that there is only a marginal, statistically insignificant negative effect on shifts in voting likelihood ($-.120$, CI: $-.262, .020$; p -value = $.095$). This finding speaks against Hypothesis 1.

Note that the pooled estimate is smaller than the smallest detectable effect based on the sample size; hence, we cannot confidently determine whether or not citizens withdraw support from their copartisan by a tiny margin. However, the marginal difference in shifting behaviors contrasts with the clear disapproval of incumbents' violation of democratic principles, suggesting that respondents are highly reluctant to express their discontent with that behavior in their vote choice. Given that shifts in voting likelihood are measured on an ordinal and not binary scale (rang-

¹³ However, the estimate for PiS voters fails statistical significance at the 95% level ($p = .055$).

¹⁴ See Appendix F1 in the online supporting information (pp. 17–18) for additional statistical tests on partisan treatment-effect heterogeneity.

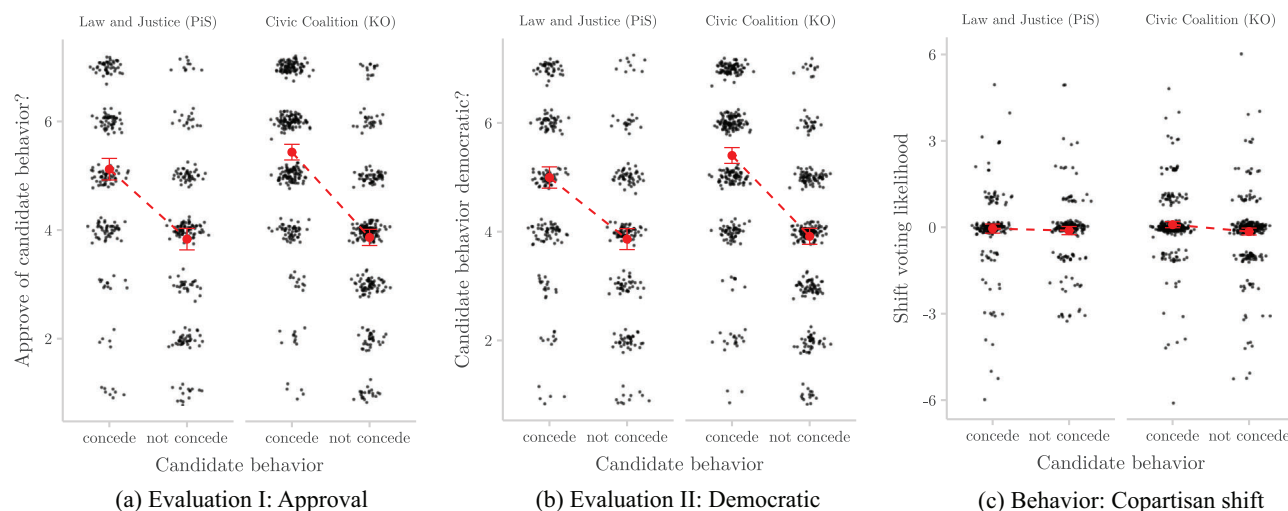


FIGURE 5 Copartisan loses: mean responses for evaluation and behavioral outcomes by treatment condition and partisan camp. *Note:* Copartisan shifts are calculated by subtracting pretreatment from posttreatment support for candidate chosen before treatment assignment. Black dots plot original data points. Bars represent 95% confidence intervals.

ing from extremely unlikely to extremely likely), it is even more remarkable that voters, if at all, only slightly withdraw some support from a copartisan who undermined electoral institutions.

In addition to shifts in voting likelihoods, I examine whether the winning candidate’s behavior affects vote switching to the rival candidate or abstaining at the next election, as well as shifts in self-reported voting likelihood for the out-partisan candidate (see Appendix E in the online supporting information, p. 17). Undemocratic elite behavior does not affect any of these outcomes. Taken together, even though many voters consider rejecting out-partisan nominees for the Electoral Commission as undemocratic and disapprove of such incumbent behavior, neither KO nor PiS supporters are willing to punish their copartisans for engaging in such conduct.

Scenario 2: Losing copartisan refuses to concede

How do voters perceive and respond to losing copartisans who do not accept defeat? As Figure 5b and Table 3 suggest, both the PiS and KO electorates are capable of detecting undemocratic behavior shown by defeated copartisans.¹⁵ A similar picture emerges when subjects are asked whether they approve of the candidate’s behavior (Panel [a] in Figure 5): both partisan groups approve of a copartisan’s refusal to concede considerably less than sustaining electoral integrity.

Does this negative evaluation translate into self-reported behavior? Figure 5c displays the mean shifts for the conditions in which the copartisan lost. The

pooled estimate indicates a slightly negative effect (−.172, CI: −.303, −.045; *p*-value = .009). When it comes to partisan differences, only KO voters, but not PiS voters, punish losing undemocratic copartisans when a copartisan refuses to concede defeat. Given the negative evaluation of the election loser’s undemocratic conduct, this result again suggests that most citizens, and PiS supporters in particular, are hesitant to withdraw support from their initially supported copartisan who refuses to concede defeat, although they dislike such conduct. Hence, the results for the losing scenario provide some support for Hypothesis 1. However, when investigating partisan groups separately, the quasi-behavioral punishment is mainly driven by KO voters. Even though the scenarios differ not only in whether the copartisan lost or won the election but also in the kind of violation of democratic principles committed, the results suggest that voters may be more willing to sanction undemocratic conduct shown by defeated candidates rather than incumbent candidates (H2).

A slightly different pattern emerges when examining PiS voters’ abstention behavior. After observing the defeated copartisan violating democratic norms, this partisan group is more likely to abstain at the next election (see Appendix E in the online supporting information, p. 17). At the same time, neither partisan group becomes more likely to vote for the alternative out-partisan candidate.

These findings indicate that PiS voters prefer not to vote at all when a copartisan undermines electoral principles, rather than casting a ballot for the out-partisan. Overall, however, despite considering candidates’ denial of defeat undemocratic and disapproving

¹⁵ Civic Coalition voters, however, rate not conceding slightly more undemocratic than Law and Justice voters (see Appendix E1 in the online supporting information, pp. 17–18).

TABLE 3 Copartisan loses: Difference-in-means between democratic and undemocratic treatment conditions.

| | Pooled | Law and Justice (PiS) | Civic Coalition (KO) |
|------------------|------------------------------|------------------------------|------------------------------|
| Approval | −1.461** (−1.629, −1.291) | −1.288** (−1.563, −1.011) | −1.570** (−1.777, −1.361) |
| Democratic | −1.349** (−1.511, −1.184) | −1.132** (−1.400, −.862) | −1.486** (−1.692, −1.280) |
| Copartisan shift | −.172** (−.303, −.045) | −.070 (−.283, .142) | −.238** (−.397, −.081) |
| N | 1142 | 451 | 691 |

Note: Pooled sample includes Law and Justice (PiS) and Civic Coalition (KO) voters. Copartisan shifts are calculated by subtracting pretreatment from post-treatment support for candidate chosen before treatment assignment. Ninety-five percent bootstrap confidence intervals in parenthesis. [†] $p < .1$; * $p < .05$; ** $p < .01$.

of it, there is no forceful rejection of undemocratic copartisan candidates.

Examining elite punishment among electoral segments

Are politically informed citizens more critical of undemocratic elite conduct and sanction this behavior more forcefully than their uninformed counterparts? And are liberal and majoritarian understandings of democracy associated with how voters evaluate and punish undemocratic copartisan politicians?

To examine potential heterogeneity in treatment effects based on political knowledge, respondents were asked to select the Electoral Commission's primary function from a list of items. As the results in Appendix F2 in the online supporting information (pp. 19–20) show, subjects who correctly identified the Commission's main purpose consider undemocratic elite behavior less democratic than those who failed the knowledge test. However, this evaluation does not translate into shifting behavior, as informed and uninformed voters withdraw support to similar degrees. Hence, politically informed voters are more capable of distinguishing between democratic and undemocratic behavior but do not sanction undemocratic copartisans more forcefully than their uninformed counterparts, thereby providing no support for Hypothesis 3.

A similar pattern can be found in the relationship between a liberal understanding of democracy and citizen evaluation of elite behavior (see Appendix C in the online supporting information, pp. 9–10 for items and measurement model and Appendix F2 pp. 21–24, for results). Respondents with a stronger commitment to liberal democracy disapprove more of undemocratic elite conduct and consider such behavior indeed as more undemocratic than democratic elite conduct but do not withdraw more support from copartisans who violated democratic principles. The more majoritarian voters are, the more positively they eval-

uate the winning candidate's behavior, irrespective of whether she sustains or attacks electoral institutions. Similarly, more majoritarian respondents do not abandon undemocratic copartisans more than democratic ones. Hence, Hypotheses 4a and 4b cannot be confirmed.

I furthermore test whether the intensity of identifying with the in- and out-party is associated with the evaluation of elite conduct and the willingness to withdraw support from undemocratic copartisan candidates (see Appendix F3 in the online supporting information, pp. 25–28). Overall, the strength of in-party identification is not correlated with voters' evaluation of politician's conduct or shifting behavior. Similarly, voters holding more favorable views of the out-party candidate are not more likely to sanction copartisans for behaving undemocratically, lending no support for Hypothesis 5. Lastly, respondents' economic status (measured by income) is also not associated with the degree to which voters withdraw support from an undemocratic copartisan candidate (see Appendix F3, pp. 29–30).

CONCLUSIONS: CITIZENS AS A DEMOCRATIC SAFEGUARD IN REPEATED ELECTIONS?

Why do voters often fail to protect democracy at the ballot box? And under which conditions are they willing to safeguard democracy and retract support from politicians who attacked democratic institutions? These questions are at the core of the growing body of literature on resistance against democratic backsliding, a regime transformation that unfolded in several countries worldwide. This article argues that citizens' ability to react only retrospectively to undemocratic behavior by political elites during the subsequent election has significant implications for both their willingness to safeguard democracy and the design of empirical studies aimed at assessing their readiness to hold undemocratic politicians accountable.

In order to contain undemocratic elite behavior in the context of repeated elections, voters need to look back at their preferred politician's behavior while in government or opposition and determine whether her conduct conflicts with their democratic principles. If they believe those principles have been breached, citizens decide if her misconduct is severe enough to withdraw support. The temporal sequence in which citizens can avert democratic backsliding is a key attribute of this safeguard mechanism, as voters can sanction undemocratic conduct at the ballot box only after they observe her behavior.

A survey experiment closely following the sequence of voting in Poland indicates that both main partisan camps (Law and Justice, or PiS, and the Civic Coalition, or KO) classify elite attacks as undemocratic and disapprove of them. Consistent with theoretical expectations, political knowledge and embracing a liberal notion of democracy increase the extent voters perceive elite attacks as undemocratic and disapprove of them. And yet, despite rejecting their copartisan politicians' conduct during their tenure, disapproval of elite conduct is mostly inconsequential for citizens' vote choices, irrespective of whether their political camp lost or won an election. However, KO voters withdraw some support from a copartisan candidate who refuses to concede but would also be reluctant to punish undemocratic behavior committed by a copartisan incumbent. PiS supporters do not show any substantive shifting behavior in either of the two scenarios.

Although democratic backsliding in Poland has only been induced by the PiS government in recent years, incumbent (PiS) and opposition (KO) partisans react mostly similarly to undemocratic copartisan conduct. At the same time, they may do so for vastly different reasons. Given PiS's previous attempts to attack democratic institutions, it appears that its voters have already priced in this kind of elite conduct: undermining the balanced membership of the Electoral Commission and not conceding defeat is not a sufficiently severe enough violation of democratic principles to revise a vote for PiS made in previous elections. By contrast, KO voters approach the experiment from a different perspective: retracting support from their copartisan candidate when she behaved undemocratically would benefit the incumbent PiS, a party that has already attacked democratic institutions. Consequently, opting to withdraw support from their undemocratic copartisan by either casting their vote for PiS or abstaining would inadvertently bolster the chances of an undemocratic out-partisan political force prevailing in elections. Even in the presence of a second opposition candidate, choosing to cast one's vote for another opposition contender may exacerbate divisions among opposition factions, bolstering

the prospects of the undemocratic incumbent during reelection campaigns.

Asymmetric democratic backsliding thus undermines the effectiveness of accountability mechanisms, as the absence of a viable electoral alternative committed to democratic governance and possessing realistic prospects of victory leaves the opposition electorate with limited options to avert elite violations of democratic principles.

The article's sequential approach and empirical findings have implications for how researchers construct experimental designs to assess citizens' willingness to sanction undemocratic partisans in repeated election cycles. Previous research has neglected this sequential aspect of voting against undemocratic politicians by either confronting survey participants with multiple candidate choices or vignette experiments without featuring reoccurring elections. This article's findings indicate that in contrast to previous work, not even a segment of the electorate—for example, moderate partisans (Graham & Svulik, 2020)¹⁶ or voters with more favorable views of out-partisan candidates (Aarslew, 2023)—would be willing to shift votes and sanction undemocratic behavior in repeated elections.

More generally, the article's findings yield rather worrying implications for the extent to which citizens may serve as a democratic safeguard in electoral democracies. Given that authoritarian-leaning incumbents—such as the PiS party in Poland—are at the core of trends of democratic backsliding in recent years (Çınar, 2021), it is concerning that voters on the winning side do not significantly defect from their initially supported candidate in a sequential voting scenario. Meanwhile, current trends in backsliding countries suggest that incumbents do not refrain from undermining even core democratic institutions. In Türkiye, the ruling AKP party already demonstrated in the 2019 Istanbul mayoral election that it is unwilling to concede without resistance and considers attacking electoral institutions a viable option (cf. Svulik, 2023). Only vote swings from undemocratic incumbents to opposition candidates would help contain the further decay of democracy in Poland, Türkiye, and elsewhere. These shifts could occur because citizens are dissatisfied with incumbents' policy performance, but most likely not because of their behavior toward democratic institutions. As this study suggests, withdrawing support from an initially endorsed politician because she subverted democracy is a price most citizens refuse to pay.


¹⁶ Note that Graham and Svulik (2020) also include citizens without any party preference in their analysis. In the study at hand, voters who would not vote for either of the candidates did not participate in the experiment. Nevertheless, the study yields insights into whether moderate *partisans* are more willing to reduce support than their more extreme counterparts.

While vote shifting between partisan camps seems unlikely to keep elites who have shown such undemocratic behavior in check, the mobilization of citizens who usually abstain in elections to turn out for the opposition may present an alternative electoral movement that could reduce the chances of an undemocratic incumbent being reelected. Notably, the 2023 Polish parliamentary election, in which the incumbent PiS party experienced heavy electoral losses, was characterized by a significant increase in turnout (New York Times, 2023). A promising direction for future research may thus be to identify the drivers of mobilization against undemocratic elites among citizens who have previously abstained from participating in elections.

ACKNOWLEDGMENTS

This research was supported by the Swiss National Science Foundation (Grant No. PZ00P1_185908) and approved by the ETH Zurich Ethics Commission (#2021-N-18). I would like to thank Lisa Anders, Moritz Bondeli, Alexa Federice, Kristian Frederiksen, Margaret Hanson, Callie Jones, Paula Jöst, Piotr Koc, Jasmin König, Barton Lee, Monika Nalepa, Nicole Olszewska, Ugur Ozdemir, Greta Schenke, Frank Schimmelfennig, Bennet Schwoon, Ronja Szczepanski, Shikhar Singh, Milan Svolik, Dimiter Toshkov, Natasha Wunsch, the audiences at the 2022 EITM Summer Institute, and the EUSA, MPSA, and WPSA Conferences, as well as three anonymous reviewers for their invaluable feedback.

ORCID

Marc S. Jacob  <https://orcid.org/0000-0001-8267-1956>

REFERENCES

- Aarslew, Laurits F. 2023. "Why Don't Partisans Sanction Electoral Malpractice?" *British Journal of Political Science* 53(2): 407–23.
- Albertus, Michael, and Guy Grossman. 2021. "The Americas: When Do Voters Support Power Grabs?" *Journal of Democracy* 32(2): 116–31.
- Arceneaux, Kevin, and Rory Truex. 2022. "Donald Trump and the Lie." *Perspectives on Politics* 21(3): 863–79.
- Ashworth, Scott, and Anthony Fowler. 2020. "Electoralates versus Voters." *Journal of Political Institutions and Political Economy* 1(3): 477–505.
- Barro, Robert J. 1973. "The Control of Politicians: An Economic Model." *Public Choice*, 19–42.
- BBC. 2020. Duda vs Trzaskowski: The fight for Poland's future. <https://www.bbc.com/news/world-europe-53339992>.
- Bermeo, Nancy. 2016. "On Democratic Backsliding." *Journal of Democracy* 27(1): 5–19.
- Bunce, Valerie. 2003. "Rethinking Recent Democratization: Lessons from the Postcommunist Experience." *World Politics* 55(2): 167–92.
- Carey, John, Katherine Clayton, Gretchen Helmke, Brendan Nyhan, Mitchell Sanders, and Susan Stokes. 2022. "Who will Defend Democracy? Evaluating Tradeoffs in Candidate Support Among Partisan Donors and Voters." *Journal of Elections, Public Opinion & Parties*, 32(1): 230–45.
- Chiopris, Caterina, Monika Nalepa, and Georg Vanberg. 2021. "A Wolf in Sheep's Clothing: Citizen Uncertainty and Democratic Backsliding." Harvard University, University of Chicago, and Duke University. https://www.researchgate.net/publication/348754770_A_WOLF_IN_SHEEP'S_CLOTHING_CITIZEN_UNCERTAINTY_AND_DEMOCRATIC_BACKSLIDING.
- Çınar, İpek. 2021. "Riding the Democracy Train: Incumbent-Led Paths to Autocracy." *Constitutional Political Economy* 32(3): 301–25.
- Cohen, Mollie J., Amy Erica Smith, Mason W. Moseley, and Matthew L. Layton. 2023. "Winners' Consent? Citizen Commitment to Democracy when Illiberal Candidates Win Elections." *American Journal of Political Science* 67(2): 261–76.
- Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, et al. 2023. *V-Dem [Country-Year/Country-Date] Dataset v13*. <https://doi.org/10.23696/vdemds23>.
- Davis, Nicholas T., Kirby Goidel, and Yikai Zhao. 2021. "The Meanings of Democracy among Mass Publics." *Social Indicators Research* 153(3): 849–921.
- Fearon, James D. 1999. "Electoral Accountability and the Control of Politicians: Selecting Good Types versus Sanctioning Poor Performance." In *Democracy, Accountability, and Representation*, edited by Adam Przeworski, Susan C. Stokes, and Bernard Manin, 55–97. Cambridge: Cambridge University Press.
- Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50:5–25.
- Frederiksen, Kristian Vrede Skaaning. 2022. "Does Competence Make Citizens Tolerate Undemocratic Behavior?" *American Political Science Review* 116(3): 1147–53.
- Gidengil, Elisabeth, Dietlind Stolle, and Olivier Bergeron-Boutin. 2022. "The Partisan Nature of Support for Democratic Backsliding: A Comparative Perspective." *European Journal of Political Research* 61(4): 901–29.
- Graham, Matthew H., and Milan W. Svolik. 2020. "Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States." *American Political Science Review* 114(2): 392–409.
- Gratton, Gabriele, and Barton E. Lee. 2023. "Liberty, Security, and Accountability: The Rise and Fall of Illiberal Democracies." *The Review of Economic Studies*.
- Grillo, Edoardo, and Carlo Prato. 2023. "Reference Points and Democratic Backsliding." *American Journal of Political Science* 67(1): 71–88.
- Grossman, Guy, Dorothy Kronick, Matthew Levendusky, and Marc Meredith. 2022. "The Majoritarian Threat to Liberal Democracy." *Journal of Experimental Political Science* 9(1): 36–45.
- Healy, Andrew, and Neil Malhotra. 2013. "Retrospective Voting Reconsidered." *Annual Review of Political Science* 16(1): 285–306.
- Helmke, Gretchen, Mary Kroeger, and Jack Paine. 2022. "Democracy by Deterrence: Norms, Constitutions, and Electoral Tilting." *American Journal of Political Science* 66(2): 434–50.
- Karp, Jeffrey A., Susan A. Banducci, and Shaun Bowler. 2003. "To Know it is to Love it?" *Comparative Political Studies* 36(3): 271–92.
- Key, Valdimir Orlando, Jr. 1966. *The Responsible Electorate*. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Kora, Sheriff, and Momodou N. Darboe. 2017. "The Gambia's Electoral Earthquake." *Journal of Democracy* 28(2): 147–56.
- Krishnarajan, Suthan. 2023. "Rationalizing Democracy: The Perceptual Bias and (Un)Democratic Behavior." *American Political Science Review* 117 (2): 474–96.
- Laebens, Melis G., and Aykut Öztürk. 2021. "Partisanship and Autocratization: Polarization, Power Asymmetry, and Partisan Social Identities in Turkey." *Comparative Political Studies* 54(2): 245–79.
- Manin, Bernard, Adam Przeworski, and Susan C. Stokes. 1999. "Introduction." In *Democracy, Accountability, and Representa-*

- tion, edited by Adam Przeworski, Susan C. Stokes, and Bernard Manin, 1–26. Cambridge: Cambridge University Press.
- Mazepus, Honorata, and Dimiter Toshkov. 2022. “Standing up for Democracy? Explaining Citizens’ Support for Democratic Checks and Balances.” *Comparative Political Studies* 55(8): 1271–97.
- Mazur, Stanisław. 2021. “Public Administration in Poland in the Times of Populist Drift.” In *Democratic Backsliding and Public Administration: How Populists in Government Transform State Bureaucracies*, edited by Michael W. Bauer, Guy Peters, Jon Pierre, Kutsal Yesilkagit, and Stefan Becker, 100–126. Cambridge University Press.
- McCoy, Jennifer, Tahmina Rahman, and Murat Somer. 2018. “Polarization and the Global Crisis of Democracy: Common Patterns, Dynamics, and Pernicious Consequences for Democratic Politics.” *American Behavioral Scientist* 62(1): 16–42.
- McCoy, Jennifer, and Murat Somer. 2019. “Toward a Theory of Pernicious Polarization and How it Harms Democracies: Comparative Evidence and Possible Remedies.” *Annals of the American Academy of Political and Social Science* 681(1): 234–71.
- New York Times. 2023. *Centrist Parties Poised to Oust Poland’s Nationalist Government*. <https://www.nytimes.com/2023/10/15/world/europe/poland-election.html>.
- Osterberg-Kaufmann, Norma, Toralf Stark, and Christoph Mohamad-Klotzbach. 2020. “Challenges in Conceptualizing and Measuring Meanings and Understandings of Democracy.” *Zeitschrift für Vergleichende Politikwissenschaft* 14(4): 299–320.
- Parku, Sharon. 2014. “Who says Elections in Ghana are ‘free and fair’?” *Afrobarometer Policy Paper*, 15.
- Pastor, Robert A. 1999. “The Role of Electoral Administration in Democratic Transitions: Implications for Policy and Research.” *Democratization* 6(4): 1–27.
- Pech, Laurent, Patryk Wachowiec, and Dariusz Mazur. 2021. “Poland’s Rule of Law Breakdown: A Five-Year Assessment of EU’s (In)Action.” *Hague Journal on the Rule of Law* 13(1): 1–43.
- Pirro, Andrea L. P., and Ben Stanley. 2022. “Forging, Bending, and Breaking: Enacting the ‘Illiberal Playbook’ in Hungary and Poland.” *Perspectives on Politics* 20(1): 86–101.
- Przeworski, Adam. 1991. *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*. Cambridge: Cambridge University Press.
- Reuters. 2021. *Polish president vetoes media bill, U.S. welcomes move*. <https://www.reuters.com/business/media-telecom/polish-president-says-he-vetoed-media-law-2021-12-27/>.
- Rychard, Andrzej. 1998. “Institutions and Actors in a New Democracy.” In *Participation and Democracy, East and West*, edited by Dietrich Rueschemeyer, Marilyn Rueschemeyer, and Björn Wittrock, 26–50. Armonk, N.Y. and London: M.E. Sharpe.
- Sadurski, Wojciech. 2019. *Poland’s Constitutional Breakdown*. Oxford: Oxford University Press.
- Sedelius, Thomas, and Olga Mashtaler. 2013. “Two Decades of Semi-Presidentialism: Issues of Intra-Executive Conflict in Central and Eastern Europe 1991–2011.” *East European Politics* 29(2): 109–34.
- Simonovits, Gabor, Jennifer McCoy, and Levente Littvay. 2022. “Democratic Hypocrisy and Out-group Threat: Explaining Citizen Support for Democratic Erosion.” *Journal of Politics* 84(3): 1807–11.
- Słomka, Tomasz. 2019. “President of the Republic of Poland: State Representative and Political Arbiter.” In *The Political System of Poland*, edited by Stanisław Sulowski and Tomasz Słomka, 177–90. Berlin and New York: Peter Lang.
- Svolik, Milan W. 2020. “When Polarization Trumps Civic Virtue: Partisan Conflict and the Subversion of Democracy by Incumbents.” *Quarterly Journal of Political Science* 15(1): 3–31.
- Svolik, Milan W. 2023. “Voting Against Autocracy.” *World Politics* 75(4): 647–91.
- Tworzecki, Hubert. 2019. “Poland: A Case of Top-Down Polarization.” *The ANNALS of the American Academy of Political and Social Science* 681(1): 97–119.
- Weingast, Barry R. 1997. “The Political Foundations of Democracy and the Rule of the Law.” *American Political Science Review* 91(2): 245–63.
- Wunsch, Natasha, Marc S. Jacob, and Laurenz Derksen. 2022. “The Demand Side of Democratic Backsliding: How Divergent Understandings of Democracy Shape Political Choice.” ETH Zurich. <https://osf.io/c64gfl/>.
- Zgut, Edit. 2022. “Informal Exercise of Power: Undermining Democracy Under the EU’s Radar in Hungary and Poland.” *Hague Journal on the Rule of Law* 14(2–3): 287–308.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jacob, Marc S. 2024. “Citizens as a democratic safeguard? The sequence of sanctioning elite attacks on democracy.” *American Journal of Political Science* 1–16.
<https://doi.org/10.1111/ajps.12847>