

“When they go low, we kick them”:

The reciprocal nature of citizens’ support for democratic transgressions^{*,†}

Short Title: Reciprocating Democratic Transgressions

Lisa Janssen[‡]

Hannah Werner[§]

Anna Kern^{**}

Abstract

This study examines how citizens respond to democratic transgressions by political opponents and whether they, in turn, support reciprocal transgressions by their own party. Drawing on psychological theories of negative reciprocity, we argue that commitment to democratic principles is relational, shaped by the (un)democratic behavior of other political actors. Through survey experiments in England (N = 4,040) and the United States (N = 4,084), we find that exposure to opponents’ transgressions more than doubles citizens’ likelihood of endorsing undemocratic actions by their own party. Deductive coding of open-ended responses (N = 907) reveals that many justify reciprocal transgressions as a way to restore fairness and protect democracy from an uneven playing field. Our findings highlight the need for an interactive approach to the study of democratic violations that captures the dynamic and complex nature of citizens’ attitudes towards undemocratic behavior.^{††}

Keywords: democratic backsliding; democratic transgressions; reciprocity; support for democracy; fairness.

* Replication files are available in the JOP Dataverse (<https://dataverse.harvard.edu/dataverse/jop>). The empirical analysis has been successfully replicated by the JOP replication analyst. Supplementary material for this article is available in the appendix in the online edition.

† This study was supported by a personal PhD fellowship from the Research Foundation Flanders (FWO; grant 11P7Y24N) and by the Belgian FNRS-FWO EOS project NOTLIKEUS (project 40007494).

‡ Ghent University, lisa.janssen@ugent.be

§ Corresponding author. University of Zurich, hannah.werner@uzh.ch

** Ghent University, anna.kern@ugent.be

†† This research involved human participants and was approved by the Institutional Review Board of Ghent University and conducted in compliance with relevant ethical guidelines.

“Michelle [Obama] always says, ‘When they go low, we go high.’ No. No. When they go low, we kick them.”

~ Eric Holder, former US Attorney General (2018)

Citizens’ willingness to condone undemocratic behavior is a defining problem in today’s democracies. A long line of scholarship emphasizes that public support is essential to democratic survival (e.g., Dahl 1956; Easton 1965; Lipset 1959). Citizens are expected to provide legitimacy to democratic institutions and constrain elites when they violate democratic norms, thereby serving as a last crucial bulwark against elite-driven autocratization (Weingast 1997; Wunsch and Gessler 2023). Yet, while many express strong support for democracy in the abstract (Wuttke et al. 2022), a growing body of research shows that citizens are willing to trade-off democratic principles for political gain (e.g., Aarslew 2023; Carey et al. 2022; Ferrer et al. 2025; Frederiksen 2022; Gidengil et al. 2022; Gidron et al. 2026; Simonovits et al. 2022). Indeed, a troubling pattern in (backsliding) democracies is that elites who undermine democratic principles rarely face meaningful electoral backlash (Graham and Svobik 2020; Wunsch and Gessler 2023). This has sparked a growing interest in why citizens fail to defend democratic institutions despite their stated commitment to them, with scholars exploring drivers such as rationalization (Krishnarajan 2023), different conceptions of democracy (Wunsch et al. 2025), and affective polarization (Voelkel et al. 2024).

In this study, we advance a different approach to explain citizens’ antidemocratic attitudes, one that foregrounds the relational nature of political behavior. Whereas existing research has mostly studied citizens’ attitudes towards democratic transgressions in isolated “clean slate” settings, we explicitly theorize and test the dependence of these attitudes on the conduct of other political actors. To that end, we develop a theoretical argument grounded in the concept of negative reciprocity: The practice of responding to negative behavior with similar negative behavior (i.e., an eye for an eye). Reciprocity has long been recognized as a

foundational feature of social life (Becker 1956; Gouldner 1960; Molm 2010). A vast body of research shows that the impulse to reciprocate is hardwired into the human brain, making it a strong and universal driver of human behavior (e.g., Chernyak et al. 2019). Negative reciprocity, in particular, reflects the belief that retaliation is a legitimate and appropriate response to perceived wrongdoing (Eisenberger et al. 2004). We extend this logic to the political domain, arguing that when political opponents violate the rules of the democratic game, citizens feel compelled to respond in kind. In other words, citizens are driven to *reciprocate* democratic violations. In this view, support for undemocratic actions is not formed in a vacuum but emerges as a reactive response to (perceived) transgressions by the other side.

In doing so, our theoretical framework complements existing research on the biased perceptions of opponents' willingness to support undemocratic practices (Braley et al. 2023; Druckman et al. 2023; Pasek et al. 2022). While this literature similarly emphasizes the interdependence of democratic commitments, it operates on a forward-looking logic: citizens are argued to preemptively adjust their own democratic commitments based on what they *expect* the other side to do. Instead, we argue that citizens look backward: Rather than anticipating undemocratic behavior, we examine how citizens react when democratic violations have actually occurred. This distinction is especially important as democratic erosion advances across the globe (Nord et al. 2026). In many countries, democratic violations are no longer anticipated but have become part of citizens' lived political reality. Theorizing on how citizens respond when such violations take place is crucial to understand when and why undemocratic attitudes gain traction amongst the general public in ongoing autocratization processes.

To test the role of reciprocity in shaping attitudes towards democratic transgressions, we conducted survey experiments in England (N = 4,040) and the US (N = 4,084). The experimental design introduces an interactive approach as it confronts respondents in the treatment group with fictional undemocratic behavior by the political opponent in vignettes. Following the treatment, respondents choose whether they think their in-party should engage

in behavior that mirrors the democratic violation or should not take any undemocratic action. We complement our experimental results with deductive coding of open questions to gain further insights into citizens' justifications for reciprocating democratic transgressions.

Our results show a clear pattern: Reciprocity is a powerful driver of support for democratic transgressions. Across both England and the US, we document large effect sizes, with respondents exposed to an opponent's democratic violation being more than *twice* as likely to endorse undemocratic acts by their own party. These findings are robust across a wide range of additional tests, including alternative operationalizations of the dependent variable, subgroup analyses, and primes designed to heighten awareness of democratic norms. Importantly, the tendency to reciprocate transgressions is not limited to highly polarized individuals or specific partisan groups. The absence of an interaction with affective polarization suggests that the impulse to reciprocate democratic violations is widespread, reflecting a broader tendency rather than a response confined to those with high partisan animosity.

In addition, we show that citizens reciprocate transgressions knowingly: Citizens do not endorse reciprocal undemocratic acts because they are simply unaware that it constitutes a breach of democratic rules. Rather, we find that informing citizens about the importance of key democratic principles does not change their impulse to reciprocate. Instead, based on respondents' self-reported motivations, we find that principles of fairness underlie the desire to respond in kind. Citizens express that mirroring the violation of the out-party allows them to restore balance within politics, essentially protecting democracy from an unequal playing field.

These findings have important implications for our understanding of citizens' commitment to democratic principles. First, they highlight that such commitment is malleable and conditional on the conduct of other parties in the political arena, underscoring the need for a relational perspective on democratic behavior. Studying these attitudes in isolation risks misinterpreting citizens' antidemocratic support in contexts where they encounter violations by the other side. This insight resonates with theoretical accounts of democratic stability that stress

the importance of mutual tolerance and shared commitments among political actors (Helmke et al. 2022; Levitsky and Ziblatt 2018; Schedler 2021; Weingast 1997). Second, by rooting our interactive approach in the well-established need for negative reciprocity, we offer a new and deeper psychological understanding of why democratic violations can ignite further undemocratic reactions. Our study highlights that many citizens experience a deeper trade-off – one between their commitment to democracy and their adherence to principles of reciprocity and fairness – that goes beyond mere self-interest as previous studies have often suggested. This insight implies that citizens do not (only) support reciprocal transgressions out of ill intent or because democratic violations have become normalized, but from a belief that reciprocation is the democratic and just thing to do.

Explaining anti-democratic attitudes

Healthy and stable democracies are predicated on the support of their citizens (Easton 1965). Scholars have long emphasized the importance of public support in maintaining democratic stability (e.g., Dahl 1956; Lipset 1959), arguing that when citizens' support for democracy erodes, so may democracy itself (Claassen 2020). Within democratic systems, citizens serve as a vital 'vertical' check: When elites violate democratic principles, they can be held accountable through the withdrawal of popular support. In that light, scholars have argued that citizens represent "the last bulwark against the systematic dismantling of checks and balances by overbearing executives" (Wunsch and Gessler 2023, p. 914).

But do citizens possess the strong, principled democratic commitments to fulfill this role? So far, the verdict is mixed. A burgeoning literature has emerged that studies an apparent conflict between citizens' strong commitment to democracy on paper (Wuttke et al. 2022; Holliday et al. 2024) and their willingness to trade-off democratic principles for political gain. Citizens are prepared to restrict the civil liberties of people they disagree with (Crawford and Pilanski 2014), they are inclined to vote for undemocratic candidates that align with their party

preferences (Graham and Svobik 2020), and they are willing to accept flawed democratic processes as long as they are on the winning side (Werner et al. 2025). This paints a paradoxical picture: Citizens profess to hold sincere democratic attitudes across a wide range of survey studies, but when push comes to shove, they are willing to condone democratic violations.

Scholars have advanced various reasons as to why people may not defend democratic principles despite a general commitment to democracy. The most prevalent one is that citizens can identify democratic transgressions but simply prioritize other goals, such as partisan victories or policy outcomes (e.g. Frederiksen 2022; Graham and Svobik 2020; Simonovits et al. 2022). They might prefer democratic over undemocratic behavior, *ceteris paribus*, but when faced with a trade-off between democratic procedures and political outcomes, they are willing to overlook transgressions. Another explanation is that citizens suffer from perceptual biases that lead them to apply double standards in their evaluation of elite behavior (Krishnarajan 2023). Through rationalization, citizens might not consciously give up on democratic norms but rather “find ways to convince themselves that they are getting their desired policy *and* democracy.” (Krishnarajan 2023, p. 474). A related argument is that different understandings of democracy may explain this conflict (Wunsch et al. 2025), as power grabs by popularly elected officials may be seen as consistent with a majoritarian definition of democracy.

We build on this literature but add a more interactive element to explain citizens’ support for undemocratic measures. We argue that the perceived acceptability of undemocratic behavior is not only based on citizens’ intrinsic interests, perceptions, or understandings, but that it is also contingent on the behavior of other parties. This interactive logic aligns with recent work showing that people’s democratic commitments are shaped by perceptions of opponents’ willingness to support undemocratic practices (Braley et al. 2023; Druckman et al. 2023; Pasek et al. 2022). Braley et al. (2023) conceptualize this dynamic as the ‘subversion dilemma’: Citizens become more willing to defect from democratic principles when they expect political opponents will do the same. While this literature highlights the fundamentally

interdependent nature of democratic commitments, it operates on a forward-looking logic centered on anticipated democratic subversion (akin to a prisoner's dilemma). Our approach is theoretically distinct in two respects. First, we shift attention from anticipation to reaction: citizens are not updating beliefs about what opponents might do in the future, but responding to what they have already done. In doing so, we highlight citizens' immediate reactionary impulse in contexts where democratic transgressions are actively unfolding. Second, and relatedly, we propose a distinct psychological mechanism grounded in negative reciprocity. We posit that when political opponents violate democratic norms, citizens become more willing to support democratic transgressions by their own side in response. In this view, support for undemocratic behavior emerges not only from partisan bias or strategic expectations about future opponent conduct, but from a backward-looking impulse to reciprocate perceived democratic wrongdoing and restore fairness after the rules of the political game were broken.

Reciprocity in the Context of Democratic Transgressions

A large body of research has shown that reciprocity is a universal and powerful driver of human behavior: Studies spanning diverse academic disciplines (e.g., psychology, biology, economics, anthropology), employing a wide range of methodological approaches, as well as covering varied cultural contexts, consistently underscore the omnipresence of reciprocal behavior (e.g., Falk et al. 2008; Fehr and Gächter 2000; Gouldner 1960; Molm 2010; Nowak and Sigmund 2005). The norm of reciprocity is so hardwired within our thinking that it is considered one of *the* defining features of social life (Molm 2010). Indeed, Becker (1956) has even referred to our species as the “homo reciprocus” (p. 1). While the concept of reciprocity is not undebated, it is widely understood as the practice of *offering a response in kind* (i.e., tit for tat, quid pro quo) (Fehr and Gächter 2000). Central to the practice of reciprocity is a pattern of mutually contingent exchange that ensures a symmetry of giving and taking in social life (Schedler

2021). Importantly, the innate human need for reciprocity is not confined to rational utility calculations and occurs even if it is not in people's self-interest (Nowak and Sigmund 2005).

Our argument regarding reciprocity in the context of democratic transgressions builds specifically on theories about *negative* reciprocity. 'Positive' reciprocity is commonly defined as a social practice that obligates people to repay favorable treatment (Eisenberger et al. 2004). 'Negative' reciprocity, in contrast, incorporates a retaliation principle in which "the emphasis is placed not on the return of benefits but on the return of injuries" (Gouldner 1960, p. 172). It is thus rooted in the widely known principle of 'an eye for an eye'. Negative reciprocity comprises a belief that retribution is a legitimized and proper way to respond to unfavorable treatment (Eder et al. 2006). Scholars have shown that negative reciprocity is deeply embedded in our social interactions and that propensities to punish harmful behavior (i.e., negative reciprocity) are even stronger than propensities to reward favorable behavior (i.e., positive reciprocity) (Chernyak et al. 2019; Shaw et al. 2019). The dynamics of negative reciprocity are documented in a plethora of contexts, ranging from counter-cheating in marital relationships (Jefferson et al. 2024) to reciprocal tariffs in interstate trade wars (Mattoo and Staiger 2019).

While negative reciprocity is thus a broadly observed human tendency, we argue that the political domain presents a particularly fertile ground for its dynamics to unfold. As a highly social and interactive arena, politics brings the norms governing interpersonal relations to the forefront. Furthermore, its competitive nature and high stakes heighten people's sensitivity to perceived inequalities, thereby increasing the propensity to reciprocate unfavorable acts (Shaw et al. 2019). Despite the theoretical appeal of reciprocity as a driving logic of political behavior (see Schedler 2021), we know surprisingly little about citizens' reciprocation tendencies in politics. Indeed, reciprocity "is a relatively underexamined phenomenon with, to date, little direct application to the political sphere" (Eder et al. 2006, p. 810).

In this paper, we apply the concept of negative reciprocity to the specific context of undemocratic behavior. Democratic principles serve as foundational rules of the political game.

When those rules are violated by political opponents, citizens can feel mistreated which activates people's innate tendency to reciprocate (Gouldner 1960). The central premise of our argument is thus that when faced with undemocratic acts of political opponents, the need for negative reciprocity compels citizens to engage in a tit-for-tat response, meaning that they come to support similar undemocratic acts by their own political camp. Importantly, negative reciprocity is a response to the *sense* of being harmed (Falk et al. 2008). Citizens do not need to witness factual undemocratic acts by political opponents to trigger these reactions: Merely perceiving the other side as having violated democratic norms (Braley et al. 2023; Krishnarajan 2023) – whether or not this is objectively true – can be sufficient to initiate a reciprocal reaction.

Anecdotal evidence at the elite-level offers illustrative support for our theory that democratic violations tend to be reciprocated (Helmke et al. 2022). A prominent example is the debate over court-packing in the US (Neal 2020). Following Republican efforts to block Obama's Supreme Court nominee in 2016 and accelerate the confirmation of a conservative justice under Trump, many Democrats began advocating for expanding the Supreme Court once back in power - explicitly framing this move as retaliatory and as “the only way to restore balance to the court and protect US democracy” (Kapur 2021). Similar tit-for-tat dynamics are evident in the use of gerrymandering. These elite-level interactions reflect a logic of negative reciprocity that mirrors the patterns we theorize at the citizen-level.

It is important to note, however, that we do not study interpersonal reciprocity among elites or horizontal retaliation between citizens. Instead, we focus on a *vertical* reciprocity: how citizens evaluate elite-level violations and how reciprocity shapes expectations for how their preferred party should respond. This dimension is particularly important given citizens' central role in protecting democracies through *vertical accountability*. By shedding light on how reciprocity norms influence this evaluative function, we contribute to understanding when and why citizens may fail to constrain undemocratic elite behavior. Again, vertical reciprocity too has clear real-world parallels. A striking example is California's so-called Election Rigging

Response Act, which was explicitly framed as a response to Republican gerrymandering in Texas and passed with 64.4% of the public vote, underscoring the importance of citizens' views in enabling the reciprocation of undemocratic practices (Jeong 2025). On this basis, we formulate the central hypothesis: *Respondents who are exposed to a democratic transgression perpetrated by a political opponent are more supportive of a (reciprocal) democratic transgression by the own party as compared to respondents who are not exposed to a democratic transgression perpetrated by a political opponent.*

Motivations for reciprocating democratic transgressions

There are different reasons why citizens might want elites to reciprocate democratic transgressions. For instance, voters may feel the need to re-establish an equilibrium in fairness (Rabin 1993). Reciprocal transgressions are then viewed as a way to remedy a perceived injustice, stemming from feeling wronged and mistreated by the opponent's prior breach of rules. When reciprocal violations are motivated by fairness, they intend to level the playing field of the democratic game. Reciprocal transgressions, from this perspective, are seen as the right and democratic thing to do: They provide an opportunity to even the score, restore balance in the democratic system, and prevent a greater harm – an unfair advantage for the other side – from occurring (for a related argument on democratic rationalization, see Krishnarajan 2023). This fairness-based motivation also highlights the democratic ambiguity of reciprocation, as actions intended to protect fairness and democratic balance themselves contribute to support for undemocratic behavior, a paradox to which we return in the conclusion.

Another potential explanation is rooted in the desire for revenge (see Riaz et al. 2024). Perceptions that the political opponent has violated democratic principles can instill feelings of anger, leading voters to seek emotional release through retaliation (Jackson et al. 2019). Under these conditions, reciprocal transgressions specifically *target* the political opponent: What

satisfies the reciprocal transgressor purely motivated by revenge is not the perceived restoration of justice within their democracy, but the suffering of the opponent for their initial provocation.

Last, exposure to a democratic defection could also initiate a process of normalization in which citizens adjust their perceptions of what constitutes appropriate and prevalent democratic conduct. Normalization is a largely subconscious process that can have a significant, but hidden, impact on citizens' beliefs (Bicchieri 2017). It can drive support for reciprocal transgressions by causing a shift in democratic norms, in which measures that were once viewed as violations of democracy may come to be seen as normal or routine practices within politics (Clayton et al. 2021). Normalization can also manifest through a change in the perceived public pressures associated with voicing undemocratic preferences, reducing the tendency for preference falsification (Valentim 2026).

We do not consider the theoretical explanations that we propose for the reciprocation of transgressions to be mutually exclusive or exhaustive. Indeed, Weingast (1997) and Helmke et al. (2022) develop theoretical models on the interdependence of democratic norms rooted in rational choice and deterrence (i.e., the expectation that transgressors will be discouraged from further violations if they foresee retaliation). Our focus on fairness, revenge, and normalization reflects extensive evidence that people reciprocate even when doing so is not in their self-interest (e.g., Nowak and Sigmund 2005), which has led scholars to argue that rational-choice models are “unambiguously refuted” as the central explanation of reciprocity (Fehr and Gächter 2000, p. 160). Importantly, however, research indicates that citizens' reasons to reciprocate are complex and challenging to study, often arising from a combination of factors (Jackson et al. 2019). Hence, we refrain from formulating hypotheses about the individual mechanisms underlying reciprocity and examine these exploratorily in the results section.

Experimental design

To empirically investigate the existence of negative reciprocity in the context of democratic transgressions, we fielded two pre-registered survey experiments in England (May 28-June 16, 2024) and the US (August 27-September 4, 2024).¹ Whereas the survey in England used quota sampling for gender and education, the survey in the US applied interlocked quotas for gender, age, and party affiliation. Prior research shows that Prolific provides high-quality data and that quota-based convenience samples yield valid and replicable insights, particularly when the focus is on experimental effects (Coppock 2019). Based on a priori power analyses using pilot data, we aimed for approximately 4,000 respondents per country, yielding well over 85% power to detect heterogeneous treatment effects. We planned for attenuated interaction effects that are half the size of the main effects documented in pilot tests. After data-cleaning, the final sample sizes are 4,040 and 4,084 respondents in England and the US, respectively.² Supplementary Information (SI) B provides more information about the composition of our samples.

England and the US are suitable cases for testing our argument as they are majoritarian democracies where two main parties typically alternate in power, allowing for a relatively straightforward design with clear in- and out-groups. England and the US, however, also differ in various important ways. Although two parties dominate English politics, it still functions as a multi-party system, with Reform UK gaining ground in recent years. We focus specifically on England rather than the UK to avoid complicating our design with regional party dynamics (e.g., SNP in Scotland). England and the US also differ in their experience with democratic transgressions and backsliding in the years leading up to our data collection: While our

¹ Both experiments were pre-registered (see SI A for more details), approved by the IRB of the host institution (anonymized for review), and an anonymized replication package is available at [OSF](https://osf.io/).

² We excluded 52 respondents in England and 141 in the US before randomization for not providing consent or failing attention checks. Another 49 respondents in England were screened out for not holding English nationality.

experiment was conducted before Trump's second term, the US had already undergone substantial democratic backsliding between 2016 and 2020. England, by contrast, had not experienced autocratization to the same degree (Coppedge et al. 2025). Examining our argument in both contexts therefore allows us to evaluate whether reciprocity tendencies hold in countries that share institutional features yet differ in recent democratic trajectories.

Our case selection also has implications for the generalizability of our findings. On the one hand, we examine reciprocity dynamics in long-standing democracies where democratic attitudes are presumably more deeply rooted (Hernández 2016), potentially offering a more conservative test of our theory. On the other hand, research shows that reciprocity depends on the number of actors involved in an exchange (Boyd and Richerson 1988), with diffusion of responsibility and free-rider problems weakening reciprocation in multiparty contexts. Insofar as both England and the US approximate two-party systems, our design may therefore be more conducive to detecting reciprocity effects than studies set in more fragmented contexts.

Procedure

After participants accepted the study on Prolific, they were directed to a web questionnaire and surveyed on their demographics, political attitudes, and level of affective polarization. Next, respondents were randomly assigned to either the treatment or control condition. The treatment group was presented with a screen with a hypothetical democratic transgression by a political opponent.³ The control group was not presented with any vignette and instead, moved on directly to answer questions on their support for democratic transgressions. Although demand-effects cannot be ruled out entirely, our between-subject design is less vulnerable to it and the open-ended answers provided no indication that respondents inferred the study's purpose.

³ We use hypothetical democratic transgressions to avoid deception, while maintaining experimental control. Hypothetical scenarios are common in experimental research and are shown to perform well at predicting real-world attitudes (Aarslew 2023; Hainmueller et al. 2015).

Moreover, this design allows us to empirically test the difference between citizens' willingness to *react* undemocratically versus their willingness to *act* undemocratically. Specifically, the control group mirrors the approach of prior studies that measure anti-democratic attitudes in an isolated setting, without exposing respondents to the behavior of political opponents. As such, it provides a meaningful benchmark for testing our core theoretical claim: That support for undemocratic behavior increases after exposure to democratic violations by opponents.

As preregistered, a random subset of the sample (i.e., 20%) was presented with an additional vignette before being assigned to the treatment or control group in which they were informed about core democratic principles in their country, matching the principle under attack in the consequent treatment text and measures. This additional vignette serves to test whether citizens' willingness to reciprocate democratic defections remains when respondents are made explicitly aware that this behavior goes against democratic principles in their country.

Stimulus

As people may react differently to violations of different democratic principles, we randomly assign respondents to one of three types of hypothetical democratic transgressions. In the procedure of designing the vignettes, we considered three criteria that are essential for a valid test of our hypotheses: Severity, realism, and targeting.

Considering the severity of the transgression, it is crucial that the democratic vignettes present respondents with a scenario that is not too incremental, since a transgression without substantive impact on democracy may not spark a reaction among citizens. On the other hand, the democratic defection should also not be too severe, since this may drive respondents to question the realism of the vignette. In the 21st century, autocratization has become an ambiguous process in which democracy is gradually broken down, often behind a legal façade (Lührmann and Lindberg 2019). In order to match this current mode of subversion and maintain realism, we aimed for transgressions that constitute a clear breach of liberal democratic

principles but that alone do not dismantle democracy. To that end, we also focus specifically on democratic transgressions by *politicians*, reflecting the contemporary trend in which democratically elected elites erode democracy from within. In addition to considerations of severity and realism, an important criterion for the vignettes is that they need to *target* the political opponent for reciprocal defections to be sensible. Negative reciprocity occurs when one feels wronged by the actions of another and decides to reciprocate against the perceived wrongdoer (Eisenberger et al. 2004). The transgression presented in the vignette must thus directly harm respondents' preferred party in order for them to feel negatively affected.

Based on these criteria, we selected the following transgressions: (1) taking action to prohibit protests by the political opponent, (2) proposing a law to restrict the voting rights of supporters of the opposing party, and (3) denying access to critical media during a press release (see SI, Table C1). These acts represent violations of liberal democratic norms rather than violations of the law per se (Ahmed 2023) and are grounded in real-world political events in both countries. SI O provides an overview of these real-life examples and shows respondents' high realism ratings of the vignettes. Given the real-world occurrence of these violations, as well as evidence that citizens overestimate opponents' willingness to subvert democracy (Braley et al. 2023), respondents in the control group may already perceive the opponent as having engaged in democratic transgressions. This pre-existing perception could lead to partial activation of reciprocity tendencies prior to the experiment, making the observed differences between control and treatment groups a more conservative estimate of the treatment effect.

Each transgression targets a key democratic principle: banning protests undermines the right to peaceful assembly, restricting voting rights violates electoral equality, and excluding media harms press freedom. Prior work shows that Americans value all three principles, though to a varying degree: Equal voting rights are among the most crucial democratic elements, while the toleration of protests is deemed somewhat less essential, and non-interference with the press even less so (Carey et al. 2019). Our data corroborate these findings (SI Table B1).

An example vignette with the corresponding dependent variable is presented in Table 1 (SI C provides further details on the vignettes and question-wordings). Which party appears as the respondent's in- or out-party in the vignette is determined by their pre-treatment responses. In both countries, respondents indicate which of the two major parties – Labour or the Conservatives, and Democrats or Republicans – they like the most. The least-liked of the two parties is treated as the respondent's political opponent in the vignettes, meaning that the treatment group sees an undemocratic act committed by that party. We focus in the vignettes on the two major parties in England because using transgressions by minor parties risks weakening the realism of the vignette and perceived credibility of the democratic threat. Nonetheless, to acknowledge the multi-party context in England, we deduce respondents' in-party from a question in which respondents rank five political parties (i.e., Labour, Conservatives, Liberal Democrats, Reform UK, and the Greens) from most to least liked.

Importantly, the credibility of the vignette might depend on whether the respondent's political opponent is in power at the time of data collection. To that end, we phrased the hypothetical undemocratic acts as initiatives or proposals - rather than fully implemented actions - so that it remains plausible for both governing and opposition parties to engage in them. Moreover, empirical tests show that treatment effects are independent of party affiliation and winner-loser status (see Figure 3 and SI F).

Table 1. Example vignette and dependent variable: Equal voting rights

Vignette and question wording	
Vignette (Treatment only)	Imagine the following: Politicians from [out-party] have taken action to pass a law that would make it more difficult for supporters of [in-party] to cast their vote during the next election. Politicians from [in-party] are now discussing how they should respond to this act
Dependent variable (Treatment and control)	Amongst politicians of [in-party], there is a discussion about how they should handle voting rules in the [UK/US]. What would you like to see them do? I) Politicians from [in-party] <i>should take action</i> to pass a law that would make it more difficult for supporters of [out-party] to cast their vote during the next election. II) Politicians from [in-party] <i>should not take action</i> to pass a law that would make it more difficult for supporters of [out-party] to cast their vote during the next election. III) Don't know

Measurements

After respondents read the vignette, we measure our main dependent variable, which is whether citizens support democratic transgressions by the own party. To that end, we create a survey item covering either the democratic principle of protesting, voting rights, or freedom of the press. Table 1 presents an example of the wording for equal voting rights. We first inform respondents that their favored party is discussing how to handle protests, voting rules, or critical media in their country. This sentence serves to reduce social desirability bias by highlighting that politicians may hold diverse views on the issue and that there is not one ‘correct’ answer. Next, we ask respondents to select which action they prefer their in-party to take. Given that reciprocity is about offering a response in kind, respondents can choose to endorse a (reciprocal) democratic transgression that mirrors the undemocratic act shown in the treatment-vignette, reject the (reciprocal) democratic transgression, or indicate that they are unsure which response they prefer. In line with the recommendations of Westwood et al. (2022), we include a don’t know option to prevent overestimation of support for undemocratic behavior (selected by 10.5% in England and 6.8% in the US). After answering this question, respondents are asked to provide a motivation for their answer in an open-ended question. This open-ended question provides a space for respondents to freely write down their justifications and is crucial in examining the reasons why citizens reciprocate transgressions.

A potential limitation of our dependent variable is that it presents respondents with a choice between endorsing a reciprocal transgression, rejecting it, or selecting “don’t know.” While this structure captures our core aim of identifying whether citizens wish to reciprocate or not, it does not encompass the full range of possible procedural or political responses. To address this concern, we supplement our analysis with more established candidate-choice measures (US only) and a scale capturing support for escalated reciprocal transgressions (see SI G and H, respectively). Across all measures, we observe consistent patterns, suggesting that our core findings are robust to alternative operationalizations.

Results

Do citizens reciprocate democratic defections? First, we present the means across treatment groups to gain insights into the overall baseline endorsement of undemocratic measures and the magnitude of the treatment effects. Figure 1 presents these descriptive results, showing the percentage of respondents who endorse a (reciprocal) democratic transgression in England and the United States across the control and treatment groups, both for the pooled results and for each vignette independently (i.e., protests, voting rights, and freedom of the press).⁴

Overall, the baseline support levels for undemocratic actions by the own party are relatively low: Among those in the control group, 11.5% in England and 10.9% in the US express a preference for their party to engage in undemocratic acts. This finding aligns with previous studies: Although citizens are often willing to tolerate undemocratic behavior for political gain (Graham and Svolik 2020)⁵, only a small fraction of the population explicitly endorses undemocratic measures when asked directly (e.g., Holliday et al. 2024; Westwood et al. 2022). While it is important to emphasize that the vast majority of citizens prefer their party to uphold democratic integrity, the fact that one in ten people express unprovoked support for democratic transgressions still speaks to a normatively troubling minority preference.

⁴ Following our preregistration, the samples differ across Figures 1–4. For the primary test of H1, we exclude respondents assigned to the democracy-information vignette. We further preregistered that, should this condition yield no substantively different results, these respondents would be included in the interaction analyses to ensure sufficient statistical power. For transparency, we report a replication of Figure 1 on the full sample in the SI D.

⁵ Direct measures and trade-off designs tap into different forms of democratic commitment: explicit endorsement of undemocratic behavior versus willingness to overlook such behavior in pursuit of partisan goals. As such, trade-off designs impose a higher threshold for democratic responses, as respondents must bear a political cost to uphold democracy, such as defecting from their preferred party. This contributes to the comparatively higher levels of tolerance for democratic transgressions observed in such studies and presumably in real-life contexts.

Turning to our first hypothesis, the expected reciprocal effect is clearly visible. As hypothesized, we find that the percentage of respondents who endorse democratic violations by the own political camp is larger when they have previously been exposed to a violation by a political opponent. More concretely, when respondents are informed about a democratic transgression by the political opponent, the percentages double from 11.5% to 22.4% in England and from 10.9% to 25.2% in the US. In sum, we find strong support for our argument on the role of reciprocity: When the opponent has broken the rules of the political game, citizens support transgressions by the own party in return. These conclusions are further supported by multinomial regression models presented in SI D, which show strong and highly significant treatment effects on respondents' support for reciprocal transgressions (results are robust to OLS specifications, see SI D). We thus confirm H1, where we expected a significant difference in citizens' support for a breach of democratic principles depending on whether they were previously exposed to a democratic transgression perpetrated by a political opponent.

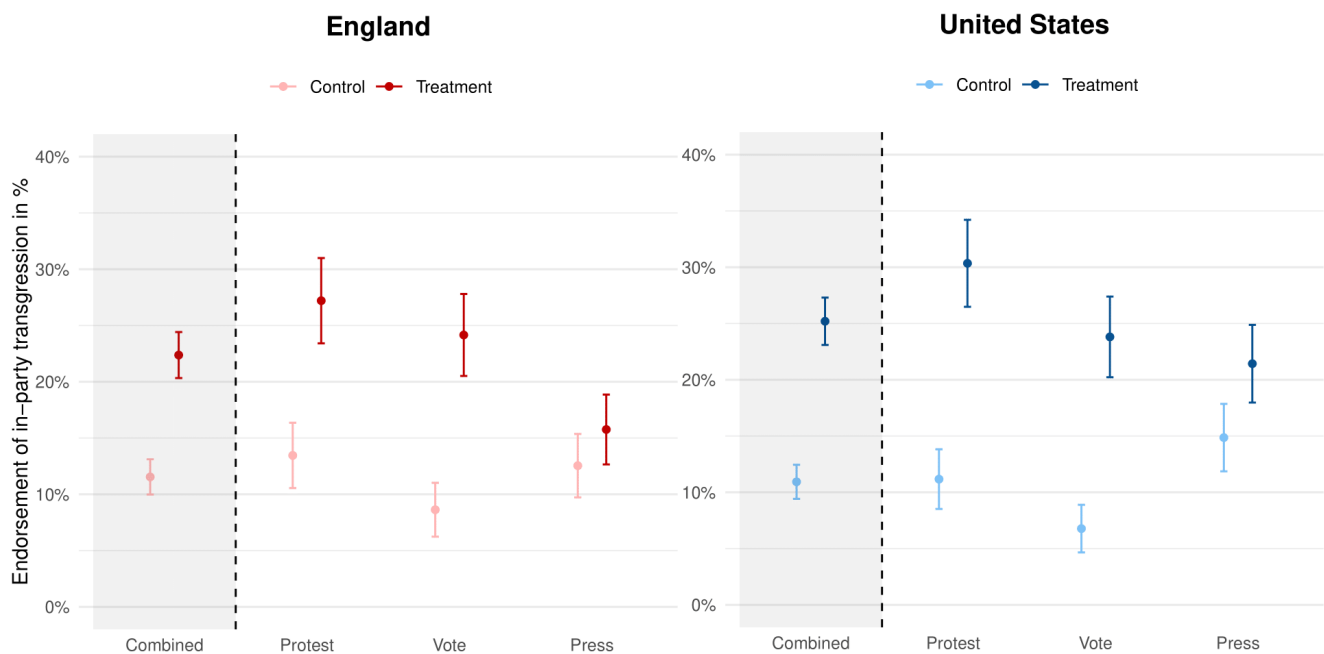


Fig 1. Percentage of respondents who support (reciprocal) transgressions.

Note: Error bars present 95% CIs. For the combined results, $N = 3,202$ in England and $N = 3,276$ in the US. As pre-registered, respondents receiving the additional vignette about core democratic principles are excluded. Model estimates are reported in SI Table D1.

Alt text: Two side-by-side dot plots with error bars, for England and the US, showing control and treatment estimates of percentage endorsement of in-party transgression across the combined sample and three vignettes, with treatment consistently higher.

Looking at the different transgression-types individually, we find strong support for reciprocal reactions for protesting and voting rights in both England and the US. On both topics, the percentage of respondents selecting an undemocratic response more than doubles for the treatment group as compared to the control group. The effect sizes are especially pronounced for violations on voting rights: In the US support for in-party transgressions even increases from 6.8% to 23.8%, representing more than a threefold rise. A potential explanation is that citizens regard voting rights as a cornerstone of democracy, making violations in this domain particularly salient (Carey et al. 2019; Ferrer et al. 2025). Moreover, elections represent a zero-sum game, where any advantage gained by one side directly comes at the expense of the other.

The results for freedom of the press, however, are less consistent. Though we again document a significantly higher willingness to engage in undemocratic acts among the treatment group in the US, the effect sizes are a bit smaller. Moreover, the effects do not reach significance in England. One possible explanation for the weaker reciprocal response on this issue is that violations of freedom of the press are perceived as less severe. Research shows that citizens view freedom of the press as a less central democratic principle compared to voting or protest rights (Carey et al. 2019). Consistent with this interpretation, we find the highest baseline support for undemocratic behavior in the control group on this topic in both countries.

Robustness checks and scope conditions

We conduct a range of additional tests to assess the robustness and scope of our model. First, we preregistered a second hypothesis that affective polarization could amplify tendencies to reciprocate transgressions. For reasons of space constraints we present an elaborate theory for this hypothesis as well as the detailed results in SI E. As illustrated in Figure 2, we document a statistically significant *direct* effect of affective polarization on respondents' endorsement of

in-party transgressions (RRR = 1.009, $p = 0.036$ in England and RRR = 1.009, $p < 0.001$ in the US). Substantively, this means that affectively polarized citizens are somewhat more supportive of in-party undemocratic measures, no matter whether they are in the control or treatment group. We, however, find no significant *interaction* effect between affective polarization and the treatment as visualized by the parallel slopes. The lack of an interaction effect indicates that tendencies to reciprocate democratic transgressions are not stronger among more affectively polarized citizens. Further robustness checks accounting for non-linearity (Janssen and Turkenburg 2025), separating in- and out-party affect, relying on horizontal social distance measures (England only), employing a mean-distance operationalization of affective polarization, using a dyadic measure of affective polarization, relying on a measure of partisan strength (US only), as well as running OLS regressions excluding ‘Don’t know’ responses yield similar results.

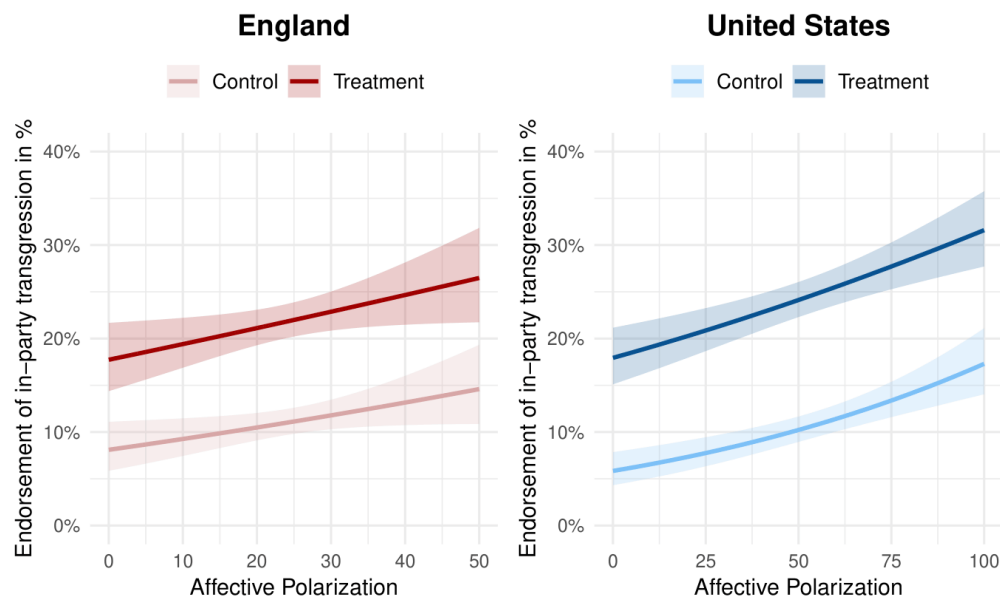


Fig 2. Support for (reciprocal) transgressions by affective polarization.

Note: Error bars present 95% CIs. $N = 3,983$ in England and $N = 4,052$ in the US. Model estimates are reported in SI Table E1. Figures disaggregated by transgression type are available in the extended SI Figure E4.

Alt text: Two side-by-side line plots, for England and the US, showing percentage endorsement of in-party transgression rising with affective polarization for both a lighter control line and a darker, higher treatment line, with roughly parallel slopes.

Second, we demonstrate in Figure 3 that reciprocity functions independently of who respondents select as the out-party. Opponents of the Labour and Conservative Party, as well as opponents of the Democratic and Republican Party are equally prone to reciprocate transgressions. However, the baseline support levels for undemocratic measures in the control group are considerably higher among opponents of the Democratic Party in the US.

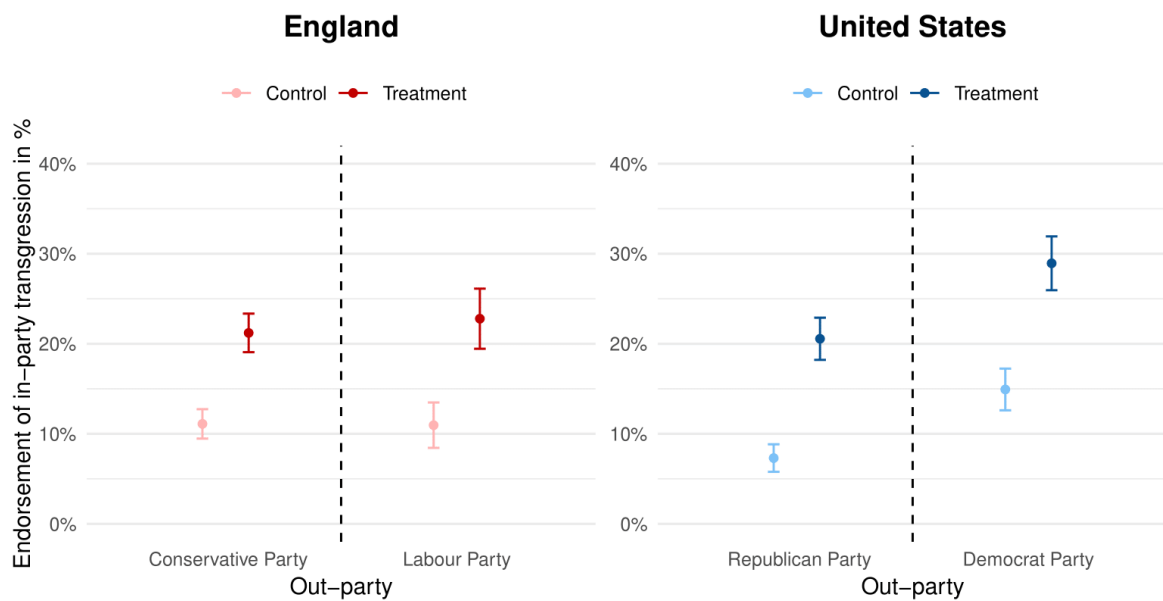


Fig 3. Support for (reciprocal) transgressions by out-party.

Note: Error bars present 95% CIs. N = 4,031 in England and N = 4,069 in the US. Model estimates are reported in SI Table F1. Figures disaggregated by transgression type are available in the extended SI Figure F1.

Alt text: Two side-by-side dot plots with error bars, for England and the US, plotting percentage endorsement of in-party transgression by out-party: Conservative and Labour in England, Republican and Democratic in the United States. Each out-party shows a lighter control estimate and a higher darker treatment estimate.

Third, we examine whether citizens not only mirror out-party violations but also support *escalated* responses. That is, undemocratic acts by their own party that exceed the severity of the opponent's initial transgression. For this, respondents rated their approval of an escalated undemocratic act on an 11-point scale (0–10) (See SI G for more details on the item wording and results). While baseline support for such escalated behavior is again low - 36.9% of English and 38.6% of American respondents strongly disapprove – we nonetheless observe significant positive treatment effects as shown in SI G. This suggests that exposure to a political

opponent's democratic violation not only increases support for direct reciprocation but can also push some citizens toward endorsing *more extreme* violations by their own party.

Fourth, we investigate whether reciprocity also holds for a different outcome variable that taps into mechanisms of electoral accountability. We find that treated respondents are more willing to vote for an undemocratic in-party candidate when the opposing party subverts democracy first, hence corroborating the findings from our main analysis (SI Table H2).

Lastly, we test whether reciprocity persists even when citizens are made explicitly aware that it goes against core democratic principles. Prior research indicates that citizens have varying understandings of democracy and frequently fail to accurately identify the actions of politicians as undemocratic (Krishnarajan 2023; Wunsch et al. 2025). Consequently, it could be argued that citizens only reciprocate the behavior of the opponent because they are unaware that these acts constitute a breach of democratic principles. Hence, we run a pre-registered analysis among a random subset of the sample (i.e., 20%) who are shown information about the importance of either protesting, voting, or freedom of the press prior to treatment-allocation (vignettes available in SI I). The democratic information provided to respondents always matches their assigned transgression (e.g., those assigned to the transgression of voting rights receive information specifically about the importance of voting). Making democratic principles explicit can have a dual effect: it may heighten citizens' democratic awareness and make them more inclined to uphold democracy, but it may also amplify perceptions of wrongdoing and the desire to reciprocate by clearly stating that the opponent's actions are undemocratic.

The results of the analysis are presented in Figure 4 (see SI I for more details). The baseline endorsement of undemocratic behavior by the own party is somewhat lower when respondents are first informed about the democratic principle in their country (i.e., 9.2% in England and 10.3% in the US), which is also reflected in the statistically significant and negative direct effect of this additional vignette (RRR = 0.785 in England, and RRR = 0.775 in the US). We, however, find no indication that the treatment effects are significantly lower when

respondents are made explicitly aware of the democratic principle. If anything, in England, we document that the treatment effects become slightly stronger with a relative risk ratio of 2.45 as compared to 2.13 for the aggregated results. These findings further underline citizens' willingness to reciprocate democratic transgressions, even when they are told about the importance of upholding the principle for democracy. So, though informing respondents about the democratic principles in their country slightly decreases their baseline support for undemocratic behavior, it does not seem to affect their tendency to reciprocate.

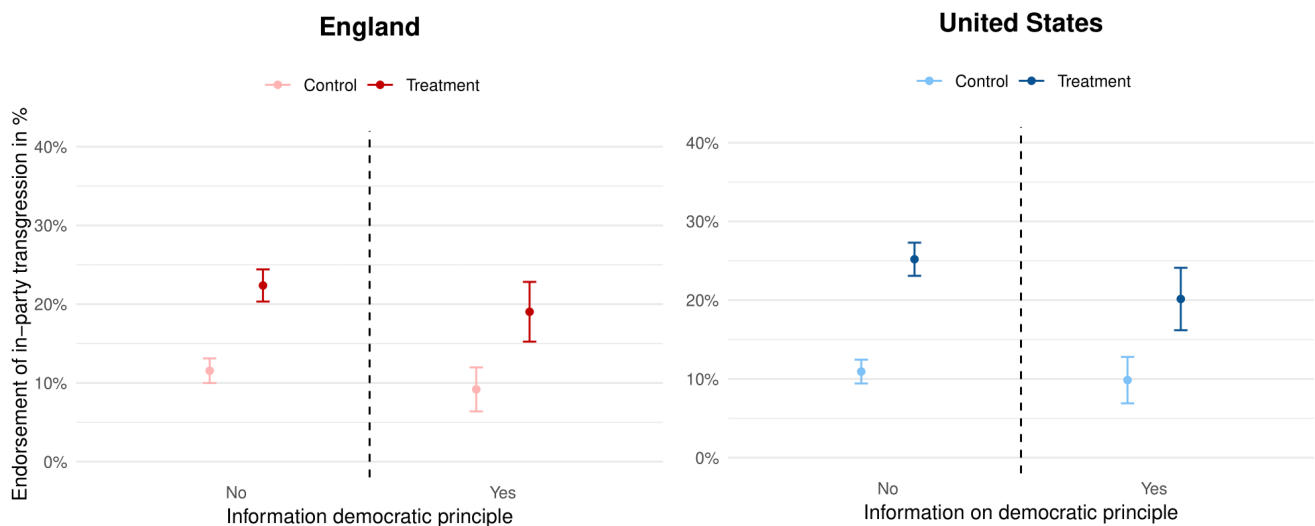


Fig 4. Support for (reciprocal) transgressions by exposure to democracy-information.

Note: Error bars present 95% CIs. Information group – N = 829 (England), N = 793 (US); No-information group – N = 3,202 (England), N = 3,276 (US). Model estimates are reported in SI Table I1. Figures disaggregated by transgression type are available in the extended SI Figure I1.

Alt text: Two side-by-side dot plots with error bars, for England and the US, plotting percentage endorsement of in-party transgression by whether respondents received information on the democratic principle, No or Yes.

Exploratory Analyses: Why do citizens reciprocate transgressions?

What drives citizens to engage in reciprocal democratic transgressions? Theoretically, we identified various potential motivations: a need to re-establish fairness within democracy, a desire to take revenge, or a process of normalization.⁶ To empirically study these motivations,

⁶ See SI K for treatment effects on fairness, revenge, and normalization using closed-ended questions.

we included open-ended questions in our survey, allowing respondents to freely articulate their justifications without the limitations imposed by closed-ended response options. In total, 907 out of 930 respondents in the treatment group who endorsed reciprocal undemocratic actions provided a written justification (England N = 424; US N = 483). As preregistered, we code these open answers using a deductive scheme implemented with ChatGPT-4 (prompt available in SI J). The model was prompted to assign each open answer to the best-fitting category: Fairness, revenge, normalization, or other. ChatGPT has been shown to perform well on text coding tasks, even outperforming crowd workers (Gilardi et al. 2023). Furthermore, we complement this analysis with qualitative abductive coding, the results of which are available in SI J and informed the development of the deductive coding scheme.

Turning to the findings, the distribution of the categories is presented in Figure 5. The most frequently cited rationale for supporting reciprocal democratic transgressions is fairness, accounting for 68% of all responses. Respondents in this category describe reciprocity as a symmetrical and justified response to the initial out-party violation. As one respondent put it: *"I feel like if the Democrats are doing it, then it is only fair for our party to do the same thing."* Relatedly, many in this category view reciprocal acts as a corrective mechanism aimed at restoring balance in the political sphere and ensuring equal treatment across parties. For example, one respondent explained: *"In the spirit of fairness and democracy, there should not be rules for one party and not the other. If Democrats restrict certain news outlets from attending, then Republicans should be allowed the same option."* General appeals to reciprocity, such as phrases like *"tit for tat"*, were also classified under this category. Further manual coding shows that fairness overwhelmingly remains the largest category, even when these abstract notions of reciprocity are excluded (see SI J for more details).

Revenge-related motivations are documented in 11% of the open answers. These responses voice a desire to punish the opponent for their undemocratic behavior through retaliation or reflect a general animosity towards the opposing party. For example, one

respondent remarked “*This is a retaliation to what they intend doing to us, so we let them have a taste of their own medicine*”. Another wrote “*Republicans are evil and they work against the good of the people that vote for them. They are extremely racist and very unfair. Therefore, it’s time for someone to treat them unfair!*”.

Notably, we do not document clear evidence of the theorized normalization process, with only 2% of the answers falling into this category. This may reflect a limitation of open answers, as certain categories may be underreported due to social desirability or because they are not easily observable to respondents. Indeed, normalization is often subconscious in nature (Bicchieri 2017), making it less likely to be explicitly articulated by respondents.

Lastly, we identify a residual “Other” category, which comprises approximately 19% of responses. This group includes a range of answers, some of which did not meaningfully engage with the question (e.g., gibberish or responses like “Don’t know”). To deepen our understanding of the motivations captured in this category and to remain open to rationales beyond our initial theoretical framework, we complement the deductive automated coding with a qualitative abductive analysis. While the overarching patterns are consistent – particularly the dominance of fairness-based justifications – the abductive approach reveals additional motivational themes (e.g., deterrence, a desire to win elections) (Results are presented in SI J).

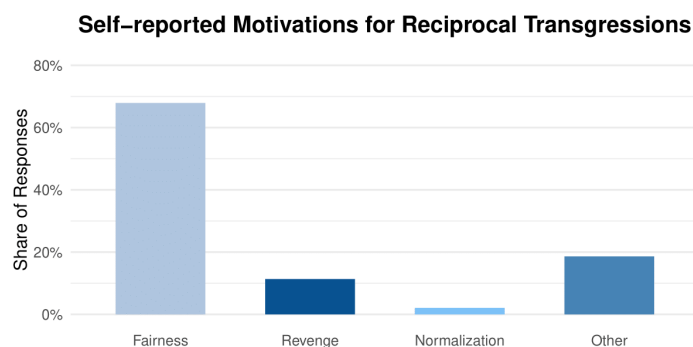


Fig 5. Self-reported motivations for reciprocal democratic transgressions.

Alt text: Bar chart of the share of open-ended responses across four motivation categories, with fairness by far the largest, followed by other and revenge, and normalization the smallest.

Discussion and Conclusion

This study has highlighted the interactive nature of citizens' democratic attitudes by introducing and testing the concept of reciprocity in the context of democratic transgressions. While previous research has largely focused on citizens' one-sided motivations in explaining support for undemocratic measures, our argument shifts the focus to the underexplored relational context of politics. By leveraging survey experiments in England and the US, we show that citizens' support for undemocratic measures is contingent on the conduct of other parties. Specifically, our study yields four key findings.

First, when citizens are exposed to democratic violations of political opponents, they become twice as likely to endorse undemocratic measures by their own party in response. In other words, we find evidence that citizens reciprocate democratic defections. The strong driving force of reciprocity in citizens' support for democratic violations suggests that (perceived) undemocratic elite behavior – even if initially condemned by citizens – can initiate a vicious cycle driven by reciprocal actions. Where prior research has already established that citizens are more likely to endorse democratic violations when they expect opponents will do the same (Braley et al. 2023; Pasek et al. 2022), our results demonstrate that such support also emerges as a direct reaction to violations that have already taken place. This has important implications for interventions aimed at improving citizens' democratic commitments. Scholars have argued that correcting biased beliefs about opponents' willingness to subvert democracy can reduce support for undemocratic practices (Braley et al. 2023; Voelkel et al. 2024), though the robustness of such interventions has been questioned (Dias et al. 2024; Druckman 2023). Our findings suggest a further boundary condition: in contexts where violations are occurring, misperception correction may prove insufficient on its own, as citizens are not only acting on biased beliefs about opponents' intentions but also reacting to actual undemocratic events.

It is worth noting, however, that despite the strong tendency to reciprocate transgressions, the vast majority of citizens oppose violations of democratic principles. This

broad rejection of undemocratic behavior occurs even when citizens are provoked by an out-party transgression. As such, our results also highlight that most citizens are committed to democratic principles: When the opposing party subverts democracy, the majority still prefers their party to maintain democratic integrity rather than to descend to reciprocal actions. Although support for undemocratic actions tends to be higher in studies involving trade-offs between democracy and partisan interests (e.g., Graham and Svobik 2020), our findings align with prior work which shows that most citizens reject democratic violations even amid high levels of partisan conflict (Holliday et al. 2024; Westwood et al. 2022).

Second, we find that the tendency to reciprocate democratic transgressions is independent of partisan affiliation and affective polarization. These findings partly resonate with recent work – which challenges the notion that affective polarization poses a fundamental threat to democracies (Broockman et al. 2023; Voelkel et al. 2024) – though we do document that baseline endorsement of undemocratic behavior is somewhat higher among polarized citizens. As such, our results paint a more nuanced picture, in which anti-democratic attitudes are not entirely independent of affective polarization, but its effects do not extend to the reciprocation of transgressions. The absence of a significant moderating effect suggests that reciprocity is a more universal impulse, one that emerges broadly across the electorate and independently of partisan animosity. This distinguishes our reciprocity account from existing work on interconnected democratic commitments, where biased beliefs about opponents' willingness to violate democracy are concentrated among strong partisans (Pasek et al. 2022).

Third, citizens do not reciprocate transgressions due to a lack of awareness about democracy: tendencies to reciprocate transgressions are strong, even when citizens are explicitly informed about the importance of a democratic principle in their country. Since reciprocation tendencies can thus not be prevented by educating citizens about the relevance of democratic principles, developing successful strategies to halt a potential spiral of reciprocal transgressions proves to be complex.

Lastly, our open question analysis provides insights into the self-reported justifications of citizens to endorse democratic transgressions. Contrary to the dominant view in the literature that some citizens somewhat selfishly prioritize their own political gains over the principles of democracy, we find that many citizens report being caught in a trade-off between competing moral principles: Commitment to democratic norms on the one hand, and universal principles of fairness on the other hand. From their perspective, mirroring the out-party violation provides an opportunity to restore the balance in politics and protect democracy on a broader scale. This observation speaks to a recent study by Krishnarajan (2023), who shows that citizens tend to rationalize undemocratic behavior as serving higher democratic goals. It also reveals a fundamental insight into the principles that citizens adhere to in politics. Though these citizens may not meet the ideal standard of principled democrats, many act from deeply held beliefs about fairness and reciprocity. They are thus ‘doing the wrong thing for the right reasons’: while their response can further undermine democratic rules, they do so by a conviction that reciprocation is the right and just response, not out of a rejection of democracy.

Turning to the normative implications, the democratic ambiguity of negative reciprocity in politics needs to be acknowledged. Though the desire to reciprocate may come from good intentions like fairness and maintaining a level playing field, reciprocating undemocratic behavior is not without risk. Our results indicate that citizens not only endorse responses in kind, but are willing to support escalated undemocratic reactions (SI G), suggesting that negative reciprocity may set in motion a self-reinforcing cycle of violations. Indeed, scholars have previously warned that reciprocity can “threaten the institutional foundations of democracy” (Schedler 2021, p. 258) and result in a “death spiral” of undemocratic behavior (Braley et al. 2023, p. 1283). At the same time, however, doing nothing in response to democratic transgressions risks granting a free pass to the transgressor. Reciprocation may also serve as a credible signal to political opponents that democratic violations will not go unanswered, potentially deterring further democratic subversion (Markovits 2026). Indeed,

game-theoretical work has shown that reciprocity can foster cooperative behavior (Güth 1995), implying that under certain conditions it may function as a stabilizing force. As such, determining what constitutes the best – and ultimately the most democracy-preserving – response represents a fundamental normative and strategic dilemma, one that political theorists have likewise debated (Bateman 2025; Malkopoulou 2026). One way out of this dilemma, as Schedler (2021) argues, lies in resistance through accountability, ostracism, mild forms of retaliation, or civil disobedience. This means that behavior aimed at inhibiting and punishing the perpetrator in the interest of democratic preservation is normatively desirable. Behavior aimed at the continued dismantling of democratic principles, such as the right to vote and protest or the freedom of the press, is not. Consequently, we consider reciprocal transgression in our study as normatively problematic, although both theoretically and empirically, citizens may have valid reasons to endorse them.

Our study also has limitations, which we encourage future research to address. The experimental design, while strong in internal validity, relies on hypothetical scenarios and therefore cannot capture all the nuances of real-world political interactions. Future studies using more naturalistic or field-experimental designs would therefore be valuable for assessing the external validity of these findings, particularly regarding how politicians' rhetoric and competing information environments shape citizens' willingness to reciprocate (see Druckman 2023). Prior research has established that elite discourse around undemocratic behavior can influence the public's perceptions (Aytaç 2026). As such, elites who emphasize the need for reciprocity and fairness may further heighten reciprocal support for undemocratic actions among their supporters. Another limitation of our study concerns the geographic scope: The focus on England and the US – both democracies with majoritarian systems – limits the generalizability of the findings to other political contexts with different party structures.

With regards to the experimental design, our study focuses specifically on tit-for-tat dynamics in response to 'targeted' democratic transgressions – cases where the out-party

directly violates democratic norms at the expense of the respondents' preferred party. While this offers a strong test case for reciprocity, it remains unclear whether similar dynamics would emerge in response to non-targeted violations, such as those affecting third parties or broader democratic institutions. In addition, our design does not allow us to disentangle whether citizens' reciprocal responses are uniquely driven by democratic violations, or whether they might also be provoked by general disliked out-party acts (such as undesirable – yet procedurally democratic – policies initiated by opponents). Relatedly, our dependent variable operationalizes reciprocity as a choice between retaliation and restraint. While we complement this measure with candidate-choice items and a scale capturing support for escalated transgressions, the design may nevertheless overlook alternative ways in which citizens may respond to democratic violations. Future research could therefore explore whether citizens also support more institutional or democracy-reinforcing responses to democratic transgressions, such as challenging these actions through courts or other democratic procedures.

Despite these limitations, our results demonstrate the far-reaching impact of negative reciprocity on human behavior, highlighting how its effects extend to citizens' democratic commitments. In doing so, we make a crucial contribution to the literature by stressing the interdependent nature of democratic attitudes and by providing novel insights into citizens' self-reported need for fairness when confronted with democratic transgressions by out-party elites. Importantly, our results call attention to the need for a more dynamic approach to studying democratic support – one that acknowledges the fundamental interactive nature of politics and human behavior.

Acknowledgements: We thank Olga Stamm for her assistance with the qualitative coding of the open-ended responses. We are especially grateful to Suthan Krishnarajan for his feedback and advice throughout various stages of the project. We also thank Jelle Koedam, Yphtach Lelkes, Markus Wagner, Stefaan Walgrave, Laurits Aarslew, Jochem Vanagt, Andres Reiljan,

Daniel Devine, Jamie Pow, Natasha Wunsch, Joep van Lit, and Bert Bakker for their helpful comments on earlier versions of the manuscript. Finally, we thank the anonymous reviewers for their constructive comments and suggestions.

References

- Aarslew, Laurits F. 2023. "Why Don't Partisans Sanction Electoral Malpractice?" *British Journal of Political Science* 53 (2): 407–23.
- Ahmed, Amel. 2023. "Is the American Public Really Turning Away from Democracy? Backsliding and the Conceptual Challenges of Understanding Public Attitudes." *Perspectives on Politics* 21 (3): 967–78.
- Aytaç, Selim E. 2026. "Framing Democracy: How Elite Discourse Shapes Citizens' Understandings of Democracy and Enables Backsliding." *World Politics* (forthcoming).
- Bateman, David A. 2025. "Democracy-Reinforcing Hardball: Can Breaking Democratic Norms Preserve Democratic Values?" *Comparative Political Studies* 59(6): 1191-1225.
- Becker, Howard P. 1956. *Man in Reciprocity*. New York: Praeger.
- Bicchieri, Cristina. 2017. *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford: Oxford University Press.
- Boyd, Robert, and Peter J. Richerson. 1988. "The Evolution of Reciprocity in Sizable Groups." *Journal of Theoretical Biology* 132 (3): 337–56.
- Braley, Alia, Gabriel S. Lenz, Dhaval Adjudah, Hossein Rahnema, and Alex Pentland. 2023. "Why Voters Who Value Democracy Participate in Democratic Backsliding." *Nature Human Behaviour* 7: 1282–93.
- Broockman, David E., Joshua L. Kalla, and Sean J. Westwood. 2023. "Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not." *American Journal of Political Science* 67 (3): 808–28.

- Carey, John M., Gretchen Helmke, Brendan Nyhan, Mitchell Sanders, and Susan Stokes. 2019. "Searching for Bright Lines in the Trump Presidency." *Perspectives on Politics* 17 (3): 699–718.
- Carey, John M., 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 and Parties* 32 (1): 230–45.
- Chernyak, Nadia, Kristin L. Leimgruber, Yarrow C. Dunham, Jingshi Hu, and Peter R. Blake. 2019. "Paying Back People Who Harmed Us but Not People Who Helped Us: Direct Negative Reciprocity Precedes Direct Reciprocity in Early Development." *Psychological Science* 30 (9): 1273–83.
- Claassen, Christopher. 2020. "Does Public Support Help Democracy Survive?" *American Journal of Political Science* 64 (1): 118–34.
- Clayton, Katherine, Nicholas T. Davis, Brendan Nyhan, Ethan Porter, Timothy J. Ryan, and Thomas J. Wood. 2021. "Elite Rhetoric Can Undermine Democratic Norms." *PNAS* (23): 1–6.
- Coppedge, Michael, et al. 2025. *V-Dem [Country-Year/Country-Date] Dataset v15*. Varieties of Democracy (V-Dem) Project. <https://doi.org/10.23696/vdemds25>.
- Coppock, Alexander. 2019. "Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach." *Political Science Research and Methods* 7 (3): 613–28.
- Crawford, Jarret T., and Jane M. Pilanski. 2014. "Political intolerance, right and left". *Political Psychology* 35: 841-51.
- Dahl, Robert A. 1956. *A Preface to Democratic Theory*. Chicago: University of Chicago Press.
- Dias, Nicholas C., Laurits F. Aarslew, Kristian Vrede Skaaning Frederiksen, Yphtach Lelkes, Lea Pradella, and Sean J. Westwood. 2024. "Correcting Misperceptions of Partisan Opponents Is Not Effective at Treating Democratic Ills." *PNAS Nexus* 3 (8): pgae304.

- Druckman, James N. 2023. "Correcting Misperceptions of the Other Political Party Does Not Robustly Reduce Support for Undemocratic Practices or Political Violence." *PNAS* 120 (37).
- Druckman, James N., et al. 2023. "Correcting Misperceptions of Out-Partisans Decreases American Legislators' Support for Undemocratic Practices." *PNAS* 120 (23): 1–3.
- Easton, David. 1965. *A Systems Analysis of Political Life*. New York: Wiley.
- Eder, Paul, Karl Aquino, Carl Turner, and Americus Reed II. 2006. "Punishing Those Responsible for the Prison Abuses at Abu Ghraib: The Influence of the Negative Reciprocity Norm (NRN)." *Political Psychology* 27 (6): 807–21.
- Eisenberger, Robert, Patrick Lynch, Justin Aselage, and Stephanie Rohdieck. 2004. "Who Takes the Most Revenge? Individual Differences in Negative Reciprocity Norm Endorsement." *Personality and Social Psychology Bulletin* 30 (6): 787–99.
- Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2008. "Testing Theories of Fairness—Intentions Matter." *Games and Economic Behavior* 62 (2): 287–303.
- Fehr, Ernst, and Simon Gächter. 2000. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives* 14 (3): 159–81.
- Ferrer, Sergi, Enrique Hernández, Enrique Prada, and Damjan Tomic. 2025. "The Value of Liberal Democracy: Assessing Citizens' Commitment to Democratic Principles." *European Journal of Political Research* 64 (4): 2018–38.
- 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.
- Gidron, Noam, Yotam Margalit, Lior Sheffer, and Itamar Yakir. 2026. "Why Masses Support Democratic Backsliding." *American Journal of Political Science* 70(2): 720-737.

- Gilardi, Fabrizio, Meysam Alizadeh, and Maël Kubli. 2023. "ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks." *PNAS* 120 (30): 1–3.
- Gouldner, Alvin W. 1960. "The Norm of Reciprocity: A Preliminary Statement." *American Sociological Review* 25 (2): 161–78.
- Graham, Matthew H., and Milan W. Svobik. 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.
- Güth, Werner. 1995. "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives." *International Journal of Game Theory* 24 (4): 323–44.
- Hainmueller, Jens, Dominik Hangartner, and Teppei Yamamoto. 2015. "Validating vignette and conjoint survey experiments against real-world behavior". *PNAS* 112 (8), 2395-400.
- 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.
- Hernández, Enrique. 2016. "Europeans' Views of Democracy: The Core Elements of Democracy." In Mónica Ferrín and Hanspeter Kriesi, eds., *How Europeans View and Evaluate Democracy*. Oxford: Oxford University Press.
- Holliday, Derek E., Shanto Iyengar, Yphtach Lelkes, and Sean J. Westwood. 2024. "Uncommon and Nonpartisan: Antidemocratic Attitudes in the American Public." *PNAS* 121 (13).
- Jackson, Joshua C., Virginia K. Choi, and Michele J. Gelfand. 2019. "Revenge: A Multilevel Review and Synthesis." *Annual Review of Psychology* 70: 319–45.
- Janssen, Lisa, and Emma Turkenburg. 2025. "Breaking Free from Linear Assumptions: Unravelling the Relationship between Affective Polarization and Democratic Support." *European Journal of Political Research* 64 (3): 1465–79.
- Jefferson, Sean G., Monica P. Vega, and Kimmerly Newsom. 2024. "Revenge Sex: Weaponizing Infidelity." *Deviant Behavior* 46 (3): 365-375.
- Jeong, Helen. 2025. "Californians Overwhelmingly Approve Prop 50." *NBC News*.

- Kapur, Sahil. 2021. "Democrats to Introduce Bill to Expand Supreme Court from 9 to 13 Justices." *NBC News*.
- Krishnarajan, Suthan. 2023. "Rationalizing Democracy: The Perceptual Bias and (Un)democratic Behavior." *American Political Science Review* 117 (2): 474–96.
- Levitsky, Steven, and Daniel Ziblatt. 2018. *How Democracies Die*. New York: Crown.
- Lipset, Seymour M. 1959. "Some Social Requisites of Democracy: Economic Development and Political Legitimacy." *American Political Science Review* 53 (1): 69–105.
- Lührmann, Anna, and Staffan I. Lindberg. 2019. "A Third Wave of Autocratization Is Here: What Is New about It?" *Democratization* 26 (7): 1095–1113.
- Malkopoulou, Anthoula. 2026. "Democratic Self-Defence as a Dirty Hands Problem." *Democratization* 1–21.
- Markovits, Daniel B. 2026. "Does Fear of Retaliation Constrain Support for Democratic Backsliding?" Unpublished manuscript.
- Mattoo, Aaditya, and Robert W. Staiger. 2019. "Understanding Trade Wars." In Meredith A. Crowley, ed., *Trade War: The Clash of Economic Systems Endangering Global Prosperity*. London: CEPR Press.
- Molm, Linda D. 2010. "The Structure of Reciprocity." *Social Psychology Quarterly* 73 (2): 119–31.
- Neal, Jeff. 2020. "Should Democrats Pack the Supreme Court?" *Harvard Law Today*.
- Nord, Marina, David Altman, Tiago Fernandes, Ana Good God, and Staffan I. Lindberg. 2026. *Democracy Report 2026: Unravelling the Democratic Era?* Gothenburg: V-Dem Institute, University of Gothenburg.
- Nowak, Martin A., and Karl Sigmund. 2005. "Evolution of Indirect Reciprocity." *Nature* 437: 1291–98.

- Pasek, Michael H., Lee-or Ankori-Kalinsky, Alex Levy-Vene, and Samantha L. Moore-Berg. 2022. "Misperceptions about Out-Partisans' Democratic Values May Erode Democracy." *Scientific Reports* 12: 16284.
- Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83 (5): 1281–1302.
- Riaz, Sascha, Daniel Bischof, and Markus Wagner. 2024. "Out-Group Threat and Xenophobic Crimes: Evidence of Local Intergroup Conflict Dynamics between Immigrants and Natives." *The Journal of Politics* 86 (4): 1146–61.
- Schedler, Andreas. 2021. "Democratic Reciprocity." *Journal of Political Philosophy* 29 (2): 252–78.
- Shaw, Alex, Anam Barakzai, and Boaz Keysar. 2019. "When and Why People Evaluate Negative Reciprocity as More Fair than Positive Reciprocity." *Cognitive Science* 43 (8): e12768.
- Simonovits, Gabor, Jennifer McCoy, and Levente Littvay. 2022. "Democratic Hypocrisy and Out-Group Threat: Explaining Citizen Support for Democratic Erosion." *The Journal of Politics* 84 (3): 1806–11.
- Valentim, Vicente. 2026. "Norms of Democracy, Staged Democrats, and Supply of Exclusionary Ideology." *Comparative Political Studies* 59 (6): 1367-1405.
- Voelkel, Jan G., et al. 2024. "Megastudy Testing 25 Treatments to Reduce Antidemocratic Attitudes and Partisan Animosity." *Science* 386 (6719): 1–11.
- Weingast, Barry R. 1997. "The Political Foundations of Democracy and the Rule of Law." *American Political Science Review* 91 (2): 245–63.
- Werner, Hannah, Eri Bertsou, and Sofie Marien. 2025. "Winners' Restraint: Do Citizens Reject Political Processes When Procedural Violations Accumulate?" *Political Behavior* 47: 1527–50.
- Westwood, Sean J., Justin Grimmer, Matthew Tyler, and Clayton Nall. 2022. "Current Research Overstates American Support for Political Violence." *PNAS* 119 (12): 1–10.

Wunsch, Natasha, and Theresa Gessler. 2023. "Who Tolerates Democratic Backsliding? A Mosaic Approach to Voters' Responses to Authoritarian Leadership in Hungary." *Democratization* 30 (5): 914–37.

Wunsch, Natasha, Marc S. Jacob, and Laurenz Derksen. 2025. "The Demand Side of Democratic Backsliding: How Divergent Understandings of Democracy Shape Political Choice." *British Journal of Political Science* 55: e39.

Wuttke, Alexander, Konstantin Gavras, and Harald Schoen. 2022. "Have Europeans Grown Tired of Democracy? New Evidence from Eighteen Consolidated Democracies, 1981–2018." *British Journal of Political Science* 52 (1): 416–28.

Biographical Statements: Lisa Janssen is a Doctoral Researcher at Ghent University, Ghent 9000, Belgium. Hannah Werner is an Assistant Professor at the University of Zurich, Zurich 8050, Switzerland. Anna Kern is an Associate Professor at Ghent University, Ghent 9000, Belgium.