Special Issue Article

Norms of Democracy, Staged Democrats, and Supply of Exclusionary Ideology

Comparative Political Studies
2026, Vol. 59(6) 1367–1405

DOI: 10.1177/00104140241283009
journals.sagepub.com/home/cps

Vicente Valentim[1,*]

Abstract

Existing literature sees attitudinal support for liberal democratic values as a crucial way to prevent the success of illiberal actors and their policies. But recent studies show that democratic support can coexist with lack of punishment for backsliding practices. To reconcile these findings, I argue that social norms generate incentives to declare that one opposes illiberal actors even if one does not—what I call *staged democrats*. Staged democrats are not a stable safeguard against illiberal politicians, because information shocks can reveal that they do not sincerely oppose illiberalism. Studying Switzerland, where referendums provide information on the sincere preferences of citizens, I collect data on referendum results, public opinion surveys, and party positions. My analyses show that if referendums reveal that the positions of far-right parties are more popular than expected, those parties become more exclusionary. Attitudes only prevent illiberal policy if they are sincerely held.

Keywords

social norms, democratic norms, democratic backsliding, preference falsification, democratic support, staged democrats

[1]University of Oxford, United Kingdom

*Current affiliation: IE University, Madrid, Spain

Corresponding Author:
Vicente Valentim, Nuffield College, University of Oxford, United Kingdom.
Email: vvalentim@faculty.ie.edu

Data Availability Statement included at the end of the article

Introduction

A number of countries worldwide have recently experienced democratic backsliding, often at the hand of far-right politicians, whose policies and tactics menace form and substance of liberal democracy.[1] Substantively, far-right policies threaten core values of liberal democracy (Bichay, 2022). According to many definitions of the democracy, the inclusion of minorities is one such value (Merkel, 2004; Mudde & Rovira Kaltwasser, 2012; Young, 2000). The policy and rhetoric of far-right politicians, to which opposition to immigration and ethnic minorities is key, is in clear tension with goals of integration (e.g., Mudde, 2007; Pirro, 2014; Rydgren, 2017; Sakki & Pettersson, 2016). In some countries, these politicians also oppose gender equality (Anduiza & Rico, 2022; Off, 2023) and the rights of sexual minorities (Bernhard, 2021; Donà, 2021; Yermakova, 2021).[2]

Even if one follows more minimalist definitions of democracy, that do not take minority inclusion as one of its core values (e.g., Capoccia, 2023; Przeworski, 1999; Schumpeter, 2013), far-right politicians still pose a threat to formal democratic procedures. In an effort to carry out their proposed policies, they often erode structures of accountability and electoral rules (Bugarič, 2015; Das, 2023; Huber & Schimpf, 2016; Mudde & Rovira Kaltwasser, 2012; Väliverronen & Saikkonen, 2021)—which represent typical legalist backsliding strategies (Capoccia, 2023).

A commonly discussed solution to thwart far-right success is the fostering of liberal democratic attitudes. Mass-level support for democracy and a rejection of illiberal actors has often been regarded as a bulwark of democracy (Almond & Verba, 1963; Easton, 1965; Linz & Stepan, 1996; Lipset & Rokkan, 1990). If democratic citizens overwhelmingly hold these attitudes, politicians implementing illiberal policy should be met with electoral punishment. Recently, however, a number of studies have shown that high attitudinal support for democratic values often coexists with an absence of such punishment (Carvajal et al., 2023; Goldman, Drutman, and Pocasangre, 2024; Graham & Svolik, 2020; Simonovits et al., 2021).

A likely explanation for this puzzle, I argue, lies on the fact that previous research has mostly overlooked the possibility that democratic citizens may engage in preference falsification when asked about their attitudes. In other words, they may declare that they support liberal democratic values insincerely, to avoid reputational costs (Kuran, 1987; Valentim, 2024a). Democracies tend to put in place social norms against preferences at odds with core liberal democratic values. I term these *norms of democracy*, to distinguish them from democratic norms—which refer, more narrowly, to norms about what *form of government* individuals are expected to prefer (Bischof et al., 2023; Goldstein, 2022). These norms entail social sanctions to those who deviate, which provide incentives for conformity *even if one privately*

disagrees with the norm. Assuming the possibility of preference falsification suggests that observing individuals act in accordance with norms of democracy is compatible with two different states of the world: one where individuals declare to support democracy because they honestly do (what I call *sincere democrats*), and another where they do so simply out of preference falsification (what I term *staged democrats*).

The core interest of this paper lies in how staged democrats affect political elites. I argue that, if illiberal politicians gather information suggesting that support for norms of democracy is, at least partly, staged, they have an incentive to put forward more counternormative policy. This means that an equilibrium where support for norms of democracy is based on preference falsification is only a fragile safeguard against policy at odds with those norms.

Empirically, I focus on norms against support for far-right actors and their policies—one of the norms of democracy that has been the most studied in previous work. My argument would suggest that information revealing that there is more support for far-right positions than typically meets the eye can impel politicians of that ideology to become more exclusionary.

The difficulty with testing this hypothesis is that it requires finding instances where politicians receive information suggesting that support for norms of democracy is staged. To overcome this issue, I leverage a unique setting in Switzerland. In this country, there are frequent referendums that invite citizens to vote on discrete policies. While anti-far-right norms can make politicians with that ideology underestimate the private support for their views, referendum results can correct those misperceptions.

To test this hypothesis, I build a dataset collecting results of more than 300 referendums over 38 years. I also collect data on surveys that ask for citizens' self-reported preferences on the topics of those referendums, and data on the position that each party took on that referendum. Then, I merge these data with information on the ideological position of parties, retrieved from the Comparative Manifesto Project (CMP) (Volkens et al., 2015).

As a preliminary step, I show that, at the individual level, the gap between preferences expressed in voting and those declared in a survey is three times higher if a far-right party endorses a given position. This suggests that some individuals are indeed staged democrats, who privately support far-right positions but do not declare that support to conform with established norms of democracy.

I then test the core implications of the argument, which pertain to how this situation affects political elites. I construct a variable called electoral surplus, which captures the average distance between the official and declared support for a party's referendum positions over a legislative period—as a measure of how much "hidden support" there is for such positions. Relying on two-way fixed effects models, I show that, the higher the electoral surplus for far-right

positions, the less those parties make positive references to minority groups in the subsequent election. Additional analyses support the interpretation that the results are driven by the information collected in the referendum: the effects are stronger in the first election in the dataset, and for referendums on culture, religion and media—the topic most clearly linked to minorities.

This paper has implications for two bodies of literature. First, it speaks to a body of literature that has looked into how political elites, their behavior, and rhetoric can erode democratic norms (Clayton et al., 2021; Levitsky & Ziblatt, 2019; Valentim, 2021; Valentim et al., 2022). This paper looks at the opposite causal nexus. It focuses on how norms can affect the behavior of politicians—concretely, the ideological positions that they run for politics with. In so doing, they suggest that the relation between norms and elite behavior is complex, and that both can affect one another. This conclusion supports the work of Ammassari (2022, 2024), and Art (2011), who have argued that social norms can impact the behavior and decisions of individuals actively involved in politics.

Second, the argument put forward here has implications for a debate on whether democratic support is declining and the implications of that trend (Foa and Mounk, 2016, 2022; Kriesi, 2020; Norris, 1999, 2011, 2017; Wuttke et al., 2022). While the empirical analyses here draw upon norms against support for far-right actors and their policies, the argument should apply in the case of different norms of democracy. This includes norms that give citizens an incentive to support democracy over other forms of government (Bischof et al., 2023; Goldstein, 2022). With that in mind, this paper qualifies a shared assumption in the literature on democratic support, according to which such support can be equated with democratic resilience. As this paper suggests, that is only the case if citizens who declare to support democracy are sincere democrats. If a significant portion of them are staged democrats, democratic support will represent at best a fragile safeguard against illiberal actors and their ideology.

The Argument for Attitudes as a Safeguard Against Illiberal Actors

A long-standing tradition in political science assumes that attitudes are a key—maybe *the* key—way of promoting democratic resilience and preventing the success of illiberal actors. The general argument is that, if commitment to a given norm is deemed desirable, then fostering attitudes in line with that norm is one of the crucial ways to ensure its persistence.

Because it is taken as a central determinant of democratic health, a large body of literature has tried to assess trends in democratic support at the citizen level. The rationale is that these trends can help assess how resilient democracies are. If researchers find that democratic support is high and steady,

the typical conclusion is that democracy is healthy (Norris, 1999, 2011, 2017). If, on the other hand, one finds that low or decreasing democratic support, the typical conclusion is that one should worry about the future of democracy and its resilience (Claassen & Magalhães, 2023; Foa and Mounk, 2016, 2022; Wike & Fetterolf, 2018).[3] Despite the ongoing debates in it, this literature tends to share a common assumption: that democratic support and democratic survival are so closely connected that the former can be taken as a measure of the latter. For simplicity, in the following I will call this the *sincere attitudes assumption*.

The sincere attitudes assumption is not found only in the literature on democratic support. A wealth of literature has tried to explain the development of far-right (Doosje et al., 2012), authoritarian (Carnevale et al., 2020), or populist (Erisen et al., 2021) attitudes. Again, the underlying assumption is that these attitudes will translate into behavior, which means that the more individuals declare to hold those attitudes, the more likely they are to support politicians with those ideologies.

Norms of Democracy, Sincere Democrats, and Staged Democrats

In this paper, I argue that the sincere attitudes assumption discussed in the previous section may be ill-founded. Common as it may be, this assumption is at odds with recent work showing that, even if many democratic citizens declare that they are supportive of democracy, only a small minority are willing to electorally punish candidates who violate democratic norms (Carvajal et al., 2023; Goldman, Drutman, and Pocasangre, 2024; Graham & Svolik, 2020; Simonovits et al., 2021).

To understand this seeming contradiction, I build on multidisciplinary work on social norms and their enforcement. Often thought of as *conditional preferences*, norms are informal rules of behavior that individuals follow because they expect that others approve of them and can sanction deviations (Bicchieri, 2017). The point of departure of my argument is that, in democracies, there are often political norms against behavior deemed at odds with core values of liberal democracy. Examples include norms against supporting far-right politicians (Harteveld et al., 2019; Lagios et al., 2022; Valentim, 2021) and authoritarian successor parties (Grzymala-Busse, 2002; Valentim, 2024a), norms disincentivizing racism and xenophobia (Benjumea, 2022; Kuklinski, Cobb, and Gilens, 1997), or norms against behavior generally associated with authoritarianism (Dinas et al., 2024; Muñoz, 2012).

I refer to these norms as *norms of democracy*, to distinguish them from democratic norms—which typically refer to norms prescribing support for democracy as a form of government (Bischof et al., 2023; Goldstein, 2022). Norms of democracy, thus defined, are a more encompassing concept. They

represent a cluster of perceptions about the acceptability of different—yet related—behaviors and attitudes, linked by the fact that they are perceived as pertaining to core liberal democratic values.[4,5] They ecompass democratic norms, but also the other aforementioned norms about issues like support for far-right parties and authoritarian-linked behavior. While these behaviors are distinct and do not necessarily always go hand-in-hand, what they have in common is the source of the political stigma against them, which lies on their tension with liberal democratic values.

Three notes are warranted about this concept. First, not all behavior that is at odds with norms of democracy is inherently anti-democratic. Support for far-right parties, for example (which the empirical sections of the paper study) may not be anti-democratic insofar as these parties may not openly oppose democracy as a form of government. What makes it in tension with a norm of democracy is that those parties—even the less extreme ones— reject core values of liberal democracy, like the protection of minorities (Mudde, 2007). Second, the specific norms that are included in the cluster of norms of democracy may vary from country to country. But, to the extent that they all generate incentives for conformity, the argument should apply equally to them. Finally, that there is a norm of democracy does not mean that those who are affected by it unanimously perceive it as democratic or desirable. Similarly, it does not mean that those who disagree with it perceive their own views as undesirable. All that is required by my argument is that individuals perceive that their own views are contrary to an established norm, such that acting on it is socially costly. This assumption is likely to hold considering that previous work has provided evidence that individuals are typically aware of the social norms associated with their views and identity (Alvarez-Benjumea & Valentim, 2024; Groenendyk et al., 2022; Pickup et al., 2022).

Deviating from norms of democracy—like deviations from any other norm—is socially costly. Alvarez-Benjumea and Valentim (2024) show evidence of this point in the context of anti-far-right norms. As their findings show, individuals who run into someone identifying as a far-right supporter are likely to call them out, insult them, avoid interacting with them, spread rumors about them, or deny help to them if asked—all of which represent the typical kind of sanctions imposed on other, non-political norms (Eriksson et al., 2021; Molho et al., 2020; Nowak & Sigmund, 2005; Ohtsuki & Iwasa, 2006; Rost et al., 2016).[6]

Under the threat of social sanctions, norm abidance becomes rational *even for those who disagree with the norm*—a phenomenon termed preference falsification (Kuran, 1995). This desire for conformity is a very established pattern, which has been supported by the findings of literature in a number of different fields. Examples include studies in social psychology (e.g., Asch, 1956; Chierchia, Piera Pi-Sunyer, and Blakemore, 2020; Song et al., 2012), anthropology (e.g., Delgado-Gaitan, 1988;

Tucker, 2012), sociology (e.g., Mackie, 1996; Scheff, 1988; Swidler, 1986; Warren, 1968), or behavioral economics (e.g., Andreoni et al., 2021; Fehr & Gachter, 2000; Goeree & Yariv, 2015). That established norms provide incentives for preference falsification, I argue, is one likely reason for the seemingly contradictory findings discussed above. Individuals may be declaring insincere attitudes, like support for norms of democracy, which is why they often do not punish actors whose actions are at odds with them.

Based on this discussion, I propose a distinction between two types of citizens, both of which act according to these norms. On the one hand there are *sincere democrats*, who follow them because they are in line with their sincere preferences. This is the type of citizen that has been assumed in most previous work. However, not all citizens who act according to norms of democracy are necessarily sincere democrats. Others may be *staged democrats*, whose preferences are not aligned with those norms, but who are aware that breaching them can come with social costs. In an effort to avoid those costs, they falsify their preferences and act as though they, too, supported established norms of democracy (for a similar point in the context of authoritarian regimes, see Baik & Shen, 2023).[7,8]

The concept of staged democrats is analytically distinct from the related concepts of social desirability bias and democratic hypocrisy. Building on previous work on preference falsification (Kuran, 1987, 1995), the concept affects a wider range of behaviors than survey answers, including the willingness to participate in protests (Patel, 2013), views expressed in political conversations (Carlson & Settle, 2022), or expressions of support for political platforms whenever they can be observed by others (Valentim, 2024a). Staged support for norms of democracy represents a specific instance of preference falsification: that which happens to comply with social norms pertaining to core liberal democratic values. But it leaves out other instances, for example, when individuals engage in preference falsification to avoid legal punishment in the context of authoritarian regimes (e.g., Baik & Shen, 2023; Hale, 2022; Jiang & Yang, 2016; Kuran, 1987). At the same time, the concept differs from that of democratic hypocrisy (Simonovits et al., 2021) because it does not refer to a situation where individuals have double standards about norm erosion depending on who engages in that erosion. Instead, it refers to a situation where, in regard to the same political stance (and without variation in the actors who sponsor it), individuals reveal different preferences in private than they are willing to declare in public.

Staged Democrats and the Prevention of Illiberal Policy

In the previous section, I have discussed how norms of democracy can lead to staged democratic support—a specific instance of preference

falsification. However, given that preference falsification is fairly established in previous work (Blair et al., 2020; Bursztyn et al., 2020a, 2020b; Dinas et al., 2024; Jiang & Yang, 2016; Valentim, 2021, 2024a), the main contribution of this paper is not in showcasing this phenomenon. Rather, it lies in highlighting how it can affect the behavior *of politicians*.

I argue that sincere and staged democrats have crucially different implications when it comes to preventing illiberal actors from putting forward exclusionary ideology. While, for the most part, the two types of citizens are observationally equivalent, there are instances when the veil of preference falsification is lifted, revealing the sincere preferences of citizens. Electoral acts represent one such instance. Since voting is private, it is unlikely that someone will suffer social sanctions for their electoral choices. For this reason, individuals are more likely to reveal their sincere preferences when casting a vote, even those preferences are at odds with established norms (Ewing, 2001; Funk, 2016; Kuran, 1995; Valentim, 2024c). Elections, or other instances where individuals feel more comfortable expressing their sincere preferences, may thus reveal a gap between those preferences and the public behavior that is typically observed. If this gap reveals that some who act in accordance with norms of democracy are staged democrats, this suggests that there is less sincere support for those norms than one would otherwise infer based on their behavior.

The goal of this paper is precisely to illustrate one instance in which staged and sincere democrats differ in their implications. I focus on one specific norm of democracy: the norm against support for far-right parties and their policies.[9] This is one of the norms of democracy that has been most commonly discussed in previous research, and one of those for which there is wider empirical support (Ammassari, 2022, 2024; Blinder et al., 2013; Bursztyn, Egorov, and Fiorin, 2020; Harteveld et al., 2019; Harteveld & Ivarsflaten, 2018; Valentim, 2021).[10] Support for these parties may, in some cases, endanger formal democratic procedures. In others it will not, but will still endanger the liberal component of democracy. Either way, there is typically a shared understanding amid democratic citizens that supporting these parties is unacceptable in a democracy (Alvarez-Benjumea & Valentim, 2024)—which means that this constitutes a norm of democracy as I have defined it.

A crucial implication of the realization that more individuals support the views of far-right politicians than is typically observable concerns the policies, rhetoric, and ideology that those politicians put forward. Previous work has provided ample evidence that elites often harbor inaccurate views of what voters think (Belchior, 2014; Broockman & Skovron, 2018; Hertel-Fernandez et al., 2019; Kertzer et al., 2019; Pereira, 2021; Walgrave et al., 2023). Importantly, this pattern is also found when it comes to

Switzerland, the case that I study in the empirical analyses of this paper (Pilet et al., 2024).

This inaccuracy in estimating voter preferences should be made even more acute when individuals engage in preference falsification—especially if one considers the sources of information that politicians trust the most when learning about the views of voters. Their most trusted source are direct conversations with citizens—a type of interaction that is likely to be heavily influenced by social norms (Walgrave & Soontjens, 2023). As such, one might expect that, if citizens stage support for a norm of democracy, far-right politicians can underestimate how much latent support there is for their ideology and policies. As a consequence, they may feel that more exclusionary ideology will not yield them wide electoral support.[11]

However, if politicians are provided with information revealing that there is more private disagreement with the norm than is typically observed, they may update their ideological platform accordingly. Previous work has shown that, when they learn about the actual distribution of voters' preferences, politicians typically update their own behavior to come closer to them (Butler & Nickerson, 2011; Hager & Hilbig, 2020; Sevenans, 2021; Chu & Recchia, 2022, but see Kalla & Porter, 2021). Crucially, one way in which they learn about those preferences is from direct democracy tools like initiatives and referenda. When democracies include these tools, they tend to make legislators moving closer to the policy positions of the electorate (Matsusaka, 2010), a finding that has also been replicated in the Swiss setting (Helfer et al., 2021; Leemann & Wasserfallen, 2016).

In line with this work, information suggesting that there is more private support for the policies of far-right actors than typically meets the eye should make politicians perceive that opposition to those policies is, at least partly, staged. This provides them with an incentive to put forward a more exclusionary ideology. Since staged democrats do not sincerely agree with the established norm, they will not incur the costs of enforcing it. They will not vote against a politician for putting forward policy at odds with an established norm of democracy. They will also not voice opposition by participating in protests or other such actions. They may even electorally *reward* the norm breaching policy of that politician, since it is aligned with their sincere preferences.[12]

This is the crucial difference between sincere and staged democrats. Most of the time, both act as though they support norms of democracy. But staged democrats only do so in settings where they can be observed, and where social sanctions are likely. In private, however, they are comfortable acting against those norms. As such, staged democrats will only contribute to keep norms of democracy in place as long as political elites are unaware of how they privately feel about those norms. Once

that information is revealed, they will no longer contribute to keeping that norm in place.[13]

Research Design

Empirically, the goal of this paper is to illustrate the different implications of sincere and staged support for norms of democracy. The difficulty with testing this hypothesis empirically is that one needs to find instances where politicians realize that the private views of citizens differ from those that they commonly express. To overcome this difficulty, I focus on the case of Switzerland, where citizens are frequently called to voice their opinions on referendums.[14] Because voting is private, electoral outcomes can provide information as to how popular established norms really are (Bursztyn, Egorov, and Fiorin, 2020; Jung & Tavits, 2021; Valentim, 2021), which could update politicians' priors. This is even more likely when it comes to results of referendums, which allow individuals to vote on specific discrete policies. This is in contrast with elections, where citizens are called to vote on a pack of policies, making it unclear what exactly is driving their vote. Moreover, referendums happen during legislative periods, which means that the information gathered from their results can help politicians define their ideology in the subsequent election. Following my argument, I would expect that higher-than-expected support for the positions of far-right politicians in referendums should increase their chances of putting forward more exclusionary ideology.

The question, then, becomes how one can measure whether support for a given policy in a referendum is higher or lower than expected. I do so by comparing referendum results to declared preferences in survey—an approach that has been followed by previous research (Funk, 2016). Unlike voting, survey interviews are not truly private behaviors because individuals need to convey their answer to an interviewer. Fearful of judgement, they may prefer to provide what they perceive to be the socially desirable answer (Zaller, 1992).[15] For these reasons, if individuals stage preferences on a given policy to follow established norms of democracy, the direction and length of those staged preferences should be observable when comparing actual referendum results to declared attitudes in surveys.

To be sure, as mentioned above, surveys are not the source of information that politicians trust the most when learning about public opinion (Walgrave & Soontjens, 2023). However, using surveys as a comparison is done primarily to allow one to measure the elusive concept of preference falsification, which is typically very hard to capture. To be sure, politicians may update their priors by directly comparing the two figures themselves—and, indeed, previous work has shown that they adapt their behavior to evidence from surveys (Hager & Hilbig, 2020).[16] However, this information update can also come about if the referendum results update the priors that they had built based on

other sources of information on citizens preferences, like daily conversations with them, protests, or focus groups. All of these behaviors have been shown to inform the perceptions that politicians form about the distribution of voters' preferences, and all of them are likely to be affected by social norms and preference falsification (Bergen & Labonté, 2020; Carlson & Settle, 2022; Opp, 2001; Walgrave & Soontjens, 2023).

To conduct the analyses, four ingredients are needed. First, data on the official referendum results. With the help of excellent research assistants, I collected such data from the `Swissvotes` platform.

The second necessary ingredient is data on the declared behavior of individuals in surveys—so that one can check whether the official electoral result represents a positive or negative update vis-a-vis the polled opinions of citizens. I rely on data with data from VOX surveys, which ask a language-stratified random sample of the Swiss population about their behavior in referendums. These surveys provide information on whether individuals voted in a referendum and, if they did, what their vote was. Also with the help of research assistants, I have collected all these surveys and calculated the percentage of respondents who report to have voted yes and no.[17]

It is important to note that the interviews in the surveys included in my sample were conducted over the phone. This means that the survey forces an interaction with the interviewer. It is a well-established regularity of survey research that this type of interview can make individuals feel that they need to manage their image and provide what they think is the socially desirable answer— much more so than in modes of interview that do not require an interaction with the interviewer, like self-administered or internet interviews (Kreuter et al., 2008; Tourangeau et al., 2000).

The third necessary ingredient is data on the official position that parties took on each referendum, so that one can assess the effect of this new information on the specific position that the party took. To make such analysis possible, I coded the position of each Swiss party on the policy being voted in all referendums in the sample.

Finally, one needs data on the ideological positions of those parties in elections—to check whether they adapt such positions to referendum results. I rely on data from the CMP project, which annotates the electoral manifestos of parties in a number of countries worldwide, in every election. Based on that, they provide variables with the party's positions on a number of issues. The CMP is a common source of data for analyses of the policy positions of parties (e.g., Abou-Chadi and Krause, 2018; Borbáth & Gessler, 2023; Düpont & Rachuj, 2022; Margalit et al., 2022). For my purposes, it presents the advantage of including all the elections in my sample and, for each of those elections, includes many parties. Other data sources, like the V-Party project (Lindberg et al., 2022) that I use for some additional analyses, have less parties and using them would significantly reduce the number of observations.

Since my hypothesis is about how vote results affect the behavior of *far-right politicians specifically*, I need to code which fall into that category. To that end, I add a dummy for far-right parties, which, following secondary literature, is coded 1 for the Swiss Democrats (Mudde, 2007), the Swiss People's party (Coffé & Voorpostel, 2010; Fitzgerald and Lawrence, 2011; Stockemer, 2012), Federal Democratic Union (Skenderovic, 2009), Freedom Party of Switzerland (Mudde, 2007), and Ticino League (Albertazzi, 2007; Bernhard, 2017).

My main outcome of interest is the extent to which each party puts forward exclusionary views in a given election. I focus on the CMP variable that measures each party's general statements on minority groups. This variable was constructed by asking the CMP coders whether the manifesto of each party makes positive mentions to minorities typically targeted by far-right discourse, such as immigrants and sexual minorities.[18,19] This type of general opposition to minorities is probably the most distinctive position of far-right parties, and one of core reasons why the views of far-right politicians are at odds with liberal democracy (Mudde, 2007).

The main independent variable of interest is the average difference between the vote for a party's position in the referendum and the declared vote for the same position in surveys. I call this variable *electoral surplus*. The intuition behind this measure is that it captures the difference between what voters privately do (as revealed in the referendums) and what they typically declare (in surveys), averaged out over the previous legislative period (four years before each election). The variable thus intends to tap the extent to which citizens are engaging in preference falsification, and the extent to which referendum results reveal that there is more private support for the general positions of a party—over a number of different policy areas—than typically meets the eye.[20] To calculate electoral surplus, I start with looking at all the referendums that take place over one legislature, and the positions that each party took on them. For each of those positions, I calculate the difference between the official vote share for that position and the vote share for that position as declared in the surveys.

After merging all datasets, I end up with a dataset (which, from now on, I call "dataset A") where each row is one party*referendum. I know the percentage of citizens who voted *yes* and *no* in each referendum, the percentage of respondents who declare to have voted *yes* and *no* in the survey, the official position of the parties on that referendum, the electoral surplus of that party's position on the referendum, whether the party is far-right or not, and that party's position on minorities in the subsequent election. This dataset includes 329 referendums that took place between 1981 (the date of the first VOX survey available) and 2019 (the date of the last Swiss election at the time of writing), and a total of 4105 party*referendum dyads. Table A.1 in the Online Appendix shows the full list of referendums.

In the elite-level analyses, the goal is to check the effect of electoral surplus on the ideology of parties in the subsequent election. This means that each unit should be one party*election, not one party*referendum. To enable these analyses, I group the data that way, meaning that each observation represents one Swiss party in one general election. The electoral surplus variable thus represents the average surplus of that party's referendum positions over the previous legislature. This second dataset (which, from now on, I call "dataset B") includes 128 observations from 16 parties in ten elections (1983–2019). Table B.2 show the descriptive statistics of the variables used in the elite-level analyses throughout the paper.

Results

Citizen Level: Staged Democrats and the Incentives to Hide Support for Far-Right Positions

As discussed above, the main contribution of this paper lies on the elite-level argument. However, a necessary precondition for that argument to hold is that should be able to find evidence of preference falsification and staged democrats with the data at hand. To check whether this precondition is empirically supported, I look at whether the positions of far-right politicians are more under-reported in surveys than the positions of other politicians. This would suggest that individuals who support policies backed by far-right parties often engage in preference falsification to follow the norm against far-right support.

To test this hypothesis, I take the 329 referendums in dataset A and reshape it into long format, so that the official and declared support for each of the positions (yes and no) appear in different rows in the dataset. Then, I regress the support for a given position on a dummy for whether that support is measured in official referendum votes or in surveys, a dummy for whether that position was supported by any far-right party, and the interaction of the two.[21] To control for differences across referendums, I include referendum fixed effects.

The results, shown in Table 1, support the staged democrats hypothesis. As the interaction coefficient for *Official vote x Any FRP supports position* shows, the gap between official and declared support for a given position is much higher if that position is backed by far-right parties. On average, a position that is backed by at least one far-right party has 4.5 percentage points more hidden vote than one that is not backed by any far-right party. This makes it almost three times higher than the gap between official and declared vote for positions that are not backed by far-right parties—which, as indicated by the coefficient for *Official vote (dummy)*, is about 1.7 percentage points. Table B.1 in the Online Appendix further shows that this gap is bigger the more far-right parties support a given position.

Table 1. Support for Referendum Positions Depending on whether It is Measured via Surveys or Official Data, and on whether That Position is Supported by Far-Right Parties.

	(1)
Official vote (dummy)	1.692**
	(0.770)
Any FRP supports position (dummy)	14.36***
	(3.517)
Official vote x any FRP supports position	4.479***
	(0.918)
Constant	37.74***
	(1.826)
N	1282

Standard errors in parentheses.
All models include referendum fixed effects.
The outcome is the support for each referendum position, be it official or declared in surveys.
*p < .10, **p < .05, ***p < .01.

This gap between private and declared preferences is the building block behind my main independent variable in the analyses of the next section: electoral surplus. As discussed in the previous section, electoral surplus captures the average distance between the official and declared support for a party's referendum positions, averaged out over the four years in a legislative period.[22]

Elite Level: Information About Private Views and Supply of Exclusionary Ideology

I now move to the analyses of the elite-side level of the argument. The goal in this section is to check whether the information provided by electoral surplus makes far-right politicians put forward more exclusionary policy in the subsequent election.

I start with plotting the correlation between electoral surplus and a party's position on minority groups in their subsequent manifesto. This is shown in Figure 1. The horizontal axis represents the mean electoral surplus for the party's position in the previous legislature. Positive values mean that, on average, a party's position had higher support in official results than it did in surveys. Negative values mean that, on average, a party's position had lower support in official results than in surveys. Each point represents one party-*election. The dotted blue line corresponds to far-right parties, while the solid black line corresponds to the remaining parties.[23]

The Figure is in line with theoretical expectation. When it comes to far-right parties, higher electoral surplus is associated with less positive references

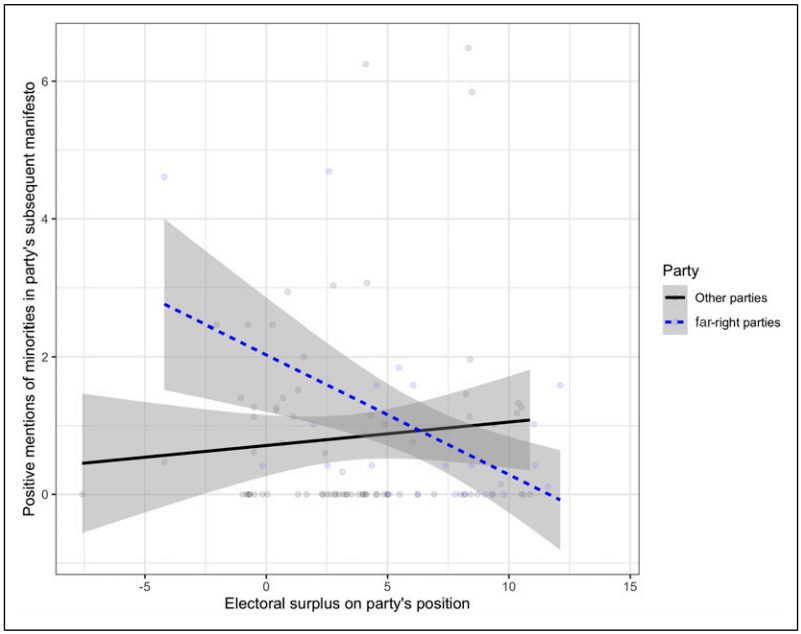

Figure 1. Information from referendums and the position of Swiss parties on minorities in the subsequent election.

to minority groups in the subsequent manifesto. This is not the case, however, when it comes to parties that are not far-right. In the case of these parties, if anything, the effect is the opposite.

While this Figure provides preliminary evidence in support of my theoretical argument, it represents a simple correlation. Many variables other than the actual electoral surplus can explain the pattern found. For these reasons, I move to analyses of fixed-effects regression models. The outcome variable is positive mentions of minority groups. As independent variables, I include the mean electoral surplus for the positions of each party in the previous legislative period. Then, I interact this variable with a dummy coded 1 for far-right parties and zero for parties that are not far right. I also add party fixed effects that control for all time-constant characteristics of the parties, as well as year fixed effects that control for trends in the outcome variable.

The results are shown in Table 2. The table includes two models: with clustered standard errors and with bootstrapped standard errors, given the small number of clusters (16) (Cameron et al., 2008). Regardless of the model one looks into, the results support the theoretical expectation. The main effect for electoral surplus is positive. This suggests that, in line with Figure 1 above, in the case of parties that are not far-right, an electoral surplus in their positions is associated—if anything—with

Table 2. Information From Referendums and the Position of Swiss Parties on Minorities in the Subsequent Election (Two-Way Fixed Effects Models).

	(1)	(2)
Mean electoral surplus in party's position	0.139	0.139
	(0.0827)	(0.0868)
Electoral surplus x far right party	−0.291**	−0.291**
	(0.110)	(0.119)
Constant	1.813***	1.813***
	(0.258)	(0.520)
Standard errors	Clustered	Bootstrapped

Standard errors in parentheses.
The outcome is a variable tapping positive mentions of minority groups by each party.
All models include party and year fixed effects.
*p < .10, **p < .05, ***p < .01.

more positive views of minority groups. The interaction effect, however, is negative and significant in both models. This suggests that, unlike what happens with the remaining parties, far-right parties react to an electoral surplus of their positions by making their policy positions *less* positive of minority groups. As shown in Table B.3, this effect is negative and significant even if one focuses only on the subsample of far-right parties.

The Online Appendix reports a number of additional robustness checks. Table B.4 shows that the results hold if one calculates electoral surplus as a proportion, instead of an absolute difference. Table B.5 presents a placebo that takes the lagged outcome as the dependent variable. Table B.6 replicates the analyses using the log-transformation of the dependent variable. Table B.7 uses different ways of dealing with standard errors. Table B.8 adds party fixed effects alone. Figures 9 and 10 present two robustness checks that check whether the results hold after sequentially removing each election and each party from the sample, respectively. I also replicate the main analyses using different outcomes. Tables B.18 and B.19 do so by looking into a battery of outcomes from the V-party project. Tables B.20 and B.21 do so by drawing upon other—if more indirect—proxies of exclusionary ideology included in the CMP. Table B.9 shows that the results are slightly stronger when one draws upon referendums that are not citizen initiatives. Table B.10 shows that the results remain very similar if one removes initiatives that were backed by political parties. Table B.12 focuses only on large parties. Table B.13 shows that the results hold using both referendums where the position of the party won and where that position lost. I also deal with the possibility of unholy alliances where far-right and left parties support the same referendum positions. Table B.14 replicates the analyses after removing the leftmost parties from the sample, while Table B.15 removes referendums where those parties supported the same positions as far-right parties.

In the Online Appendix, I also look into the effect on parties that are not far right. Table B.17 shows that, when it comes to these parties, the effect of electoral surplus depends on their ideological position. For more leftist parties, the effect is positive: these parties react to the information provided by that surplus by becoming more *positive* of minorities. However, as parties become more rightist, that effect becomes increasingly negative.

Finally, in the Online Appendix I also deal with a potential alternative explanation. It could be that the results are simply driven by an over-representation of voters in surveys (Sciarini & Goldberg, 2017), or by sampling bias (Lowe & McCormick, 1955) or non-response bias (Voogt & Van Kempen, 2002). These biases can create a distance between declared vote shares and official vote shares without it necessarily meaning that preference falsification is taking place.[24] This point may be particularly worrisome particularly important given that, in Switzerland, turnout is historically low and the sets of voters who show up to vote on a referendum depend on the issue being voted on (Lutz, 2007). To address these concerns, in Table B.16 I replicate the analyses using a different independent variable. Instead of electoral surplus, I use the raw vote share for each party's position in the referendums. The logic is that the higher that vote share is, the more likely it is to be perceived as *surprisingly high*. The results hold.

Plausibility Checks

I now conduct two plausibility checks which refer to additional implications of the argument. First, if the effect is, indeed, driven by the information that politicians gather from the referendum, it should be stronger in the first election in the sample. In subsequent elections, parties should have less additional information to learn, and the effect should become weaker.

To test this hypothesis, I replicate the main analyses of Table 2 on two subsets of the data: one including only the first election in the sample (in 1983), and another including the remaining elections. In the model for the first election, standard errors are robust, because each party is included in the sample only once. In the model for the remaining election, standard errors are clustered by party. Figure 11 in the Online Appendix shows that the results remain very similar if one uses bootstrapped standard errors instead. Moreover, in the analyses of 1983, one cannot include fixed effects because there is only one election. It is also important to note that splitting the sample this way reduces the number of observations significantly, especially for the analyses of the first election. With this being said, this is still an implication of the argument, on which one should provide evidence.

The results of these analyses are shown in Figure 2. The figure shows two coefficients—one for each of the subsamples described above. Both refer to the interaction effect between a party being far-right and electoral

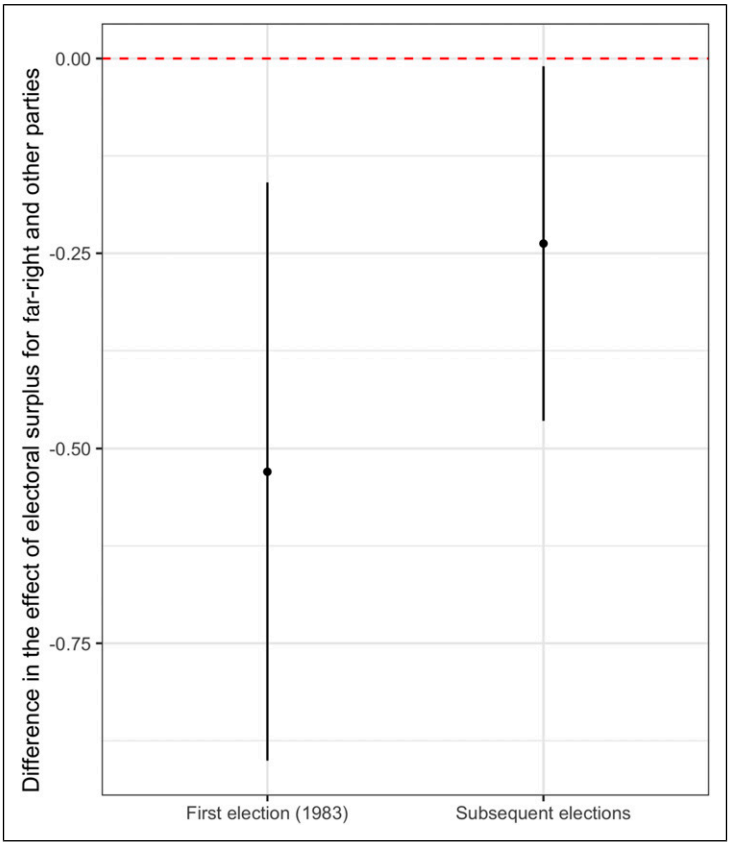

Figure 2. Comparting the effect of the information from referendums in the first election in the sample and in the subsequent ones. *Notes.* The outcome is a variable tapping positive mentions of minority groups by each party. Coefficients represent the interaction effect between electoral surplus and the dummy for far-right parties, which is the main coefficient of interest in Table 2. Negative values mean that electoral surplus makes far-right parties put forward less positive views of minority groups, when compared to other parties. Vertical lines denote 90% confidence intervals.

surplus—the main coefficient of interest in Table 2. In line with the argument, the Figure shows that the effect is clearly stronger in the first election than in the remaining ones. Table B.11 in the Online Appendix replicates these analyses in the form of a three-way interaction between electoral surplus, the dummy for far-right parties, and year of election.

As a second plausibility check, I look at how the effect varies depending on the topic of the referendum. The analyses of Table 2 include referendums on all topics. This is done in an effort to remain agnostic about

what constitutes a far-right topic. I assume that any position supported by far-right parties can, in principle, become construed as a far-right issue in the referendum campaign. In so doing, the results capture the effect of learning that the *general position* of far-right parties is more privately supported than one might have previous assumed. However, if the results are driven by referendums providing information that makes politicians realize that there is more support for far-right policy than usually meets the eye, these effects should be stronger when that information comes from referendums on topics that are more closely related to minorities.

To test this hypothesis, I replicate the analyses split by referendum topic. Taking advantage of the fact that `Swissvotes` classifies the topic associated with each referendum, I generate new variables that indicate the mean electoral surplus of each parties position—only in referendums that fall under each specific topic.[25] Then, I replicate the analyses of Table 2 replacing the variable that indicates electoral surplus with a variable that indicates electoral surplus in referendums on each of these topics.

The results are shown in Figure 3. Entries again represent the interaction coefficient between the electoral surplus of a party's position and a dummy for far-right parties. The Figure includes two facets, which mirror the two models in that table. As one might expect, the strongest effects are from referendums

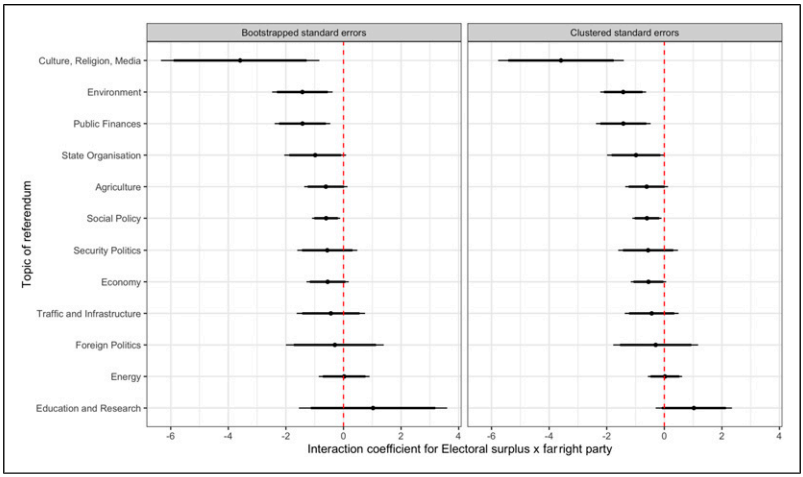

Figure 3. Effect of information from referendums, depending on the topic of the referendum. *Notes.* The outcome is a variable tapping positive mentions of minority groups by each party. Coefficients represent the interaction effect between electoral surplus and the dummy for far-right parties, which is the main coefficient of interest in Table 2. Negative values mean that electoral surplus makes far-right parties put forward less positive views of minority groups, when compared to other parties.

on the topic most clearly linked to minorities—those on culture, religion and media,. Conversely, the two topics on which there is most clearly no effect are two topics clearly not associated with minorities—energy, and education and research.

Conclusion

A citizen-level consensus in favor of democracy is arguably the best safeguard against politicians with views at odds with democracy. If voters have strong attachment to democratic values and punish those who reject them, it is very costly for politicians to put forward exclusionary policies or engage in backsliding practices. This paper has argued, however, that oftentimes what looks like a consensus in favor of democratic values may not actually be one. Established norms of democracy provide citizens with an incentive to act as though they agree with the norm when, in fact, they do not. Observing that citizens behave as though they support democracy is thus compatible with two states of the world: one where they are truly supportive of norms of democracy; or one where they stage their support for those norms to avoid social sanctions. The two states of the world are often observationally equivalent, but their implications are not. The latter provides more incentives to put forward counternormative policy than the former.

My argument does away with a puzzle in previous research on democratic support and democratic backsliding. A large body of research has found that existing levels of democratic support do not suggest that citizens are overwhelmingly turning away from democracy (Kriesi, 2020; Norris, 1999, 2011, 2017; Wuttke et al., 2022). At the same time, however, another body of research has found that many citizens are not often willing to punish those who breach democratic norms and conventions—especially when those are carried out by their preferred party (Graham & Svolik, 2020; Simonovits et al., 2021; Svolik, 2022)—or are willing to trade off democratic institutions for other benefits, like economic ones (Neundorf et al., 2024). The argument in this paper suggests that one possible explanation for these patterns is that some of these citizens may constitute what I have termed *staged democrats*. They do not strongly support norms of democracy; they simply behave as though they did to avoid social sanctions. As such, they have no incentive to change their vote choice to punish politicians who breach those norms.

The findings also speak to the potential pernicious effects of direct democracy acts. While such acts may increase responsiveness, they can also push far-right politicians to run for election on platforms that are less inclusive of minorities. Under that light, the conclusions of this paper echo previous concerns about how direct democracy may lead to normatively

undesirable outcomes—especially when the policies that are being voted on are related to minorities (e.g., Donovan, 2012). More broadly, the results also align with arguments about how institutional features that seem clearly democratic—like referendums or party primaries (Levitsky & Ziblatt, 2019)—can bring about troublesome outcomes for liberal democracy.

What do these findings mean for democratic resilience and preventing the success of illiberal actors? First and foremost, they suggest that norms generating preference falsification are only a weak form of prevention (Ogunye, 2021). In the short term, these norms can mask support for illiberal actors, make them underestimate how much latent support there is for policies at odds with liberal democratic values, and lower incentives to provide them. However, because this equilibrium is based on preference falsification, these norms can change very fast, and information revealing the sincere preferences of voters can undo their effect.

Second, the findings have implications for diagnoses of democratic resilience. The paper suggests that democratic citizens often have an incentive to stage support for norms of democracy. It is thus challenging to assess the true prevalence of supporter for these norms based on self-declared views—like traditional survey answers. Variation in those survey answers—both across time and across space—may reflect changes in how comfortable individuals feel expressing counternormative views, more than they reflect an actual change in their preferences. As such, an assessment of the prevalence of undemocratic views has to devise strategies that are able to do away with the preference falsification that can be taking place—a point previously made about other counternormative views like racism (Kuklinski, Cobb, and Gilens, 1997; Kuklinski et al., 1997).

The findings also have implications for the design of policies and interventions aimed at promoting democratic resilience. Such interventions should be careful to generate sincere instead of staged democratic support. While it may be hard to predict ex ante what type of interventions will generate each type of support, those based on social pressure and peer punishment alone seem more likely to generate preference falsification. This type of interventions can work in the short-term, and are potentially more effective than the absence of any intervention. However, policies that are to be effective in the long-term should try to persuade citizens and change their private views—by addressing the actual preferences of citizens, and not simply their translation into behavior (Ogunye, 2021).

At the same time, the findings come with some limitations and scope conditions. One limitation lies on the reliance on a single case. A key peculiarity of the Switzerland is the strength of the main far-right in the country (the SVP), one of the most successful in Europe. On the one hand, the electoral strength of this party could make the findings shown here a conservative

estimate because it leaves less room for misperceptions about the likely success of far-right policies. On the other hand, the direct democracy setting means that there are more opportunities for politicians to learn about the sincere preferences of voters, which might make these effects more likely to come about. Ultimately, a truly definitive answer to the question of how much these findings travel to other cases can only be obtained by replication. Future research can add to the conclusions of this paper by finding similar designs to the one proposed here in other countries.

Acknowledgments

I would like to thank Patricia Funk for kindly sharing data with me, which helped collecting some of the data on referendums. I am extremely thankful to Angela Odermatt and Peter Bristow for their excellent research assistance. I would also like to thank Larry Bartels, Carles Boix, Moritz Bondeli, Giovanni Capoccia, Mihail Chiru, Rafaela Dancygier, Sara Wallace Goodman, Gretchen Helmke, Stathis Kalyvas, Isabela Mares, Tali Mendelberg, Angela Odermatt, Temi Ogunye, Milan Svolik, Katerina Tertytchnaya Klaudia Wegschaider, and Kurt Weyland for helpful comments on early versions of the manuscript. The paper has benefitted from helpful presentations at two editions of the Conference *Back from the Brink* at Yale and Oxford, at the Conference on *Democratic Frontsliding* at Princeton University and at the 2023 Annual Meeting of the American Political Science Association. I am thankful to participants for their comments. Finally, I am thakful to the editors and anonymous reviewers at Comparative Political Studies for their thorough and constructive engagement with earlier versions of the paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Vicente Valentim https://orcid.org/0000-0003-4837-4258

Data Availability Statement

Replication materials and code can be found at Valentim (2024b).

Supplemental Material

Supplemental material for this article is available online.

Notes

1. It should be noted, however, that politicians of other ideologies can also threaten democratic survival (e.g., Cella et al., 2023). However, at least in Europe, democracy typically erodes from the right (Svolik et al., 2023).
2. It should be noted, however, that in other contexts these politicians sometimes portray themselves as strategically inclusive of these groups (Moffitt, 2017; Turnbul-Dugarte & Ortega, 2023; Yermakova, 2021).
3. It should be noted that a third strand of studies comes to a more midway view. This literature argues that there some worrying trends, but these are specific to some cohorts and countries. As such, they do not represent a widespread crisis of democracy in Western countries (Kriesi, 2020; Wuttke et al., 2022).
4. The idea that perceptions of acceptability about different related attitudes and behaviors are formed and move in tandem builds on previous work on political norms (Valentim, 2024c).
5. While norms can be formed against *preferences*, those preferences will only be known if translated into behavior. This means that, eventually, norms will always be enforced by sanctioning some behavior—even if that behavior is simply the expression of a given preference (Alvarez-Benjumea & Valentim, 2024; Kuran, 1987, 1995; Noelle-Neumann, 1984; Valentim, 2024a, 2024c).
6. In the Swiss case concretely, which I study in the empirical sections of the paper, voters are also likely to be aware of the costs of breaching these norms. For example, after the far-right party SVP scored a good electoral result in 2023, it promised to push for "less political correctness" (The Guardian, 2023)—an expression that the far-right often uses to refer to norms against some of their views (Rydgren, 2017). That the party chose opposition to this phenomenon as a key issue in their post-election communication suggests that there is a perception in Swiss society that far-right views are still likely to be met with social sanctions.
7. This typology considers only individuals who do declare to support norms of democracy, because the goal here is to problematize the assumption in previous research that such declared support is sincere. Of course, a society is composed both of these individuals and of those openly breach norms of democracy.
8. The concepts of sincere and staged democrats represent ideal types. They are treated as a binary distinction here for ease of exposition and analysis. To be sure, the distinction between staged and sincere democrats is likely to be a matter of degree. Some individuals can stage their preferences more often, or more intensely. At the same time, there may be variation in the types of policies that are affected by this phenomenon, and there may be over-time variation in the extent to which a given citizen is likely to stage their preferences. Future research may provide further depth to these analyses by digging deeper into some of these questions.
9. Following recent literature, I conceive of far-right parties as a group consisting of both radical-right and extreme-right parties (e.g., Mudde, 2019; Rooduijn et al., 2023).

Radical-right parties are *radical* in that they favor profound transformations in liberal democratic systems, and *right* in that they believe in a natural order with inequalities. Extreme-right parties differ in that they openly oppose democracy as a form of government (Mudde, 2007, p. 31). However, since both these party families breach established norms of democracy (Valentim, 2024c), the theoretical reasoning here applies to both. That is the reason why I focus on this umbrella term rather than on radical- or extreme-right parties.

10. Figures 4 and 5 in the Online Appendix suggest that this norm extends to the Swiss setting, which I draw upon in the empirical analyses.
11. It is often the case that elites perceive voters to be more conservative than they actually are (Broockman & Skovron, 2018; Pilet et al., 2024)—although Belchior (2014) finds a bias in the opposite direction. However, my argument should hold regardless of elites having a conservative bias, a liberal one, or neither. What is crucial for the argument is that politicians gather receive information suggesting that, on a specific topic, voters were more supportive of far-right-supported policies than anticipated. Such information updates can take place if one assumes a distribution of the error in elite perceptions such that sometimes they are surprised by the share of the vote that those policies receive. This can happen even if their overall bias is, on average, in the conservative direction. Moreover, it is plausible to think that, even despite having an overall conservative bias, elites may underestimate the extent to which there are *far-right views specifically* in their society, given how strongly these are affected by social norms. In other words, politicians may perceive that there are more voters in the conservative spectrum than there really are, but underestimate how many of them are at the rightmost end of the distribution. This interpretation is in line with recent work showing that European elites are worse at representing the view of voters on cultural issues—especially migration—in which they are significantly more liberal (Günther, 2024). This suggests that there is room for their priors to be updated and to learn that the citizenry is more conservative *on these topics specifically*.
12. To be sure, the probability of pushing back against a politician engaging in norm-eroding practices is not the same for all individuals—even for those who are sincere democrats. Moving away from the status quo should be more costly and threatening for potential losers (Goodman, 2014). However, the argument here is that, if citizens overwhelmingly support norms of democracy in the ways that many survey questions suggest (Svolik, 2022), politicians should have no incentive to put forward ideology, rhetoric, and policy that breaches those norms. The reason why citizens are permissive of these practices, I argue, is that not all of them sincerely support norms of democracy.
13. The effect of the revelation of this information should be felt even if it does not consist of a truly exogenous shock. Even if parties strategically supported referendums to learn about the private views of citizens on a given topic, there should nevertheless be some uncertainty about the actual outcome of that vote. The more

that outcome suggests that preference falsification is going on, the more it should affect the subsequent behavior of politicians. That being said, Table B.10 in the Online Appendix shows that the results are very similar when one focuses only on referendums that are not sponsored by political parties.

14. To be sure, Switzerland is often lauded as a country with a strong democratic system. As such, one might wonder why a study on how attitudes can safeguard liberal democracy should draw upon this country. First, while Switzerland has had formal democratic procedures for long, it has been severely lacking in some democratic aspects. For example, women only won the right to vote in federal elections in 1971, and in the Canton of Appenzell Innerrhoden women suffrage was not implemented until 1990 when it was mandated by a decision from the Federal Supreme Court. At the same time, while the country does not seem to be in immediate risk of becoming an autocracy, in recent years some trends have led to concerns over the quality of liberal democracy in the country, especially when it comes to the inclusion of minorities. In a noteworthy example, in 2009 the country approved a ban on minarets following a referendum vote. Concerns about the protection and safety of minorities in the country have also been conveyed by the Minority Rights Group, who have recently expressed worries about rising levels of xenophobia and Islamophobic rhetoric in the country's public sphere (Minority Righs Group, 2023). This is an important point considering that the specific norm of democracy studied in the empirical sections of the paper (support for far-right parties and their policies) is very closely connected to the protection of minorities (Mudde, 2004).
15. To be sure, not all individuals vote in each referendum. This is especially the case in Switzerland, where voter turnout is comparatively very low (Blais, 2014). This, however, should not threaten the validity of the design because surveys ask respondents about their turnout in the referendum. I have only taken into account respondents who report to turn out in each referendum.
16. In the empirical analyses, I calculate the declared support for a policy by using respondents who turn out as the denominator. While this may raise concerns about whether politicians really do compare figures only among survey respondents who report to turn out, this decision was supported by three lines of reasoning. First, as noted in this passage, the comparison of survey and electoral behavior is done as a proxy of the comparison between what individuals privately think against what they are willing to declare to others. As such, the most direct comparison should be one of like with like: comparing declared support for a policy among those who declare to turn out with actual support for that policy among those who actually do turn out. In so doing, one effectively calculates the declared support for a given policy in a way that is equivalent to how the official support for that policy (the point of comparison) is also calculated. Second, even if we assume that politicians may learn directly from this comparison, results of public opinion polls from voting acts are typically reported with the share of those who declare to turn out as the denominator. Exit polls, for example, will report the vote share for

parties as adding up to 100, by focusing only on voters who turn out. Finally, the same approach has been used by previous research using a similar design to look into preference falsification in the survey responses of voters (Funk, 2016, p. 445). For this reasons, using the share of those who declare to turn out as the denominator seems to be the best approach.

17. I will be using these surveys to collect data on declared support for the specific policies of far-right parties, not support for those parties *in general*. However, previous work has shown that support for policies and ideological views associated with radical-right parties is itself typically regarded as counternormative—in much the same way as support for the parties per se (Blinder et al., 2013; Ivarsflaten, Blinder, and Ford, 2010; Valentim, 2024c). Moreover, the argument of the paper should apply to specific policies: it is by getting information about the sincere support for such policies that parties can revise their ideology in subsequent elections.
18. At the time of writing, the CMP did not provide a measure tapping negative mentions to those minorities, which is why I focus on *positive* mentions alone.
19. I focus on this outcome because opposition to minorities is one of the key ideological positions that far-right parties in different countries share (Mudde, 2004). However, focusing on positive mentions of minorities might raise concerns that the effects are driven by a blurring mechanism, rather than by parties talking more negatively about underprivileged groups. More generally, one might fear that the results are driven by the choice of the specific outcome variable to focus on. To address these concerns, Tables B.20, B.21, B.18 and B.19 in the Online Appendix replicate the analyses drawing upon other—if potentially more imperfect—measures of exclusionary views, which themselves come from different projects (CMP and V-Party). These tables show that the results hold even on outcomes that are more "negative" in nature—such as negative mentions of migrants in Tables B.20 and B.21, opposition to pluralism in Tables B.18 and B.19. That the results can be replicated using these clearly negative outcomes seems incompatible with the effects are being driven by a blurring mechanism.
20. To be sure, the reported behavior in surveys may, itself, be affected by the outcome of the referendum (Funk, 2016). However, to the extent that this may be the case, it should mean that my measure underestimates electoral surplus; but it should not affect comparisons across values of the measure. To address this concern more directly, in Table B.16 I replicate the analyses using not electoral surplus, but the overall vote share for a party's position as the main independent variable.
21. The coefficient for *Official vote* is a dummy indicating whether a given vote share is official (as opposed to declared in a survey). It is *not* the overall support for a given position, which is the outcome in these analyses.
22. Figure 7 in the Online Appendix shows the average electoral surplus for far-right parties and the remaining parties.
23. Figure 8 in the Online Appendix replicates this Figure using a loess instead of a linear model fit.

24. However, to the extent that politicians may learn from the very comparison of survey results against official results, this difference might still have the hypothesized effect—even if its origin is due to reasons other than the ones hypothesized here.
25. The electoral surplus for referendums on other topics is thus coded as 0. This is done to reflect the reasoning here, according to which referendums on other topics as treated as though no information had been learned from it.

References

Abou-Chadi, T., & Werner, K. (2018). The causal effect of radical right success on mainstream parties' policy positions: A regression discontinuity approach. *British Journal of Political Science*, *50*(3), 829–847. https://doi.org/10.1017/S0007123418000029

Albertazzi, D. (2007). Addressing 'the People': A comparative study of the Lega Nord's and Lega dei Ticinesi's political rhetoric and styles of Propaganda. *Modern Italy*, *12*(3), 327–347. https://doi.org/10.1080/13532940701633791

Almond, G. A., & Verba, S. (1963). *The civic culture: Political attitudes and democracy in five nations*: Princeton University Press.

Alvarez-Benjumea, A., & Valentim, V. (2024). The enforcement of political norms. *British Journal of Political Science*, 1–24. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/enforcement-of-political-norms/D11DFF765140B4E4608D99BA373B22BF (visited on 07/30/2024).

Ammassari, S. (2022). It depends on personal networks: Feelings of stigmatisation among populist radical right members. *European Journal of Political Research*, *62*(3), 723–741. https://doi.org/10.1111/1475-6765.12568

Ammassari, S. (2024). Deterrent or stimulus? How perceived societal stigma affects participation in populist radical right parties. *Political Studies*, Online First, *00323217241234527*. https://doi.org/10.1177/00323217241234527 (visited on 06/20/2024).

Andreoni, J., Nikiforakis, N., & Siegenthaler, S. (2021). Predicting social tipping and norm change in controlled experiments. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(16), Article e2014893118. https://doi.org/10.1073/pnas.2014893118

Anduiza, E., & Rico, G. (2022). Sexism and the far-right vote: The individual dynamics of gender Backlash. *American Journal of Political Science*, *68*(2), 478–493. https://doi.org/10.1111/ajps.12759

Art, D. (2011). *Inside the radical right: the development of anti-immigrant parties in Western Europe*. Google-Books-ID: 9dUTXJakqLoC. Cambridge University Press.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. https://doi.org/10.1037/h0093718

Baik, J., & Shen, X. (2023). Retrieving true preference under authoritarianism. https://voices.uchicago.edu/baik/files/2023/03/WritingSample2_PreferenceFalsificationMethod.pdf. (visited on 11/29/2023).

Belchior, A. M. (2014). Explaining MPs' perceptions of voters' positions in a party-mediated representation system: Evidence from the Portuguese case. *Party Politics*, *20*(3), 403–415. https://doi.org/10.1177/1354068811436046

Benjumea, A. Á. (2022). Uncovering hidden opinions: Social norms and the expression of xenophobic attitudes. *European Sociological Review*, *39*(3), 449–463. https://doi.org/10.1093/esr/jcac056

Bergen, N., & Labonté, R. (2020). "Everything is perfect, and we have no problems": Detecting and limiting social desirability bias in qualitative research. *Qualitative Health Research*, *30*(5), 783–792. https://doi.org/10.1177/1049732319889354

Bernhard, L. (2017). Three faces of populism in current Switzerland: Comparing the populist communication of the Swiss people's party, the Ticino League, and the Geneva Citizens' movement. *Swiss Political Science Review*, *23*(4), 509–525. https://doi.org/10.1111/spsr.12279

Bernhard, M. (2021). Democratic backsliding in Poland and Hungary. *Slavic Review*, *80*(3), 585–607. https://www.cambridge.org/core/journals/slavic-review/article/democratic-backsliding-in-poland-and-hungary/8B1C30919DC33C0BC2A66A26BFEE9553

Bicchieri, C. (2017). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bichay, N. (2022). Populist radical-right junior coalition partners and liberal democracy in Europe. *Party Politics*, *30*(2), 236–246. https://doi.org/10.1177/13540688221144698

Bischof, D., Lars Allinger, T., & Corre Juratic, M. L. E., Kristian Vrede Skaaning Frederiksen. (2023). *(Mis-)Perceiving support for democracy: The role of social norms for democracies*. https://osf.io/dpq7w/

Blair, G., Coppock, A., & Moor, M. (2020). When to worry about sensitivity bias: A social reference theory and evidence from 30 years of list experiments. *American Political Science Review*, *114*(4), 1297–1315. https://doi.org/10.1017/S0003055420000374. https://www.cambridge.org/core/journals/american-political-science-review/article/when-to-worry-about-sensitivity-bias-a-social-reference-theory-and-evidence-from-30-years-of-list-experiments/3B922CC54881BA694E4D4B07FD286060

Blais, A. (2014). Why is turnout so low in Switzerland? Comparing the attitudes of Swiss and German citizens towards electoral democracy. *Swiss Political Science Review*, *20*(4), 520–528. https://doi.org/10.1111/spsr.12116. https://www.academia.edu/download/52987632/2014-Why_is_Turnout_So_Low_in_Switzerland.pdf

Blinder, S., Ford, R., & Ivarsflaten, E. (2013). The better angels of our nature: How the antiprejudice norm affects policy and party preferences in great Britain and Germany. *American Journal of Political Science*, *57*(4), 841–857. https://doi.org/10.1111/ajps.12030. https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12030

Borbáth, E., & Gessler, T. (2023). How do populist radical right parties differentiate their appeal? Evidence from the media strategy of the Hungarian Jobbik party. *Government and Opposition*, *58*(1), 84–105. https://doi.org/10.1017/gov.2021.28. https://www.cambridge.org/core/product/identifier/S0017257X21000282/type/journal_article

Broockman, D. E., & Skovron, C. (2018). Bias in perceptions of public opinion among political elites. *American Political Science Review*, *112*(3), 542–563. https://doi.org/10.1017/S0003055418000011. https://www.cambridge.org/core/journals/american-political-science-review/article/bias-in-perceptions-of-public-opinion-among-political-elites/2EF080E04D3AAE6AC1C894F52642E706

Bugarič, B. (2015). A crisis of constitutional democracy in post-communist Europe: "Lands in-between" democracy and authoritarianism. *International Journal of Constitutional Law*, *13*(1), 219–245. https://doi.org/10.1093/icon/mov010. https://academic.oup.com/icon/article-abstract/13/1/219/689918

Bursztyn, L., Egorov, G., & Fiorin, S. (2020a). From extreme to mainstream: The erosion of social norms. *The American Economic Review*, *110*(11), 3522–3548. https://doi.org/10.1257/aer.20171175. https://www.aeaweb.org/articles?id=10.1257/aer.20171175

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020b). Misperceived social norms: Women working outside the home in Saudi Arabia. *The American Economic Review*, *110*(10), 2997–3029. https://doi.org/10.1257/aer.20180975. https://www.aeaweb.org/articles?id=10.1257/aer.20180975

Butler, D. M., & Nickerson, D. W. (2011). Can learning constituency opinion affect how legislators vote? Results from a field experiment. *Quarterly Journal of Political Science*, *6*(1), 55–83. https://doi.org/10.1561/100.00011019. https://sites.temple.edu/nickerson/files/2017/07/Butler_Nickerson.QJPS11.pdf

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427. https://www.jstor.org/stable/40043157

Capoccia, G. (2023). *Countering illiberalism in liberal democracies: Actors, strategies, temporalities*. Working Paper.

Carlson, T. N., & Settle, J. E. (2022). What goes without saying: Navigating political discussion in America. *English* (New edition). Cambridge University Press.

Carnevale, A. P., Smith, N., Dražanová, L., Gulish, A., & Campbell, K. P. (2020). *The role of education in taming authoritarian attitudes*. Tech. Rep. Georgetown university center on education and the workforce ERIC number: ED609008. Georgetown University Center on Education and the Workforce. https://eric.ed.gov/?id=ED609008

Carvajal, H., Cox, L., & Garbiras-Diaz, N. (2023). Easier said than done: Citizens' stated and revealed democratic commitment in Latin America. Working paper.

Cella, L., Çınar, I., Stokes, S., & Uribe, A. (2023). Building tolerance for backsliding by trash-talking democracy: Theory and evidence from Mexico. Working paper.

Chierchia, G., Pi-Sunyer, B. P., & Blakemore, S.-J. (2020). Prosocial influence and opportunistic conformity in adolescents and young adults. *Psychological Science*, *31*(12), 1585–1601. https://doi.org/10.1177/0956797620957625

Chiru, M., Giugal, A., Gutoiu, G., & George, S. (2024). The electoral returns of party-sponsored citizen legislative initiatives. *The Journal of Politics*.

Chu, J. A., & Recchia, S. (2022). Does public opinion affect the preferences of foreign policy leaders? Experimental evidence from the UK parliament. *The Journal of Politics*, *84*(3), 1874–1877. https://doi.org/10.1086/719007

Claassen, C., & Magalhães, P. C. (2023). Public support for democracy in the United States has declined generationally. In *Public opinion quarterly* (Vol 83, Issue 3, pp. 719–732). Oxford University Press.

Clayton, K., Davis, N. T., Nyhan, B., Porter, E., Ryan, T. J., & Wood, T. J. (2021). Elite rhetoric can undermine democratic norms. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(23), Article e2024125118. https://doi.org/10.1073/pnas.2024125118

Coffé, H., & Voorpostel, M. (2010). Young people, parents and radical right voting. The case of the Swiss people's party. *Electoral Studies. Special Symposium: Voters and Coalition Governments*, *29*(3), 435–443. https://doi.org/10.1016/j.electstud.2010.03.015. https://www.sciencedirect.com/science/article/pii/S0261379410000326

Das, S. (2023). *Democratic backsliding in the world's largest democracy.* SSRN Scholarly Paper. https://papers.ssrn.com/abstract=4512936

Delgado-Gaitan, C. (1988). The value of conformity: Learning to stay in school. *Anthropology & Education Quarterly*, *19*(4), 354–381. https://doi.org/10.1525/aeq.1988.19.4.05x0922l

Dinas, E., Martínez, S., & Valentim, V. (2024). Social norm change, political symbols, and expression of stigmatized preferences. *The Journal of Politics*, *86*(2), 488–506. https://www.journals.uchicago.edu/doi/abs/10.1086/726951

Donà, A. (2021). Radical right populism and the backlash against gender equality: The case of the Lega (Nord). *Contemporary Italian Politics*, *13*(3), 296–313. https://doi.org/10.1080/23248823.2021.1947629

Donovan, T. (2012). Direct democracy and campaigns against minorities. *Minnesota Law Review*, *97*, 1730. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/mnlr97&section=46&casa_token=3db9_RBev5EAAAAA:boXm-ki8D0RHMXoiwPRXiTfbzXaI90zuASSTiEYsvFt1dJLV6qouFz9rDkRBvMTpMQ2jFqpT

Doosje, B., van den Bos, K., Loseman, A., Feddes, A. R., & Mann, L. (2012). "My in-group is superior!": Susceptibility for radical right-wing attitudes and behaviors in Dutch youth. *Negotiation and Conflict Management Research*, *5*(3), 253–268. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1750-4716.2012.00099.x

Düpont, N., & Rachuj, M. (2022). The ties that bind: Text similarities and conditional diffusion among parties. *British Journal of Political Science*, *52*(2), 613–630.

https://doi.org/10.1017/s0007123420000617. https://www.cambridge.org/core/product/identifier/S0007123420000617/type/journal_article

Easton, D. (1965). *A systems analysis of political life*. Wiley Publishing.

Eriksson, K., Strimling, P., Gelfand, M., Wu, J., Abernathy, J., Akotia, C. S., Aldashev, A., Andersson, P. A., Andrighetto, G., Anum, A., Arikan, G., Aycan, Z., Bagherian, F., Barrera, D., Basnight-Brown, D., Batkeyev, B., Belaus, A., Berezina, E., Björnstjerna, M., Blumen, S., & Van Lange, P. A. M. (2021). Perceptions of the appropriate response to norm violation in 57 societies. *Nature communications*, *12*(1), 1481. https://doi.org/10.1038/s41467-021-21602-9

Erisen, C., Guidi, M., Martini, S., Toprakkiran, S., Isernia, P., & Littvay, L. (2021). Psychological correlates of populist attitudes. *Political Psychology*, *42*(S1), 149–171. https://onlinelibrary.wiley.com/doi/abs/10.1111/pops.12768

Ewing, G. (2001). Altruistic, egoistic, and normative effects on curbside recycling. *Environment and Behavior*, *33*(6), 733–764. https://doi.org/10.1177/00139160121973223

Fehr, E., & Gachter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, *90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980

Fitzgerald, J., & Duncan, L. (2011). Local cohesion and radical right support: The case of the swiss people's party. *Electoral Studies*, *30*(4), 834–847. https://www.sciencedirect.com/science/article/pii/S026137941100103X

Foa, R., & Mounk, Y. (2022). *The end of the consolidation paradigm: A response to our critics*. Johns Hopkins University Press.

Foa, R. S., & Mounk, Y. (2016). The danger of deconsolidation: The democratic disconnect. *Journal of democracy*, *27*(3), 5–17. https://doi.org/10.1353/jod.2016.0049

Funk, P. (2016). How accurate are surveyed preferences for public policies? Evidence from a unique institutional setup. *The Review of Economics and Statistics*, *98*(3), 442–454. https://doi.org/10.1162/REST_a_00585

Goeree, J. K., & Yariv, L. (2015). Conformity in the lab. *Journal of the Economic Science Association*, *1*(1), 15–28. https://doi.org/10.1007/s40881-015-0001-7

Goldman, J., Lee, D., & Pocasangre, O. (2024). Democracy hypocrisy: Examining America's fragile democratic convictions. Democracy Fund Report. https://democracyfund.org/idea/democracy-hypocrisy/

Goldstein, D. (2022). *The social foundations of democratic norms*. Working Paper.

Goodman, S. W. (2014). *Immigration and membership politics in Western Europe*. Cambridge University Press. https://www.cambridge.org/core/books/immigration-and-membership-politics-in-western-europe/6EEE2D381322F0DCFEBE11AD20025BFD

Graham, M. H., & Svolik, M. W. (2020). Democracy in America? Partisanship, polarization, and the robustness of support for democracy in the United States. *American Political Science Review*, *114*(2), 392–409. https://doi.org/10.1017/S0003055420000052

Groenendyk, E., Kimbrough, E. O., & Pickup, M. (2022). How norms shape the nature of belief systems in mass publics. *American Journal of Political Science*, *67*(3), 623–638. https://doi.org/10.1111/ajps.12717

Grzymala-Busse, A. M. (2002). *Redeeming the communist past: The regeneration of communist parties in East Central Europe*. Cambridge University Press.

Günther, L. (2024). Political Representation Gaps and Populism. *SSRN Working Paper*.

Hager, A., & Hilbig, H. (2020). Does public opinion affect political speech? *American Journal of Political Science*, *64*(4), 921–937. https://doi.org/10.1111/ajps.12516

Hale, H. E. (2022). Authoritarian rallying as reputational cascade? Evidence from Putin's popularity surge after Crimea. *American Political Science Review*, *116*(2), 580–594. https://doi.org/10.1017/S0003055421001052

Harteveld, E., Dahlberg, S., Kokkonen, A., & van der Brug, W. (2019). Social stigma and support for the populist radical right: An experimental study. *Scandinavian Political Studies*, *42*(3-4), 296–307. https://doi.org/10.1111/1467-9477.12153

Harteveld, E., & Ivarsflaten, E. (2018). Why women avoid the radical right: Internalized norms and party reputations. *British Journal of Political Science*, *48*(2), 369.

Helfer, L., Wäspi, F., & Varone, F. (2021). Does direct democracy enhance politicians' perceptions of constituents' opinions? Evidence from Switzerland. *Swiss Political Science Review 27*(4), 695–711. https://doi.org/10.1111/spsr.12495

Hertel-Fernandez, A., Mildenberger, M., & Stokes, L. C. (2019). Legislative staff and representation in Congress. *American political Science Review*, *113*(1), 1–18. https://www.cambridge.org/core/journals/american-political-science-review/article/legislative-staff-and-representation-in-congress/D7735FCF39B843B9F3269FD39362FD66 (visited on 06/25/2024).

Huber, R. A., & Schimpf, C. H. (2016). A drunken guest in Europe? *Zeitschrift für Vergleichende Politikwissenschaft*, *2*(10), 103–129. https://doi.org/10.1007/s12286-016-0302-0

Ivarsflaten, E., Scott, B., & Ford, R. (2010). The anti-racism norm in Western European immigration politics: Why we need to consider it and how to measure it. *Journal of Elections, Public Opinion and Parties*, *20*(4), 421–445. https://doi.org/10.1080/17457289.2010.511805

Jiang, J., & Yang, D. L. (2016). Lying or believing? Measuring preference falsification from a political purge in China. *Comparative Political Studies*, *49*(5), 600–634. https://doi.org/10.1177/0010414015626450

Jung, J.-H., & Tavits, M. (2021). "Do referendum results change norm perceptions and personal opinions?" en. *Electoral Studies*, *71*, 102307. https://doi.org/10.1016/j.electstud.2021.102307

Kalla, J. L., & Porter, E. (2021). Correcting bias in perceptions of public opinion among American elected officials: Results from two field experiments. *British Journal of Political Science*, *51*(4), 1792–1800. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/correcting-bias-in-perceptions-of-public-opinion-among-american-elected-officials-results-from-two-field-experiments/ACD664051FDA8FD407B79058FD578F78

Kertzer, J. D., Busby, J., Monten, J., Tama, J., & Kafura, C. (2019). *Elite misperceptions and the domestic politics of conflict*. Tech. Rep. Working paper. https://jkertzer.sites.fas.harvard.edu/Research_files/NATO-Misperceptions-Web.pdf (visited on 06/25/2024).

Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys: The effects of mode and question sensitivity. *Public Opinion Quarterly*, *72*(5), 847–865. https://doi.org/10.1093/poq/nfn063

Kriesi, H. (2020). Is there a crisis of democracy in Europe? *Politische Vierteljahresschrift*, *61*(2), 237–260. https://cadmus.eui.eu/handle/1814/70137 (visited on 04/06/2023).

Kuklinski, J. H., Cobb, M. D., & Gilens, M. (1997a). "Racial attitudes and the" new South. *The Journal of Politics*, *59*(02), 323–349. https://doi.org/10.2307/2998167. https://www.jstor.org/stable/2998167

Kuklinski, J. H., Sniderman, P. M., Knight, K., Piazza, T., Tetlock, P. E., Lawrence, G. R., & Mellers, B. (1997b). Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science*, *41*(2), 402–419. https://doi.org/10.2307/2111770

Kuran, T. (1987). Preference falsification, policy continuity and collective conservatism. *The Economic Journal*, *97*(387), 642–665. https://doi.org/10.2307/2232928

Kuran, T. (1995). *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press.

Lagios, N., Meon, P.-G., & Tojerow, I. (2022). *Is demonstrating against the far right worth it? Evidence from French presidential elections*. SSRN Scholarly Paper. https://papers.ssrn.com/abstract=4233804 (visited on 06/28/2024).

Leemann, L., & Wasserfallen, F. (2016). The democratic effect of direct democracy. *American Political Science Review*, *110*(4), 750–762. https://www.cambridge.org/core/journals/american-political-science-review/article/democratic-effect-of-direct-democracy/4CA3ECDD249268E630F564721E8B7DFE (visited on 06/25/2024).

Levitsky, S., & Ziblatt, D. (2019). *How democracies die* (Reprint edition). Broadway Books.

Lindberg, S. I., Düpont, N., Higashijima, M., Kavasoglu, Y. B, Marquardt, K. L, Bernhard, M., Döring, H., Hicken, A., Laebens, M., Medzihorsky, J., Neundorf, A., Reuter, O. J, Ruth—Lovell, S., Wiesehomeier, K. R, Wright, N., Alizada, N., Bederke, P., Gastaldi, L., & Brigitte, Seim. (2022). *V–Dem party coding units V2*. Varieties of Democracy (V–dem) Project.

Linz, J. J., & Stepan, A. (1996). *Problems of democratic transition and consolidation: Southern Europe, south America, and post-communist Europe*. jhu Press.

Lipset, S. M., & Rokkan, S. (1990). Cleavage structures, party systems, and voter alignments. In P. Mair (Ed), *The West European party system* (pp. 91–138). Oxford University Press.

Lowe, F. E., & McCormick, T. C. (1955). Some survey sampling biases. In: *Public opinion quarterly 19.3* (pp. 303–315). Oxford University Press. https://academic.oup.com/poq/article-abstract/19/3/303/1860598 (visited on 06/29/2024).

Lutz, G. (2007). Low turnout in direct democracy. *Electoral Studies*, *26*(3), 624–632. https://www.sciencedirect.com/science/article/pii/S0261379406001065 (visited on 06/29/2024. https://doi.org/10.1016/j.electstud.2006.10.008

Mackie, G. (1996). Ending footbinding and infibulation: A convention account. *American Sociological Review*, *61*(6), 999–1017. https://doi.org/10.2307/2096305

Margalit, Y., Slough, T., & Ting, M. M. (2022). After defeat: How governing parties respond to electoral loss. *Political Science Research and Methods*, *10*(4), 739–758. https://doi.org/10.1017/psrm.2021.62. https://www.cambridge.org/core/product/identifier/S2049847021000625/type/journal_article

Matsusaka, J. G. (2010). Popular control of public policy: A quantitative approach. *Quarterly Journal of political science*, *5*(2), 133–167. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6992be64941ae707f3ae4371ee8b1f2b14538b61 (visited on 06/25/2024).

Merkel, W. (2004). Embedded and defective democracies. *Democratization*, *11*(5), 33–58. https://doi.org/10.1080/13510340412331304598

Minority Righs Group. (2023). *Switzerland: Current issues*. Tech. rep. https://minorityrights.org/country/switzerland/ (visited on 06/28/2024).

Moffitt, B. (2017). Liberal illiberalism? The reshaping of the contemporary populist radical right in Northern Europe. *Politics and Governance*, *5*(4), 112–122. https://doi.org/10.17645/pag.v5i4.996, https://www.cogitatiopress.com/politicsandgovernance/article/view/996

Molho, C., Tybur, J. M., Van Lange, P. A. M., & Balliet, D. (2020). Direct and indirect punishment of norm violations in daily life. *Nature communications*, *11*(1), 3432. https://www.nature.com/articles/s41467-020-17286-2 (visited on 03/01/2021).

Mudde, C. (2004). The populist zeitgeist. *Government and Opposition*, *39*(4), 541–563. https://doi.org/10.1111/j.1477-7053.2004.00135.x

Mudde, C. (2007). *Populist radical right parties in Europe*. Cambridge University Press.

Mudde, C. (2019). *The Far Right Today*. Google-Books-ID: aD25DwAAQBAJ. John Wiley & Sons.

Mudde, C., & Rovira Kaltwasser, C. (2012). *Populism in Europe and the Americas: Threat of corrective for democracy?* Cambridge University Press. https://works.bepress.com/cas_mudde/70/ (visited on 01/19/2017).

Muñoz, J. (2012). *La construcción política de la identidad española: ?del nacionalcatolicismo al patriostismo democrático?* CIS.

Neundorf, A., Dahlum, S., Vrede Skaaning Frederiksen, K., & Öztürk, A. (2024). *The appeal of electoral autocracy: Assessing citizens' revealed societal preferences*. https://osf.io/agrjq (visited on 06/28/2024).

Noelle-Neumann, E. (1984). *The spiral of silence*. The University of Chicago Press. https://www.press.uchicago.edu/ucp/books/book/chicago/S/bo3684069.html (visited on 03/12/2018).

Norris, P. (1999). Critical citizens. *Global support for democratic governance*. Oxford University Press.

Norris, P. (2011). *Democratic deficit: Critical citizens revisited*. Cambridge University Press.

Norris, P. (2017). Is Western democracy backsliding? Diagnosing the risks. *The Journal of Democracy*, *28*(2).

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*, 1291–1298. https://doi.org/10.1038/nature04131

Off, G. (2023). Gender equality salience, backlash and radical right voting in the gender-equal context of Sweden. *West European Politics*, *46*(3), 451–476. https://doi.org/10.1080/01402382.2022.2084986

Ogunye, T. (2021). Are norms necessarily normative.

Ohtsuki, H., & Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, *239*(4), 435–444. https://doi.org/10.1016/j.jtbi.2005.08.008

Opp, K.-D. (2001). Social networks and the emergence of protest norms. In M. Hechter, & K.-D. Opp (Eds.), *Social norms* (pp. 243–273). Russel Sage foundation.

Patel, D. S. (2013). Preference falsification, revolutionary coordination, and the Tahrir square model. Working Papers. https://www.academia.edu/download/39234887/ProceedingsPatelE_08.21.13.pdf (visited on 06/28/2024).

Pereira, M. M. (2021). Understanding and reducing biases in elite beliefs about the electorate. *American Political Science Review*, *115*(4), 1308–1324. https://www.cambridge.org/core/journals/american-political-science-review/article/understanding-and-reducing-biases-in-elite-beliefs-about-the-electorate/FBD98EFF932D264B1D489FEE9C5A8872 (visited on 06/25/2024).

Pickup, M., Kimbrough, E. O., & De Rooij, E. A. (2022). Expressive politics as (costly) norm following. *Political Behavior*, *44*(4), 1611–1631. https://doi.org/10.1007/s11109-020-09667-6

Pilet, J.-B., Sheffer, L., Helfer, L., Varone, F., Vliegenthart, R., & Walgrave, S. (2024). Do politicians outside the United States also think voters are more conservative than they really are? *American Political Science Review*, *118*(2), 1037–1045. https://doi.org/10.1017/s0003055423000527. https://www.cambridge.org/core/journals/american-political-science-review/article/do-politicians-outside-the-united-states-also-think-voters-are-more-conservative-than-they-really-are/D21A9077EE2435F2B910394378E96450

Pirro, A. L. P. (2014). Populist radical right parties in central and eastern Europe: The different context and issues of the prophets of the patria. *Government and Opposition*, *49*(4), 600–629. https://www.cambridge.org/core/journals/government-and-opposition/article/populist-radical-right-parties-in-central-and-eastern-europe-the-different-context-and-issues-of-the-prophets-of-the-patria/D17567E9BC3A862DAD112FE1155AFCC6 (visited on 10/09/2023).

Przeworski, A. (1999). Minimalist conception of democracy: A defense. In I. Shapiro, & C. Hacker-Cordón (Eds), *Democracy's values*. Google-Books-ID: 8dBT8EeM7dgC. Cambridge University Press.

Romarri, A. (2020). Do far-right mayors increase the probability of hate crimes? Evidence from Italy. *Evidence from Italy*.

Rooduijn, M., Pirro, A. L. P., Halikiopoulou, D., Froio, C., Van Kessel, S., De Lange, S. L., Mudde, C., & Paul, T. (2023). The PopuList: A database of populist, far-left, and far-right parties using expert-informed qualitative comparative classification (EiQCC). *British Journal of Political Science*, *54*(3), 1–10. https://doi.org/10.1017/S0007123423000431. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/populist-a-database-of-populist-farleft-and-farright-parties-using-expertinformed-qualitative-comparative-classification-eiqcc/EBF60489A0E1E3D91A6FE066C7ABA2CA

Rost, K., Stahel, L., & Frey, B. S. (2016). Digital social norm enforcement: Online firestorms in social media. *PLoS One*, *11*(6), Article e0155923. https://doi.org/10.1371/journal.pone.0155923

Rydgren, J. (2017). Radical right-wing parties in Europe: What's populism got to do with it? *Journal of Language and Politics*, *16*(4), 485–496. https://doi.org/10.1075/jlp.17024.ryd

Sakki, I., & Pettersson, K. (2016). Discursive constructions of otherness in populist radical right political blogs. *European Journal of Social Psychology*, *46*(2), 156–170. https://doi.org/10.1002/ejsp.2142

Scheff, T. J. (1988). Shame and conformity: The deference-emotion system. *American Sociological Review*, *53*(3), 395–406. https://www.jstor.org/stable/2095647 (visited on 06/25/2024).

Schumpeter, J. A. (2013). *Capitalism, socialism and democracy*. Routledge. https://books.google.com/books?hl=es&lr=&id=MRg5crpAOBIC&oi=fnd&pg=PR2&dq=Capitalism,+Socialism,+and+Democracy&ots=oK2wUbhnw1&sig=-gVkMNQBkIgak75GaUsSTy_3A3A (visited on 10/10/2023).

Sciarini, P., & Goldberg, A. C. (2017). Lost on the way: Nonresponse and its influence on turnout bias in postelection surveys. *International Journal of Public Opinion Research*, *29*(2), 291–315. https://academic.oup.com/ijpor/article-abstract/29/2/291/2981868 (visited on 06/29/2024).

Sevenans, J. (2021). How public opinion information changes politicians' opinions and behavior. *Political Behavior*, *43*(4), 1801–1823. https://doi.org/10.1007/s11109-021-09715-9

Simonovits, G., McCoy, J., & Littvay, L. (2021). Democratic hypocrisy and out-group threat: Explaining citizen support for democratic erosion. *The Journal of Politics*, *84*(3), 1806–1811. https://doi.org/10.1086/719009

Skenderovic, D. (2009). *The radical right in Switzerland: Continuity and change, 1945-2000*. Berghahn Books.

Song, G., Ma, Q., Wu, F., & Li, L. (2012). The psychological explanation of conformity. *Social Behavior and Personality: An International Journal*, *40*(8), 1365–1372. https://doi.org/10.2224/sbp.2012.40.8.1365

Stockemer, D. (2012). The Swiss radical right: Who are the (new) voters of Swiss peoples' party? *Representation*, *48*(2), 197–208. https://doi.org/10.1080/00344893.2012.683490

Svolik, M. (2022). *Commitment to democracy: Revealed-preference evidence from seven European countries*. Prague. https://coms.events/epsa-2022/data/abstracts/en/abstract_0340.html?zoom_highlight=svolik?zoom_query=svolik&zoom_and=0&presenters=&pres_type=-1&room=-1&date=-1&zoom_sort=0 (visited on 04/06/2023).

Svolik, M. W., Avramovska, E., Lutz, J., & Milačić, F. (2023). Europe, democracy erodes from the right. *Journal of democracy*, *34*(1), 5–20. https://muse.jhu.edu/pub/1/article/875795 (visited on 10/09/2023).

Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review*, *51*(2), 273–286. https://www.jstor.org/stable/2095521 (visited on 03/05/2020).

The Guardian. (2023). *Switzerland's populists promise 'less political correctness' after election win*. The Guardian. https://www.theguardian.com/world/2023/oct/23/switzerland-populists-svp-promise-less-political-correctness-after-election-win (visited on 06/24/2024).

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press. https://books.google.com/books?hl=es&lr=&id=bjVYdyXXT3oC&oi=fnd&pg=PR11&dq=The+Psychology+of+Survey+Response&ots=Z-4nGH3ByX&sig=VM29Qry2HDRQILhvEAWqi8WmJPY (visited on 10/10/2023).

Tucker, B. (2012). Do risk and time experimental choices represent individual strategies for coping with poverty or conformity to social norms? Evidence from rural southwestern Madagascar. *Current Anthropology*, *53*(2), 149–180. https://doi.org/10.1086/664569

Turnbul-Dugarte, S. J., & Ortega, A. L. (2023). Instrumentally inclusive: The political psychology of homonationalism. *American Political Science Review*, *118*(3), 1360–1378. https://doi.org/10.1017/S0003055423000849. https://www.cambridge.org/core/journals/american-political-science-review/article/instrumentally-inclusive-the-political-psychology-of-homonationalism/1D9425F6FA20F34B0918018275A507A8

Valentim, V. (2021). Parliamentary representation and the normalization of radical right support. *Comparative Political Studies*, *54*(14), 2475–2511. https://doi.org/10.1177/0010414021997159

Valentim, V. (2024a). Political stigma and preference falsification: Theory and observational evidence. *The Journal of Politics*, *86*(4), 1382–1402. https://doi.org/10.1086/729966

Valentim, V. (2024b). *Replication data for: Norms of democracy, staged democrats, and supply of exclusionary ideology*. https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/JFLKEU (visited on 07/30/2024).

Valentim, V. (2024c). The normalization of the radical right: A norms theory of political supply and demand. *Oxford studies in democratization*. Oxford University Press.

Valentim, V., Ziblatt, D., & Elias, D. (2022). *How mainstream politicians Erode norms*. https://osf.io/mjbnf/

Väliverronen, E., & Saikkonen, S. (2021). Science communicators intimidated: Researchers' freedom of expression and the rise of authoritarian populism. *Journal of Science Communication*, *20*(4), 1–8.

Volkens, A., Lehmann, P., MatthieÃŸ, T., Merz, N., Regel, S., & Werner, A. (2015). The manifesto data collection. *Manifesto project (MRG/CMP/MARPOR)*. Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Voogt, R. J. J., & Van Kempen, H. (2002). Nonresponse bias and stimulus effects in the Dutch national election study. *Quality and Quantity*, *36*(4), 325–345. https://doi.org/10.1023/A:1020966227669

Walgrave, S., Jansen, A., Sevenans, J., Soontjens, K., Pilet, J.-B., Brack, N., Varone, F., Helfer, L., Vliegenthart, R., Meer, T. V. D., Breunig, C., Bailer, S., Sheffer, L., & Loewen, P. J. (2023). Inaccurate politicians: Elected representatives' estimations of public opinion in four countries. *The Journal of Politics*, *85*(1), 209–222. https://www.journals.uchicago.edu/doi/full/10.1086/722042 (visited on 06/25/2024).

Walgrave, S., & Soontjens, K. (2023). How politicians learn about public opinion. *Research & Politics*, *10*(3), 20531680231200692. https://doi.org/10.1177/20531680231200692

Warren, D. I. (1968). Power, visibility, and conformity in formal organizations. *American Sociological Review*, *33*(6), 951–970. https://www.jstor.org/stable/2092686 (visited on 06/25/2024).

Wike, R., & Fetterolf, J. (2018). Liberal democracy's crisis of confidence. *Journal of Democracy*, *29*(4), 136–150. https://doi.org/10.1353/jod.2018.0069

Wuttke, A., Gavras, K., & Schoen, H. (2022). Have Europeans grown tired of democracy? New evidence from eighteen consolidated democracies, 1981–2018. *British Journal of political Science*, *52*(1), 416–428. https://www.cambridge.org/core/journals/british-journal-of-political-science/article/have-europeans-grown-tired-of-democracy-new-evidence-from-eighteen-consolidated-democracies-19812018/27BFAF99039CD7B495EFED4042BB93DF (visited on 04/06/2023).

Yermakova, O. (2021). "PiS vs LGBT: The "othering" of the LGBT movement as an element of populist radical right party discourse in Poland. *Sprawy narodowościowe* (Vol. 53, pp. 1–28). Instytut Slawistyki Polskiej Akademii Nauk. https://www.ceeol.com/search/article-detail?id=1021508 (visited on 10/09/2023).

Young, I. M. (2000). *Inclusion and democracy*. Oxford University Press.

Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Author Biography

Vicente Valentim is an assistant professor of political science at IE University. His work studies the formation, enforcement, and erosion of political norms. His first book "The Normalization of the Radical Right: A Norms Theory of Political Supply and Demand" was published by Oxford University Press in September 2024.