

mores

MORAL EMOTIONS IN POLITICS

Cracks in the
System

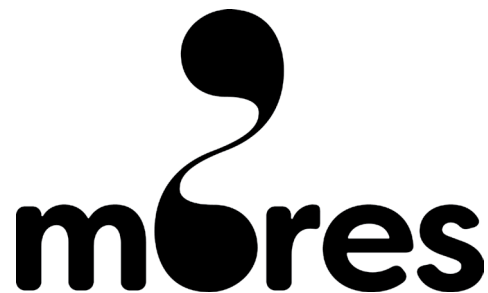
Christian Unger/ via Shutterstock



Funded by
the European Union

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

WORKING PAPER SERIES



MORAL EMOTIONS IN POLITICS: HOW THEY UNITE, HOW THEY DIVIDE

Paper title: Cracks in the System. A democratic inoculation intervention that enhances recognition of authoritarian weaknesses

Authors: Gábor Orosz¹, János Salamon^{2,3}, Zsolt Enyedi^{4,5}, Stephen Whitefield⁶, Péter Krekó⁷

Contact: contact@mores-horizon.eu

Information: MORES Working Paper Series, 5

Download the research data: <https://openarchive.tk.mta.hu/664/>

1 Flinders University Institute for Mental Health and Wellbeing, College of Education, Psychology and Social Work, Flinders University, SA, Australia

2 Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary

3 Department of Ergonomics and Psychology, Budapest University of Technology and Economics, Budapest, Hungary

4 Department of Political Science, Central European University, Austria

5 CEU Democracy Institute, Hungary

6 Department of Politics and International Relations, University of Oxford, Oxford, UK

7 Institute of Psychology, ELTE Eötvös Loránd University, Budapest, Hungary



Funded by
the European Union

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.



Table of Contents

1. Introduction	03
1.1 Present study	05
2. Method	07
2.1 Participants	07
2.2 Procedure	07
2.3 Primary outcomes	09
2.4 Secondary outcomes	10
2.5 Moderators	10
3. Results	11
3.1 Primary, preregistered outcomes	11
3.2 Secondary outcomes, preregistered hypothesis testing	13
4. Discussion	16
References	21
Supplemental Material	24



Abstract

Modern authoritarian systems often consolidate through gradual, hard-to-detect shifts, yet can collapse rapidly once cracks in regime stability become visible. Building on current democratic inoculation research—and drawing on gradual norm internalisation and mechanisms of drastic and rapid changes—we developed two brief (<15 min), narrative-based interventions designed to heighten citizens’ sensitivity to authoritarian rhetoric; based on past experiences with democratic transgressions. In a preregistered survey experiment with 1,175 Hungarian adults, matched to national demographics among regular internet users, participants were randomly assigned to one of three conditions: (1) Gradual Build-Up, illustrating the slow erosion of rule of law and normalisation of democratic backsliding; (2) Cracks, emphasising early warning signals of authoritarian weakening (e.g., elite defection, propaganda fatigue); or (3) a matched non-political control. All interventions employed expert opinions and historical examples, paired with a “saying-is-believing” self-persuasion task, and combined with the logic of the “democratic inoculation” mechanisms. Results revealed a consistent pattern: the Cracks intervention reliably decreased agreement with authoritarian statements and increased recognition of authoritarian tone, concern about authoritarianism, moral outrage towards the perpetrator of a democratic transgression, and perceived democratic threat, with reasonable effect sizes. It also improved recognition of early instability cues and heightened negative moral affect towards the government following an existing restrictive legal proposal. The effects of the Gradual Build-Up intervention were weaker and less consistent across outcomes. Importantly, partisanship, populist attitudes, and cognitive reflection did not moderate any intervention effects, suggesting a broad effect on sensitivity to authoritarian cues and moral emotions that can reach people on both sides of the political spectrum.

Keywords: Democratic backsliding, Authoritarian rhetoric, Democratic inoculation, Wise interventions, Self-persuasion



1. Introduction

This research explores why democratic backsliding often goes unnoticed until authoritarian systems begin to ‘crack’

Despite global expansion and the increasing sophistication of authoritarian governance, citizens living under democratic backsliding often struggle to interpret the meaning of political signals. Incremental legal changes, media narratives changing gradually, and strategic ambiguity make it difficult to distinguish routine political practices from genuine threats (or opportunities) of regime change (Abels et al., 2024). As a result, even politically attentive citizens may remain uncertain, disengaged, or psychologically inert in the face of authoritarian consolidation. This raises a central psychological question: can subtle, low-cost interventions generate small but meaningful interpretive shifts or reframing moments that help citizens make sense of authoritarian politics in real time? Short, precisely targeted interventions have been shown to increase democratic knowledge, support, and participation across diverse contexts—from Tunisia’s post-authoritarian transition (Finkel et al., 2023) to Turkey’s electoral authoritarian regime (Öztürk et al., 2025). Large cross-national experiments similarly demonstrate that messages emphasising civic rights and institutional checks reliably boost willingness to defend democracy (Eroğlu et al., 2025). Importantly, converging evidence from social psychology shows that even minimal framing shifts—such as construing political action as an identity rather than a behaviour—can meaningfully alter civic engagement and real-world participation (Bryan et al., 2011). Building on this literature, our study examines two theoretically grounded mechanisms for meaningful interpretational shifts in a context where informational autocracy has been consolidating for nearly fifteen years.

In contrast to twentieth-century authoritarian regimes that typically consolidated power through overt coercion and rapid institutional takeover, contemporary authoritarian leaders pursue incremental institutional change, strategically employing legalism and propaganda to generate legitimacy and maintain political support (Scheppele, 2018; Guriev & Treisman, 2023). Democratic backsliding and illiberal consolidation in the 21st century typically unfold through incre-

A brief narrative inoculation makes authoritarian rhetoric easier to spot—especially when it highlights regime fragility

mental, legally framed steps that erode checks and balances while preserving an electoral façade (Bermeo, 2016; Levitsky & Ziblatt, 2018; Linz, 2000; Schedler, 2013; Varol, 2015). Such micro-shifts accumulate slowly, fostering gradual norm internalisation and the “creeping normality” of democratic erosion (Bugarič, 2019; Greskovits, 2015; Labanino & Dobbins, 2023), in which each individual deviation appears tolerable when evaluated in isolation (Scheppele, 2013). Over time, citizens may also exhibit authoritarian threat habituation: repeated low-intensity violations elicit diminishing concern, making the broader trajectory of institutional decline harder to perceive. In hybrid regimes like Hungary, stealth authoritarian strategies (e.g., constitutional engineering, media capture, the politicisation of oversight bodies) reshape the informational environment in ways that further obscure long-term patterns (Enyedi, 2018; Krekó & Enyedi, 2018; Scheppele, 2018). Psychological research helps explain why these dynamics so often go unnoticed: people tend to process everyday political events at a concrete, low-level construal, limiting their ability to “zoom out” and detect structural, holistic change (Trope & Liberman, 2010), and they struggle to generate historical-analogy inferences that would situate present developments within well-known authoritarian trajectories. In short, citizens often fail to see the forest for the trees. These mechanisms together suggest an interpretative bottleneck: even attentive citizens may lack the perspective-taking bandwidth needed to recognise slow authoritarian build-up in real time. Therefore, we suggest that an intervention capable of providing this perspective or psychological distance by “inoculating” voters with past experiences can help people connect isolated actions into a coherent and evolving authoritarian pattern. This mindset shift can make them more sensitive to signals of incremental authoritarian consolidation and evoke moral emotional reactions from them.

At the same time, even in modern illiberal democracies, it is rather rare to see an inverse incremental political change towards re-democratisation; instead, the collapse of illiberal regimes often occurs through sudden, non-linear shifts. Under authoritarian constraints, citizens tend to update beliefs only incrementally until an unexpected event happens: belief change is driven by prediction error, such that surprising events generate disproportionately large updates. The Rescorla–Wagner learning framework (Rescorla & Wagner, 1972) predicts precisely this pattern. Cascade and tipping-point models similarly demonstrate how visible cues such as elite defection, propaganda breakdown, mass protest, or public ridicule can rapidly shift collective expectations (Kuran, 1991; Lohmann, 1994). Recent theories of informational autocracies argue that regimes reliant on curated competence and narrative control become surprisingly fragile once citizens realise that their private doubts are widely shared (Guriev & Treisman, 2022), breaking the pluralistic ignorance. These frameworks imply that highlighting salient, surprising signals of authoritarian weakness should produce stronger psychological updating than focusing on slow erosion.

The Gradual Build-Up intervention was constructed to counteract the cognitive tendencies that make slow authoritarian erosion difficult to detect



*“ Even attentive citizens
may lack the perspective-
taking bandwidth
needed to recognise slow
authoritarian build-up in
real time ”*

Péter Krekó

MORES Co-investigator,
ELTE Eötvös Loránd University,
Budapest

creeping normality, threat habituation, and low-level construal of day-to-day political events (e.g., Marcus et al., 2000; Trope & Liberman, 2010). Historically informed personal testimonies illustrate how censorship, loyalty-based appointments, and subtle fear cues accumulate incrementally, mapping onto theories of gradual norm internalisation and the psychological difficulty of generating high-level, pattern-based inferences. Participants are then prompted to “zoom out” by personalising and articulating the core insight in their own words, leveraging saying-is-believing processes to disrupt habituated interpretations and consolidate a more synthetic understanding of institutional decline.

In contrast, the Cracks intervention applies theories explaining why sudden, diagnostic signals of political change can trigger disproportionately large psychological updates. Drawing on prediction-error learning and Rescorla–Wagner dynamics (1972), and cascade models in which elite defection, public ridicule, or propaganda breakdown act as high-surprise cues (Kuran, 1991; Lohmann, 1994), the narratives present vivid moments when authoritarian regimes become visibly fragile. By situating these moments within historical patterns documented in comparative politics (Slater, 2010; Gunitsky, 2015), the module fosters both critical awareness and a sense of latent political possibility. The accompanying reflective task encourages participants to integrate these unexpected signals into their interpretive framework, strengthening the belief-updating processes that these theories predict.

In line with our preregistration, we expected the *gradual build-up* intervention to heighten sensitivity to authoritarian rhetoric relative to the control condition

1.1 Present study

Our intervention design directly operationalizes these dual mechanisms of authoritarian perception through fact- and narrative-based exercises grounded in the traditions of both the wise interventions (Walton, 2014; Walton & Wilson, 2018) and the democratic inoculation experiments (e.g., Hradický, 2024). The current project, combining survey-experimental and narrative-based intervention methods, tests whether brief, historically grounded democratic inoculation exercises can strengthen two complementary forms of democratic awareness: (1) recognition of the subtle, incremental mechanisms through which authoritarian regimes consolidate power, and (2) recognition of the early cracks and high-diagnostic signals that reveal structural weakness and potential democratic re-opening.

In line with our preregistration (<https://osf.io/h9dnm/overview>), we expected the Gradual Build-Up intervention to heighten sensitivity to authoritarian rhetoric relative to the Control condition (H1). Specifically, participants in the Gradual Build-Up condition were predicted to (H1a) classify authoritarian statements more accurately, (H1b) perceive them as more concerning, (H1c) express stronger moral outrage, (H1d) judge them as more

threatening to democracy, and (H1e) show greater disagreement with the rhetoric. We further predicted that the Cracks intervention would increase recognition of instability cues in authoritarian systems (H2), such as elite defection, protest coordination, and declining propaganda effectiveness.

We preregistered direct comparisons between the two intervention conditions: the Gradual Build-Up module was expected to outperform Cracks on all authoritarian-rhetoric sensitivity outcomes (H3a–H3e), whereas Cracks was expected to outperform Gradual Build-Up on detecting regime-instability signals (H4). Secondary hypotheses anticipated that each intervention, compared to the control condition, would reduce support for a restrictive media-and-NGO law (H5a–H5b) and increase negative moral emotions towards the government proposing it (H6a–H6b), without positing differences between the two intervention conditions. Together, these predictions evaluate whether brief, theoretically anchored narrative interventions can generate the kinds of cognitive and affective reframing that support democratic awareness in an informational autocracy.



*“ We were struck
by how much more
compelling participants
found the mechanisms of
democratic breakdown
than those of democratic
buildup; even though it is
the latter, not the former,
that generations have
been taught in history
classrooms ”*

Gábor Orosz

MORES contributor,
Flinders University, Adelaide



2. Method

Testing narrative-based democratic inoculation through historically grounded psychological interventions in Hungary

2.1 Participants

Data were collected online in Hungary through a professional survey company. Participants were adults (18+) recruited to approximate the national distribution of gender, age, education, and settlement type among regular internet users. After providing informed consent, participants were randomly assigned to one of three conditions. All outcome measures were administered in the same session, immediately after the intervention. Two attention checks were embedded; exclusions followed preregistered rules (duplicate entries, two failed attention checks, and completion times more than ± 3 standard deviations from the mean). The final sample consisted of 1,175 adults. The gender distribution (48.8% men; 51.2% women) closely matched national benchmarks and did not differ across conditions ($\chi^2(2) = 0.47, p = .39$). Participants ranged from 18 to 69 years old ($M = 43.1, SD = 12.7$). Educational attainment reflected the Hungarian internet-using population: 18.2% primary or less, 48.9% secondary, 32.9% tertiary. Residential status also aligned with national patterns: 24.0% in Budapest, 26.3% in county seats, 31.7% in other towns, 18.0% in villages. Hungary provides a theoretically relevant context, as democratic backsliding and informational authoritarian practices have consolidated for more than a decade (Krekó & Enyedi, 2018), making it an appropriate setting for testing perception-focused inoculation interventions.

2.2 Procedure

Participants completed the study online in a single session. After providing informed consent, they were randomly assigned to one of three conditions: a non-political Control condition, the Gradual Build-Up condition, or the Cracks condition. Figure 1 presents the study flow (see on page 8).

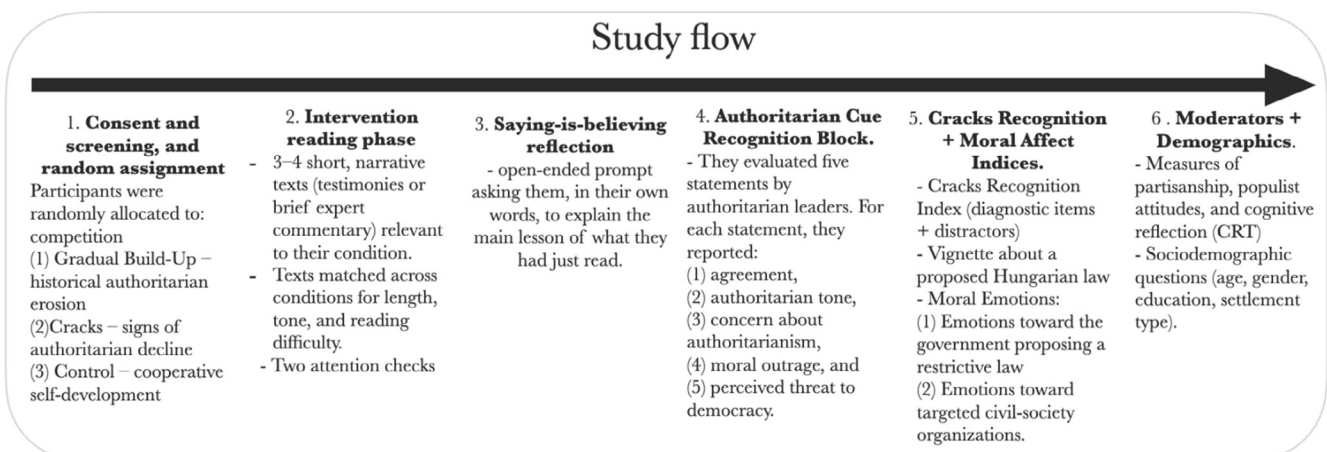
In all conditions, participants first completed a brief reading task. In the two intervention conditions, these readings consisted of historically grounded narrative testimonials accompanied by short expert commentaries. The Gradual Build-Up materials illustrated how authoritarian systems can consolidate power incrementally through small, legally framed changes and

the normalisation of subtle fear cues. The Cracks materials depicted early, high-diagnostic signals of authoritarian weakening—such as elite defection, propaganda fatigue, public ridicule, and emerging protest coordination. Condition materials were matched in text length, tone, and reading difficulty; a full side-by-side description of all intervention components is provided in Supplemental Table S1. Following the reading task, all participants completed a saying-is-believing self-persuasion exercise in which they explained the main idea of the passage “to a younger person.” This component was included to promote self-generated meaning-making and to increase psychological engagement with the material.

The control condition presented participants with a brief, non-political module about constructive, self-developmental forms of competition. The readings drew expert opinions and everyday stories illustrating how competition can foster personal growth, intrinsic motivation, and collaborative skill-building when framed as a challenge rather than a zero-sum struggle. Participants then completed a structurally similar, thematically neutral reflection task used in the other conditions, describing the main insight they saw in healthy competition. This condition was intentionally designed to match the interventions in text length, tone, narrative style, and cognitive engagement, while remaining fully unrelated to politics, democracy, or authoritarianism.

Participants then completed the Authoritarian Cue-Recognition Block, evaluating five authoritarian-style leader statements on multiple dimensions (agreement, perceived authoritarian tone, concern about authoritarianism, moral outrage, and perceived democratic threat). Next, they completed the Cracks Recognition Index, rating the extent to which 20 political signals (7

Figure 1 – The study flow of the intervention



diagnostic items, 13 distractors) reflected potential instability in an authoritarian system.

In the final section, participants read a vignette describing a proposed Hungarian law that would restrict the activities of foreign-funded civil society organisations and independent media. They reported their agreement with the proposed law, and their emotional reactions towards (a) the government actors proposing the law and (b) the targeted civil-society and media organisations.

Measures of partisanship, populist attitudes, cognitive reflection, and demographic variables (age, gender, education, settlement type) were administered at the end of the session. All primary and secondary outcomes were assessed immediately after the intervention.

2.3 Primary outcomes

The measures we report here were assessed immediately after the intervention.

Evaluating Authoritarian Statements. Participants evaluated five political leader statements that reflected core elements of authoritarian rhetoric (moralising, delegitimizing dissent, asserting exclusive national representation). An example item is: “Independent courts must not stand in the way of the nation’s interests.” For each statement, participants responded on five dimensions: Agreement with the message (5 items; $\alpha = .78$), indexing acceptance of authoritarian framing; perceived authoritarian tone (5 items; $\alpha = .86$), assessing whether the statements sounded punitive or norm-eroding; concern about authoritarianism (5 items; $\alpha = .86$), capturing perceived democratic danger; moral outrage (5 items; $\alpha = .80$), measuring anger, indignation, or moral disgust elicited by the rhetoric; perceived threat to democracy (5 items; $\alpha = .85$), reflecting judgments that such rhetoric endangers democratic institutions. Items were averaged within each scale; higher scores indicate stronger endorsement of the focal construct. Full item texts appear in the Supplemental Materials (SM2).

Cracks Recognition Index. Participants completed a 20-item measure developed for this study to assess sensitivity to early signs of authoritarian system instability, informed by work on informational autocracies and cascade dynamics (Guriev & Treisman, 2022; Kuran, 1991; Levitsky & Way, 2010). The scale included seven diagnostic items reflecting recognised indicators of regime weakening (e.g., elite defection, inconsistent official communication, declining propaganda credibility, protest coordination) and thirteen

distractor items that do not reliably signal systemic erosion (e.g., celebrity criticism, new social programmes). For each statement, participants rated the extent to which it reflected “cracks in an authoritarian system” on a 6-point scale (1 = does not indicate cracks at all, 6 = clearly indicates cracks). The Cracks Recognition Index was computed as the means of the seven diagnostic items. Internal consistency was high ($\alpha = .88$). Item development



“ We found that narrative interventions can generate the kind of cognitive and affective reframing that support democratic awareness in an informational autocracy ”

János Salamon

MORES Researcher,
ELTE Eötvös Loránd University,
Budapest

procedures and full item set are reported in Supplemental Materials (SM3).

2.4 Secondary outcomes

Agreement with and Moral–Emotional Reactions to a Democratic Norm Violation. Participants read a vignette describing a proposed Hungarian law that would impose strict regulatory constraints on foreign-funded civil society organisations and independent media, called “Bill on the Transparency of Public Life”; Number of publication: T/11923; Date of submission: 2025–05–13 23:38 . The text summarised the law’s content, supporters, and critics; the full wording appears in Supplemental Materials (SM4). Participants evaluated their agreement with the proposed law (10-point scale: 1 = not at all, 10 = completely), and their emotional reactions toward: (1) the government actors proposing the law, and (2) the targeted civil-society organisations and media outlets. They rated eight discrete emotions (anger, disgust, contempt, shame, respect, pride, empathy, indignation) on a 7-point scale (1 = not at all, 7 = very strongly). Two indices were created: (1) Government Negativity Index: mean of negative emotions directed at the government (anger, disgust, contempt, shame, indignation; $\alpha = .96$) and (2) Civil Society Positivity Index: mean of positive emotions towards civil-society actors (respect, pride, empathy; $\alpha = .91$). Higher scores indicate stronger negative or positive affects towards the respective targets.

Participants completed a saying-is-believing self-persuasion exercise to boost psychological engagement

2.5 Moderators

Cognitive reflection was assessed with three items (e.g., “If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?”). The items were dichotomously coded as either correct (coded as 1) or incorrect (coded as 0), with a higher mean indicating a heightened willingness to think analytically (Frederick, 2005 via Shenhav et al., 2012).

Partisanship. Pro-governmental versus pro-opposition attitudes were measured with the following question: “If you had to choose between the government and the opposition, which side would you prefer to vote for? The government/The opposition” (see Orosz et al., 2024).

Populist attitudes. Twelve items were used to assess populist attitudes using a newly developed scale (Enyedi et al., in prep). An example item is: “Our society needs protection from international influence rather than greater openness.” The scale demonstrated adequate internal consistency ($\alpha = .70$).



3. Results

Intervention signalling authoritarian fragility produce stronger democratic recalibration than gradual-erosion framing



Randomization Checks. Gender ($p_{\text{cracks}}=.725$, $p_{\text{build-up}}=.187$), settlement type ($p_{\text{cracks}}=.157$, $p_{\text{build-up}}=.899$), and partisanship ($p_{\text{cracks}}=.249$, $p_{\text{build-up}}=.618$) were balanced across conditions. Age ($p_{\text{cracks}}=.714$, $p_{\text{build-up}}=.047$) and education ($p_{\text{cracks}}=.285$, $p_{\text{build-up}}=.002$) showed statistically detectable differences (participants in the Built-up condition were slightly younger and somewhat less educated compared to the control condition). In the case of the Cracks intervention, there was no difference between any pre-intervention variables; however, in the case of the Built-up condition, covariate-adjusted models were run as robustness checks. The overall pattern of results does not change after controlling for these demographic variables.

3.1 Primary, preregistered outcomes

1. Authoritarian Cue Recognition. We measured five dimensions of reactions to authoritarian-style political statements:

Agreement. We conducted an OLS regression predicting agreement with the authoritarian political statements from condition, controlling for age and education. In line with the preregistered analysis plan and to ensure consistency across contrasts, Bonferroni-adjusted tests were applied to all pairwise condition comparisons. Participants in the Cracks condition reported significantly lower agreement with authoritarian statements than those in the Control condition, $b = -0.30$, $SE = 0.07$, $t(1170) = -3.94$, $p < .001$, $SMD = -0.29$. Estimated marginal means indicated a drop from 2.82 ($SE = 0.05$) in the Control group to 2.53 ($SE = 0.06$) in the Cracks group. Participants in the Gradual Build-Up condition did not differ significantly from the Control condition, $b = -0.10$, $SE = 0.07$, $t(1170) = -1.48$, $p = .419$, with an estimated mean of 2.72 ($SE = 0.05$). In contrast to our predictions, direct comparison between the two intervention arms showed that agreement was significantly lower in the Cracks condition than in the Gradual Build-Up condition, $b = -0.19$, $SE = 0.07$, $t(1170) = -2.62$, $p = .027$, $SMD = -0.19$. Overall, only the Cracks intervention reduced agreement with authoritarian rhetoric relative to the control narrative.

Perceived authoritarian tone also varied by condition. Relative to the control, participants in the Cracks condition rated the authoritarian messag-

“ Participants in the ‘cracks’ condition reported significantly stronger negative emotions towards the government proposing restrictive measures against the media and NGOs ”

Péter Krehó

MORES Co-investigator,
ELTE Eötvös Loránd University,
Budapest

es as more authoritarian in style ($M_s = 3.89$ vs. 4.21), $b = 0.32$, $SE = 0.09$, $t(1170) = 3.64$, $p < .001$, $SMD = 0.27$. The Built-up condition did not differ significantly from the control condition ($M = 4.03$), $b = 0.15$, $SE = 0.08$, $t(1170) = 1.75$, $p = .243$, and authoritarian tone ratings did not differ significantly between the Cracks and the Built-up conditions, $b = -0.18$, $SE = 0.09$, $t(1170) = -2.05$, $p = .121$.

Concern about authoritarian tendencies showed a similar pattern. Participants assigned to the Cracks condition expressed notably higher concern than those in the Control group ($M_s = 4.42$ vs. 4.08), $b = 0.34$, $SE = 0.09$, $t(1170) = 3.90$, $p < .001$, $SMD = 0.29$. The Built-up condition produced only a small, non-significant increase relative to Control ($M = 4.19$), $b = 0.11$, $SE = 0.08$, $t(1170) = 1.38$, $p = .503$. In contrast to our preregistered hypothesis, concern ratings in Cracks were also significantly higher than in the Built-up condition, $b = -0.22$, $SE = 0.08$, $t(1170) = -2.68$, $p = .023$, $SMD = -0.19$.

Moral outrage towards the authoritarian statements also differed across conditions, as it was preregistered. Participants in the Cracks condition reported substantially stronger outrage than those in the Control group ($M_s = 3.75$ vs. 3.36), $b = 0.39$, $SE = 0.09$, $t(1170) = 4.38$, $p < .001$, $SMD = 0.32$. The Built-up condition produced a smaller but still significant increase relative to Control ($M = 3.56$), $b = 0.20$, $SE = 0.08$, $t(1170) = 2.40$, $p = .0497$, $SMD = 0.17$. Outrage levels were not significantly higher in Cracks than in the Built-up condition, $b = -0.19$, $SE = 0.09$, $t(1170) = -2.18$, $p = .089$.

Perceived threat to democracy was also elevated in the Cracks condition (in line with the preregistration). Participants exposed to the Cracks narrative rated the political situation as more threatening than those in the Control group ($M_s = 4.32$ vs. 3.97), $b = 0.34$, $SE = 0.08$, $t(1170) = 4.11$, $p < .001$, $SMD = 0.30$. The Built-up condition did not differ significantly from the Control condition ($M = 4.12$), $b = 0.15$, $SE = 0.08$, $t(1170) = 1.86$, $p = .191$. In contrast to our pre-registered hypothesis, perceived threat was significantly higher in the Cracks condition than in the Built-up condition, $b = -0.20$, $SE = 0.08$, $t(1170) = -2.43$, $p = .046$, $SMD = -0.17$.

Although preregistered hypotheses predicted significant effects for the Gradual Build-Up condition relative to Control, these effects were weak or non-significant. Instead, the Cracks condition showed consistent and significant differences from Control across outcomes. The implications of this pattern are discussed below.

2. Recognition of cracks in authoritarian systems.

In line with the preregistration, OLS regression predicting the Cracks Recognition Index from condition (controlling for age and education) showed a significant effect of the Cracks intervention. Participants in the Cracks condition more strongly recognised signs of regime instability than those in the Control condition ($M_s = 4.70$ vs. 4.43), $b = 0.27$, $SE = 0.07$, $t(1170) = 3.87$, $p < .001$, $SMD = 0.28$. By contrast, the Gradual Build-Up condition did not differ from Control ($M = 4.46$), $b = 0.02$, $SE = 0.06$, $t(1170) = 0.38$, $p = 1.00$. A direct comparison between the two intervention arms indicated higher crack recognition in the Cracks than in the Gradual Build-Up condition, $b = -0.24$, $SE = 0.07$, $t(1170) = -3.62$, $p < .001$, $SMD = -0.26$.

3.2 Secondary outcomes, preregistered hypothesis testing

Agreement with the Proposed Law

No significant differences were found between the intervention conditions and the control condition regarding agreement with the proposed law (Cracks vs. Control: $p = .240$; Build-Up vs. Control: $p = 1.00$).

Government Negativity

Government-directed negative affect also differed across conditions. Participants in the Cracks condition reported significantly stronger negative emotions towards the government proposing the restrictive law than those in the Control group ($M_s = 4.35$ vs. 3.80), $b = 0.55$, $SE = 0.15$, $t(1170) = 3.60$, $p = .001$, $SMD = 0.27$. The Gradual Build-Up condition also showed elevated negativity relative to Control ($M = 4.17$), $b = 0.37$, $SE = 0.15$, $t(1170) = 2.55$, $p = .033$, $SMD = 0.18$. The difference between the two intervention conditions was not significant, $p = .649$.

Civil Society Positivity

Positive effects towards civil society and independent media did not differ significantly across conditions. Participants in the Cracks condition reported similar evaluations than those in the Control group ($M_s = 3.48$ vs. 3.19), $b = 0.29$, $SE = 0.14$, $t(1170) = 2.05$, $p = .121$. A nearly identical non-significant effect was identified in the Gradual Build-Up condition ($M = 3.44$), $b = 0.26$, $SE = 0.13$, $t(1170) = 1.92$, $p = .165$. The two intervention conditions did not differ from each other, $p = 1.000$.

Exploratory Moderation Analyses

We conducted exploratory analyses testing whether the effects of the Cracks and Gradual Build-Up interventions were moderated by three theoretically relevant individual differences: partisanship, populism, and cognitive reflection (CRT). Across all preregistered primary and secondary outcomes—including agreement with authoritarian statements, perceived authoritarian tone, concern, moral outrage, perceived threat, cracks-recognition, government negativity, and civil-society positivity—none of the interaction terms reached significance (all $p_s > .05$ for partisanship, except for its interaction with the Gradual Build-Up condition predicting agreement with the proposed Hungarian law, $p = .027$; all $p_s > .05$ for populism, except for its interaction with the Cracks condition predicting agreement with authoritarian statements, $p = .042$; all $p_s > .05$ for CRT). In other words, most of the intervention effects were statistically similar for pro-government and anti-government respondents, for individuals higher or lower in populist attitudes, and for participants with different levels of analytical reasoning. These patterns suggest that the Cracks intervention generally exerted broad, non-selective effects that did not systematically depend on ideological commitments or cognitive style.



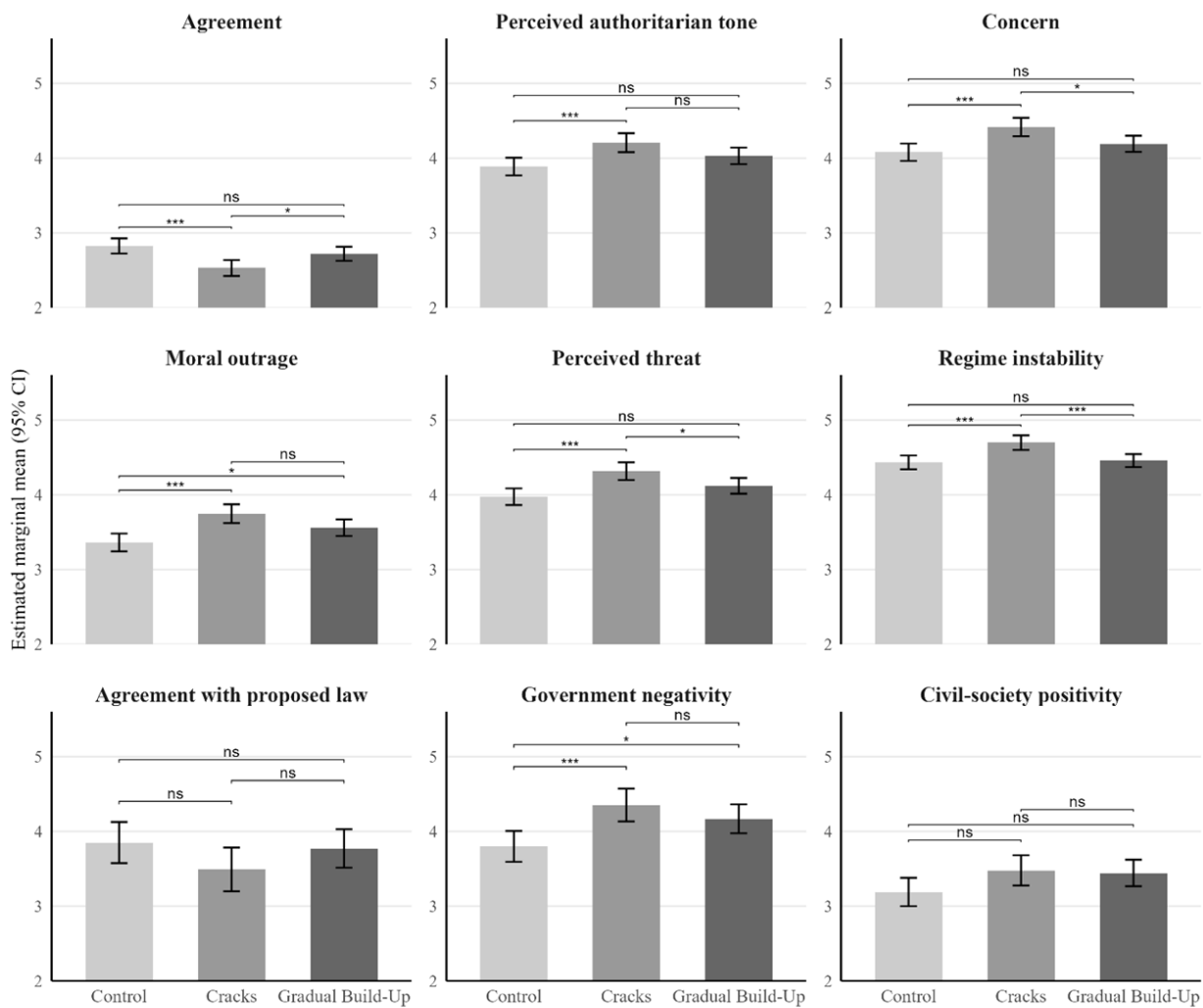
Table 1 – Summary of Intervention Effects on Authoritarian Cue Recognition Outcomes

OUTCOME	CONTRAST CRACKS VS. CONTROL	P VALUE	SMD	CONTRAST BUILD-UP VS. CONTROL	P VALUE	SMD
Agreement	↓ lower agreement	< .001	-0.29	n.s.	0.419	-0.10
Authoritarian tone	↑ higher auth. tone	< .001	0.27	n.s.	0.243	0.15
Concern about authoritarianism	↑ stronger concern	< .001	0.29	n.s.	0.503	0.10
Moral outrage towards the auth. leader	↑ stronger outrage	< .001	0.32	↑ stronger outrage	0.0497	0.17
Perceived threat to democracy	↑ stronger threat	< .001	0.30	n.s.	0.191	0.13
Cracks Recognition Index	↑ stronger crack recognition	< .001	0.28	ns.	1.000	0.03
Agreement with the Proposed Law	ns.	0.240	-0.13	ns.	1.000	-0.03
Government Negativity (Moral emotions)	↑ stronger gov. negativity	< .001	0.27	↑ stronger gov. negativity	0.033	0.18
Civil Society Positivity (Moral emotions)	ns.	.121	0.15	ns.	0.165	0.13

Notes. SMD = standardised mean difference from models with standardised outcomes. A positive SMD reflects higher scores in the intervention condition relative to control; a negative SMD reflects lower scores.



Figure 2 – Effects of narrative-based democratic inoculation interventions across outcomes



Notes. Bars depict estimated marginal means with 95% confidence intervals. Horizontal brackets indicate pairwise comparisons between conditions within each outcome; significance levels are denoted as $p < .05$ (*), $p < .01$ (**), and $p < .001$ (***), based on Bonferroni-adjusted tests. Outcome scales differ by measure: most outcomes range from 1 to 6, government negativity and civil-society positivity range from 1 to 7, and agreement with the proposed law ranges from 1 to 10. Higher values indicate stronger endorsement, perception, or emotional intensity on the respective outcome scales.



4. Discussion

Prediction error, democratic agency, and the psychological visibility of authoritarian decline. A way forward?

The present findings demonstrate that brief, psychologically precise, narrative-based democratic inoculation interventions can shift how citizens living in an informational autocracy interpret authoritarian rhetoric and perceive signs of regime fragility. Across all preregistered outcomes, the Cracks intervention produced a coherent pattern: it reduced agreement with authoritarian statements, increased perceived authoritarian tone, heightened concern about authoritarianism and moral outrage, elevated perceived democratic threat, and improved recognition of early, high-diagnostic instability cues. These changes emerged after a single, short online module. By contrast, the Gradual Build-Up intervention (despite using the same structure, delivery format, length, and reflective depth) generated weaker and less consistent effects. Although the preregistered hypotheses emphasised the Gradual Build-Up intervention as the primary driver of authoritarian cue recognition, the observed pattern suggests that, in contexts of long-standing democratic backsliding, instability-focused cues may provide stronger psychological leverage than erosion-focused explanations. Understanding this asymmetry illuminates the cognitive, affective, and informational mechanisms through which citizens recognise authoritarian cues and make sense of politics in an informational autocracy.

The first contribution is clarifying why slow authoritarian consolidation is psychologically difficult to detect. Political science establishes that democratic backsliding typically proceeds through incremental, legally framed steps that preserve an appearance of normalcy (Bermeo, 2016; Linz, 2000; Schedler, 2013; Varol, 2015). Each micro-shift appears tolerable when evaluated narrowly. Construal level theory suggests that citizens typically process day-to-day political events at a concrete, low-level construal, which inhibits the recognition of long-term, gradual structural changes (Trope & Liberman, 2010).

Meanwhile, historical-analogy reasoning (about authoritarian trajectories) places substantial cognitive demands that people often fail to meet

Democratic resilience may depend less on teaching facts than on cultivating interpretive skill and contingency

(Neustadt & May, 1986). In informational autocracies like Hungary, these perceptual tendencies are compounded by structural opacity: media capture, constitutional engineering, and politicised oversight bodies systematically obscure the visibility of institutional decline (Enyedi, 2018; Krekó & Enyedi, 2018; Scheppele, 2018). Together, these frameworks point to an interpretive obstacle: even politically attentive citizens may struggle to “zoom out” and integrate isolated events into a coherent sense of authoritarian consolidation.

The Gradual Build-Up intervention attempted to intervene at this bottleneck by presenting historically grounded testimonies illustrating censorship, loyalty-based appointments, and subtle fear cues. The saying-is-believing reflection offered an opportunity to participants to reorganise these examples into a higher-level pattern. Yet the intervention produced relatively modest effects (mainly elevating moral outrage and government-directed negativity). One plausible explanation is that citizens living through long-term backsliding are already saturated with narratives of erosion in Hungary. After more than a decade of institutional decline, incremental framing may no longer feel novel or diagnostically useful. Therefore, participants may have lacked the motivational or cognitive leverage necessary to reorganise familiar erosion narratives into a new interpretive framework. Moreover, abstract characterizations of authoritarian build-up may be challenging to connect to personally observable cues, limiting the intervention’s psychological traction. In line with this interpretation, classic work on expert judgement suggests that although historical analogies can help organise political experience, they are often more misleading than informative for prediction when extended beyond surface similarity (Tetlock, 2005).

These patterns align with broader arguments that “awareness-raising” alone rarely shifts political behaviour (Graham & Svobik, 2020). As recent theoretical work notes (Kertzer & Zeitzoff, 2017), re-democratisation faces two major psychological obstacles: insufficient understanding of how autocratization operates and apathy or fatalism about whether change is possible. Awareness-focused interventions may address the former but leave the latter intact. When citizens learn more about gradual authoritarianism without being shown a plausible pathway out of the situation, the new information may feel inert—important but not actionable. Our results echo this logic: deeper understanding produced small, diffuse benefits, but did not meaningfully change perceptions of regime fragility or democratic threat sensitivity. In other words, awareness without activation may be insufficient for meaningful perceptual or motivational change (Geddes et al., 2014).

In contrast to the limited traction of the Gradual Build-Up module, the Cracks intervention operated on a distinct set of psychological mechanisms by directly focusing on salient, high-diagnostic signals. Computational models of political cognition conceptualise this in terms of prediction error: belief change is greatest when observed events meaningfully deviate from what is expected. The classic Rescorla–Wagner framework formalises this dynamic, predicting that learning is proportional to the discrepancy between expected and actual outcomes (Rescorla & Wagner, 1972).



“ Our results offer theoretically grounded evidence for why instability-focused inoculation may outperform erosion-focused inoculation ”

János Salamon

MORES Researcher,
ELTE Eötvös Loránd University,
Budapest

The instability cues highlighted in the Cracks narratives (such as elite defection, propaganda inconsistency, public ridicule, and protest coordination) can be interpreted as types of events that generate large prediction errors. In line with these arguments, cascade and tipping-point models (Kuran, 1991; Lohmann, 1994) show that such signals recalibrate perceived common knowledge: when prominent elites dissent or when symbols of state authority become objects of mockery, individuals infer that their private doubts may be widely shared, triggering rapid shifts in expectations about regime stability. Modern theories of informational autocracies (Guriev & Treisman, 2022) further argue that because such regimes rely heavily on curated competence and narrative control, even small cracks in this façade can have outsized psychological impact. Importantly, these cues are concrete and personally recognizable. To the extent that multiple such instability cues have become more visible in the informational environment in recent years, the Cracks intervention may have resonated with patterns that participants already perceive or anticipate, further amplifying its psychological impact. Unlike abstract discussions of erosion, elite splits, ridicule, and coordination failures can be mapped directly onto everyday experience, enhancing psychological salience and facilitating the transfer of the narrative lens to contemporary political judgments.

We suppose that the self-persuasive saying-is-believing exercise amplified these mechanisms (Aronson, 1999). Several wise-intervention studies demonstrated that when individuals restate a principle in their own words, they integrate it into their interpretive framework, producing durable shifts in meaning (see Walton, 2014; Walton & Wilson, 2018; Walton, 2025). In the Cracks condition, this reflective component likely encouraged participants to consolidate the insight that authoritarian power is contingent and internally fragile. As a result, after the self-persuasive exercise, participants evaluated authoritarian rhetoric through a lens that became sensitive not only to the collapse of the authoritarian systems as a primary effect (crack-recognition), but also to the recognition of authoritarian rhetoric in general, with its affective (e.g., concerns and moral emotions) and cognitive consequences (e.g., agreement).

This interpretation is consistent with recent arguments that focusing citizens on the endings of authoritarian regimes—rather than on their gradual consolidation—may generate more robust democratic orientations (Graham & Svobik, 2020). Thinking about collapse can provide hope, agency, and a sense that the status quo is not permanent. These motivational ingredients may be especially important in long-standing authoritarian environments, where resignation and habituation are widespread. In our study, participants exposed to instability cues appeared more alert, more emotionally responsive, and more willing to interpret political statements through a democratic lens. Thus, activation—imagining how authoritarianism weakens

Why ‘cracks’ works: instability cues feel concrete, diagnostic, and action-relevant—making them cognitively and emotionally powerful

or ends—may be a more powerful psychological lever than comprehension alone (Kertzer & Zeitzoff, 2017; Geddes et al., 2014).

A second contribution concerns the ideological symmetry of the effects. The interventions showed little evidence of systematic interaction with partisanship, populist attitudes, or cognitive reflection. Such symmetry is relatively uncommon in political communication, where persuasion often depends on identity-protective cognition. Two isolated interactions nevertheless reached statistical significance: partisanship interacted with the Gradual Build-Up condition in predicting agreement with the proposed law, and populist attitudes interacted with the Cracks condition in predicting agreement with authoritarian statements. These effects were outcome-specific and confined to measures that directly invited evaluative judgments, and they did not generalise across the broader pattern of largely non-selective intervention effects. This suggests that identity-related moderators may surface primarily when interventions engage explicit agreement rather than broader interpretive or affective processes.

One explanation is that the interventions, especially the Cracks module, did not instruct participants on what to believe but reorganised how they interpret political cues. The team spent a lot of time designing the intervention materials to avoid or minimise provoking reactance from any political side. Structural features of authoritarianism we highlighted in terms of elite fracture, narrative incoherence, and performance breakdown might not be strongly provocative for either conservative or liberal voters, neither those who support populist views nor those who reject it, and neither those who prefer in-depth analytic thinking nor those who make less cognitive efforts. This pattern aligns with recent cross-national inoculation and civic-education research showing broad, non-selective effects (Finkel et al., 2023; Öztürk et al., 2025; Yameogo et al., 2025; Ovádek & Pavone, 2025). In sum, the present results (among other current examples) suggest that well-designed reframing interventions can reduce the role of partisan motivated reasoning by engaging meaning-making processes that are not tightly bound to ideological identity.

Several limitations qualify for these conclusions. All measures were collected immediately after the intervention; future work should test durability, generalisation to novel political content, and behavioural consequences such as selective exposure, information sharing, or resistance to manipulation. The Cracks Recognition Index, although reliable, requires further validation, including factor structure, test–retest reliability, and comparisons with expert-coded indicators of regime fragility. Finally, the study was conducted in a single national context characterised by extensive backsliding; assessing the portability of these interventions in other hybrid regimes and at-risk democracies will be crucial.

Despite these limitations, the present findings offer theoretically ground-

The findings suggest that awareness explains; activation moves. Fragility frames may activate citizens' democratic attention more effectively

ed evidence for why instability-focused inoculation may outperform erosion-focused inoculation. Gradual build-up requires abstraction, perspective-taking, and historical reasoning. At the same time, instability cues operate through prediction error, cascade logic, and disruptions to authoritarian narrative coherence, triggering sharper, more drastic changes in perception. The Cracks intervention effectively harnessed these mechanisms by presenting vivid, historically grounded, culturally legible examples of regime fragility and by requiring participants to articulate their meaning. This combination of prediction-error salience and self-generated reframing produced systematic recalibrations in how citizens interpreted authoritarian rhetoric. Authoritarian systems slowly and silently spread in classic democracies; however, these invisible changes can become visible through interventions that highlight their potential cracks.

Taken together, the results suggest that strengthening democratic resilience may require more than increasing citizens' understanding of authoritarianism. It may require cultivating a sense of possibility: that authoritarian power can weaken, that cracks can appear, and that the trajectory of decline is neither fixed nor irreversible. Interventions that activate this sense of agency may therefore be especially promising tools in contexts of entrenched backsliding.



“ The deeper question is not simply whether we can prepare citizens to act, but whether we can identify when civic action contributes to democratic resilience and when it legitimises backsliding ”

Gábor Orosz

MORES contributor,
Flinders University, Adelaide

References

- Abels, C. M., Huttunen, K. J. A., Hertwig, R., & Lewandowsky, S. (2024). Dodging the autocratic bullet: enlisting behavioural science to arrest democratic backsliding. *Behavioural Public Policy*, 1-28.
- Aronson, E. (1999). The power of self-persuasion. *American Psychologist*, 54(11), 875–884. <https://doi.org/10.1037/h0088188>
- Bermeo, N. (2016). On democratic backsliding. *Journal of Democracy*, 27(1), 5–19. <https://doi.org/10.1353/jod.2016.0012>
- Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences of the United States of America*, 108(31), 12653–12656. <https://doi.org/10.1073/pnas.1103343108>
- Enyedi, Z. (2018). Democratic backsliding and academic freedom in Hungary. *Perspectives on Politics*, 16(4), 1067–1074. <https://doi.org/10.1017/S1537592718002165>
- Eroğlu, M. H., Finkel, S. E., Neundorf, A., Öztürk, A., & Rascón Ramírez, E. G. (2025). Choosing democracy over party? How civic education can mitigate the anti-democratic effects of partisan polarization. *British Journal of Political Science*, 55, e65, 1–10. <https://doi.org/10.1017/S000712342500016X>
- E. Finkel, S., Neundorf, A., & Rascón Ramírez, E. (2024). Can online civic education induce democratic citizenship? Experimental evidence from a new democracy. *American Journal of Political Science*, 68(2), 613–630. <https://doi.org/10.1111/ajps.12765>
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- Geddes, B., Wright, J., & Frantz, E. (2014). Autocratic breakdown and regime transitions: A new data set. *Perspectives on Politics*, 12(2), 313–331. <https://doi.org/10.1017/S1537592714000851>
- Graham, M. H., & Svobik, M. W. (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. <https://doi.org/10.1017/S0003055420000052>
- Greskovits, B. (2015). The hollowing and backsliding of democracy in East Central Europe. *Global Policy*, 6(S1), 28–37. <https://doi.org/10.1111/1758-5899.12225>
- Gunitsky, S. (2015). Corrupting the cyber-commons: Information politics and authoritarian resilience. *Perspectives on Politics*, 13(1), 42–54. <https://doi.org/10.1017/S1537592714003120>
- Guriev, S., & Treisman, D. (2022). *Spin dictators: The changing face of tyranny in the 21st century*. Princeton University Press.
- Hradický, J. M. (2024). Democratic Inoculation in Slovakia and Serbia:

- Do Politics Learn from the Experience of Democratic Backsliding?. *Studia Politica Slovaca*, 17(1), 36–58. <https://doi.org/10.31577/SPS.2024-1.3>
- Kertzer, J. D., & Zeitzoff, T. (2017). A bottom-up theory of public opinion about foreign policy. *American Journal of Political Science*, 61(3), 543–558. <https://doi.org/10.1111/ajps.12314>
- Krekó, P., & Enyedi, Z. (2018). Explaining Eastern Europe: Orbán's laboratory of illiberalism. *Journal of Democracy*, 29(3), 39–51. <https://doi.org/10.1353/jod.2018.0043>
- Kuran, T. (1991). Now out of never: The element of surprise in the East European Revolution of 1989. *World Politics*, 44(1), 7–48. <https://doi.org/10.2307/2010422>
- Johnston, M. (2005). *Syndromes of corruption: Wealth, power, and democracy*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511490965>
- Labanino, R., & Dobbins, M. (2023). Multilevel venue shopping amid democratic backsliding in new European Union member states. *Politics and Governance*, 11(1), 65–79. <https://doi.org/10.17645/pag.v11i1.5882>
- Levitsky, S., & Way, L. A. (2010). *Competitive authoritarianism: Hybrid regimes after the Cold War*. Cambridge University Press.
- Levitsky, S., & Ziblatt, D. (2018). *How democracies die*. Crown.
- Linz, J. J. (2000). *Totalitarian and authoritarian regimes*. Lynne Rienner Publishers.
- Lohmann, S. (1994). The dynamics of informational cascades: The Monday demonstrations in Leipzig, East Germany, 1989–91. *World Politics*, 47(1), 42–101. <https://doi.org/10.2307/2950679>
- Orosz, G., Faragó, L., Paskuj, B., & Krekó, P. (2024). Strategies to combat misinformation: Enduring effects of a 15-minute online intervention on critical-thinking adolescents. *Computers in Human Behavior*, 159, 108338. <https://doi.org/10.1016/j.chb.2024.108338>
- Orosz, G., Faragó, L., Paskuj, B., Rakovics, Z., Sam-Mine, D., Audemard, G., Samy, M., Modéliar, A., & Krekó, P. (2024). Softly empowering a prosocial expert in the family: Lasting effects of a counter-misinformation intervention in an informational autocracy. *Scientific Reports*, 14, Article 11763. <https://doi.org/10.1038/s41598-024-61232-x>
- Orosz, G., Paskuj, B., Faragó, L., & Krekó, P. (2023). A prosocial fake news intervention with durable effects. *Scientific Reports*, 13, 3958. <https://doi.org/10.1038/s41598-023-30867-7>
- Ovádek, M., & Pavone, T. (2025). The memory engine: Do citizens learn from past democratic backsliding episodes? Draft paper and pre-analysis plan.
- Öztürk, A., Finkel, S. E., Neundorf, A., & Rascón Ramirez, E. (2025). *Mobilizing for democracy under authoritarianism: The power of social media campaigns*. Unpublished manuscript.
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century-Crofts.

- Schedler, A. (2013). *The politics of uncertainty: Sustaining and subverting electoral authoritarianism*. Oxford University Press.
- Scheppele, K. L. (2013). Not your father's authoritarianism: The creation of the 'Frankenstate.' *Newsletter of the European Politics and Society Section of the American Political Science Association*, 23(Winter), 5–9.
- Scheppele, K. L. (2018). Autocratic legalism. *The University of Chicago Law Review*, 85(2), 545–584. <https://chicagounbound.uchicago.edu/uclrev/vol85/iss2/2>
- Shenhav, A., Rand, D. G., & Greene, J. D. (2012). Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, 141(3), 423–428. <https://doi.org/10.1037/a0025391>
- Slater, D. (2010). *Ordering power: Contentious politics and authoritarian leviathans in Southeast Asia*. Cambridge University Press.
- Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.
- Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, 117(2), 440–463. doi: <https://doi.org/10.1037/a0018963>
- Varol, O. O. (2015). Stealth authoritarianism. *Iowa Law Review*, 100(4), 1673–1742.
- Van der Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2), 1600008. <https://doi.org/10.1002/gch2.201600008>
- Walton, G. M. (2014). The new science of wise psychological interventions. *Current Directions in Psychological Science*, 23(1), 73–82. <https://doi.org/10.1177/0963721413512856>
- Walton, G. M., & Wilson, T. D. (2018). Wise interventions: Psychological remedies for social and personal problems. *Psychological Review*, 125(5), 617–655. <https://doi.org/10.1037/rev0000115>
- Walton, G. M. (2025). *Ordinary magic: The science of how we can achieve big change with small acts*. Harmony.
- Yameogo, S., Neundorf, A., & Ozturk, A. (2025). Promoting democracy in the context of terrorism: Experimental evidence from Burkina Faso [Pre-print]. OSF. https://osf.io/preprints/osf/b4hju_v1



Supplemental Material

Table 2 – The content of the interventions

QUESTION	GRADUAL BUILD-UP CONDITION	CRACKS (SIGNS OF AUTHORITARIAN DECLINE CONDITION
Framing	Highlights how authoritarian systems often consolidate power gradually through small, legally framed steps. Emphasises normalisation, subtle fear cues, and the slow narrowing of democratic space. After reading the historical testimonials and expert commentary, participants were asked to give advice on how to explain this gradual, often hard-to-notice process to young people in simple, accessible language.	Frames authoritarian regimes as structurally fragile, with decline often signalled by early signs such as elite defection, ridicule, contradictory messaging, and emerging protest coordination. After reading the testimonials and expert commentary, participants were asked to give advice on how to explain these early “cracks” to young people so they can recognise when an authoritarian system is beginning to weaken.
Example testimonials/ Expert commentary	<p><i>“... the attack on democracy begins slowly and gradually. For most citizens, it is barely noticeable at first. Elections are still held. Opposition politicians still sit in parliament. Independent newspapers still appear. Each step seems small—none appears to be something that would truly endanger democracy.”</i></p> <p>— Levitsky & Ziblatt (2018), <i>How Democracies Die</i></p>	<p><i>“Authoritarian regimes often find themselves in a contradictory position: while they strive to suppress the opposition, their own communication becomes increasingly inconsistent, reactive, and opaque. This paradox is often a precursor to a political breaking point.”</i></p> <p>— Howard Johnston (2005). “Talking the Walk: Speech Acts and Resistance in Authoritarian Regimes.” In <i>Repression and Mobilisation</i>, p. 151.</p>
Saying-is-believing self-persuasion	The saying-is-believing task asks participants to restate the core idea of the intervention message as advice to someone else (e.g., a younger person). This self-generated explanation encourages participants to actively interpret the content rather than passively consume it, which strengthens personal relevance, deepens understanding, and promotes longer-term integration of the message.	
Example messages participants wrote	<i>If it's about politics, try to get information from multiple sources, and compare the information you have obtained. Think, weigh things, and never keep your opinion to yourself. Oppression can exist today as well, only in a much more refined form. For example, when someone is laughed at, stigmatised, or put at a disadvantage because they think differently, or because their religion, identity, or political views are different. Do not be afraid of being excluded—true, it is not a grateful role, but there is a need for people who dare to think differently and, by doing so, show others a path towards a better direction</i>	<i>“I think that a political system starts to tip over when people who strongly support a certain politician suddenly begin to have a completely different point of view. In my family there were people who held the current party in high regard and others who supported the opposition party, but the one who supported the current party suddenly stopped talking about it even at family gatherings, because he also felt that what is happening now is not good and everything is turning in a worse and worse direction. I can only bring his example in this regard, because he held an important position in the party. This means that inside, this party had started to become unstable. They no longer stick to what they promised, and when they introduce something beneficial, like a loan, they only announce the rules and criteria one or two days before. Because until the point when it was invented everything was fine, and afterwards something must have happened which made it necessary to introduce these criteria.”</i>