Behavioural Public Policy (2024), 1–28
doi:10.1017/bpp.2024.43

ARTICLE

Dodging the autocratic bullet: enlisting behavioural science to arrest democratic backsliding

Christoph M. Abels[1], Kiia Jasmin Alexandra Huttunen[2], Ralph Hertwig[3] and Stephan Lewandowsky[1,2]

[1]Department of Psychology, University of Potsdam, Potsdam, Germany; [2]School of Psychological Science, University of Bristol, Bristol, United Kingdom and [3]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany
Corresponding author: Christoph M. Abels; Email: christoph.maximilian.abels@uni-potsdam.de

(Received: 11 September 2023; Revised: 9 June 2024; Accepted: 24 July 2024)

Abstract

Despite a long history of research on democratic backsliding, the process itself – in which the executive branch amasses power and undermines democratic processes and institutions – remains poorly understood. We seek to shed light on the underlying mechanisms by studying democratic near misses: cases in which a period of autocratic governance is quickly reversed or full backsliding is prevented at the last minute. Building on the literature on near misses in sociotechnical systems such as nuclear power plants, we adapt the drift-to-danger model to the study of democratic systems. Two key findings emerge: First, democratic backsliding is often triggered by political elites pushing the boundaries of their power by violating norms, which are crucial yet vulnerable safeguards for democracy. Second, democratic backsliding is unpredictable and non-linear, being driven by the interaction between political elites and the public. Norm-violating elites may feel legitimized by a supportive public that sees norm violations as justified. At the same time, political elites may signal that norm-violating behaviour is acceptable, potentially leading the public to adopt anti-democratic beliefs and behaviours. We identify risk factors that make norm violations more likely and outline behavioural sciences-based interventions to address these violations.

Keywords: democratic backsliding; drift-to-danger model; elite norm violations; near misses

Introduction

When democracies fail, they rarely crash and burn in an instant. In most cases, their demise is slow. Failing democracies drift through a period of backsliding, in which the executive branch amasses power and undermines democratic processes and institutions. In some cases, a period of autocratic governance is quickly reversed or a full backsliding is prevented at the last minute. These ‘near misses’ (Ginsburg and Huq,

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

2018) are at the heart of our investigation. While near misses are a comparatively new concept in democracy studies, the field of human factors has long distinguished between accidents and near-accidents in sociotechnical systems such as oil rigs (Jones *et al.*, 1999). We adapt the drift-to-danger model of sociotechnical accidents developed by Rasmussen (1997) to conceptualize democratic instability. This approach helps to understand how democratic systems can gradually erode, often in plain sight. It highlights that incremental deviations from a liberal democratic equilibrium follow a non-linear dynamic: Once a tipping point or threshold has been reached, reversing democratic backsliding becomes extremely difficult or impossible, and the transition to an authoritarian regime can be swift (e.g., Hitler's establishment of a one-party dictatorship and a repressive police state within months of his appointment as Chancellor; Weber, 2022).

In this article, we analyse several near misses to identify both enabling risk factors and protective interventions that transcend the particularities of each near-miss episode (Lührmann *et al.*, 2020). By synthesising insights and methods from the behavioural sciences that can be recruited to strengthen democratic systems and prevent backsliding, the article serves as a conceptual review with empirical aspects. First, we adapt the drift-to-danger model and apply it to democratic near misses, identifying elite norm violations as a key cause of backsliding in all cases. Second, we draw on experimental, survey and empirical data to examine the consequences of elite norm violations on public attitudes and behaviours. Third, we outline classes of behavioural science-based interventions suitable for addressing risk factors identified as facilitating or amplifying elite norm violations.

Democracy and its erosion

Discriminating between democratic and authoritarian regimes is becoming increasingly difficult, as most states hold elections (Lührmann *et al.*, 2018) and have learned to mimic various other attributes of liberal democracies. Only a few regimes (e.g., Belarus, Iraq, North Korea and Russia) are openly authoritarian, relying on a repressive security apparatus and coercion to control their citizens. 'In the modern era, authoritarian wolves rarely appear as wolves. They are now clad, at least in part, in sheep's clothing' (Varol, 2015, p. 1677). To illustrate, the democratic system in Hungary – an European Union (EU) member state – has been seriously eroded by measures such as gerrymandering, hijacking of state institutions, constitutional changes that weaken democratic checks and balances, and control of the media and public discourse (Polyák, 2019; Szelényi, 2022).

Following Lindberg *et al.* (2014), we conceptualize democracy in terms of five core components: electoral, liberal, participatory, deliberative and egalitarian. These components form the basis of the democracy scores assigned by the Varieties of Democracy (V-Dem) project (e.g., Lindberg *et al.*, 2014; Boese *et al.*, 2021). The *electoral* component captures the idea that leaders' responsiveness is achieved through a system of competition and accountability, ensured by regular free and fair elections. *Liberal* refers to the protection of individual and minority rights against a tyranny of the majority. *Participatory* means that citizens' active political participation in all political processes is encouraged – for example, through engagement in political parties and civil society

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

organizations and direct democracy. *Deliberative* means that decision-making should be based on respectful and reasonable dialogue in pursuit of the public good. Finally, democracy should be *egalitarian* and strive to distribute resources such as education and health equitably.

Frequently, countries fail to ensure several of these core components of democracy, leading to incomplete democracies, hybrid systems and autocratic types of governance. Complete breakdowns or reversals of democracy are not as common as they used to be (Boese *et al.*, 2021). Most symptoms of reversal are subtle, and backsliding processes take time (Haggard and Kaufman, 2021). In some cases, a transient period of democratic backsliding is reversed. Instances in which democracies are exposed to social, political or economic forces that could catalyse backsliding, but manage to overcome these forces and avoid a full and lasting backslide to autocratic governance can be understood as *near misses*. A near miss is defined as a 'case in which a country 1) experiences a deterioration in the quality of initially well-functioning democratic institutions, without fully sliding into authoritarianism, but then, 2) within a time frame of a few years, at least partially recovers its high-quality democracy' (Ginsburg and Huq, 2018, p. 17).

Democracy is in a tough spot globally. At the time of writing in 2024, the world is almost evenly divided between democratic (91) and autocratic states (88), with 71% of people living in autocracies, up from 48% in 2013 (Nord *et al.*, 2024). Citizens in 60 countries, making up around 45% of the world's population, are being asked to cast their votes in elections in 2024. The majority of these elections (52%) are being held in countries in which democracy is declining (Nord *et al.*, 2024). Although it may seem unlikely that established democracies will experience substantial backsliding, countries such as the US and UK have recently shown early signs of democratic erosion. In the US, in particular, the last few years have seen a steady deterioration of norms and practices crucial for maintaining democracy. Although the country has experienced tumultuous periods before (e.g., the Watergate scandal), four problematic developments now coincide for the first time: political polarization, conflicts over in-group membership, high levels of social and economic inequality and excessive use of executive power (Mettler and Liebermann, 2020).

The effects of political polarization are particularly salient, as cooperation between the two parties in US Congress has become increasingly difficult, with members of Congress willing to break with established norms (e.g., denying a sitting president the hearings required to fill a vacant Supreme Court seat; Kar and Mazzone, 2016). During his presidency and even more so as the Republican candidate for the 2024 Presidential election, Donald Trump also repeatedly attacked the judiciary and the rule of law (Freedom House, 2019). Beyond that, elements of the Republican party challenged the legitimacy of the 2020 presidential election, with various attempts to overturn the result and keep Trump in office (Helderman, 2022), culminating in a violent insurrection on 6 January 2021 (Haslam *et al.*, 2023).

Political elites in the UK have also shown disregard for democratic norms. A case in point was the unlawful prorogation of parliament (i.e., ending of the parliamentary session) in September 2019. This move was largely seen as an attempt by then Prime Minister Boris Johnson to avoid parliamentary scrutiny of his government's Brexit plans and to prevent parliament from thwarting a hard Brexit. Johnson argued that

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

the goal was to give his government time to prepare for the next parliamentary session (Hadfield, 2019). However, the Supreme Court decided that the decision was 'unlawful because it had the effect of frustrating or preventing the ability of Parliament to carry out its constitutional functions without reasonable justification.'[1]

Modeling democratic near misses

Although established democracies like the US and UK appear resilient on the surface, these recent developments give cause for concern. This is because the processes at the heart of backsliding – which starts gradually with slow incremental change before suddenly switching to change that is difficult to reverse – remain poorly understood (Bermeo, 2016; Waldner and Lust, 2018; Wiesner *et al.*, 2019; Haggard and Kaufman, 2021; Wunsch and Blanchard, 2022).

We therefore turn to the field of human factors, which has a long history of studying accidents in complex sociotechnical systems such as nuclear power plants or oil rigs, to provide a conceptual lens through which to study democratic near misses. The term 'near miss' is widely used here to distinguish accidents, which result in injury or damage, from incidents without such detrimental outcomes (Jones *et al.*, 1999). A near miss thus refers to any event that could have caused substantial damage but was prevented 'by only a hair's breadth' (Reason, 2016, p. 118). To cite Jones *et al.* (1999), a near miss is 'an unintended incident which, under different circumstances, could have become an accident' (p. 63). In addition to highlighting risk factors, analyses of near misses can draw attention to the safety layers that contribute to preventing an adverse event (Gnoni *et al.*, 2022). Ginsburg and Huq (2018) have also previously discussed near misses in the context of democracy, arguing that they can help to identify the economic, political and social conditions that 'can repel a threat to participatory governance once such a threat has arisen' (p. 17).

We argue that sociotechnical systems share similarities with democratic systems and can therefore help to understand the non-linear process of erosion underlying democratic backsliding. In particular, we identify the drift-to-danger model developed by Rasmussen (1997) as a valuable framework to study democratic backsliding.

The drift -to-danger model

Rasmussen (1997) argued that any system is shaped by objectives and constraints to which individuals must adhere in order for the system to work effectively. Nevertheless, various degrees of freedom remain. Individuals interpret this leeway and develop strategies to balance the effort they invest and the demands of the system. If the operating conditions change, these strategies will be modified. If, for example, a factory increases its employees' workload without hiring more staff, employees might neglect safety protocols to meet the new demands. According to Rasmussen, this will likely result in 'systematic migration toward the boundary of functionally acceptable performance' (p. 189). If this boundary is irreversibly crossed, an error or an accident may

[1] https://www.supremecourt.uk/cases/docs/uksc-2019-0192-summary.pdf

occur. Where exactly the boundary lies is inherently difficult to identify; accidents are often the only source of information on its position (Cook and Rasmussen, 2005).

In most systems, boundary transgressions are anticipated and addressed by adding several safety layers – also known as defences-in-depth – to the system's design (for an overview, see Marsden, 2022). These safety layers, which are ideally independent, guard against each others' breakdown, absorb violations, and thus maintain system stability even when failures such as human error or a malfunctioning alert system occur (Reason, 2016). Many accidents discussed in the human factors literature, such as the partial meltdown of a nuclear reactor at the Three Mile Island power plant, can be attributed to multiple failures in complex systems, in which both human operators and technology contributed to the accident (Perrow, 1984).

Rasmussen (1997) argued that while these multiple safety layers initially help the system to maintain its operations, the absence of a feedback signal – that is, visible negative effects of a transgression – prevents the necessary changes in behaviour. As the safety layers wear down over time, the system becomes unable to manage the strain, and the gradual build-up of transgressions eventually results in accidents. Additionally, the number of safety layers increases the overall complexity of the system, which in turn increases the level of risk (Marsden, 2022). Rasmussen's model thus describes a system whose gradual erosion is not directly visible, but becomes apparent only when the system breaks down under pressure. Rasmussen identified the absence of an overarching monitoring or coordination layer with sufficient understanding of the entire organization to identify deviations and respond accordingly as a major flaw of the defences-in-depth approach (Rasmussen, 1997).

It should be noted here that not all transgressions go unnoticed. Operators may choose to ignore them if deviating from rules and norms has become accepted practice in the organization. The explosion of the Challenger space shuttle in 1986 is a case in point (Perrow, 1996). NASA engineers had repeatedly deviated from their goal of zero failures prior to the explosion, as previous deviations had not resulted in an accident. Crucially, damage to critical components (i.e., the O-ring seals in the booster rockets) had been discovered in tests and flights preceding the accident (Rogers *et al.*, 1986).

In summary, according to the drift-to-danger model, complex sociotechnical systems are designed to be fault tolerant. This makes them resistant to human or technical error. However, while safety layers can tolerate small faults, they fail catastrophically once the compounding of multiple small faults reaches a threshold. A near miss happens when those faults can be contained and reversed before the threshold is crossed. We think of democracy as a similar system of largely independent safety layers – checks and balances as well as legal and informal norms – designed to protect the system against disruptions. Rasmussen's criticism about the absence of a coordination layer that monitors safety layers and identifies deviations from rules and procedures (Rasmussen, 1997) also applies to democratic systems. Although a substantial number of democratic institutions (e.g., courts, the media, parliament, government and civil society) implement a defences-in-depth approach, the lack of an overarching coordination layer creates systemic vulnerabilities. Furthermore, democracies are much more complex and dynamic than technical systems. Modelling studies show that the dynamic demands of political (e.g., voters, parties, politicians, lobbyists) and economic (e.g., budget constraints, inflation) factors introduce additional risks by pushing the system to operate at the limits of acceptable strain (Eliassi-Rad *et al.*, 2020; Morrison and Wears, 2022; Wiesner *et al.*, 2023). Like technical systems, democracies can absorb small deviations from the ideal operational practice. If violations are normalized, however, their effects can accumulate, eventually leading to non-linear and irreversible system changes.

In the following, we analyse five cases in which such catastrophic failures have been successfully averted. We examine the lessons that can be drawn from considering these cases through the lens of the drift-to-danger model and the behavioural sciences generally.

Historical analysis of democratic near misses

Our review of the literature[2] identified five cases of democratic near misses, presented in Table 1: Finland (1930), the UK (1930s), Spain (1981), Colombia (2010), Sri Lanka (2015) and South Korea (2017). The cases demonstrate that the erosion of democracy often begins with political elites pushing the boundaries of their power, and that – consistent with the drift-to-danger model – democratic erosion begins gradually (Rasmussen, 1997). Like frogs in a pan of slowly heating water, those who protect democracy often fail to see the risks to the system until it is almost too late. The sudden and unexpected collapse of democracy in Chile in 1973 serves as a clear illustration of this process. Consequently, Chile, while not being a near miss itself, is highlighted as a special case in Table 1.

Figure 1 illustrates the drift-to-danger model as applied to democratic backsliding. As political elites repeatedly violate norms, democracy slowly drifts towards autocracy. Public or behavioural interventions can reduce the prevalence or severity of such norm violations, thereby slowing the drift. The safety layers designed to slow or prevent democratic backsliding are also subject to protective and erosive forces. Risk factors such as misinformation, populism and polarization can undermine them; behavioural science interventions can strengthen them. The number of safety layers and the point at which a layer fails are difficult to predict. If at least one safety layer holds, full backsliding can be prevented, leading to a near miss. However, if all layers fail, the threshold to autocracy will be reached, endangering core democratic principles (e.g., freedom of speech, protection of minority rights). Thus, any norm violation is inherently problematic, as it remains unpredictable when and if a violation will push democracy over the edge.

Elite norm violations in near misses

The near misses presented in Table 1 reveal complex non-linear patterns of interacting factors; however, elite violations of democratic norms emerge as a core driver of democratic backsliding in all cases. These norm violations do not necessarily breach constitutional boundaries, indicating that constitutional and other legal provisions

[2] We searched Google Scholar for relevant journal articles using the keywords: ‘near misses’ AND ‘democratic backsliding’, ‘near misses’ AND ‘backsliding’, and ‘democratic near misses’. More details on case selection are presented in the Appendix.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Table 1. Selected historical cases of democratic near misses

Country	Year	Context	Outcome
Finland	1930	The Lapua Movement, a nationalist political group, emerged in Finland after farmers attacked a communist youth parade in the village of Lapua, leading to violent clashes. Gaining support particularly in rural areas and among conservative and nationalist groups, the movement sought to establish a strong, authoritarian state. Their tactics included political violence, intimidation, and the kidnapping and arrest of communist politicians on charges of treason. The conservative president and parts of the military sympathised with the movement. In 1932, the movement's supporters gathered in Mäntsälä, demanding a new "patriotic government" and threatening violence if their demands were not met (Ginsburg and Huq, 2018).	Key military personnel did not join the insurrection. Judges imposed harsh sentences on those involved in the Mäntsälä incident. Members of the conservative party who had benefited electorally from the Lapua Movement started to perceive it as a threat. A newly formed cross-party "lawfulness front" split the Agrarian Union that had supported the movement, and served as a counter-movement. A centre-left coalition was elected in 1937, ending Finland's drift to autocracy (Ginsburg and Huq, 2018).
United Kingdom	1930s	Member of Parliament Oswald Mosley established the British Union of Fascists (BUF) in an attempt to establish a totalitarian regime in the UK. The BUF gained popularity as the country struggled with deep economic depression. Emerging autocracies in Italy and Germany were seen as examples of powerful modern governments. When the movement became increasingly affiliated with antisemitism and violence, its popularity decreased (Ewing and Gearty, 2001).	In 1934, a rally at Olympia revealed the true character of the BUF and its close ties to the police. Attended by thousands, including members of the BUF's paramilitary wing, the "Blackshirts", the event turned violent when they clashed with anti-fascist protesters, leading to hours of unrest. Although the BUF was never officially dissolved, public support waned after the incident, and the Public Order Act of 1936 addressed the threat it posed – for example, by banning political uniforms (Cullen, 1993).
Spain	1981	The election of a new prime minister in February 1981 was interrupted by 200 civil guardsmen who seized control of the parliament building and held members hostage for 18 hours. The insurrectionists, trying to stop Spain's increasing democratization, demanded the appointment of a conservative general as the new prime minister (Levitsky and Ziblatt, 2023).	The coup failed due to the intervention of King Juan Carlos I and political elites who denounced it (Maxwell, 1991). Additionally, a rally of more than one million people in Madrid united politicians from all camps, ultimately helping to maintain democracy (Levitsky and Ziblatt, 2023).
Colombia	2010	President Alvaro Uribe sought to extend his presidency beyond constitutional limitations. In 2004, Congress amended the constitution to permit his re-election. It later emerged that this amendment had been facilitated through bribes, illegal wiretapping of the Supreme Court, and intimidation of journalists. In 2010, Uribe tried to secure a third term in office (Ginsburg and Huq, 2018).	The Constitutional Court rejected the amendment to allow a third term, arguing that it would have given the president excessive authority through selecting key judiciary figures (e.g., attorney general, Supreme Court members, chief prosecutor). Judges were committed to upholding the constitution (Ginsburg and Huq, 2018).

(*Continued*)

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Table 1. (*Continued.*)

Country	Year	Context	Outcome
Sri Lanka	2015	Mahinda Rajapaksa's 2005 election win allegedly involved voter suppression through a deal with the Tamil Tigers, an armed separatist group. Once in office, he aggressively pursued the group, declaring victory over them a year before his re-election. Rajapaksa's presidency was characterised by widespread corruption, including nepotism, erosion of legal institutions, and attacks on journalists. In 2010, Rajapaksa amended the constitution to allow a third term (Ginsburg and Huq, 2018).	When Rajapaksa called snap elections in 2015, his former health minister, Maithripala Sirisena, ran against him, winning the support of a coalition consisting of almost 50 parties. Rajapaksa's attempts to annul the election by declaring a state of emergency were thwarted by security and judicial leaders. Upon entering office, Sirisena began to undo the measures taken by Rajapaksa to weaken democracy (Ginsburg and Huq, 2018).
South Korea	2017	Democratic erosion began in 2008 e.g., through undermining press freedom, electoral rights and freedom of academic and cultural expression. In 2010, a scandal revealed illegal government surveillance of citizens, particularly journalists. In the run-up to the 2012 elections, the National Intelligence Service covertly posted online comments favouring the presidential candidate of the ruling party. Park Geun-hye was elected but was later implicated in a corruption scandal involving abuse of power, misuse of state funds and pressuring journalists. Further, artists critical of the government were excluded from government support programmes (Laebens and Lührmann, 2021).	There were no electoral consequences in 2012 as the public was unaware of the violations. The president's involvement in corruption was revealed in 2016, sparking protests and calls for her removal. This public outcry was the catalyst forcing parliament to impeach Park Geun-hye. In 2018, she was sentenced to prison on corruption charges (Laebens and Lührmann, 2021).
Chile	1973	Chile is a special case, as it primarily illustrates the sudden decline of democratic governance. Before Salvador Allende became president with a narrow plurality in September 1970, Chile's public policy was largely a result of bargaining between the governing party and the opposition. Allende deviated from this practice, involving third parties like courts and the army to resolve conflicts. This change, coupled with economic decline and widespread protests against Allende's "Chilean Road to Socialism" in 1972 (Navia and Osorio, 2019), led the opposition to pass a resolution in August 1973 accusing the government of constitutional violations, effectively undermining its legitimacy (for a detailed account, see Goldberg, 1975; Steenland, 1974).	In response to the protests against his policies, Allende invited military officers to join his cabinet in November 1972. While the political elite did not unequivocally support this move, around 70% of citizens saw it as positive for democracy (Herrera and Morales, 2023). However, since at least November 1972, elements within the military had started planning to oust the President, driven in part by growing anti-Marxist sentiments in the armed forces (Goldberg, 1975). On 11 September 1973, General Augusto Pinochet seized power and established a military dictatorship, which lasted until 1988, when the transition back to democracy began.

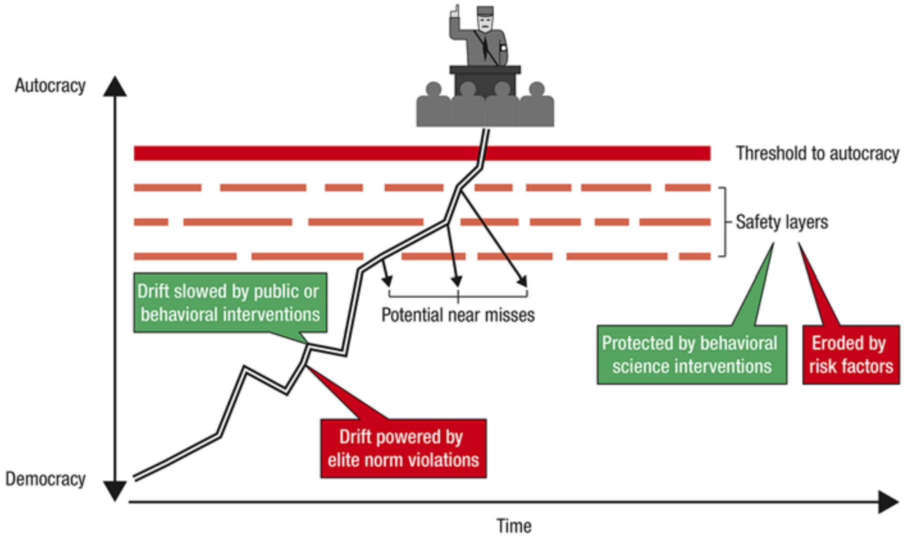
Autocracy
Threshold to autocracy
Safety layers
Drift slowed by public or behavioral interventions
Potential near misses
Protected by behavioral science interventions
Eroded by risk factors
Drift powered by elite norm violations
Democracy
Time

Figure 1. Illustration of the drift-to-danger model applied to democratic backsliding. The solid black line represents a gradual drift toward autocracy. Elite norm violations are a principal driver of this drift and can be opposed by behavioural countermeasures. The threshold to autocracy (solid red horizontal line) is protected by a number of safety layers (thin red lines) that can be undermined by risk factors and strengthened by behavioural science interventions. If at least one safety layer holds, making it possible to reverse the drift, a near miss occurs.

alone may be insufficient. Gaps and ambiguities in any constitution, no matter how well designed, leave room for interpretation and exploitation (Levitsky and Ziblatt, 2018). In a stable democracy, societal and political norms fill these gaps, ensuring the smooth operation of the system by governing elite and party behaviours and their interaction with the public. Political norms act as crucial, but often unspoken, safety layers that can arrest a drift to danger. However, they can become risk factors once eroded (e.g., if mainstream parties renege on the agreement not to form coalitions with extremist parties), changed (e.g., if violations become 'normal') or ignored (e.g., even if some political actors and institutions assume them to be still operative).

Given this critical role of elite norm violations in democratic near misses, we identify two areas in which behavioural science insights can safeguard democracy. The first involves direct interaction with political elites: Is the public willing to tolerate norm violations or does it punish such violations? Can politicians' behaviour be shaped by pro-democratic interventions? The second involves the broader societal risk factors that may facilitate elite norm violations by eroding safety layers. Can those risk factors be mitigated by behavioural interventions? We examine both areas in turn.

Democratic norms, elite norm violations and behavioural science

Two particularly important democratic norms are mutual toleration and institutional forbearance (Levitsky and Ziblatt, 2018). The norm of mutual toleration states that each party accepts the other's right to compete for power and govern, as long as they adhere

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

to the democratic process. Rivals are not seen as existential threats and politicians are collectively willing to agree to disagree. However, the openness of democracies can be exploited by bad-faith actors such as extremist political organizations. Indeed, Hitler described the strategy behind the rise of National Socialism as using the democratic process to destroy democracy (Weber, 2022).

Institutional forbearance means exercising restraint in situations where actions are legal but against the spirit of the constitution. In the US, there is a 200-year-old tradition of the sitting president to nominate a Supreme Court replacement even in a presidential election year, symbolising cooperation between president and Senate (Kar and Mazzone, 2016). The Republican-led Senate broke this norm when it refused to hold hearings for President Obama's nominee.

Elite attacks against these norms pose a major threat to democratic stability. They can be especially damaging in segments of the public where support for democratic norms and emancipatory values is already low (Kromphardt and Salamone, 2021).

Elite norm violations can be amplified by the fact that norms are not static, but change over time, sometimes rapidly. Such changes can imperil democracy without any obvious breaches of rules or laws. For example, Bursztyn et al. found that the widespread social norm against overt expression of racism and xenophobia unravelled quickly after Donald Trump's election. Study participants evidently interpreted his victory as a sign of widespread, hitherto hidden, anti-immigrant sentiment and became more willing to express such views (Bursztyn *et al.*, 2020). Thus, elite norm violations can systematically change the normative power of social norms, effectively giving 'mainstream' endorsement to behaviours previously considered unacceptable. Over time, such violations can undermine the existing norm, making the violation the new normal.

Second, elite norm violation can be enabled by a public that is unwilling to punish transgressions. For example, voters in Colombia did not oppose Uribe's measures to expand presidential powers (Posada-Carbó, 2011). Similar developments can be witnessed in the US at present. Political polarization is such that behaviours previously considered unacceptable have become normalized (Mettler and Liebermann, 2020). During the 2016 US presidential campaign, empowered by a supportive base, Donald Trump invited a foreign adversary, Russia, to find and release emails from Hillary Clinton's private server (Parker and Sanger, 2016). He later instructed Attorney General Jeff Sessions to shut down an investigation into his campaign's ties to Russia. When Sessions refused, Trump fired him (Baker *et al.*, 2018). Republicans in Congress, with few exceptions, stood by Trump, even after his 2023 indictment for mishandling confidential documents and attempts to overturn the election. Two-thirds of Republican voters supported his renewed candidacy and were ready to vote for him regardless of whether he was convicted (Montanaro, 2023).

Accepting the results of fair and free elections is, of course, a crucial norm in a democracy. Evidence suggests that rhetoric undermining this principle by claiming widespread electoral fraud in the US reinforced such beliefs (Clayton *et al.*, 2021). Additionally, emotional responses to norm violations by out-group elites (e.g., anger among Democrats over Republican actions or vice versa) tend to decrease over time, suggesting a desensitization effect of repeated norm violations (Clayton *et al.*, 2021).

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Overall, research provides conflicting evidence about people's willingness to punish norm-violating elites. On the one hand, only a small fraction of US citizens put democratic principles above their partisan identification when voting (Graham and Svolik, 2020). Thus, violations often go unpunished. Similarly, there is evidence that misinformation spread by politicians rarely affects individuals' feelings towards them (Swire *et al.*, 2017; Swire-Thompson *et al.*, 2020). Swire-Thompson *et al.* (2020) concluded that: 'Liking a politician has the unfortunate side effect of blinding us to their falsehoods' (p. 31). In fact, aggrieved groups can see politicians' lies as a 'symbolic challenge' to an illegitimate establishment. Repeatedly spreading false information may also pay off for elites: Attempts to correct the falsehoods may become less effective due to people becoming habituated to the lies (T. Koch, 2017) and political opponents becoming desensitized (Clayton *et al.*, 2021). Norm violations can also set examples of seemingly acceptable behaviour that partisans may then adopt (Bicchieri *et al.*, 2022).

On the other hand, there is some evidence that politicians are sensitive to the potential costs of having statements publicly corrected (e.g., by fact checkers). An experimental study found that the threat of reputational damage reduced the likelihood that US lawmakers would make inaccurate statements (Nyhan and Reifler, 2015). Specifically, Nyhan and Reifler randomly assigned state legislators to a treatment or control condition ahead of state elections. Legislators in the treatment condition received a letter reminding them that their public statements were subject to fact checking and that false statements carried a reputational cost. During the campaign, legislators in the treatment condition were found to be generally more accurate than their counterparts in the control condition. These findings could not be replicated in a more recent study (Ma *et al.*, 2023), however, perhaps because of the numbing effects of the post-truth world ushered in with the election of Donald Trump (Lewandowsky *et al.*, 2017).

Tsipursky and colleagues tested an intervention for politicians aimed at raising the benefits of committing to the truth and punishing the spread of misinformation. Politicians were invited to take a Pro-Truth Pledge consisting of three components (share, honour and encourage the truth), each designed to reduce misinformation sharing (Tsipursky *et al.*, 2018a, 2018b). It involved, for instance, sharing sources to allow others to verify information, defending others attacked for sharing factual information, and asking peers to stop using unreliable sources. The pledge seems to have had a beneficial effect, increasing signers' sharing of truthful information on Facebook four weeks after taking it (Tsipursky *et al.*, 2018b). These encouraging results are consistent with the finding that both voters and donors prefer candidates with pro-democratic positions (Carey *et al.*, 2022).

Risk factors enabling elite norm violation

Elite norm violations do not take place in a vacuum. They can be enabled or amplified by risk factors that erode democratic safety layers (Figure 1). Table 2 presents a selection of risk factors relevant from a behavioural sciences perspective that emerged from our analysis of near misses and the drift-to-danger model, categorized according to the five V-Dem core components of democracy. Before turning to the V-Dem core

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Table 2. Factors that undermine democracy by the five V-Dem components

Component	Example	Definition	Threat
Liberal	Right-wing populism	Populism can be understood as an ideology that divides society into "the pure people" and "the corrupt elite" (Rovira Kaltwasser, 2017, p. 491), and expects politicians to implement the "general will of the people" (for a review, see Kaltwasser, 2012; Mudde, 2017). Although politics "for the people" can be beneficial for society, the characteristics of right-wing populism – such as vilification of marginalized groups and impatience with deliberation – can have adverse consequences for democracy (Mansbridge and Macedo, 2019).	Right-wing populism is frequently at odds with liberal democracy, as it emphasises popular sovereignty and majority rule, each of which is said to serve "the pure people" instead of "corrupt elites", thereby potentially undermining existing checks and balances (Mudde and Rovira Kaltwasser, 2018; Rovira Kaltwasser, 2017). In particular, right-wing populism poses a threat to democratic accountability by considering courts and the judiciary as impediments to rulers' ability to exercise the presumed will of the people (Aytaç *et al.*, 2021).
Electoral	Voter disenchantment	Refers to disengagement from political life – for example, by deciding not to vote. C. M. Koch et al. (2023) have argued that disengagement can also result in people voting for populist parties as a way to express their dissatisfaction with the current political system.	As political disengagement is more frequent in socioeconomically disadvantaged populations, elections further increase divides between socioeconomic groups – and thus economic inequality and feelings of dissatisfaction with politics (Gallego, 2010; C. M. Koch *et al.*, 2023; Schaub, 2021).
Participatory	Misinformation	Incorrect information, frequently disseminated with the intent to mislead (Ecker *et al.*, 2022), threatens the epistemic potential of democratic decision-making as it hinders access to accurate information (Brown, 2018). For instance, voting decisions based on inaccurate information can have numerous harmful consequences at a societal level (e.g., Lewandowsky *et al.*, 2017; Pantazi *et al.*, 2021).	Misinformation in social media increases political cynicism in non-partisans. This can create partisan echo chambers and lead to wider political disenchantment, with detrimental effects on democracy. While the effect of misinformation exposure is more pronounced in non-partisans, it is independent of whether people actually believe the misinformation they: Mere exposure to it increases cynicism (Lee and Jones-Jang, 2024).

(*Continued*)

Table 2. (*Continued.*)

Component	Example	Definition	Threat
Participatory	Conspiracy theories	Are understood as an attempt to make sense of events by proposing a secret plot between powerful individuals or organizations that aims to accomplish sinister ends through continuing deception of the public (Douglas and Sutton, 2008; Goertzel, 1994). Conspiracy theories are often vague, are not inherently true or false, and tolerate internal contradictions (Wood *et al.*, 2012).	Beliefs in conspiracy theories are associated with increased political cynicism (Swami, 2012), decreased interpersonal trust (Frenken and Imhoff, 2023) and trust in authorities (Goertzel, 1994), decreased political participation (Ardèvol-Abreu *et al.*, 2020), and increased willingness to harm the state and its representatives (Imhoff *et al.*, 2021).
Deliberative	Social and political polarization	Describes a situation where significant differences in opinions result in clashes of views between segments of the public. In the context of political polarization, society splits into opposing camps along partisan lines. The divisions go beyond political debates and extend to social relationships and how people interact with one another on a daily basis (Boese *et al.*, 2021).	In highly polarized societies, each side considers only their views to be correct and sees the other political camp(s) as threats to the nation. This can ultimately result in a willingness to use all means possible to defend their side's interests (Somer *et al.*, 2021).

components, we consider a more domain-general insight about risk perception and behaviour.

A neglected but potentially important risk factor is a lack of personal experience with the implications of autocratic rule. Personal experience with a risk – be it a macroeconomic shock such as the Great Depression (Malmendier and Nagel, 2011), a period of hyper-inflation (Malmendier and Nagel, 2016), a catastrophic natural hazard such an earthquake (Wachinger *et al.*, 2013) or a global pandemic such as COVID-19 (Dryhurst *et al.*, 2020) – has been found to influence people's perception of risk more generally. For instance, a recent analysis of more than 15,000 people in Germany found that those who had contracted coronavirus consistently rated the likelihood of infection higher than those without such experience. Media coverage also influenced risk judgements, but to a lesser extent (Schulte-Mecklenbeck *et al.*, 2024). Similarly, in an international survey of 24 countries, personal experience of global warming predicted the willingness to endorse specific mitigation actions (Broomell *et al.*, 2015).

People tend to learn about risks either from personal experience or from description (Hertwig and Wulff, 2022). Ample evidence from psychology and economics indicates that people's propensity to take risks in the future depends on lessons taught by past experiences. For example, one of the probabilistic outcomes of unprotected sex is contracting a sexually transmitted infection (STI). When base rates of STIs are low, as is typically the case, not contracting an STI is the likelier outcome of having unprotected sex. At the current rates of disease in Europe, a person would need to have sex with at least 15 (randomly selected) people to reach a 50% probability of encountering a partner with syphilis, chlamydia or gonorrhea (in 2016; see Ciranka and Hertwig, 2023). Therefore, most people (especially adolescents) who have unprotected sex do not contract STIs, with one likely consequence being that many do not learn to protect themselves.

In contrast, people who do experience rare events with negative outcomes such as contracting an STI are more risk averse. This 'hot-stove effect' (Denrell, 2007) gives rise to a powerful behavioural bias that prevents them from repeating the behaviour associated with the adverse outcome – a cat that has sat on a hot stove lid once is unlikely to do so again.

These behavioural regularities have implications for the efficacy of warnings about risks in general and democratic decline in particular. Democracies may warn their citizens about the potential consequences of behaviours such as elite norm violations, but these warnings compete with the everyday experience of a still-functioning democracy (Hertwig and Wulff, 2022) and may thus go unheeded. Indeed, safe experiences can undermine the effectiveness of warnings in various domains (see Barron *et al.*, 2008). This dynamic may also help explain why early warnings about the risks of climate change were relatively ineffective (see Hertwig and Wulff, 2022; E. U. Weber, 2006; E. U. Weber and Stern, 2011).

This is especially problematic when the probability of a catastrophic event is low but increases over time. Hertwig and Wulff (2022) used the example of Mount Vesuvius to illustrate this dynamic – the volcano described as 'Europe's ticking time bomb' (Barnes, 2011, p. 140). Around 600,000 people live in the Red Zone that would be at highest risk in the event of an eruption. Yet neither expert warnings (e.g., Mastrolorenzo *et al.*, 2006) nor financial incentives (e.g., Barberi *et al.*, 2008) have persuaded them to leave the danger zone. Hertwig and Wulff (2022) argued that this can be attributed to the residents' long-lasting 'all-clear experience' (p. 641): The last violent eruption occurred in 1944. People who have never experienced an eruption behave as if they *underweight* the probability of one occurring.

Experience is a powerful teacher of risks, causing people to both overweight risk (once experienced) and underweight it (after a sequence of safe experiences). Simulations – e.g., of earthquakes, investment risks and old age – can provide tangible demonstrations of the impact of potential risks without exposing individuals to actual harm (Hertwig and Wulff, 2022). Available simulations like the Swiss Seismological Service's Earthquake Games[3] or role-playing simulations on transitions to democracy (Jiménez, 2015) or civil–military relations during mass uprisings (Harkness and DeVore, 2021) can provide the blueprint for interventions that simulate life and risks in an autocracy. Citizens who 'experience' the risk of democratic decline may be better calibrated to address its threats and prospective losses. Interventions could take the form of online games, virtual reality simulations or interactive museum exhibits. For example, the House of Terror in Budapest[4] and the Museum of Occupations and Freedom Fights in Vilnius[5] illustrate the brutal realities of living under an oppressive regime.

[3] See http://www.seismo.ethz.ch/en/knowledge/miscellaneous/earthquake-games/

[4] https://www.terrorhaza.hu/en/

[5] http://www.genocid.lt/muziejus/en/

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

There is increasing evidence that such simulations are more effective than description-based interventions. A study on COVID-19 vaccination found that people exposed to an interactive risk–ratio simulation were more likely to get vaccinated and tend to have a better understanding of the benefit-to-harm ratio (Wegwarth *et al.*, 2023) than people presented with the same information in a conventional text-based format. It is difficult to directly target elites with interventions; using insights from behavioural science to make citizens more sensitive to the risks of democratic backsliding seems a promising approach to bolster democratic resilience.

The liberal, participatory and deliberative components of the V-Dem taxonomy appear especially vulnerable to erosion because – relative to the electoral and egalitarian components – they rely more on norm commitment than on legislation or regulation. We next show how the behavioural sciences can inform measures to counter the risks identified in Table 2, thus reinforcing safety layers against norm violations by elites. The allocation of risks to the V-Dem taxonomy is not clear cut, as some risks such as misinformation can affect more than one aspect of democracy, for instance liberal and participatory. Yet, for the sake of analytical clarity, we categorize each risk under the aspect of democracy it most directly impacts. This approach allows for a more precise identification of how specific threats undermine democratic functions and facilitates targeted responses.

For example, misinformation primarily threatens the participatory aspect of democracy by distorting public opinion and voter behaviour, which undermines the legitimacy of electoral processes and the responsiveness of political representatives. While it also has implications for liberal democracy by potentially eroding trust in institutions and the rule of law, its most direct and immediate impact is on the quality and inclusiveness of public participation.

Right-wing extremism under the banner of populism

Addressing right-wing extremism cloaked in populism from a behavioural perspective is easier said than done: Its proponents appeal to emotion – in particular anger and outrage (Gerbaudo *et al.*, 2023) – and use rhetorical strategies that are difficult to counter, while sidestepping policy debate. According to Kayam (2023, p. 277), three of Trump's main strategies are 'make it simple, make it negative, and make it "Twitty"', which means using ad populum and ad hominem appeals. This makes conventional argumentation difficult, if not impossible. Less conventional approaches include using satire to shine a light on the shortcomings of right-wing populist policies (e.g., nationalist solutions to global problems). Using a large-scale survey methodology, Boukes and Hameleers (2020) examined how the Dutch satirical show, *Zondag met Lubach*, influenced people's willingness to vote for a populist party after the show targeted its lack of identifiable policy positions. The results revealed that the show reduced support for both the party and its leader, and that the decline was particularly strong among citizens inclined to vote for populist parties. Satire has been shown to work in other contexts as well, such as fact checking (Boukes and Hameleers, 2023): Satirical corrections reduce belief in misinformation, but also lead to greater polarization than plain fact-based corrections (Boukes and Hameleers, 2023). Humour and satire thus constitute an effective tool to counter populism; however, they should not be deployed without great care.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

According to Rovira Kaltwasser (2017), fighting right-wing populism by depicting its proponents as villains and its opponents as heroes is ineffective. Such framing fosters polarization and can create a populism vs anti-populism divide that may unwittingly align precisely with the division that right-wing extremists seek to create. Instead, reminding people of the value of deliberation and group norms may help close the social divide (e.g., Cialdini and Goldstein, 2004; Mansbridge and Macedo, 2019; Kendall-Taylor and Nietsche, 2020; Pantazi *et al.*, 2022).

Another issue of concern for the behavioural sciences is the post-truth communication frequently employed by right-wing populists. Post-truth phenomena such as misinformation and conspiracy theories exploit existing societal chasms as well as individual beliefs about the government and political elites (Waisbord, 2018; Uscinski *et al.*, 2022). We discuss interventions to counter mis- and disinformation in the next section.

Once right-wing populists are in government, interventions become even more difficult and can backfire, as discussed by Schlipphak and Treib (2017) using the cases of Austria in the early 2000s and Hungary under Victor Orban. In both cases, EU interventions (e.g., sanctions) did not reduce public support for the government; on the contrary, support increased over time. Schlipphak and Treib (2017) argued that this effect can be attributed to successful blame deflection. The politicians framed the EU's actions as an illegitimate intervention from 'outside', creating an 'us' vs 'them' juxtaposition, and thus de-legitimizing the interventions. A better approach would be for the EU to build a coalition with domestic actors, intervening only when oppressed domestic groups ask for help. Instead targeting an entire country, sanctions should focus on actual offenders, such as political elites and high-ranking officials. Institutionally, an independent supervisory body could be established to conduct 'open, independent and impartial' (Schlipphak and Treib, 2017, p. 362) assessments of the state of democracy. We propose that behavioural scientists could support this body by developing guidelines and designing evidence-based measures for cases in which legal interventions are insufficient – for example, when norms are threatened.

Misinformation and conspiracy theories

The proliferation of false or misleading information and conspiracy theories through media and digital platforms – especially social media – is a global phenomenon that can have detrimental effects on public welfare (e.g., health), responses to global crises (e.g., pandemics, climate change) and the stability of democracies (Lewandowsky, Smillie, *et al.*, 2020b; Lorenz-Spreen *et al.*, 2022). It is influenced by media conglomerates and online platforms, but also by individual and collective behaviours (Lazer *et al.*, 2018; Lewandowsky, *et al.*, 2017). A range of behavioural sciences-based interventions have been proposed to target behaviours in the digital world. Recent reports by Ecker *et al.* (2022), Kozyreva *et al.* (2024) and van der Linden *et al.* (2023) have examined the evidence for these interventions. They can be divided into individual-level interventions such as inoculation (which seeks to build people's competence at discerning manipulative information), media-literacy tips, warnings and fact-check labels, debunking and accuracy and social norm nudges. 'Friction' can be introduced to slow information processing and encourage more careful analysis. Users can be taught to use lateral reading strategies, that is to leave the initial source and

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

open new tabs to search for more information about the person or organization behind a website or social media post and the claims made. There is considerable evidence that those techniques work even in the wild. For example, Roozenbeek *et al.* (2022) showed that YouTube users benefited from brief information videos that boosted their ability to distinguish manipulative information from high-quality information.

Although helpful, such individual-focused interventions are insufficient to address the scale of the misinformation problem. Systemic interventions are also needed for online content (e.g., regulatory legislation of platforms), algorithms (e.g., automated tools for content moderation) and business models (e.g., supporting reliable news media).

Issues surrounding content moderation, including the balance between safeguarding freedom of expression and minimizing risks to public health, have polarized debate, particularly in the US (see the ongoing legal dispute about the First Amendment and its impact on social media companies; Zakrzewski, 2023). In perhaps the first behavioural science study of people's preferences around content moderation, Kozyreva *et al.* (2021) found that, under specific circumstances, a majority of US respondents would remove misinformation-based social media posts on election denial, anti-vaccination, Holocaust denial and climate change. Respondents were more likely to remove posts that contained potentially dangerous misinformation or if the information had been circulated multiple times by the person. They were more reluctant to suspend accounts than to remove posts. In general, however, the US public does not categorically oppose content moderation of harmful content.

Importantly, the cognitive and behavioural sciences have already contributed to EU regulations (Kozyreva, *et al.*, 2023) by, for instance, designing and testing interventions, informing the design of regulations and revealing and documenting people's preferences (e.g., for content moderation).

Voter disenchantment

Voter apathy or lack of engagement with elections presents a problem in many liberal democracies. Communication campaigns are often seen as a relatively simple solution; however, the evidence for their effectiveness is limited (Haenschen, 2023). Behavioural science can contribute by helping to build a culture where participation and active choice is valued. Even subtle changes in wording can be enough to increase people's motivation to vote. For example, framing voting as a facet of personal identity rather than a behaviour – using phrases such as 'being a voter' rather than 'voting' – increases people's likelihood to vote (Bryan *et al.*, 2011).

Other interventions discussed to increase voter turnout include creating a pre-commitment device in form of a registry for people who commit to vote, with small penalties being imposed for failing to vote (Pedersen *et al.*, 2023). In contexts with mandatory voting or voter registration (e.g., Australia, Belgium), highlighting the role of the negative monetary effects of non-voting can be an effective nudge (Kölle *et al.*, 2017). However, these interventions are difficult to implement, often only suitable in certain contexts, and raise ethical questions about the validity of consent if citizens have to opt out from interventions such as the pre-commitment system.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Prompting people to consciously consider when, where and how they will cast their vote can also help to increase voter turnout (Gollwitzer, 1999; Gollwitzer and Sheeran, 2006). Nickerson and Rogers (2010) reported a substantial increase of voter turnout of 9.1 percentage points in single-eligible voter households. For households with two or more voters, however, the intervention reduced turnout by 1.5 percentage points, probably because these households would discuss voting and make a plan anyway. Anderson *et al.* (2018) found that, in addition to making a voting plan, individuals benefit from relevant, clear information material about the election. Furthermore, the salience of norms that represent a group's collective values influences voting intentions, regardless of how close or significant a group (e.g., friends, family) is to the individual (French Bourgeois and de la Sablonnière, 2023).

Polarization

Affective polarization is strongly correlated with democratic backsliding. Even within democracies, it can reduce accountability, freedom, deliberation and rights (Svolik, 2019; Orhan, 2022) and increase the chance of elite norm violations. Interventions to mitigate polarization can target information processing, beliefs or social relations. Information processing interventions seek to change individual reasoning patterns that guide the interpretation of information. Examples include addressing cognitive rigidity, which is associated with intergroup hostility and ideological extremism (Zmigrod, 2020), and intra-individual conflict, which can lead to paradoxical (Bar-Tal *et al.*, 2021) or counterfactual thinking (Epstude and Roese, 2008).

Findings on the distorting effects of in- and out-group perceptions suggest that polarization can be mitigated by addressing specific beliefs (Mackie, 1986). In politics, polarization results in partisan animosity, defined as 'negative thoughts, feelings or behaviours towards a political outgroup' (Hartman *et al.*, 2022, p. 1194). In a large-scale study, Voelkel *et al.* (2023) tested 25 interventions designed to reduce polarization. Only six showed lasting results. These interventions took various approaches: highlighting that most Democrats and Republicans reject polarization; showing that positive social connection across party lines is possible despite political disagreement; making national identity salient; correcting misperceptions about outpartisans' support for undemocratic actions and their tendency to dehumanize the other party; and creating sympathetic personal narratives.

Polarization interventions targeting social relations are informed by evidence on intergroup contact (Pettigrew, 1998; Pettigrew and Tropp, 2006). They use contact with out-group members to humanize outpartisans and create a more realistic image of their thinking and behaviour. Interventions include improving people's dialogue skills, enabling a constructive debate despite political differences and facilitating positive contact between partisans – for example, by highlighting what both groups have in common (Hartman *et al.*, 2022).

Recent work by Voelkel *et al.* (2023) and Broockman *et al.* (2022) cautions that while depolarization interventions reliably reduce affective polarization, they do not appear to be successful in reducing anti-democratic attitudes, such as support for partisan violence. It appears that once a society becomes so divided that political identity overtakes

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

social identity, members of the other political camp may be perceived as a threat to the nation, thus legitimizing all means possible to defend one's own interests.

Limitations and expansion

We restricted our analysis to factors falling within the realm of the behavioural sciences (Table 2) that offer scope for countermeasures. However, numerous other systemic factors may also facilitate democratic backsliding, such as economic inequality (Siripurapu, 2022) and the design of the online information environment (Lewandowsky *et al.*, 2020b). In the future, climate change may also impact the stability of democratic systems, as authoritarian policies to address the climate crisis become more likely in the most affected areas (Mittiga, 2022). Although the importance of such systemic factors must not be underestimated, they do not negate the role of the behavioural sciences.

Conclusion: behavioural science against democratic backsliding

Near misses in sociotechnical systems are adverse events that could have caused damage to people and/or property but were prevented by means of safety layers (Jones *et al.*, 1999). In democracies, near misses are understood as situations in which political systems either managed to withstand a drift towards autocracy (Figure 1) or briefly became autocratic before returning to democratic governance (Ginsburg and Huq, 2018). The present analysis of democratic near misses identified factors that have successfully prevented democratic decline in the past (as it is common in the field of safety science; see Gnoni *et al.*, 2022) and can therefore inform future interventions to increase democratic stability.

Inspired by Rasmussen (1997), we adapted the drift-to-danger model to democratic near misses. Within this framework, backsliding is enabled by the confluence of various factors: Populism, misinformation and polarization collude in eroding the safety layers that would otherwise prevent political elites from violating the norms essential to keep a democracy functioning. Elite norm violations – and the public response to them – are at the heart of all historical near misses we analysed. Indeed, this is the first crucial insight from our analysis: Democratic backsliding is closely tied to elite norm violations, but the role of the public in condoning or opposing those violations is far more variable.

The second insight concerns the non-linearity of the drift underlying democratic backsliding. Some violations can be absorbed, but democracy's breaking point might at any point be just one safety layer away. This non-linearity underscores the importance of protecting all democratic norms and calling out all violations, as the downstream effects cannot be predicted. The drift-to-danger model helps to understand how gradual declines in democracy can suddenly turn catastrophic and irreversible. While the model is not testable in itself, it suggests hypotheses – for example, that the failure of a single safeguard does not critically influence the overall trajectory of backsliding or that the exact tipping point is difficult to predict.

Although our analysis of near misses was limited to past cases, we did briefly explore the implications of that analysis within the current situation in the UK and US. The

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

US system of checks and balances has served as a model for many other democracies. Yet, its safety layers seem to be eroding in several areas. Society is highly polarized across all levels, from political leaders to citizens. Elite norm violations have become more frequent, culminating in Donald Trump disputing the legitimacy of the 2020 presidential elections.

It is, however, important to emphasize that cases such as the January 6 insurrection in the US and the prorogation of parliament in the UK represent individual episodes within a sequence of events that can ultimately contribute to democratic decline. They do not singularly constitute a near miss (from the perspective of the drift-to-danger model, no single event should cause a near miss). Yet, while the outcomes of antidemocratic actions are unpredictable, examining the intentions of the actors involved can reveal their willingness to undermine democracy in the absence of checks and balances. The intentions of political elites matter, as Levitsky and Ziblatt (2023) have shown for what they call semi-loyal democrats. Members of this group do not actively harm democracy but fail to defend it in times of polarization and crisis. By turning a blind eye to the autocratic acts of ideological allies, they enable antidemocratic extremists. History provides several examples of this mechanism. In 1934, violent rioters tried to occupy the French parliament, leading to the resignation of the centrist prime minister. Many conservative politicians did not condemn the insurrection; some even praised the rioters as 'heroes and patriots'. This lack of condemnation allowed the insurrectionists' ideas to enter mainstream conservative thought, including a preference for Hitler over the socialist prime minister – although French conservatives were historically anti-German (Levitsky and Ziblatt, 2023). Therefore, even in the absence of clear and predictable tipping points, investigating established backsliding patterns, especially in political elites' intentions and behaviour, can help to understand and prevent democratic breakdown. As Svolik *et al.* (2023) put it: 'To diagnose the vulnerabilities of contemporary democracies, we must therefore ask: When faced with a choice between democracy and partisan loyalty, policy priorities, or ideological dogmas, who will put democracy first?' (p. 6).

Our framework suggests various avenues for future research. One area involves exploring how to counteract complacency toward elite norm violations – for example, by testing interventions such as autocracy simulations and their effectiveness in raising awareness of the risk of democratic decline. Another is to examine the effects of nostalgic feelings for autocratic regimes such as the German Democratic Republic, especially in times of political or economic crisis. In these situations, nostalgia could bias people in favour of the autocratic regime, making them more critical of democracy (see for instance Neundorf *et al.*, 2020). Further research is necessary to investigate how nostalgia impacts democratic backsliding and to identify mitigating strategies.

To date, research into how the behavioural sciences can strengthen democracy and prevent backsliding is scarce. Druckman (2024)'s recent study on democratic backsliding from a psychological perspective also stresses the need to look beyond the structural factors frequently discussed in political science. Yet while Druckman (2024) also identifies elites as important actors in the backsliding process – along with social movements, interest groups and campaign organizations – his framework does not address either the process of backsliding itself or potential interventions to stop it.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Future research could therefore take a more process-oriented perspective, like the drift-to-danger model, and emphasize the role of non-elite actors in facilitating norm violations, a topic only briefly explored in this article.

Behavioural interventions are just one tool in the toolbox of forces working together to stop democratic backsliding. Most of these behavioural interventions are aimed at the public, aiming to increase awareness of the risks of democratic decline and to boost resilience to manipulation and false information (see also Herzog and Hertwig, 2025). When it comes to elites, however, these interventions seem largely ineffective in influencing those willing to push the norms of acceptable behaviour. Yet, as documented by our historical analysis, some political elites do stand up for democracy. For example, the UK Supreme Court blocked Boris Johnson's attempt to prorogue parliament, and the Georgia's Republican Secretary of State in Georgia, Brad Raffensperger, resisted Donald Trump's pressure 'to find 11,780 votes' (Shear and Saul, 2021). These examples highlight that elite resistance and push back can interrupt, at the least for the moment, democratic backsliding.

Funding statement. SL, KH, and CA are supported by an Advanced Grant from the European Research Council (101020961 PRODEMINFO) to SL. SL and RH received support from the Volkswagen Foundation (large grant 'Reclaiming individual autonomy and democratic discourse online: How to rebalance human and algorithmic decision making') and from the European Union's Horizon 2020 grant 101094752 (SoMe4Dem). SL also received funding from UK Research and Innovation through the EU Horizon replacement funding grant number 10049415 and was supported by a Fellowship from the Academy for International Affairs NRW during part of this project. The authors thank Susannah Goss very much for editing the paper.

Competing interests. The authors declare no competing interests.

References

Anderson, C. D., P. J. Loewen and R. M. McGregor (2018), Implementation intentions, information, and voter turnout: an experimental study, *Political Psychology*, **39**(5): 1089–1103.

Ardèvol-Abreu, A., H. Gil de Zúñiga and E. Gámez (2020), The influence of conspiracy beliefs on conventional and unconventional forms of political participation: the mediating role of political efficacy, *British Journal of Social Psychology*, **59**(2): 549–569.

Aytaç, S. E., A. Çarkoğlu and E. Elçi (2021), Partisanship, elite messages, and support for populism in power, *European Political Science Review*, **13**(1): 23–39.

Baker, P., K. Benner and M. D. Shear (2018, November 7), Jeff Sessions is forced out as attorney general as Trump installs loyalist. *The New York Times*. https://www.nytimes.com/2018/11/07/us/politics/sessions-resigns.html

Barberi, F., M. S. Davis, R. Isaia, R. Nave and T. Ricci (2008), Volcanic risk perception in the Vesuvius population, *Journal of Volcanology and Geothermal Research*, **172**(3–4): 244–258.

Barnes, K. (2011), Volcanology: Europe's ticking time bomb, *Nature*, **473**(7346): 140–141.

Barron, G., S. Leider and J. Stack (2008), The effect of safe experience on a warnings' impact: sex, drugs, and rock-n-roll, *Organizational Behavior and Human Decision Processes*, **106**(2): 125–142.

Bar-Tal, D., B. Hameiri and E. Halperin (2021), 'Paradoxical thinking as a paradigm of attitude change in the context of intractable conflict', in B. Gawronski (ed), *Advances in Experimental Social Psychology*, volume 63, Cambridge, MA: Academic Press, 129–187.

Bermeo, N. (2016), On democratic backsliding, *Journal of Democracy*, **27**(1): 5–19.

Bicchieri, C., E. Dimant, S. Gächter and D. Nosenzo (2022), Social proximity and the erosion of norm compliance, *Games and Economic Behavior*, **132**: 59–72.

Boese, V. A., A. B. Edgell, S. Hellmeier, S. F. Maerz and S. I. Lindberg (2021), How democracies prevail: democratic resilience as a two-stage process, *Democratization*, **28**(5): 885–907.

Boukes, M. and M. Hameleers (2020), Shattering populists' rhetoric with satire at elections times: the effect of humorously holding populists accountable for their lack of solutions, *Journal of Communication*, **70**(4): 574–597.

Boukes, M. and M. Hameleers (2023), Fighting lies with facts or humor: comparing the effectiveness of satirical and regular fact-checks in response to misinformation and disinformation, *Communication Monographs*, **90**(1): 69–91.

Broockman, D. E., J. L. Kalla and S. J. Westwood (2022), Does Affective Polarization Undermine Democratic Norms or Accountability? Maybe Not, *American Journal of Political Science*, **67**(3): 808–828.

Broomell, S. B., D. V. Budescu and H-H. Por (2015), Personal experience with climate change predicts intentions to act, *Global Environmental Change*, **32**: 67–73.

Brown, É. (2018), Propaganda, misinformation, and the epistemic value of democracy, *Critical Review*, **30**(3–4): 194–218.

Bryan, C. J., G. M. Walton, T. Rogers and C. S. Dweck (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.

Bursztyn, L., G. Egorov and S. Fiorin (2020), From extreme to mainstream: the erosion of social norms, *American Economic Review*, **110**(11): 3522–3548.

Carey, J., K. Clayton, G. Helmke, B. Nyhan, M. Sanders and S. 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): 230–245.

Cialdini, R. B. and N. J. Goldstein (2004), Social influence: compliance and conformity, *Annual Review of Psychology*, **55**(1): 591–621.

Ciranka, S. and R. Hertwig (2023), Environmental statistics and experience shape risk-taking across adolescence, *Trends in Cognitive Sciences*, **27**(12): 1123–1134.

Clayton, K., N. T. Davis, B. Nyhan, E. Porter, T. J. Ryan and T. J. Wood (2021), Elite rhetoric can undermine democratic norms, *Proceedings of the National Academy of Sciences of the United States of America*, **118**(23): 1–6.

Cook, R. and J. Rasmussen (2005), "Going solid": a model of system dynamics and consequences for patient safety, *Quality and Safety in Health Care*, **14**(2): 130–134.

Cullen, S. M. (1993), Political violence: the case of the British Union of Fascists, *Journal of Contemporary History*, **28**(2): 245–267.

Denrell, J. (2007), Adaptive learning and risk taking, *Psychological Review*, **114**(1): 177–187.

Douglas, K. M. and R. M. Sutton (2008), The hidden impact of conspiracy theories: perceived and actual influence of theories surrounding the death of Princess Diana, *Journal of Social Psychology*, **148**(2): 210–222.

Druckman, J. N. (2024), How to study democratic backsliding, *Political Psychology*, **45**(S1): 3–42.

Dryhurst, S., C. R. Schneider, J. Kerr, A. L. J. Freeman, G. Recchia, A. M. van der Bles, D. Spiegelhalter and S. van der Linden (2020), Risk perceptions of COVID-19 around the world, *Journal of Risk Research*, **23**(7–8): 994–1006.

Ecker, U. K. H., S. Lewandowsky, J. Cook, P. Schmid, L. K. Fazio, N. Brashier, P. Kendeou, E. K. Vraga and M. A. Amazeen (2022), The psychological drivers of misinformation belief and its resistance to correction, *Nature Reviews Psychology*, **1**(1): 13–29.

Eliassi-Rad, T., H. Farrell, D. Garcia, S. Lewandowsky, P. Palacios, D. Ross, D. Sornette, K. Thébault and K. Wiesner (2020), What science can do for democracy: a complexity science approach, *Humanities and Social Sciences Communications*, **7**(1): 8–11.

Epstude, K. and N. J. Roese (2008), The functional theory of counterfactual thinking, *Personality and Social Psychology Review*, **12**(2): 168–192.

Ewing, K. and C. A. Gearty (2001), *The Rise and Fall of Facism*, Oxford: Oxford University Press

Freedom House. (2019), Freedom in the world 2019: democracy in retreat. *In Freedom House*. https://freedomhouse.org/sites/default/files/Feb2019_FH_FITW_2019_Report_ForWeb-compressed.pdf.

French Bourgeois, L. and R. de la Sablonnière (2023), Realigning individual behavior with societal values: the role of planning in injunctive-norm interventions aimed at increasing voter turnout, *Analyses of Social Issues and Public Policy*, **23**(1): 155–173.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Frenken, M. and R. Imhoff (2023), Don't trust anybody: conspiracy mentality and the detection of facial trustworthiness cues, *Applied Cognitive Psychology*, **37**(2): 256–265.

Gallego, A. (2010), Understanding unequal turnout: education and voting in comparative perspective, *Electoral Studies*, **29**(2): 239–248.

Gerbaudo, P., C. C. De Falco, G. Giorgi, S. Keeling, A. Murolo and F. Nunziata (2023), Angry posts mobilize: emotional communication and online mobilization in the Facebook pages of western European right-wing populist leaders, *Social Media + Society*, **9**(1).

Ginsburg, T. and A. Huq (2018), Democracy's near misses, *Journal of Democracy*, **29**(4): 16–30.

Gnoni, M. G., F. Tornese, A. Guglielmi, M. Pellicci, G. Campo and D. De Merich (2022), Near miss management systems in the industrial sector: a literature review, *Safety Science*, **150**: 105704.

Goertzel, T. (1994), Belief in conspiracy theories, *Political Psychology*, **15**(4): 731–742.

Goldberg, P. A. (1975), The politics of the Allende overthrow in Chile, *Political Science Quarterly*, **90**(1): 93–116. https://www.jstor.org/stable/2148700

Gollwitzer, P. M. (1999), Implementation intentions: strong effects of simple plans, *American Psychologist*, **54**(7): 493–503.

Gollwitzer, P. M. and P. Sheeran (2006), Implementation Intentions and goal achievement: a meta-analysis of effects and processes, *Advances in Experimental Social Psychology*, **38**(06): 69–119.

Graham, M. H. and M. W. Svolik (2020), Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States, *American Political Science Review*, **114**(2): 392–409.

Hadfield, A. (2019, August 28), Boris Johnson suspends parliament: what does it mean for Brexit and why are MPs so angry? *The Conversation*. https://theconversation.com/boris-johnson-suspends-parliament-what-does-it-mean-for-brexit-and-why-are-mps-so-angry-122574

Haenschen, K. (2023), The conditional effects of microtargeted Facebook advertisements on voter turnout, *Political Behavior*, **45**(4): 1661–1681.

Haggard, S. and R. Kaufman (2021), The anatomy of democratic backsliding, *Journal of Democracy*, **32**(4): 27–41.

Harkness, K. A. and M. R. DeVore (2021), Teaching the military and revolutions: simulating civil–military relations during mass uprisings, *PS: Political Science and Politics*, **54**(2): 315–320.

Hartman, R., W. Blakey, J. Womick, C. Bail, E. J. Finkel, H. Han, J. Sarrouf, J. Schroeder, P. Sheeran, J. J. Van Bavel, R. Willer and K. Gray (2022), Interventions to reduce partisan animosity, *Nature Human Behaviour*, **6**(9): 1194–1205.

Haslam, S. A., S. D. Reicher, H. P. Selvanathan, A. M. Gaffney, N. K. Steffens, D. Packer, J. J. Van Bavel, E. Ntontis, F. Neville, S. Vestergren, K. Jurstakova and M. J. Platow (2023), Examining the role of Donald Trump and his supporters in the 2021 assault on the U.S. Capitol: a dual-agency model of identity leadership and engaged followership, *The Leadership Quarterly*, **34**(2): 101622.

Helderman, R. S. (2022, February 9), All the ways Trump tried to overturn the election—and how it could happen again. *The Washington Post*. https://www.washingtonpost.com/politics/interactive/2022/election-overturn-plans/

Herrera, M. and M. Morales (2023), Public opinion, democracy, and the armed forces: Chile before the 1973 military coup, *Social and Education History*, **12**(2): 160–192.

Hertwig, R. and D. U. Wulff (2022), A description–experience framework of the psychology of risk, *Perspectives on Psychological Science*, **17**(3): 631–651.

Herzog, S. M. and R., Hertwig (2025). Boosting: Empowering citizens with behavioral science. *Annual Review of Psychology*, 76.

Imhoff, R., L. Dieterle and P. Lamberty (2021), Resolving the puzzle of conspiracy worldview and political activism: belief in secret plots decreases normative but increases nonnormative political engagement, *Social Psychological and Personality Science*, **12**(1): 71–79.

Jiménez, L. F. (2015), The dictatorship game: simulating a transition to democracy, *PS: Political Science and Politics*, **48**(02): 353–357.

Jones, S., C. Kirchsteiger and W. Bjerke (1999), The importance of near miss reporting to further improve safety performance, *Journal of Loss Prevention in the Process*, **12**(1): 59–67.

Kaltwasser, C. R. (2012), The ambivalence of populism: threat and corrective for democracy, *Democratization*, **19**(2): 184–208.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Kar, R. B. and J. Mazzone (2016), The Garland Affair: what history and the constitution really say about President Obama's powers to appoint a replacement for Justice Scalia, *New York University Law Review Online*, **91**: 53–114. https://nyulawreview.org/online-features/the-garland-affair-what-history-and-the-constitution-really-say-about-president-obamas-powers-to-appoint-a-replacement-for-justice-scalia/ [2 November 2024].

Kayam, O. (2023), Trump's Rhetorical Way to Presidency, A. Akandeed, *U.S. Democracy in Danger*, 277–292, Cham: Springer Nature Switzerland.

Kendall-Taylor, A. and C. Nietsche (2020), *Combating populism: a toolkit for liberal democratic actors*. https://www.cnas.org/publications/reports/combating-populism

Koch, C. M., C. Meléndez and C. Rovira Kaltwasser (2023), Mainstream voters, non-voters and populist voters: what sets them apart?, *Political Studies*, **71**(3): 893–913.

Koch, T. (2017), Again and again (and again): a repetition-frequency-model of persuasive communication, *Studies in Communication and Media*, **6**(3): 218–239.

Kölle, F., T. Lane, D. Nosenzo and C. Starmer (2017), *Nudging the electorate: what works and why?* http://hdl.handle.net/10419/200439www.econstor.eu

Kozyreva, A., P. Lorenz-Spreen, R. Hertwig, S. Lewandowsky and S. M. Herzog (2021), Public attitudes towards algorithmic personalization and use of personal data online: evidence from Germany, Great Britain, and the United States, *Humanities and Social Sciences Communications*, **8**(1): 117.

Kozyreva, A., P. Lorenz-Spreen, S. M. Herzog, U. K. H. Ecker, S. Lewandowsky, R. Hertwig, A. Ali, J. Bak-Coleman, S. Barzilai, M. Basol, A. J. Berinsky, C. Betsch, J. Cook, L. K. Fazio, M. Geers, A. M. Guess, H. Huang, H. Larreguy, R. Maertens and S. Wineburg (2024), Toolbox of individual-level interventions against online misinformation, *Nature Human Behaviour*, **8**: 1044–1052.

Kozyreva, A., L. Smillie and S. Lewandowsky (2023), Incorporating psychological science into policy making, *European Psychologist*, **28**(3): 206–224.

Kromphardt, C. D. and M. F. Salamone (2021), "Unpresidented!" or: what happens when the president attacks the federal judiciary on Twitter, *Journal of Information Technology and Politics*, **18**(1): 84–100.

Laebens, M. G. and A. Lührmann (2021), What halts democratic erosion? The changing role of accountability, *Democratization*, **28**(5): 908–928.

Lazer, D., M. Baum, J. Benkler, A. Berinsky, K. Greenhill, M. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, C. Sunstein, E. Thorson, D. Watts and J. Zittrain (2018), The science of fake news, *Science*, **359**(6380): 1094–1096.

Lee, S. and S. M. Jones-Jang (2024), Cynical nonpartisans: the role of misinformation in political cynicism during the 2020 U.S. presidential election, *New Media & Society*, **26**(7): 4255–4276.

Levitsky, S. and D. Ziblatt (2018), *How Democracies Die: What History Reveals About Our Future*, New York: Penguin Random House

Levitsky, S. and D. Ziblatt (2023, September 8), Democracy's Assassins Always Have Accomplices. *The New York Times*. https://www.nytimes.com/2023/09/08/opinion/trump-republicans-spain-brazil.html

Lewandowsky, S., J. Cook, U. K. H. Ecker, D. Albarracín, M. A. Amazeen, P. Kendeou, D. Lombardi, E. J. Newman, G. Pennycook, E. Porter, D. G. Rand, D. N. Rapp, J. Reifler, J. Roozenbeek, P. Schmid, C. M. Seifert, G. M. Sinatra, B. Swire-Thompson, S. van der Linden and M. S. Zaragoza (2020a), *The Debunking Handbook 2020*.

Lewandowsky, S., U. K. H. Ecker and J. Cook (2017), Beyond misinformation: understanding and coping with the "post-truth" era, *Journal of Applied Research in Memory and Cognition*, **6**(4): 353–369.

Lewandowsky, S., L. Smillie, D. Garcia, R. Hertwig, J. Weatherall, S. Egidy, R. E. Robertson, C. O'Connor, A. Kozyreva, P. Lorenz-Spreen, Y. Blaschke and M. Leiser (2020b), *Technology and democracy: understanding the influence of online technologies on political behaviour and decision-making*, Luxembourg: Publications Office of the European Union. 10.2760/709177

Lindberg, S. I., M. Coppedge, J. Gerring, J. Teorell, D. Pemstein, E. Tzelgov, Y. T. Wang, A. Glynn, D. Altman, M. Bernhard, S. Fish, A. Hicken, M. Kroenig, K. McMann, P. Paxton, M. Reif, S. E. Skaaning and J. Staton (2014), V-Dem: a new way to measure democracy, *Journal of Democracy*, **25**(3): 159–169.

Lorenz-Spreen, P., L. Oswald, S. Lewandowsky and R. Hertwig (2022), A systematic review of worldwide causal and correlational evidence on digital media and democracy, *Nature Human Behaviour*, **7**(1): 74–101.

Lührmann, A., K. L. Marquardt and V. Mechkova (2020), Constraining governments: new indices of vertical, horizontal, and diagonal accountability, *American Political Science Review*, **114**(3): 811–820.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Lührmann, A., M. Tannenberg and S. I. Lindberg (2018), Regimes of the world (RoW): opening new avenues for the comparative study of political regimes, *Politics and Governance*, **6**(1): 60–77.

Ma, S., D. Bergan, S. Ahn, D. Carnahan, N. Gimby, J. McGraw and I. Virtue (2023), Fact-checking as a deterrent? A conceptual replication of the influence of fact-checking on the sharing of misinformation by political elites, *Human Communication Research*, **49**(3): 321–338.

Mackie, D. M. (1986), Social identification effects in group polarization, *Journal of Personality and Social Psychology*, **50**(4): 720–728.

Malmendier, U. and S. Nagel (2011), Depression babies: do macroeconomic experiences affect risk taking?*, *The Quarterly Journal of Economics*, **126**(1): 373–416.

Malmendier, U. and S. Nagel (2016), Learning from inflation experiences, *The Quarterly Journal of Economics*, **131**(1): 53–87.

Mansbridge, J. and S. Macedo (2019), Populism and democratic theory, *Annual Review of Law and Social Science*, **15**(1): 59–77.

Marsden, E. (2022), *The Defence in Depth Principle: A Layered Approach to Safety Barriers*, Risk Engineering. https://risk-engineering.org/concept/defence-in-depth [22 May 2024].

Mastrolorenzo, G., P. Petrone, L. Pappalardo and M. F. Sheridan (2006), The Avellino 3780-yr-B.P. catastrophe as a worst-case scenario for a future eruption at Vesuvius, *Proceedings of the National Academy of Sciences*, **103**(12): 4366–4370.

Maxwell, K. (1991), Spain's transition to democracy: a model for Eastern Europe?, *Proceedings of the Academy of Political Science*, **38**(1): 35–49.

Mettler, S. and R. C. Liebermann (2020), *Four Threats: The Recurring Crises of American Democracy*, New York: St. Martin's Griffin

Mittiga, R. (2022), Political legitimacy, authoritarianism, and climate change, *American Political Science Review*, **116**(3): 998–1011.

Montanaro, D. (2023, April 5), Most Republicans would vote for Trump even if he's convicted of a crime, poll finds. *NPR*. https://www.npr.org/2023/04/25/1171660997/poll-republicans-trump-president-convicted-crime

Morrison, J. B. and R. L. Wears (2022), Modeling Rasmussen's dynamic modeling problem: drift towards a boundary of safety, *Cognition, Technology and Work*, **24**(1): 127–145.

Mudde, C. (2017), Populism: an ideational approach, C. R. Kaltwasser, P. Taggart, P. O. Espejo and P. Ostiguyeds, *The Oxford Handbook of Populism*, Oxford: Oxford University Press.

Mudde, C. and C. Rovira Kaltwasser (2018), Studying populism in comparative perspective: reflections on the contemporary and future research agenda, *Comparative Political Studies*, **51**(13): 1667–1693.

Navia, P. and R. Osorio (2019), Attitudes toward democracy and authoritarianism before, during and after military rule. The case of Chile, 1972–2013, *Contemporary Politics*, **25**(2): 190–212.

Neundorf, A., J. Gerschewski and R.-G. Olar (2020), How do inclusionary and exclusionary autocracies affect ordinary people?, *Comparative Political Studies*, **53**(12): 1890–1925.

Nickerson, D. W. and T. Rogers (2010), Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making, *Psychological Science*, **21**(2): 194–199.

Nord, M., M. Lundstedt, D. Altman, F. Angiolillo, C. Borella, T. Fernandes, L. Gastaldi, A. G. God, N. Natsika and S. I. Lindberg (2024), *Democracy Report 2024: Democracy Winning and Losing at the Ballot*. https://www.v-dem.net/documents/43/v-dem_dr2024_lowres.pdf

Nyhan, B. and J. Reifler (2015), The effect of fact-checking on elites: a field experiment on U.S. state legislators, *American Journal of Political Science*, **59**(3): 628–640.

Orhan, Y. E. (2022), The relationship between affective polarization and democratic backsliding: comparative evidence, *Democratization*, **29**(4): 714–735.

Pantazi, M., S. Hale and O. Klein (2021), Social and cognitive aspects of the vulnerability to political misinformation, *Political Psychology*, **42**(S1): 267–304.

Pantazi, M., K. Papaioannou and J. W. van Prooijen (2022), Power to the people: the hidden link between support for direct democracy and belief in conspiracy theories, *Political Psychology*, **43**(3): 529–548.

Parker, A. and D. E. Sanger (2016, July 27), Donald Trump calls on Russia to find Hillary Clinton's missing emails. *The New York Times*. https://www.nytimes.com/2016/07/28/us/politics/donald-trump-russia-clinton-emails.html

Pedersen, V. M. L., J. D. Thaysen and A. Albertsen (2023), Nudging voters and encouraging pre-commitment: beyond mandatory turnout, *Res Publica*, **30**(2): 267–238.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Perrow, C. (1984), *Normal Accidents: Living with High-Risk Technologies*, Princeton University Press, Princeton.

Perrow, C. (1996), *The Challenger Launch Decision: Risky Technology, Culture and Deviance at NASA*, Chicago: University of Chicago Press

Pettigrew, T. F. (1998), Intergroup contact theory, *Annual Review of Psychology*, **49**(1): 65–85.

Pettigrew, T. F. and L. R. Tropp (2006), A meta-analytic test of intergroup contact theory, *Journal of Personality and Social Psychology*, **90**(5): 751–783.

Polyák, G. (2019), Media in Hungary: three pillars of an illiberal democracy, E. Połońska and C. Becketteds, *Public Service Broadcasting and Media Systems in Troubled European Democracies*, 279–303, Cham: Springer International Publishing.

Posada-Carbó, E. (2011), Latin America: Colombia after uribe, *Journal of Democracy*, **22**(1): 137–151. https://muse.jhu.edu/article/412899

Rasmussen, J. (1997), Risk management in a dynamic society: a modelling problem, *Safety Science*, **27**(2–3): 183–213.

Reason, J. (2016), *Managing the Risks of Organizational Accidents*, London: Routledge.

Rogers, W. P., N. A. Armstrong, D. C. Acheson, E. E. Covert, R. P. Feynman, R. B. Hotz, D. J. Kutyna, S. K. Ride, R. W. Rummel, J. F. Sutter, A. B. C. Walker, A. D. Wheelon and C. E. Yeager (1986), *Report to the President by the Presidential Commission on the Space Shuttle Challenger Accident*.

Roozenbeek, J., S. van der Linden, B. Goldberg, S. Rathje and S. Lewandowsky (2022), Psychological inoculation improves resilience against misinformation on social media, *Science Advances*, **8**34.

Rovira Kaltwasser, C. (2017), ‘Populism and the Question of How to Respond to It’, in C. R. Kaltwasser, P. Taggart, P. O. Espejo and P. Ostiguy (eds), *The Oxford Handbook of Populism*volume 1, Oxford: Oxford University Press, 489–508.

Schaub, M. (2021), Acute financial hardship and voter turnout: theory and evidence from the sequence of bank working days, *American Political Science Review*, **115**(4): 1258–1274.

Schlipphak, B. and O. Treib (2017), Playing the blame game on Brussels: the domestic political effects of EU interventions against democratic backsliding, *Journal of European Public Policy*, **24**(3): 352–365.

Schulte-Mecklenbeck, M., G. G. Wagner and R. Hertwig (2024), How personal experiences shaped risk judgments during COVID-19, *Journal of Risk Research*, **27**(3): 438–457.

Shear, M. D. and S. Saul (2021, January 3), Trump, in taped call, pressured Georgia official to ‘find’ votes to overturn election. *The New York Times*. https://www.nytimes.com/2021/01/03/us/politics/trump-raffensperger-call-georgia.html

Siripurapu, A. (2022), *The U.S. Inequality Debate*, Council on Foreign Relations, https://www.cfr.org/backgrounder/us-inequality-debate

Somer, M., J. L. McCoy and R. E. Luke (2021), Pernicious polarization, autocratization and opposition strategies, *Democratization*, **28**(5): 929–948.

Steenland, K. (1974), The coup in Chile, *Latin American Perspectives*, **1**(2): 9–29. https://www.jstor.org/stable/2633976

Svolik, M. W. (2019), Polarization versus Democracy, *Journal of Democracy*, **30**(3): 20–32.

Svolik, M. W., E. Avramovska, J. Lutz and F. Milaèiæ (2023), In Europe, democracy erodes from the right, *Journal of Democracy*, **34**(1): 5–20.

Swami, V. (2012), Social psychological origins of conspiracy theories: the case of the Jewish conspiracy theory in Malaysia, *Frontiers in Psychology*, 3.

Swire, B., A. J. Berinsky, S. Lewandowsky and U. K. H. Ecker (2017), Processing political misinformation: comprehending the Trump phenomenon, *Royal Society Open Science*, **4**(3): 160802.

Swire-Thompson, B., U. K. H. Ecker, S. Lewandowsky and A. J. Berinsky (2020), They might be a liar but they’re my liar: source evaluation and the prevalence of misinformation, *Political Psychology*, **41**(1): 21–34.

Szelényi, Z. (2022), How Viktor Orbán built his illiberal state. *The New Republic*. https://newrepublic.com/article/165953/viktor-orban-built-illiberal-state

Tsipursky, G., F. Votta and J. A. Mulick (2018a), A psychological approach to promoting truth in politics: the Pro-Truth Pledge, *Journal of Social and Political Psychology*, **6**(2): 271–290.

Tsipursky, G., F. Votta and K. M. Roose (2018b), Fighting fake news and post-truth politics with behavioral science: the Pro-Truth Pledge, *Behavior and Social Issues*, **27**(1): 47–70.

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Uscinski, J., A. Enders, A. Diekman, J. Funchion, C. Klofstad, S. Kuebler, M. Murthi, K. Premaratne, M. Seelig, D. Verdear and S. Wuchty (2022), The psychological and political correlates of conspiracy theory beliefs, *Scientific Reports*, **12**(1): 1–12.

van der Linden, S., D. Albarracín, L. Fazio, D. Freelon, J. Roozenbeek, B. Swire-Thompson and J. van Bavel (2023), Using psychological science to understand and fight health misinformation, *APA Consensus Statement*, November, https://www.apa.org/pubs/reports/health-misinformation

Varol, O. O. (2015), Stealth authoritarianism, *Iowa Law Review*, **100**(4): 1673–1742. https://ilr.law.uiowa.edu/sites/ilr.law.uiowa.edu/files/2023-02/ILR-100-4-Varol.pdf

Voelkel, J. G., J. Chu, M. N. Stagnaro, J. S. Mernyk, C. Redekopp, S. L. Pink, J. N. Druckman, D. G. Rand and R. Willer (2023), Interventions reducing affective polarization do not necessarily improve anti-democratic attitudes, *Nature Human Behaviour*, **7**(1): 55–64.

Wachinger, G., O. Renn, C. Begg and C. Kuhlicke (2013), The risk perception paradox—implications for governance and communication of natural hazards, *Risk Analysis*, **33**(6): 1049–1065.

Waisbord, S. (2018), The elective affinity between post-truth communication and populist politics, *Communication Research and Practice*, **4**(1): 17–34.

Waldner, D. and E. Lust (2018), Unwelcome change: coming to terms with democratic backsliding, *Annual Review of Political Science*, **21**: 93–113.

Weber, E. U. (2006), Experience-based and description-based perceptions of long-term risk: why global warming does not scare us (yet), *Climatic Change*, **77**(1–2): 103–120.

Weber, E. U. and P. C. Stern (2011), Public understanding of climate change in the United States, *American Psychologist*, **66**(4): 315–328.

Weber, T. (2022), *Als Die Demokratie Starb: Die Machtergreifung der Nationalsozialisten—Geschichte Und Gegenwart*, Freiburg: Herder

Wegwarth, O., U. Mansmann, F. Zepp, D. Lühmann, R. Hertwig and M. Scherer (2023), Vaccination intention following receipt of vaccine information through interactive simulation vs text among covid-19 vaccine–hesitant adults during the omicron wave in Germany, *JAMA Network Open*, **6**(2): e2256208.

Wiesner, K., S. Bien and M. C. Wilson (2023), *The hidden dimension in democracy* (V-Dem Working Paper). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4453098

Wiesner, K., A. Birdi, T. Eliassi-Rad, H. Farrell, D. Garcia, S. Lewandowsky, P. Palacios, D. Ross, D. Sornette and K. Thébault (2019), Stability of democracies: a complex systems perspective, *European Journal of Physics*, **40**(1): 014002.

Wood, M. J., K. M. Douglas and R. M. Sutton (2012), Dead and alive: beliefs in contradictory conspiracy theories, *Social Psychological and Personality Science*, **3**(6): 767–773.

Wunsch, N. and P. Blanchard (2022), Patterns of democratic backsliding in third-wave democracies: a sequence analysis perspective, *Democratization*, **30**(2): 278–301.

Zakrzewski, C. (2023, July 4), Judge blocks U.S. officials from tech contacts in First Amendment case. *The Washington Post*. https://www.washingtonpost.com/technology/2023/07/04/biden-social-lawsuit-missouri-louisiana/

Zmigrod, L. (2020), The role of cognitive rigidity in political ideologies: theory, evidence, and future directions, *Current Opinion in Behavioral Sciences*, **34**: 34–39.

Appendix

Cases were selected for this analysis through a literature search on Google Scholar conducted between 5 and 14 July 2023. Table A1 presents the number of papers returned for each search query.

Table A1. Results for Google Scholar search queries

Search query	Number of papers returned
‘near misses’ AND ‘democratic backsliding’	85
‘near misses’ AND ‘backsliding’	223
‘democratic near misses’	6

https://doi.org/10.1017/bpp.2024.43 Published online by Cambridge University Press

Further articles published at a later stage were included subsequently. The present analysis was limited to cases that were strongly documented as near misses in the literature (e.g., Colombia, Sri Lanka) or can historically be understood as a near miss (e.g., UK, in the 1930s).

Although various violations of democratic norms in several Western countries have been reported in the media (e.g., the prorogation of parliament in the UK; the attempt of former President Donald Trump to stop the peaceful transfer of power on 6 January 2021), these cases can be understood as instances of democratic backsliding rather than near misses, and are therefore not discussed in our historical analysis.

Cite this article: Abels CM, Huttunen KJA, Hertwig R and Lewandowsky S (2024), 'Dodging the autocratic bullet: enlisting behavioural science to arrest democratic backsliding', *Behavioural Public Policy*, 1–28. https://doi.org/10.1017/bpp.2024.43