

Politics as Usual? Measuring Populism, Nationalism, and Authoritarianism in U.S. Presidential Campaigns (1952–2020) with Neural Language Models

Sociological Methods & Research

2022, Vol. 51(4) 1721–1787

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00491241221122317

journals.sagepub.com/home/smr



Bart Bonikowski , **Yuchen Luo** ,
and **Oscar Stuhler** 

Abstract

Radical-right campaigns commonly employ three discursive elements: anti-elite populism, exclusionary and declinist nationalism, and authoritarianism. Recent scholarship has explored whether these frames have diffused from radical-right to centrist parties in the latter's effort to compete for the former's voters. This study instead investigates whether similar frames had been used by mainstream political actors prior to their exploitation by the radical right (in the U.S., Donald Trump's 2016 and 2020 campaigns). To do so, we identify instances of populism, nationalism (i.e., exclusionary and inclusive definitions of national symbolic boundaries and displays of low and high national pride), and authoritarianism in the speeches of Democratic and Republican presidential nominees between 1952 and 2020. These frames are subtle, infrequent, and polysemic, which makes their measurement difficult. We overcome this by leveraging the affordances of neural language models—in particular, a robustly optimized variant of bidirectional encoder

Department of Sociology, New York University, New York, NY, USA

Corresponding Author:

Bart Bonikowski, New York University.

Email: bonikowski@nyu.edu

representations from Transformers (RoBERTa) and active learning. As we demonstrate, this approach is more effective for measuring discursive frames than other methods commonly used by social scientists. Our results suggest that what set Donald Trump's campaign apart from those of mainstream presidential candidates was not the invention of a new form of politics, but the combination of negative evaluations of elites, low national pride, and authoritarianism—all of which had long been present among both parties—with an explicit evocation of exclusionary nationalism, which had been articulated only implicitly by prior presidential nominees. Radical-right discourse—at least at the presidential level in the United States—should therefore be characterized not as a break with the past but as an amplification and creative rearrangement of existing political-cultural tropes.

Keywords

populism, nationalism, radical politics, cultural frames, Transformer models, BERT, machine learning, active learning, computational text analysis, natural language processing

Scholars have identified populism, exclusionary and declinist nationalism, and authoritarianism as the constitutive components of radical-right politics (Mudde 2007; Mudde and Kaltwasser 2013). Empirical research has shown these elements to be present both on the supply and demand sides of political culture; that is, radical-right voters hold nationalist, populist, and authoritarian attitudes (Ivarsflaten 2008; Whitehead and Perry 2020; Akkerman, Mudde, and Zaslove 2014; Hetherington and Weiler 2009) and radical-right politicians rely on nationalist, populist, and authoritarian frames to mobilize those attitudes at the ballot box (Hawkins 2009; Lamont, Park, and Ayala-Hurtado 2017; Rooduijn 2014). What is less clear, however, is the degree to which these features are unique to radical parties and leaders. This question has motivated studies of discursive diffusion from the radical right to the political mainstream, whereby center-right (and sometimes center-left) parties mimic the messaging of radical-right parties in order to compete for the latter's voters (Mudde 2004; Rooduijn, de Lange, and van der Brug 2014). With U.S. presidential campaigns as our empirical context, we instead examine the historical antecedents of radical-right discourse: the long-standing use of populist, nationalist, and authoritarian frames by mainstream candidates and the recent exploitation—and adaptation—of these frames by the radical right (in particular, Donald Trump).¹

To do so, however, we must overcome a methodological challenge faced by scholars of political discourse: populist, nationalist, and authoritarian frames are rare, polysemic, and sensitive to candidates' linguistic idiosyncrasies and period effects. This makes them difficult to measure using computational methods, like topic models or traditional supervised machine learning classifiers. Instead, scholars have studied these frames using more conventional approaches, such as dictionary methods (Bonikowski and Gidron 2016; Rooduijn and Pauwels 2011) and hand coding (Hawkins 2009), which have considerable disadvantages, especially when working with large textual corpora.

We propose an alternative solution that takes advantage of recent innovations in natural language processing. Specifically, we rely on robustly optimized bidirectional encoder representations from Transformers (RoBERTa), a deep neural network method developed by researchers at Google and Facebook (Devlin et al. 2018; Liu et al. 2019). This large-scale language model, pretrained on 160 gigabytes of text data, represents terms in a corpus in the form of context-sensitive embeddings (i.e., dense vector representations), whereby each instantiation of a given term is assigned a unique embedding vector (in contrast to traditional embeddings that generate one vector per unique term). We fine-tune this model with an additional classification layer trained on our own human-annotated data (Sun et al. 2019) in order to assign documents to the categories representing our frames of interest. We then perform two rounds of active learning: in each round, the model's predicted probabilities of category assignment point us to high-uncertainty boundary texts, which are annotated by human coders and returned to the training set with the objective of further improving the classifier's performance (Dor et al. 2020; for an empirical application in recent political culture research, see Dai and Kustov 2022). Combining RoBERTa fine-tuning with active learning allows us to achieve a highly accurate classification of our focal frames, which—as we show—would not have been possible with traditional machine learning methods.

Having identified the constitutive components of radical-right politics—populism, two dimensions of nationalism (exclusionary criteria of national belonging and expressions of low national pride), and authoritarianism—in 71,808 paragraphs from presidential candidates' political speeches, we examine the prevalence of these frames in the campaigns of both parties. We supplement this with an analysis of two dimensions of nationalism that are seldom used by radical-right candidates: expressions of inclusive national boundaries and high national pride. Should these two frames be omitted by

mainstream campaigns as well, this could be indicative of further similarities between mainstream and radical-right candidates.

Our results show that populism and low national pride were not unique to Donald Trump's radical-right discourse; on the contrary, these frames have been commonplace in U.S. presidential politics, typically among political challengers. In contrast, the 2016 and 2020 Trump campaigns' explicit appeals to an exclusionary conception of American nationhood were unusual: no other party nominee for the highest office had been as unabashed in portraying immigrants and ethno-religious minorities as illegitimate members of the nation. This does not imply, however, that U.S. presidential campaigns had been uniformly pluralistic in the past: on the contrary, we find that authoritarianism, particularly in the form of law-and-order and tough-on-crime discourse, and the *strategic avoidance* of inclusive nationalism had both served as subtly coded substitutes for exclusionary nationalism in previous elections.

The contribution of our study is twofold. Substantively, we show that scholars should be more cautious in drawing sharp analytical boundaries between radical-right and mainstream political actors. Parties and candidates occupy positions in the same political field, and thus devise strategies in relation to one another. This can involve competitive differentiation, but also the adoption, amplification, and recombination of existing political frames. If there is a novel element in contemporary radical-right politics, it consists not of populism, as some scholars and pundits assume, but of its fusion with explicit exclusionary nationalism (Mudde and Kaltwasser 2013; Rydgren 2017). Of course, by acknowledging the existence of discursive continuities between mainstream and radical campaigns, we are not suggesting that the two types of politics carry comparably high risks for liberal democracy; they clearly do not. Rather, our argument is that by legitimating the use of morally and emotionally charged frames, mainstream actors may create—however unintentionally—the conditions of possibility for those frames' subsequent exploitation by the radical right.

Our methodological contribution is to demonstrate to scholars of culture—in sociology and beyond—that novel natural language processing methods carry tremendous potential for the measurement of otherwise elusive phenomena. The relatively rare, polysemic, and variable frames in our study had previously been difficult to capture at scale because of the inadequacy of traditional machine learning methods and the shortcomings of dictionary-based approaches. Our successful deployment of neural language models and active learning offers a promising new solution to this vexing problem (see also Ren and Bloemraad 2022). Moreover, it demonstrates that supervised

machine learning should occupy a central place in sociologists' text analysis toolkits (Molina and Garip 2019), alongside exploratory methods like topic models (DiMaggio, Nag, and Blei 2013) and geometric reductions of embeddings-based vector spaces (Kozłowski, Taddy, and Evans 2019; Stoltz and Taylor 2019). When the objects of analysis are known and the corpus is large, few methods compare in their accuracy, efficiency, and intuitiveness.

The Radical Right Versus the Mainstream

The radical right has gained increased prominence in contemporary democratic politics over the past three decades. This development has been primarily fueled by the electoral successes and mainstreaming of new parties, but in some countries, radical actors have captured traditional center-right parties (e.g., Donald Trump and the Republican Party in the United States) and single-issue parties have successfully reconfigured national institutions before disappearing from the political scene (e.g., UKIP and Brexit in the United Kingdom).

Across these varied cases, the radical right is seen by most scholars as a distinct party family that differs in fundamental ways from the center-right and center-left (typically grouped under the category of "mainstream" or "moderate" politics), and the populist or socialist left (which, with some variation across country cases, tend to be conceptualized as the "radical left"). The differences manifest themselves in party organization and policy (Golder 2016; Ivarsflaten 2008), but also in discourse. Radical-right parties consistently rely on populist, ethno-nationalist, declinist, and authoritarian claims. They vilify elites, immigrants, and ethnic, racial, and religious minorities (Ivarsflaten 2008; Mudde 2007; Rydgren 2017), typically in the context of dire assessments of the present state of the nation, and promise to take any measures necessary to restore the ethnoracial majority's dominant economic, political, and cultural status (Gest, Reny, and Mayer 2018).

The assumption that the radical right is categorically distinct from traditional party families, however, has been undermined by two developments. First, some radical-right parties had started off as mainstream parties and gradually drifted rightward under stable leadership, as in Viktor Orbán's Fidesz in Hungary and Jarosław Kaczyński's Law and Order Party in Poland (Vahudova 2020). The difference between the two party families may therefore be more appropriately understood as continuous rather than discrete. Second, in recent years, some mainstream parties have experimented with exclusionary discourse reminiscent of the radical right in an effort to

compete for the latter's voters (Mudde 2004; Rooduijn, de Lange, and van der Brug 2014), thus challenging the notion of an impermeable boundary between the two camps. Emmanuel Macron's *La République En Marche!* in France, Mark Rutte's VVD in the Netherlands, and Mette Frederiksen's Social Democrats in Denmark are three prominent examples.

Despite these developments, the sharp distinction between "pure" mainstream politics—ostensibly free of populism, nationalism, and authoritarianism—and the radical right has remained. From this perspective, the diffusion of these discursive frames into the mainstream is reducible to a process of radicalization, whereby traditional center-left and center-right parties' categorical purity is increasingly polluted by the corrosive influence of the radical right. Mainstream parties may not be passive in this process, according to this argument, but the direction of influence is clear: it flows from the periphery to the center.

Without dismissing the possibility of extreme-to-mainstream diffusion, our study challenges the overly sanguine depiction of traditional center-left and center-right parties. Though social inclusion, technocratic rationality, and commitment to liberal democratic stability may characterize some mainstream parties some of the time, these traits stand in tension with historical examples of mainstream parties' moral vilification of the political establishment, anti-immigrant discourse, racial dog-whistle politics, targeted use of state power against sundry internal "enemies," and—in the case of the United States—active maintenance of racial segregation and political exclusion. This tension raises the possibility that in addition to having been influenced in recent years by the radical right, mainstream parties may have long employed framing strategies that served as precursors to the radical right's populist, nationalist, and authoritarian agenda. We explore whether this has been the case in the United States, focusing on the political speeches of Democratic and Republican presidential candidates between 1952 and 2020.

Why would it matter if mainstream parties used discursive frames that have since become mainstays of radical-right politics? First, this would suggest a degree of cultural continuity between these party families. Rather than having invented a new form of politics—or reclaimed a toned-down version of early-twentieth-century fascism (Finchelstein 2019)—the radical right may have exploited and recombined cultural tropes readily available in mainstream political discourse. This is not to say that such cultural adoption, if operative, was deliberate or that past presidential campaigns offered the sole source of discursive templates on which the radical right has drawn. Rather, the mainstream use of such tropes can be interpreted as evidence of their broad diffusion and legitimacy in political culture.

Second, it is possible that populist, nationalist, and authoritarian messaging by mainstream political campaigns may have contributed to increased electoral opportunities for the radical right. By instrumentally stoking public fears of elites, minorities, or immigrants, mainstream candidates raise their constituents' expectations about forthcoming solutions to the threats ostensibly posed by these groups. If such promises routinely go unfulfilled, frustration and political disaffection can mount, and voters may have greater incentives to defect to more radical candidates and parties (Bonikowski and Ziblatt 2020). In this sense, mainstream parties' deployment of faux-radicalism is a high-risk gamble (cf. Dai and Kustov 2022): it may be electorally effective in the short run, but in the long run, it can undermine those parties' positions by facilitating successful challenges from the periphery.

Defining Populism, Nationalism, and Authoritarianism

The concepts of populism, nationalism, and authoritarianism have become ubiquitous in recent years, but confusion about their meaning persists; some definitional clarification is therefore warranted. By populism, we are referring to a form of moral claims-making that juxtaposes a fundamentally corrupt elite with the virtuous people and promises to restore political power to the latter (Taggart 2000). Which elites are singled out by populist claims varies with context and the ideology of the speaker, but typical targets include politicians, bureaucrats, corporate executives, the wealthy, intellectuals, and labor union leaders (Mudde 2007; Mudde and Kaltwasser 2017). These groups are portrayed as having forsaken the interests of the people in the self-serving pursuit of power, money, and status. Because elites control prominent institutions, including the government, populist claims often involve systemic critiques of institutional practices, including those at the heart of representative democracy. Although populism is one of the central features of radical-right politics, it is not inherently tied to any particular political ideology (Laclau 2005). Indeed, it has been a common trope among politicians of the radical left as well—in Latin America (Levitsky and Loxton 2013; Roberts 1995; Weyland 2001), Western Europe (March 2012), and the United States (Goodwyn 1978; Kazin 1998)—and also among center-left and center-right parties, such as the U.S. Democrats and Republicans (Bonikowski and Gidron 2016; Fahey 2021; Dai and Kustov 2022).

Nationalism, in our use of the term, involves the articulation of distinct conceptions of nationhood, typically in an effort to either highlight the

nation's present-day virtues or to offer a critique of the nation's decline and an alternative vision for its future. Individuals' nationalist dispositions are multidimensional and involve not only beliefs about criteria of legitimate membership in the nation (typically either subjective affinities or ascriptive characteristics, like ancestry and native birth), but also aspects of national heritage and institutional practice that are worthy of collective pride or shame (e.g., egalitarian principles, economic achievements, or the quality of democratic governance) (Huddy and Khatib 2007). Past research has found exclusionary conceptions of the nation's symbolic boundaries and low levels of national pride in particular to be associated with support for radical-right candidates, including Donald Trump (Bonikowski, Feinstein, and Bock 2021; Manza and Crowley 2018; Reny, Collingwood, and Valenzuela 2019). Our analyses focus on the discursive counterparts to these attitudes (i.e., exclusionary nationalism and expressions of low national pride). In addition, we examine two alternative frames rarely found on the radical right: evocations of inclusive criteria of national belonging and high national pride. We do so to confirm whether the use of these frames distinguishes mainstream from radical politics, but also to investigate the possibility that their absence from mainstream campaigns may represent coded appeals to ethnonationalist and disaffected voters.²

The third component of radical-right politics, authoritarianism, refers to the targeted use of state power against alleged domestic enemies—from “Communist infiltrators” and “violent rioters” to criminals, terrorists, and undocumented immigrants—in a manner that undermines liberal rights regimes and democratic norms and institutions (Levitsky and Ziblatt 2018; Scheppele 2018; Vahudova 2020). Such measures are often justified as responses to acute crises that ostensibly threaten the nation's stability. Although illiberal and undemocratic governance strategies tend to be implemented only once radical parties (on both the left and the right [De La Torre 2016; Weyland 2013]) gain access to power, authoritarianism is often signaled in electoral campaigns as well, in order to demonstrate the lengths to which candidates or parties are willing to go in the pursuit of their constituents' interests.

The specific content of authoritarian rhetoric is often influenced by its interdependence with populist and nationalist claims. It is the latter that frame elites or minority groups as threatening, thereby providing justification for extraordinary measures that challenge the institutional status quo. Populism and nationalism are themselves mutually complementary as well: nationalism can help fill in the content of “the people” in populist claims,

whereas populism can pinpoint which powerful factions stand in the way of the nation's flourishing. It is because of these elective affinities that populism, nationalism, and authoritarianism are often observed together on the radical right. Whether this is also the case in mainstream parties' discourse is an empirical question that we take up in this article.

What do populism, exclusionary and inclusive nationalism, evocations of low and high national pride, and authoritarianism look like in presidential speeches? Let us consider these frames in turn.

The following two passages illustrate the use of populist claims in campaign discourse. The first example comes from the 2016 Trump campaign and the second from Barack Obama's 2008 campaign. That we found such discourse in Obama's speeches in itself provides preliminary evidence that populism is not limited to right-wing or radical politics.³

It's going to be a victory for the people. A victory for the everyday citizen whose voice hasn't been heard. It will be a win for the voters, not the pundits, not the journalists, not the lobbyists, not the global special interests funding my opponent's campaign.

Finally, the American people must be able to trust that their government is looking out for all of us—not the special interests that have set the agenda in Washington for eight years, and the lobbyists who run John McCain's campaign. I've spent my career taking on lobbyists and their money, and I've won.

Both passages articulate a moral critique of the political establishment as having been captured by elites who have abandoned the interests of the people. They also frame the people as the sole legitimate source of political power and urge them to regain their rightful control of the state.

Next we turn to exclusionary and inclusive nationalism. The following quote from Donald Trump's 2016 campaign is an example of the former. It conflates undocumented migrants, Syrian refugees, and "criminal aliens" as a uniform danger to America, thereby placing these groups outside of the boundaries of legitimate national belonging.

Expanding President Obama's unconstitutional executive amnesty, including instant work permits for millions of illegal workers. Freeing even—there go your jobs—freeing even more criminal aliens by expanding Obama's non-enforcement directives. And this is to me, the beauty of them all. Obama has allowed thousands and thousands and thousands of people to come in,

Syrians from the Middle East. She wants an increase of 550 percent in Syrian refugees into our country.

Such exclusionary rhetoric stands in contrast with inclusive nationalist claims—for instance, civil rights frames that condemn segregation and discrimination or more generic pluralist depictions of the nation that emphasize the virtues of racial, ethnic, and religious diversity. These two variants of inclusive nationalism are captured, respectively, in the following two quotes, the first from Hubert Humphrey's 1968 campaign and the second from the 1976 Jimmy Carter campaign.

I submit that this election year must bring a national referendum on this issue of human rights and we must turn away from the old era that all of you know too well, of segregation and discrimination which never gave us much except trouble [...] [A]nd we must turn away from the old era when an American soldier who fought in an integrated Army would come home into a neighborhood of segregation and slumism. And we must choose a new day when that man, when he gets home, can have a job that is a meaningful job, worthy of his talents, worthy of his service to his country, and he must have training for that job and he must have equality before the law and in his neighborhood and amongst his associates.

We can have an America that provides excellence in education to my child and your child and every child. We can have an America that encourages and takes pride in our ethnic diversity, our religious diversity, our cultural diversity—knowing that out of this pluralistic heritage has come the strength and the vitality and the creativity that has made us great and will keep us great.

The other dimension of nationalism we consider is national pride. The following two passages express negative assessments of America's present-day conditions, contrasted with the nation's former glory and the promise of its brighter future. The first is from John F. Kennedy in 1960 and the second from Donald Trump in 2016. We categorize these as evocations of low pride, because they mobilize the audience's disappointment at the nation's ostensible decline.

I don't think an administration which has presided over three recessions in the last eight years, which is now presiding over the lessening of the U.S. position around the world, which has permitted the U.S. image to fade as a vital society, our most important asset, which has in that way damaged the cause of freedom,

I cannot believe that any young man or woman who looks to the future can possibly decide to sit down and sit still and look back with Mr. Nixon and the Republican party which has always opposed progress.

We've lost 70,000 factories since China's entry into the World Trade Organization. Another Bill and Hillary backed disaster. We are living through the greatest job theft in the history of the world. More jobs have been stolen from our country, so stupidly we let them go. We let our companies go so foolishly. We don't know what we're doing. A Trump administration is going to renegotiate NAFTA, stand up to the foreign cheating, and stop the jobs from leaving our country, and have jobs come back in the other direction.

Whereas some candidates decry the decline of the nation, others praise its virtues, ascribing them to the core principles of the American creed. This expression of high national pride is illustrated in the following quote from Adlai Stevenson's 1952 campaign.

America is a great, a strong, a wise, and most of all a good country. And I believe with all my heart that by these qualities, we can and we will safely in God's good time win our way to a peaceful world.

Finally, we turn to authoritarianism. This frame took on distinct forms over time. Its early manifestations in our data stressed the need to eliminate Communist sympathizers ostensibly lurking within the U.S. government, as in this 1952 quote from Dwight Eisenhower:

We are opposing [Communism] at home where its agents and converts seek to undermine our society and corrupt our government. As I have repeatedly said, the federal government must use all its resources to expose and identify Communistic activity, to keep Communists out of places of responsibility in our society, and to protect our institutions from Communist espionage, sabotage and subversion.

Subsequent iteration of authoritarianism, most pronounced in the late 1960s, focused on "law and order" and the threats to social stability posed by urban riots (a reference to racial justice and anti-war protests, which intensified in the spring of 1968 following the assassination of Martin Luther King, Jr.) (Wasow 2020). This is captured by the following two passages from Richard Nixon's 1968 campaign.

It is because millions of Americans who have not been violating the law, millions of Americans who haven't been engaging in violence, millions of Americans who haven't been shouting in the streets—now you are coming out. The forgotten Americans are being heard, and you are going to continue to be heard. It is because this great group of Americans—they are good people, decent people.

[W]hen we find [...] in city after city in this country that convicted murderers, convicted rapists, are turned free, confessed murderers, I mean, and confessed rapists, are turned free after they confess their crime because of a technicality, then I say that our courts in their decisions have gone too far in weakening the peace forces as against the criminal forces in this country.

The “law and order” rhetoric popularized by Nixon persisted in the form of punitive “tough on crime” discourse that accompanied the War on Drugs and welfare policy retrenchment of the 1980s and 1990s (Mendelberg 2017). The following 1988 quote from George H. W. Bush captures this discursive strategy.

Deep differences on crime—my opponent let murderers, who had not even served enough time to be eligible for parole, out on weekend furloughs. I want to keep them behind bars. I think some crimes—such as drug kingpins killing police officers—are so horrible they deserve the death penalty. My opponent opposes the death penalty.

Finally, in Donald Trump's campaign, authoritarianism manifested itself in draconian anti-immigrant policies (conflated with anti-terrorism and drug crackdowns), threats to imprison the Democratic nominee, Hillary Clinton, and Trump's refusal to commit to a peaceful transfer of power were he to lose the election. The first variant is illustrated in this quote from the 2016 Trump campaign:

That's why I was so happy what we did to annihilate the enemy the other day. So happy. Because we're dealing against a very dishonest system. But Hillary, so important, wants to have a radical, and this is very radical, immigration. She wants to radicalize immigration where you have people pouring in. Remember this, the border patrol agents, 16,500 gave me their endorsement. Last week, ICE, ICE, these are great people, you don't hear great things because they're not allowed to do their job.

Hypotheses

The foregoing discussion leads us toward several expectations for our analyses. Because Donald Trump's 2016 and 2020 campaigns are widely perceived in the United States as marking a radical break with mainstream electoral politics, we use them as our main point of comparison.

First, we expect that the main components of radical-right politics—populism, exclusionary and low-pride nationalism, and authoritarianism—were indeed highly prevalent in Donald Trump's campaign discourse. Were this not the case, our characterization of Trump as a radical-right candidate would be questionable. With respect to populism and authoritarianism, both of which are unidimensional concepts in our theoretical framework, the predictions are simple: we expect Trump to be among the candidates most likely to rely on both frames.

Hypothesis 1a: Populist and authoritarian claims were more prevalent in Donald Trump's 2016 and 2020 presidential campaign speeches than in most mainstream campaigns between 1952 and 2020.

Unlike populism and authoritarianism, nationalism is a multidimensional phenomenon. Past research has demonstrated that the Trump campaign was especially adept at mobilizing voters who held exclusionary conceptions of legitimate national belonging and low levels of national pride (Bonikowski, Feinstein, and Bock 2021; Lamont, Park, and Ayala-Hurtado 2017; Sides, Tesler, and Vavreck 2018). It is these two aspects of nationalism that we would expect to be the most prevalent among Donald Trump's speeches.

Hypothesis 1b: Exclusionary nationalist claims were more prevalent in Donald Trump's 2016 and 2020 presidential campaign speeches than in most mainstream campaigns between 1952 and 2020.

Hypothesis 1c: Claims evoking low levels of pride in the nation were more prevalent in Donald Trump's 2016 and 2020 presidential campaign speeches than in most mainstream campaigns between 1952 and 2020.

Having specified our predictions for the Trump campaigns, we now turn to the other candidates covered by our data between 1952 and 2020. In line with previous research on populism in U.S. presidential elections (Bonikowski and Gidron 2016; Dai and Kustov 2022; Fahey 2021)—and the suggestive evidence revealed by our reading of sample speeches—we expect populism to

be found throughout the time series in the speeches of candidates from both parties.

Hypothesis 2a: Populist claims were used by both Democratic and Republican mainstream candidates throughout the time series.

In light of the long history of racist and nativist politics in the United States (Bonilla-Silva 2001; Higham 1955; McVeigh and Estep 2019; Smith 1997), it is reasonable to expect mainstream candidates, and not just Donald Trump, to have relied on exclusionary nationalist frames as well. This tendency, however, may have been muted by concerns over alienating voters who hold more egalitarian attitudes. Indeed, historical evidence suggests that presidential candidates seeking to mobilize racial resentment have typically done so using coded, implicit language, particularly when speaking to the general electorate (Sides and Hopkins 2015). We are therefore cautious in formulating strong expectations regarding exclusionary nationalism. To the degree that we find instances of such claims at all, they are likely to be limited to Republican campaigns, given the protracted partisan polarization around civil rights in the United States.

Hypothesis 2b: Exclusionary nationalist claims were employed by mainstream Republican candidates, but less frequently so than populism, authoritarianism, and low pride.

As suggested by the literature on implicit racial claims and dog-whistle politics, coded language can be a powerful substitute for explicitly exclusionary claims, allowing candidates to mobilize white racial grievance while maintaining plausible deniability against charges of outright racism (López 2015; Mendelberg 2017; Wetts and Willer 2019). A particularly salient feature of ethno-nationalist politics in the United States has been the stoking of moral panic about social disorder and violent crime. Considering the deeply exclusionary connotations of law-and-order and tough-on-crime authoritarianism, we expect this frame to be particularly prevalent among Republicans, more so than explicit exclusionary nationalism.

Hypothesis 2c: Authoritarian claims were used more frequently by mainstream Republican candidates than by mainstream Democratic candidates.

Hypothesis 2d: Mainstream candidates relied on authoritarian claims more frequently than on exclusionary claims.

Another way to prime exclusionary nationalism without evoking it explicitly is to purposefully avoid its opposite: references to social inclusion. In light of the historical struggle between elective and ascriptive definitions of nationhood in the U.S. (Smith 1997) and the low symbolic costs associated with boilerplate evocations of American egalitarianism, a candidate's choice not to engage in inclusive discourse is likely to be a strategic one. This is all the more so when inclusion is a central component of an opponent's electoral repertoire. Under such conditions, the avoidance of inclusive depictions of America can itself send a powerful message to the candidate's ethno-nationalist supporters. We expect this pattern to be especially prevalent among Republicans.

Hypothesis 2e: Inclusive nationalist claims were employed more frequently by mainstream Democratic candidates than by mainstream Republican candidates.

The final component of nationalism evoked in Donald Trump's discourse is national pride. As is often the case among radical-right candidates, Trump painted a dismal picture of U.S. society and politics, arguing that the nation has abandoned its principles, and that he alone could restore its greatness (Mercieca 2020). By evoking low levels of national pride (i.e., portraying the nation in a negative light), candidates are able to appeal to disaffected voters, while also reinforcing their populist critiques of ostensibly morally corrupt elites. As with the other radical-right frames, we expect mainstream candidates to have routinely relied on such discourse. We do not foresee much variation in this strategy across the two parties.

Hypothesis 2f: Low levels of national pride were evoked by both Democratic and Republican mainstream candidates throughout the time series.

The above predictions also suggest a set of patterned relations between populism, authoritarianism, and the multiple dimensions of nationalism. Considering our argument that low-pride claims can support populist discourse, we expect to observe a positive correlation between these two frames. In contrast, low pride and high pride should be negatively associated, since they are logical opposites of one another. Finally, our prediction that authoritarianism can serve as a veiled form of ethno-nationalism suggests a negative correlation between authoritarianism and inclusive conceptions of nationhood.

Hypothesis 3a: Low national pride and populism were positively correlated among mainstream campaigns.

Hypothesis 3b: Low national pride and high national pride were negatively correlated among mainstream campaigns.

Hypothesis 3c: Authoritarianism and inclusive nationalism were negatively correlated among mainstream campaigns.

If high and low pride are in fact negatively correlated and are frequently used by both parties, this raises the question of which candidates are most likely to rely on one frame or the other. Given the opposite incentives to present the nation in positive or negative light depending on whether a candidate's party is currently in power, we expect to find variation within each party based on its incumbency status. Negative evaluations of the nation should be more likely among challenger candidates, who have strong incentives to criticize the status quo and to lay the blame for the nation's failures on the incumbent party and its constituents (Crabtree et al. 2020). A similar pattern should be observed for populism, which is a common electoral strategy of outsiders (Bonikowski and Gidron 2016; Dai and Kustov 2022; Fahey 2021).

Hypothesis 4a: Low national pride and populism were used more frequently by challenger candidates than by nominees of the incumbent party.

The opposite should be true of evocations of high national pride: this should be a strategy of incumbents rather than challengers, since it is the former who have strong incentives to celebrate the nation's accomplishments and to present the previous administration as having governed effectively (Crabtree et al. 2020).

Hypothesis 4b: High national pride was used more frequently by incumbent party candidates than by challengers.

Our expectations for the effects of incumbency on populism and expressions of national pride suggest a final hypothesis concerning differences between Donald Trump's 2016 and 2020 campaigns. Thus far, we have grouped both campaigns as instantiations of Trump's radical-right politics, following the common assumption in the literature that populism, nationalism, and authoritarianism are attributes of *actors* (especially politicians and parties, but also

individual voters). We know from past research, however, that the degree to which a given candidate relies on a particular discursive strategy often varies with context, sometimes generating large differences between that candidate's successive campaigns (Bonikowski and Gidron 2016; Dai and Kustov 2022; Fahey 2021). Motivated by this tradition, our study treats populism, nationalism, and authoritarianism as attributes of *claims* not of actors, allowing for within-candidate variation in their use. If we are correct about populism and low pride being especially common among political challengers, we should expect Donald Trump's 2020 reelection campaign to make less frequent use of these frames than did his 2016 campaign:

Hypothesis 4c: Donald Trump's 2020 campaign, when he ran as an incumbent, featured fewer references to populism and low national pride than his 2016 campaign, when he ran as a challenger.

To test the above hypotheses, we examine the paragraph-level prevalence of the six frames within the speeches of both parties' nominees between 1952 and 2020. In the next section, we describe our data and measurement strategy, which constitutes the main methodological contribution of our study.

Presidential Campaign Data, 1952–2020

Our analyses rely on a corpus of presidential campaign speeches by Democratic and Republican Party nominees, beginning with the 1952 contest between Dwight Eisenhower and Adlai Stevenson and concluding with the 2020 election featuring Donald Trump and Joe Biden. The data are a composite of two sources: the Annenberg/Pew Archive of Presidential Campaign Discourse (Annenberg 2000), which covers the 1952–1996 elections, and speeches from the UC Santa Barbara American Presidency Project (UCSB 2021), spanning the 2000–2020 elections. The Annenberg data include nomination acceptance speeches and all available campaign speeches delivered between September 1 and Election Day, whereas the UCSB corpus includes speeches from nomination acceptance to Election Day. To harmonize the data, we restricted the UCSB corpus to the speech range covered by the Annenberg archive. We also removed all passages that contained town hall discussions, Q&A sessions, and utterances by speakers other than the presidential candidate. Two Republican campaigns are missing from the data: Barry Goldwater's 1964 campaign and George W. Bush's 2000 campaign. The final corpus consists of 2,956 speeches

split into 71,808 paragraphs (we justify treating paragraphs as the primary units of analysis in the methods section).

In addition to the text data, we construct variables for campaign and speech attributes. These include speech length, the candidate's party, whether the candidate represented the incumbent or challenger party, and the date on which the speech was given. Table 1 presents the descriptive statistics for the corpus.

Measuring Political Frames in Texts

Previous Approaches

The measurement of populist, nationalist, and authoritarian frames in political speeches presents a methodological challenge. First, these phenomena tend to be rare: while Bonikowski and Gidron (2016), for instance, estimate that populist language can be found in about 15 percent of political speeches in the Annenberg corpus, populist content may only appear in a small portion of any speech, such as a single paragraph or sentence. The low prevalence of these frames makes their measurement akin to searching for a needle in a haystack.

Second, these frames are often articulated using ambiguous language that relies heavily on multivocal terms. For example, populist people-centrism may consist of little more than the use of first-person plural pronouns (Rooduijn and Pauwels 2011). Racial dog-whistles are even more difficult to identify, given their purposeful use of coded signals embedded in seemingly innocuous discourse (López 2015). As a result, the differences between these frames and their non-populist, non-nationalist, and non-authoritarian counterparts are often subtle—based on tone, use of morally laden terms, and implicit referents.

Third, because our interest is in frames rather than substantive topics, the variation in the thematic content of positive cases is likely to be high, spanning domains such as the economy, immigration, bureaucratic administration, the welfare state, and national security. In addition, framing language may itself be specialized to each of these substantive areas, making the identification of generic keywords difficult.

When seeking to systematically measure these types of elusive frames in large collections of text, scholars have been limited to several manual and computational approaches, each of which has its own affordances and limitations. Among the most common of these have been human coding, dictionary-based classification, topic modeling, exploratory use of word embeddings, and supervised machine learning.

Table 1. Descriptive Statistics, 1952–2020 Speech Corpus.

	Full corpus					Republicans					Democrats				
	N	Mean	Min	Max		N	Mean	Min	Max		N	Mean	Min	Max	
	speeches	paras	words	date	(m/d)	speeches	paras	words	date	(m/d)	speeches	paras	words	date	(m/d)
All	2,956	24	2,184	05/21	11/07	1,329	26	2,263	05/21	11/07	1,627	23	2,120	07/14	11/07
Challenger	1,370	21	1,890	07/11	11/07	513	24	2,009	07/11	11/07	857	20	1,819	07/14	11/07
Incumbent	1,586	27	2,438	05/21	11/07	816	27	2,423	05/21	11/07	770	27	2,454	07/21	11/07
1952	412	15	1,656	07/11	11/03	228	14	1,570	07/11	11/03	184	17	1,763	07/21	11/03
1956	116	21	1,811	08/18	11/05	42	19	1,514	08/23	11/05	74	23	1,980	08/18	11/05
1960	448	14	1,855	07/15	11/07	136	25	3,070	07/28	11/07	312	9	1,325	07/15	11/07
1964	111	27	2,254	08/27	11/02						111	27	2,254	08/27	11/02
1968	209	30	2,835	08/08	11/04	58	32	3,115	08/08	10/30	151	30	2,727	08/29	11/04
1972	153	17	1,752	07/14	11/06	58	18	1,964	08/23	11/06	95	16	1,622	07/14	11/06
1976	178	14	1,307	07/15	11/01	128	13	1,083	08/19	11/01	50	19	1,880	07/15	11/01
1980	157	19	1,809	07/17	11/03	68	17	1,313	07/17	11/03	89	20	2,188	08/14	11/03
1984	172	25	1,920	07/19	11/05	107	25	2,026	08/23	11/05	65	25	1,745	07/19	11/05
1988	104	25	1,779	07/21	11/07	40	30	2,032	08/18	11/01	64	23	1,620	07/21	11/07
1992	201	27	2,462	07/16	11/02	125	25	2,155	08/20	11/02	76	31	2,967	07/16	11/02
1996	187	32	2,811	08/15	11/04	77	33	2,417	08/15	11/04	110	31	3,088	08/29	11/04
2000	4	44	3,042	08/17	11/01						4	44	3,042	08/17	11/01
2004	162	46	3,599	07/29	11/01	134	49	3,781	09/02	11/01	28	34	2,730	07/29	10/29
2008	85	27	2,368	08/28	11/03	33	22	1,614	09/04	11/03	52	31	2,847	08/28	11/03
2012	134	35	2,675	08/30	11/05	50	33	2,487	08/30	11/02	84	37	2,786	09/06	11/05
2016	69	43	3,115	07/21	11/07	32	51	2,882	07/21	11/07	37	35	3,318	07/28	11/07
2020	54	57	3,696	05/21	11/02	13	93	8,896	05/21	10/17	41	45	2,047	08/20	11/02

Human coding has the advantage of being sensitive to the subtleties of framing and context and should, in principle, yield high classification accuracy (Hawkins 2009). Its primary limitation is that it is time- and labor-intensive. With small corpora, this method works well, but for corpora with thousands of documents composed of tens of thousands of paragraphs, its use becomes unfeasible. Moreover, this approach has other, less obvious disadvantages. For instance, annotator fatigue can lead to errors during longer coding sessions, which are often necessary for larger collections of texts. At the same time, the sensitivity to linguistic subtlety that is a virtue of this approach can also be its liability: as each text's meaning is interpreted by coders with different personal experiences, degrees of knowledge, and levels of linguistic competence, idiosyncratic biases can affect classification outcomes.

The simplest computational alternative to hand coding is keyword-based classification, also known as the dictionary method. This consists of identifying a collection of terms (i.e., a dictionary or lexicon) that represent the concept of interest and counting their occurrence in the corpus, resulting in either a continuous or binary measure of the concept's prevalence in each document (Rooduijn and Pauwels 2011). The advantage of dictionary methods is their ability to handle large corpora, as well as their transparency: readers of a dictionary-based study can easily review the keywords and evaluate their face validity. The downside, however, is in the keywords themselves. It is difficult to come up with an exhaustive list of terms that cover all instantiations of a given phenomenon, a problem that is compounded by limitations of human cognition (King, Lam, and Roberts 2017). This approach also assumes that a particular keyword can serve as an unambiguous indicator of the concept of interest. The inherent polysemy of language, whereby the meaning of a word can vary widely with context, makes this assumption unjustified in many contexts (DiMaggio 2015). Finally, it is often difficult to quantify the proportion of false negatives in the corpus that were missed by the dictionary (Nelson et al. 2021). Hand-coding a randomly sampled test set and applying the dictionary to it provides one solution, but this becomes less feasible with highly skewed data where the phenomenon of interest is rare—as is the case with the frames analyzed in this study.

Some social scientists have also relied on unsupervised text analysis methods to infer latent themes in text. The most popular among these methods is topic modeling, which represents semantic themes as probability distributions over the feature vocabulary (DiMaggio, Nag, and Blei 2013; Karell and Freedman 2019; Light and Cunningham 2016). Closely related to this are approaches that construct word co-occurrence networks and then

cluster them using community detection algorithms (Fuhse et al. 2020; Rule, Cointet, and Bearman 2015; Hoffman et al. 2018). These methods have the virtue of identifying patterns in text without the need for labor-intensive hand-coding, while enabling the direct inspection of topics or feature communities for purposes of validation. However, the inductive logic of both approaches makes them more suitable for exploratory analysis than for the measurement of frames chosen *a priori* based on theoretical expectations, as is the case in our study. Whereas topic models tend to identify the most commonly occurring themes in a corpus, we are interested in rarely occurring frames.

Alongside topic models and network-based approaches, there has been growing interest among political scientists and sociologists in unsupervised or semi-supervised methods based on word embeddings: the representation of terms or documents as dense vectors in multidimensional space, whereby geometric proximity can be interpreted as similarity in meaning (Mikolov et al. 2013; Stoltz and Taylor 2021). Embeddings can be used for comparing the meaning of a term across theoretically relevant subcorpora (Bonikowski, Feinstein, and Bock 2021), locating texts along latent dimensions of meaning anchored by pairs of opposite terms (Kozlowski, Taddy, and Evans 2019; Nelson 2021), or detecting themes with embeddings-based dictionaries (Stoltz and Taylor 2019).⁴ Whether embeddings-based approaches are effective for measuring themes depends on the degree to which the latter are unambiguously marked in the corpus by distinctive lexical features. As we have argued above, this is often not the case for the kinds of frames we study here.

Finally, one of the most common approaches—particularly outside of social science—for measuring textual phenomena is supervised machine learning: the use of a training set of labeled documents to teach a classifier to detect patterns in the data that discriminate between the categories of interest and to then apply those machine-learned classification rules to unlabeled texts (Evans and Aceves 2016; Nelson et al. 2021). The quality of the classification can then be directly evaluated on a held-out test set of documents.

Supervised classifiers have had widespread application in a variety of disciplines and industries, but they are only beginning to make their way into sociology (Molina and Garip 2019). This may be partly a consequence of remaining skepticism about the value of deductive coding compared to the inductive exploration of texts (DiMaggio 2015; Lee and Martin 2015), but also the fact that supervised machine learning methods have performed poorly in measuring the kinds of subtle frames that are often of interest to sociologists (Bonikowski and Gidron 2016). Recent advances in the use of embeddings as inputs into classification algorithms, along with the growing

popularity of embeddings models pretrained on large corpora, may alter this calculus. Nonetheless, as we will demonstrate, the performance of even these enhanced classifiers leaves much to be desired in our empirical case.

Instead of these traditional methods, we employ a novel supervised learning approach: the use of neural language models pretrained on massive textual corpora, which are then fine-tuned for a specific local task—in our case, the classification of sections of political speeches. The model we use, RoBERTa, generates contextual embeddings that uniquely encode every instance of a given word, enabling a much richer representation of meaning than is possible with standard embedding models. We further combine this approach with *active learning*, a procedure that iteratively improves classifier performance through additional targeted hand-coding. Together, these techniques allow us to identify populism, nationalism, and authoritarianism in political texts with accuracy unmatched by other supervised methods—without being encumbered by the fixed-keyword constraints typical of dictionary methods, the labor intensiveness of hand coding, or the indeterminacy of unsupervised methods.⁵

RoBERTa and Active Learning

Among the most promising recent innovations in natural language processing (NLP) is the development—and integration—of neural language models and *transfer learning*. The core idea of transfer learning is that knowledge obtained from one task can be transferred to another task. In most instances, the goal is to obtain a better feature representation for one's data than would be possible with a locally trained model. For an intuitive sense of the utility of transfer learning for classification, consider Figure 1, which illustrates four possible options (among others) for representing a three-word text sequence (or pseudo-document).

In Approach 1, we turn the text sequence into a count vector over the feature vocabulary of the corpus. This is among the simplest ways to represent a document. Applying this to the full corpus produces a document-term matrix, which serves as input into a classifier that assigns each document to a predefined category (e.g., populist or not). All we know about each document is which terms (or features) it contains and how often they occur, but we lack richer information regarding those terms' immediate context (and therefore meaning). Especially when working with small text corpora—the typical situation in most social science applications—such document vectors are likely to be sub-optimal representations of the documents' content.

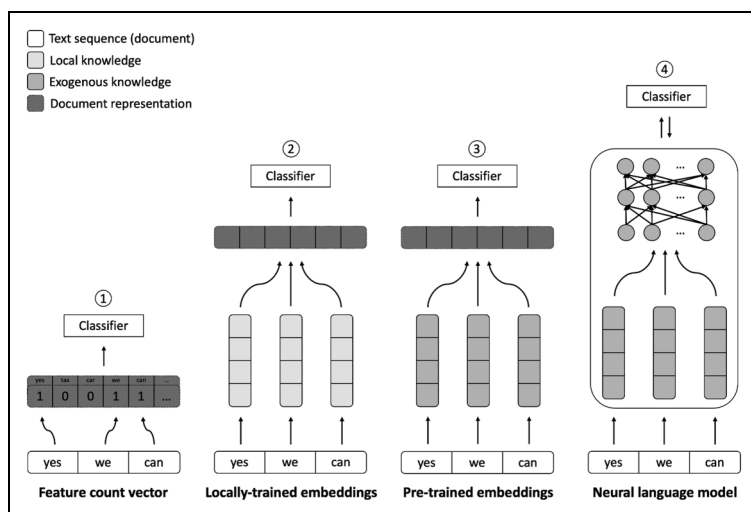


Figure 1. Four common language representations in NLP.

Note: Model architecture is simplified here for illustrative purposes.

In Approach 2, illustrated in the second panel of Figure 1, we first convert a document into a set of locally trained embeddings for each word. Word embeddings are dense vectors that encode the meaning of a given feature (e.g., word or document) relative to other features in the corpus. These vectors are learned from the features' co-occurrence in higher units of text using a variety of techniques (e.g., Dumais 2004; Mikolov et al. 2013; Pennington, Socher, and Manning 2014; for an application-oriented review, see Rodriguez and Spirling 2022). In a second step, we generate a representation of our pseudo-document by averaging the word embeddings of its constituent terms. Instead of raw word counts (as in Approach 1), this representation is informed by knowledge about the relational meaning of each term, which is likely to lead to a more informative representations of the document's content. Finally, these document-level embeddings are entered into a classifier. A drawback this approach shares with the use of feature count vectors (i.e., Approach 1), however, is that its effectiveness depends on corpus size, with larger corpora allowing for more robustly trained word embeddings. Furthermore, to classify documents, we must average the embedding vectors across all the terms in each document, thereby discarding word order and syntax, a central feature of language.

Negation, for instance, lies beyond the reach of this approach. Finally, every term is represented by a single embedding vector, which ignores the fact that words often take on distinct meanings in different contexts (i.e., their meaning is polysemic).

In Approach 3, we engage in a form of transfer learning: we represent our document using *pretrained* word embeddings. High-quality embeddings trained on millions of texts are freely available online. While they may not be semantically adapted to the local context, they allow us to bring in rich exogenous knowledge about the meaning of terms and, especially when working with small text corpora, to generate informative representations of documents with relatively little effort. Even though this approach overcomes the corpus size limitations associated with locally trained embeddings, however, it continues to suffer from the other two drawbacks of Approach 2: the need for document-level vector averaging and the inability to account for polysemy.

Neural language models, illustrated in the fourth panel of Figure 1, provide a powerful alternative to these approaches. To further elucidate the advantages of this approach, we turn to a brief overview of language models and their unique affordances for natural language processing.

Language models are designed to solve seemingly simple tasks like the prediction of a masked word in a sentence conditional on observed words surrounding it or the prediction of the next word in a left-to-right sequence (Jensen et al. 2021).⁶ When trained to perform such tasks on large corpora, the resulting models are able to account for extensive variation in word sequence structures, approximating the semantic content of a natural language. Language models have been constructed using a wide range of architectures, from simple *n*-gram lookups (Brown et al. 1992) to recurrent neural networks (RNNs) and long-short-term memory networks (LSTMs) (Sundermeyer, Schlüter, and Ney 2012). While effective, these methods have suffered from non-scalability (*n*-grams), short-term memory biased toward proximate words (RNNs), non-parallelizability (RNNs), and computationally expensive model complexity (LSTMs). Many of these problems have been overcome in recent years by the advent of a new deep neural language modeling approach: the Transformer.

A key insight of Transformer models is that a mechanism known as attention is sufficient on its own to produce highly robust predictive models (Vaswani et al. 2017). In simplest terms, attention is a process through which supervised models learn which features in a sequence are most relevant for predicting a masked target feature. A deep neural network with a large number of “self-attention heads” is able to prioritize those relevant features regardless of their distance from the target. This makes the models

computationally efficient, but also unencumbered by problems of short-term memory that had limited prior architectures (Jurafsky and Martin 2020: Chapter 9). When trained on massive text corpora, large Transformer-based models (i.e., those with many layers and hundreds of parameters) have achieved unprecedented performance on a wide range of natural language tasks (Minaee et al. 2021).

The BERT model family, widely perceived by the NLP community as a major methodological breakthrough, complements the attention mechanism of Transformers with two additional innovations: contextuality and bidirectionality (Devlin et al. 2018). Initially popularized by the ELMo model (Peters et al. 2018), a precursor to BERT, the ability to track the relative positions of words during training has enabled the generation of contextual embeddings, that is, embeddings that uniquely encode every distinct instance of a given word conditional on its context, instead of assigning a single embedding to a given word. This feature makes it possible to directly account for polysemic meanings. For instance, the term “bank” is initially mapped to an embedding vector. As this vector passes through the self-attention layers, it is re-weighted based on the terms that appear in its context, so that its representation may differ substantially if it is preceded by “river” in one instance and followed by “robber” in another (Ethayarajh 2019).

The second feature of BERT, illustrated in the “river” example, is bidirectionality—or more accurately, non-directionality—the fact that BERT predicts a masked word conditional on all words surrounding it in both directions (while keeping track of word position and relevance via attention), rather than processing them sequentially from left to right or right to left (Artetxe et al. 2022). This makes the model less prone to sequential path-dependence. In addition to these innovations, BERT also supplements the masked-word prediction training task with next-sentence prediction (Devlin et al. 2018). The result is a powerful neural language model that can be adapted for a variety of analytical tasks—including classification—through transfer learning, the approach we employ in this study.⁷

To classify discursive frames in our corpus, we use a model developed by Liu et al. (2019), known as the robustly optimized BERT pretraining approach (RoBERTa).⁸ This is an extension of the original BERT model that relies on the same general architecture but expands the training data to 63 million English news articles from CommonCrawl, 38 gigabytes of textual data from the OpenWebText corpus, and a 31-gigabyte subset of CommonCrawl data known as Stories. In addition, the pretraining tasks for RoBERTa are further optimized (e.g., dynamic instead of static masking, changes in the input representations for next-sentence

prediction, etc.), yielding enhanced performance. The model has 12 layers and a total of about 125 million parameters, all of which encode generic knowledge of the English language.

A simple way to make use of RoBERTa for our classification problem via transfer learning would be to first pass text sequences (e.g., paragraphs) from our corpus through the “frozen” model. One of the model’s layers could be used to generate document embeddings that would then serve as input into a traditional classifier (this two-step process of generating document representations and using them in a classifier would follow a similar logic to Approaches 1 through 3 in Figure 1). A more effective solution, however, is to fine-tune the RoBERTa model itself to the local corpus and, simultaneously, to the classification task. This is done by adding a classification layer on top of the pretrained model with output neurons for the different classes (e.g., populist and non-populist paragraphs), without the need for the intermediate step of encoding the documents themselves in vector form (hence the absence of the horizontal bar in the fourth diagram of Figure 1). Using human-annotated data, the model is then trained for a few additional epochs, during which the model parameters are adapted via gradient descent to specialize in the classification task at hand (Zhou and Srikumar 2021). Metaphorically speaking, rather than teaching a model how to speak English and how to identify, say, populism at the same time—as would have been the case had we relied on raw word frequency vectors or locally trained embeddings—we are teaching a model that already speaks English how to identify populism.

Because this is still a difficult task, we also employ a procedure known as hybrid active learning (Dai and Kustov 2022; Dor et al. 2020; Kazerouni et al. 2020; Settles 2009), which is particularly useful for highly skewed class proportions and small training sets: rather than fine-tuning the model with only a single batch of pre-labeled texts, we train the algorithm iteratively and add newly annotated data throughout the process. At the conclusion of each training iteration, we identify 160 documents (in our case, paragraphs) with the highest entropy—that is, the greatest classification uncertainty—and combine them with another 40 paragraphs randomly drawn from the unlabeled corpus. We then hand-code these 200 documents, return the newly labeled data to the training set, and run the classifier again. We repeat this process twice for each classification task (i.e., for each sub-dimension of populism, nationalism, and authoritarianism).

To evaluate the performance of our approach, we first compare the results of the initial round of RoBERTa classification (prior to the introduction of active learning) to those produced by random forest classifiers, a widely accepted supervised machine learning method; we then examine performance improvements to the initial RoBERTa model generated by successive rounds

of active learning. In an effort to ensure a fair comparison between the RoBERTa and random forest models, we run four variants of the latter with input in the form of, respectively, (1) raw word frequencies transformed into term-frequency-inverse-document-frequency (tf-idf) scores; (2) locally-trained word embeddings estimated with the word2vec algorithm and averaged at the document level (Mikolov et al. 2013); (3) locally-trained document-level embeddings generated with the doc2vec algorithm (Le and Mikolov 2014); and (4) pretrained word2vec embeddings averaged at the document level. These approaches correspond roughly to the alternatives illustrated in Figure 1. Hyperparameters of all the models are iteratively optimized to maximize model performance. The resulting baselines represents best-case scenarios for context-independent supervised machine learning using the random forest approach. This allows us to evaluate our claim that such traditional methods fare less well than neural language models at classifying the kinds of rare, polysemic, and vague frames that are frequently of interest to political and cultural sociologists.

The Hand-coding Process

The fine-tuning of the RoBERTa model to the political speech corpus—and the training of the random forest classifiers—requires a collection of documents that are labeled along each of the six non-mutually-exclusive dimensions of interest in our study. The first necessary decision is what unit of analysis constitutes a document. Past research on populist discourse has typically relied on the binary classification of entire speeches (e.g., Bonikowski and Gidron 2016; Hawkins 2009; Jagers and Walgrave 2007; Rooduijn and Pauwels 2011), but this approach is unlikely to work in a supervised machine learning framework: considering the infrequent occurrence of the frames in question, chances are that the signal-to-noise ratio in long documents would be too low to enable a classifier to distinguish positive from negative cases.

Just as importantly, binary coding of speeches ignores theoretically relevant variation in the intensity of populist, nationalist, and authoritarian claims (Dai and Kustov 2022). A speech featuring a single moral critique of elites, for instance, is qualitatively different in its implications for political mobilization than a speech that repeatedly reasserts the same point. This suggests that we should be measuring our six frames at a sub-speech level. Conversely, however, these framing strategies are unlikely to be fully encapsulated in single sentences. It often takes a few separate utterances to fully express the opposition between the elites and the people, to evoke particular conceptions of the nation's symbolic boundaries, or to articulate a negative

evaluation of the nation, particularly when juxtaposed with an idealized bygone era or imagined future (Bonikowski and Stuhler 2022). Consequently, we settled on paragraphs as the relevant units of analysis for the hand-coding and machine classification of our corpus.⁹

To fine-tune the classifier, we randomly drew 2,224 paragraphs from the corpus and labeled them independently for each dimension (i.e., in principle, each paragraph could have been coded as an instance of multiple discursive frames or none at all).¹⁰ To minimize coding idiosyncrasies, we generated a set of labeling guidelines for each dimension and gradually developed coding expertise by collectively discussing potentially ambiguous cases (prior to commencing the actual coding process). Every paragraph was then labeled by two coders and any resulting disagreements were adjudicated by a third coder.

These careful steps, however, did not result in unanimous agreement on an unambiguous ground truth. On the contrary, inter-coder reliability analyses yielded Cohen's kappa values of .66 (populism), .71 (exclusionary nationalism), .82 (high price), .83 (low pride), .84 (inclusive nationalism), and .9 (authoritarianism). This is important for two reasons. It demonstrates that the frames in question are inherently ambiguous, with a large number of boundary cases that defy straightforward classification. At the same time, this ambiguity affects our expectations for the performance of the classifier: the realistic goal is not to achieve perfect performance, but a balance between precision and recall that does not fare much worse than humans.

The difficulty of coding the frames also highlights a rarely acknowledged feature of the active learning method employed in our study: its ability to improve not only the algorithm's classification performance but also the analysts' understanding of the concepts being classified (cf. Evans and Aceves 2016). Recall that active learning flags paragraphs that prove particularly challenging to the classifier and returns them to the analyst for further labeling (in our case, alongside another small batch of paragraphs randomly drawn from the unlabeled data). In carrying out this procedure, it became apparent that the high-entropy cases identified by the algorithm revealed empirical patterns that the human coders had not accounted for in the first round of training. This was most likely not due to labeling errors but rather the fact that the random samples of paragraphs selected for the initial training set did not cover all possible variants of the frames in question. Having recognized this, we coded the additional paragraphs identified in the active learning process, returned them to the training set, and began the training and active learning process anew, now covering a wider range of positive cases. The ability of the algorithm to help identify additional variants of a given frame suggests that supervised machine learning need not be thought of solely as

a rote application of human coding rules to unlabeled data—when combined with active learning, it can also lead to insights in a manner reminiscent of the abductive logic of inquiry (Tavory and Timmermans 2014).

Table 2 summarizes the analytical steps we took to classify each of the six political frames (i.e., populism, exclusionary nationalism, low national pride, authoritarianism, inclusive nationalism, and high national pride). Steps 7 through 10 represent the active learning component of the workflow.

Results

Classifier Performance

We begin by reporting performance statistics for the fine-tuned RoBERTa classifier compared with the conventional random forest approach. The

Table 2. Analytical steps for labeling data, RoBERTa fine-tuning, active learning, and corpus classification.

-
1. Randomly sample 2,224 paragraphs from the corpus. Then, follow the steps below for each each of the six non-mutually-exclusive frames.
 2. Annotate each paragraph (by two independent annotators, with disagreements adjudicated by a third annotator).
 3. Identify suitable RoBERTa fine-tuning hyperparameters through a manual search.
 4. Generate 25 random splits in the labeled data, with 80 percent of each split assigned to a training set and 20 percent to a test set.
 5. Run separate RoBERTa models on the 25 splits and use each test set to evaluate model performance.
 6. Retain the best model (using area under the precision-recall curve—i.e., PR-AUC) to predict classification probabilities for each paragraph in the unlabeled data.
 7. Generate a new 200-paragraph sample from the unlabeled data with 20 percent of the paragraphs drawn randomly and 80 percent drawn based on high entropy (i.e., paragraphs with classification probability closest to 0.5).
 8. Annotate the 200 paragraphs (by two independent annotators, with disagreements adjudicated by a third annotator).
 9. Add the annotated paragraphs to the training sets of the 25 random splits described in step 4.
 10. Repeat steps 5–9, evaluating on the same test sets. In our case, two rounds of hand-coding sufficed (resulting in three runs of the RoBERTa classifiers).
 11. In the final round of models, use all 25 train-test splits to generate independent predictions for the whole corpus. Average these predictions for each paragraph and assign a label according to a .5 probability cutoff.
-

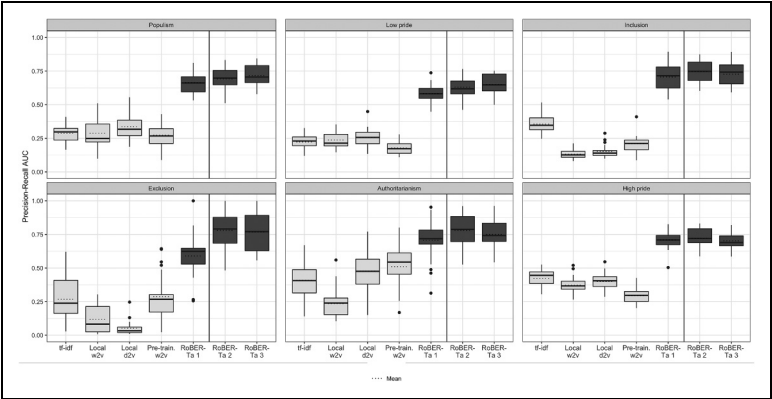


Figure 2. PR-AUC performance of RoBERTa classifier compared to random forest models.

results are visualized in Figure 2, with separate box plots for the four random forest models (in light gray) and the first round of RoBERTa classification (in dark gray). The plots also visualize the performance improvements to RoBERTa classification yielded by successive rounds of active learning (in dark gray, to the right of the vertical lines). The box plots represent the distribution of classifier performance metrics over 25 random splits. Random forest models were run on the same initial train-test splits as the RoBERTa model.¹¹

We evaluate model performance using the area under the precision-recall curve (PR-AUC) rather than the more common area under the receiver operator characteristics curve (ROC-AUC). The latter metric is less appropriate for highly imbalanced data with few positive cases, because it takes into account true negatives, which are likely to be numerous solely by virtue of the dominant class's high prevalence in the corpus (Saito and Rehmsmeier 2015). PR-AUC, in contrast, is more sensitive to the accurate classification of the low-prevalence class by focusing on the proportion of true positives in the data without taking into account false negatives.¹²

When evaluating PR-AUC performance it is important to keep in mind two constraints on the absolute values of the metric. First, the area under the curve should be interpreted relative to the prevalence of the positive class in the corpus. This is so, because an unskilled classifier that assigned paragraphs to the classes at random would yield a PR-AUC equivalent to the class prevalence rather than to 0.5, as would be the case with

ROC-AUC. The corpus class proportions are listed in Table 3 and range from 0.5 percent of paragraphs for exclusionary nationalism to 8.8 percent for high national pride. Second, note that human coders’ agreement on the appropriate paragraph labels was substantial but far from perfect, which is likely to result in an upper bound for classifier performance that is well below the theoretical maximum of 1.

The differences between the random forest (RF) classifiers and RoBERTa are stark. For five out of the six frames, the performance of the RoBERTa-based classifier far exceeds all four RF specifications. The contrast is largest for inclusive nationalism where the best RF approach—based on tf-idf—generates an average PR-AUC of 0.36, compared to 0.71 for the initial iteration of the RoBERTa models. Even for the frame with the smallest performance difference, authoritarianism, the RoBERTa classifier outperforms the best RF model (based on pre-trained word2vec embeddings) by 0.19. These results suggest not only that RoBERTa represents a breakthrough when it comes to the classification of rare, polysemic, and vague political frames, but also that traditional machine learning classifiers, at least those based on random forest algorithms, are often not up to the task at all (the latter’s performance may be acceptable for authoritarianism, but not for the other five dimensions in our data).¹³

The other conclusion from Figure 2 is that active learning can further improve the fine-tuning of RoBERTa to local corpora. For all six dimensions, the iterative labeling of high-entropy paragraphs resulted in performance improvements of the RoBERTa classifier, with the largest gains for the exclusionary nationalism and authoritarianism dimensions. The two frames that saw little improvement from active learning were inclusion and high pride,

Table 3. Classifier performance and proportion of positive cases in all paragraphs.

	Inclusion	Exclusion	Authoritarianism	Populism	High pride	Low pride
Precision	0.71	0.85	0.66	0.68	0.63	0.64
Recall	0.76	0.78	0.72	0.61	0.72	0.55
Accuracy	0.97	1	0.98	0.98	0.95	0.96
Prevalence in training set	0.048	0.01	0.027	0.029	0.076	0.055
Predicted corpus prevalence	0.051	0.005	0.023	0.026	0.088	0.042

likely due to their relatively higher prevalence in the data and less ambiguous semantic structure.

Finally, we make predictions for the whole corpus with each of our 25 models from the final round of active learning and average the predicted probabilities for each paragraph, assigning labels on the basis of a .5 cutoff. This type of ensemble learning is based on the intuition that aggregating the predictions of multiple independently trained classifiers is likely to yield better classification results than any individual classifier alone. We expect that this is also true in our case and therefore, that the classification performance in our overall corpus is somewhat better than the results reported in Figure 2. However, the high imbalance of the frames in the corpus makes the annotation of an additional, sufficiently sized test set too costly, precluding us from directly estimating improvements in classification performance added by this additional step.

Trends in Populism, Nationalism, and Authoritarianism Across 34 Presidential Campaigns, 1952–2020

We now move on to the central substantive question of our study: were populist, nationalist, and authoritarian claims present in U.S. presidential campaigns between 1952 and 2020 and if so, how did their use vary over time? The time series for the six discursive frames is presented in Figure 3.

The first thing to notice is the outsized prevalence of populism, exclusionary nationalism, and authoritarianism—three discursive frames typically associated with radical-right discourse—in the 2016 and (to a lesser degree) 2020 Trump campaigns. This pattern is consistent with popular accounts of Trumpism as predicated on xenophobic, racist, anti-elite, illiberal, and anti-democratic claims, as well as with its characterization as a watershed in contemporary U.S. radical-right politics. That Trump's reliance on exclusion and authoritarianism in both campaigns—and on populism in 2016—exceeded all other Republican and Democratic campaigns provides supporting evidence for Hypotheses 1a and 1b.

Hypothesis 1c further predicted that Donald Trump's speeches were particularly likely to mobilize low levels of national pride through negative assessments of the state of the nation. This was based on the intuition that low pride shares affinities with populist condemnations of elite corruption and on the results of past research that identified low levels of national pride as significant demand-side predictors of Trump support (Bonikowski, Feinstein, and Bock 2021). This expectation is confirmed in the third panel of Figure 3: low pride was a particularly common feature of the 2016

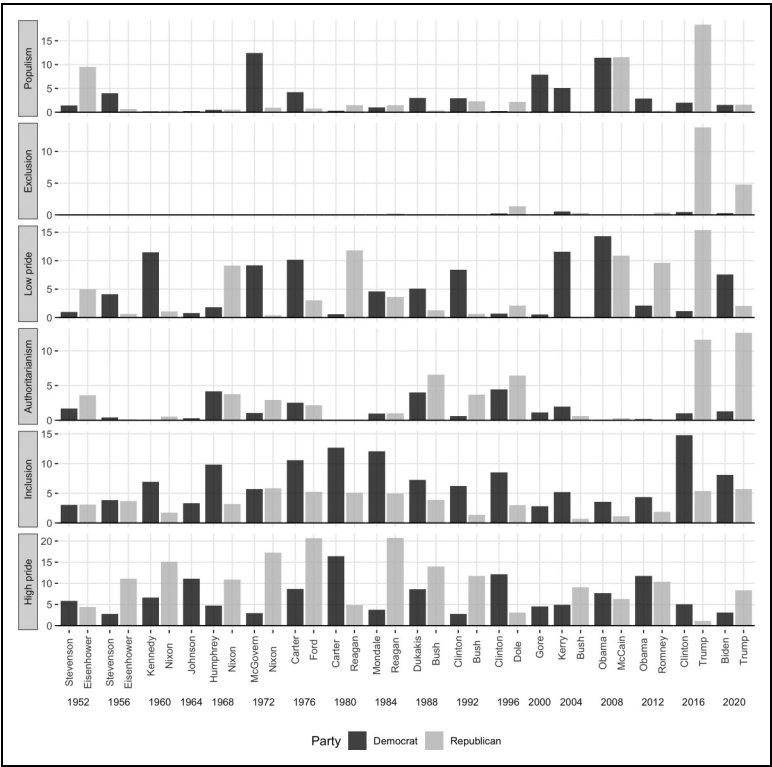


Figure 3. Share of populist, nationalist (i.e., those featuring exclusion, inclusion, and low and high pride), and authoritarian paragraphs by campaign, 1952–2020. *Note:* The y-axis shows each campaign’s mean share of frame-specific paragraphs per speech (in percents).

Trump campaign, roughly on par with populism, exclusionary nationalism, and authoritarianism, and Trump used this frame more frequently than any other candidate in our data (that this was not the case in 2020 is a finding to which we will return below).

Next, we consider the five hypotheses concerning mainstream political discourse between 1952 and 2020. The overarching conclusion from Figure 3 is that most of the radical-right frames used by Donald Trump were not unique to his campaign. Though perhaps not in the same combinations and not with the same intensity, mainstream candidates routinely

relied on similar claims-making strategies throughout the 68 years covered by our data.¹⁴

We begin with populism. In line with past research on the topic (Bonikowski and Gidron 2016; Dai and Kustov 2022; Fahey 2021), populism appears to be a relatively common feature of both Democratic and Republican campaigns. On the Republican side, the greatest share of populist paragraphs is found in the speeches of John McCain in 2008 and Dwight Eisenhower in 1952.¹⁵ Among Democrats, populism was most prevalent in George McGovern's 1972 campaign and Barack Obama's 2008 campaign.¹⁶ These findings provide support for Hypothesis 2a.

Hypothesis 2b concerned the use of exclusionary nationalism by mainstream Republican candidates. Contrary to our expectations—and in contrast to the other constitutive components of radical-right politics—the evocation of exclusionary nationalism appears to have been largely limited to the Trump campaign. We find no evidence of other campaigns, whether Democratic or Republican, relying on this frame. This finding should not be interpreted, however, as dispositively indicative of the absence of racist or xenophobic discourse in presidential campaigns prior to 2016. Instead, it may be the case that Donald Trump explicitly articulated claims that had been implicit in the speeches of prior presidential nominees.

How can we detect implicit exclusionary claims in mainstream campaigns? Given the highly coded nature of such discourse, its measurement using automated methods is difficult. Nonetheless, our analysis provides two forms of suggestive evidence, as reflected in Hypotheses 2d and 2e, respectively: mainstream candidates' reliance on authoritarian claims and their failure to invoke inclusive conceptions of American identity. As illustrated in Figure 3, authoritarianism—that is, discourse that threatens the use of state power against domestic “enemies,” advocates for punitive law-and-order policies, and violates liberal democratic norms—was prevalent throughout our time series. Since this frame is one of the three components of radical-right politics, this pattern is in line with our general argument about the precedents for radicalism in mainstream discourse. Furthermore, as predicted in Hypothesis 2d, reliance on authoritarianism is greater among Republican than Democratic candidates. Prior to Donald Trump, the largest share of authoritarian paragraphs was found in the 1988 and 1992 George H. W. Bush campaigns, the 1996 Bob Dole campaign, and the 1968 Richard Nixon campaign. At relatively lower levels, authoritarianism was also observed in the 1968, 1976, 1988, and 1996 Democratic campaigns. These findings are in line with historical accounts of law-and-order,

tough-on-crime, and anti-welfare claims serving as implicit racial cues in U.S. politics (Carter 1996; Gilens 2009; Lieberman 2022; López 2015; Mendelberg 1997, 2017; Piliawsky 1989).

In addition to authoritarian dog whistles (López 2015; Wetts and Willer 2019), our results reveal another pattern of indirect appeals to ethno-nationalism, one that does not depend on explicit exclusionary claims, but rather, on the absence of references to American ethnoracial pluralism. As Figure 3 demonstrates, inclusive nationalism is a particularly common frame in presidential candidates' speeches (second in prevalence only to expressions of national pride). This is likely so because egalitarianism is a dominant feature of the American national creed (Lipset 1990) and is therefore a low-risk mode of political claims-making. Even though all candidates in our data made use of inclusive nationalism, however, they did not do so uniformly. Republicans were considerably less likely than Democrats to rely on inclusive nationalism throughout the time series and the 2004, 2008, and 2012 elections featured particularly low levels of inclusive discourse by both parties.¹⁷ The former pattern is consistent with Hypothesis 2e.

The lower prevalence of inclusion among Republican candidates could be interpreted as a matter of emphasis: both parties rely on inclusive nationalism but Republicans do not prioritize it as highly as Democrats. This conclusion is plausible, but it is complicated by another trend in the data: the election-specific gap between the two parties' use of inclusive nationalism shifts in magnitude between elections. In other words, in a number of elections, Democrats placed strong emphasis on inclusion and Republicans did not follow suit. Since political appeals to American inclusiveness are generally anodyne, it is reasonable to view a candidate's decision not to echo an opponent's egalitarian discourse as a strategic choice. Reluctance to do so may serve as a subtle signal to voters about the two campaigns' contrasting views of America's symbolic boundaries.

This contrast is further highlighted in Figure 4, which visualizes the election-specific differences in the use of inclusive nationalism by the two parties. Not surprisingly, the largest difference is observed in 2016, between the Trump and Clinton campaigns. This is followed by the 1980 and 1984 elections that pitted Ronald Reagan against Jimmy Carter and Walter Mondale, respectively, and the 1968 contest between Nixon and Humphrey. When considered alongside the results for authoritarianism, these findings suggest that *implicit* exclusionary cues have been particularly prevalent among four Republican presidents: Nixon, Reagan, G. H. W. Bush, and Dole.¹⁸ Again, this is consistent with historical accounts of coded racial appeals in U.S. presidential politics (Maxwell and Shields 2019; Mendelberg 2017).

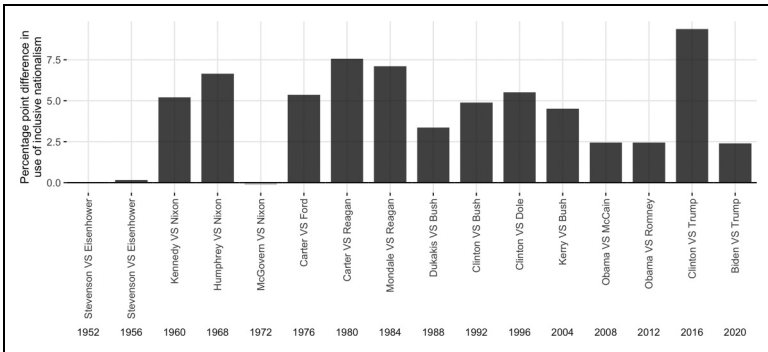


Figure 4. Election-specific difference in use of inclusive nationalist paragraphs between competing campaigns.

Note: The y-axis shows the difference in the average percentage of inclusive nationalist paragraphs per speech between pairs of campaigns. Positive numbers indicate a higher use of inclusive nationalism by the Democratic candidate compared to the Republican candidate.

The next discursive frame we examine consists of expressions of low national pride. Because popular conceptions of American nationhood characterized by low pride have been associated with support for Donald Trump and Bernie Sanders in the 2016 election (Bonikowski, Feinstein, and Bock 2021) and low pride has logical affinities with populism, we hypothesized that this campaign style should be found among mainstream candidates from both parties. Indeed, consistent with Hypothesis 2f, this is the pattern we observe in Figure 3. Evocations of low pride appear to be common across elections, account for a greater proportion of speech content than populism or authoritarianism, and are observed in similar proportions, on average, among candidates from both parties.

There is another, more subtle pattern in the low-pride time series as well: in most elections, only one of the two candidates appears to have relied on low-pride frames. It is possible that this is a function of incumbent-challenger dynamics, an intuition that is further supported by the time series for expressions of high pride: like low pride, high pride features prominently in mainstream campaigns and displays a similar zero-sum pattern across candidates within the same elections. It appears then that both low pride and high pride oscillate from campaign to campaign depending on the candidates' positions within the political field. We return to this in a subsequent analysis.

Associations Between Populism, Nationalism, and Authoritarianism Among Mainstream Candidates, 1952–2020

Thus far, we have focused on the prevalence of each of the six frames in isolation from one another. The empirical patterns we have identified, however, also suggest systematic relationships between the frames. We have argued that populism and low pride share the tendency to morally condemn the current state of the nation; this suggests a positive correlation between them. Also, low and high pride both appear to display a zero-sum logic, whereby only one campaign within any election is likely to emphasize each of these frames. That, along with the logical contradiction between low and high pride, suggests that they should exhibit a substitutive rather than complementary relationship within campaigns. If so, we would expect low and high pride to be negatively correlated.

We examine these possibilities in Figure 5 using principal component analysis (PCA) and pairwise correlations. Both are based on the campaign-level prevalence of the frames shown in Figure 3. For the PCA, we normalize the dimensions to ensure that each carries equal weight in the analysis. Because we are interested in how the frames are used in mainstream politics, the analysis does not include the 2016 and 2020 Trump campaigns.¹⁹ We also omit exclusionary nationalism because of its near-absence from mainstream campaigns.

The first component in the PCA accounts for 38.3 percent of the variance (see Appendix C for additional details) and distinguishes populism and low pride from high pride and authoritarianism. The former two frames are highly correlated ($r = .47$) and both are negatively associated with authoritarianism and high pride, though the latter appear to be unassociated ($r = .02$). The second component covers considerably less variance and is mainly associated with inclusion, which is negatively correlated with populism ($r = -.21$) and high pride ($r = -.09$). The third component (shown in Appendix C) captures an almost equal share of variance and is strongly associated with authoritarianism.

Consistent with the previous results, Democratic campaigns tend to be associated with the inclusion pole of the second PCA dimension. Democrats also lean toward the low pride and populism pole of the first dimension, whereas Republicans appear to make more frequent use of high-pride and authoritarianism. The edges connecting campaigns from the same election show that the latter pattern holds true for 10 out of the 15 campaign pairs included in the analysis. Among several notable exceptions are Carter's 1980 campaign, which relied heavily on high pride while his opponent,

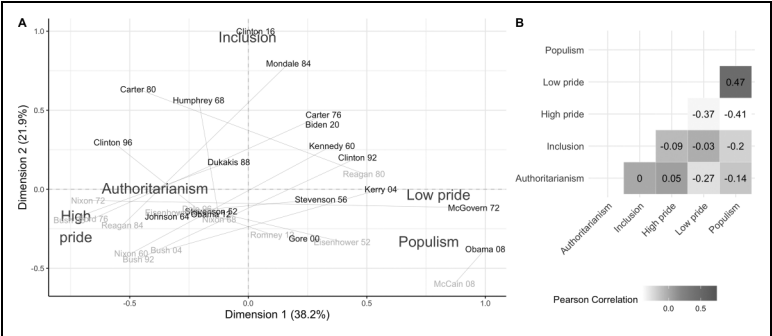


Figure 5. Principal component and correlation analysis of discursive frames and campaigns; mainstream candidates only, 1952–2020.
Note: Panel A shows the first two dimensions of a principal component analysis of campaigns and frame prevalence. The frames were normalized to ensure that each carries equal weight in the analysis. Axes represent loadings for frames and scores for campaigns. The latter were standardized by the maximum absolute loading across campaigns to fit into the same figure. Edges connect campaigns from the same election. Panel B shows the correlations between the five remaining discursive frames.

Reagan, made frequent use of low pride, as well as McCain’s 2008, Eisenhower’s 1952, and Romney’s 2012 campaigns, all of which featured low-pride claims. In subsequent regression models we will further investigate whether party affiliation is associated with the use of each frame, controlling for other speech and candidate attributes (to foreshadow our findings, this is the case for inclusion and high pride, but not the other frames).

An additional insight revealed by Figure 5 is that two of the frames typically associated with radical-right politics—populism and authoritarianism—do not frequently co-occur in mainstream campaigns. In fact, they are negatively correlated at -0.14. This suggests that even though populism and authoritarianism were frequently used independently of one another prior to the 2016 election, one of Donald Trump’s innovations consisted of fusing them in his campaign discourse. It appears that it is such creative recombination, as well as the explicit mobilization of exclusionary nationalism, that set Trump apart from his mainstream predecessors.

These intuitions are further substantiated in Figure 6, which visualizes the configurations of frames employed by the eight campaigns that were least dissimilar from Trump 2016 in their overall reliance on populism, authoritarianism, low and high pride, and inclusive nationalism. What is striking is that no

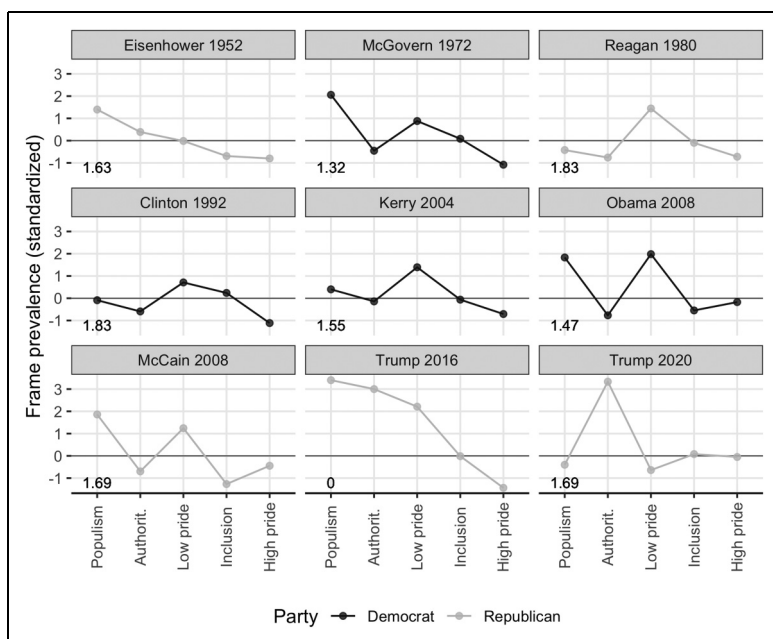


Figure 6. Standardized frame prevalence for campaigns with most similar discursive profiles to Trump 2016.

Note: The y-axis shows each campaign's standardized frame prevalence. The eight campaigns were selected based on greatest overall similarity to the 2016 Trump campaign (also included for reference) across the five frames shown on the x-axis (exclusionary nationalism was omitted due to its absence in mainstream campaigns). The campaigns are listed in chronological order; their overall difference from the 2016 Trump campaign is indicated in the bottom-left of each panel.

other campaign—including Trump's own 2020 bid for reelection—featured the same combination of frames observed in Trump's speeches during the 2016 election. These frames are certainly visible in prior mainstream campaigns—as in, for instance, Eisenhower's use of populism and authoritarianism in 1952, McGovern's reliance on populism and low pride and avoidance of high-pride frames in 1972, or McCain's combination of populism with low levels of inclusive nationalism in 2008—but while these campaigns drew on—or conspicuously avoided—two or three of the frames at a time, only Trump in 2016 simultaneously exploited as many as four of them and did so at such high levels of intensity. These findings suggest that mainstream

campaigns may have sowed the seeds of populist, nationalist, and authoritarian discourse, but it took a radical politician with a penchant for norm breaking and an instinctive ability to please a reactionary crowd (Karakaya and Edgell Forthcoming; McVeigh and Estep 2019) to combine these frames in a new way—and to capture and further radicalize the country's center-right party in the process.

Predictors of Populist, Nationalist, and Authoritarian Discourse Among Mainstream Candidates, 1952–2020

Next, we turn to a regression analysis that predicts the speech-level prevalence of each frame based on the attributes of the campaigns. This allows us to test the hypotheses concerning the effects of incumbency status and partisan affiliation on political discourse. We also include an independent variable measuring the timing of a speech within a campaign because a candidate's incentives to appeal to a broad audience rather than the party base may increase with proximity to Election Day. We interact speech timing with incumbency status, since challengers and incumbents may respond to such incentives in different ways (Bonikowski and Gidron 2016). The results, presented in Table 4, are based on negative binomial regression models with standard errors clustered at the campaign level. All models control for speech length. Because our interest is in the predictors of frame use by mainstream candidates, we exclude Donald Trump's 2016 and 2020 campaigns from the analysis (their inclusion, however, makes little substantive difference in the results, as shown in Appendix F).²⁰

Consistent with our earlier findings, there is clear evidence that Democrats are more likely to make use of inclusive nationalism. Conversely, Republicans appear to be more likely to engage in high-pride and authoritarian discourse, but neither of these differences are statistically significant at the $p < 0.05$ level. Neither populism nor low national pride are associated with party, further confirming that these frames are sufficiently flexible to express varied ideological claims (Mudde 2007). Furthermore, in line with Hypotheses 4a and 4b, populism, low pride, and high pride are associated with party incumbency status: populism and low pride are more likely to be used by challengers, who benefit from critiquing the political establishment and disparaging the state of the nation, whereas high-pride claims are more common among candidates from the incumbent party who have incentive to tout their party's accomplishments and frame them as expressions of the nation's virtues. This confirms our previous observation that expressions

Table 4. Negative binomial regressions of speech-level frame prevalence on speech attributes, 1952–2020; mainstream candidates only.

	Populism (1)	Authorit. (2)	Inclusion (3)	Low pride (4)	High pride (5)
Republican	0.19 (0.42)	0.45 (0.41)	−0.93*** (0.20)	−0.10 (0.24)	0.32+ (0.17)
Incumbent party	−1.28** (0.39)	0.11 (0.40)	0.12 (0.19)	−1.68*** (0.26)	0.80*** (0.17)
Weeks to election (centered)	0.03 (0.03)	−0.02 (0.05)	0.03 (0.04)	0.004 (0.01)	0.05* (0.02)
Incumbent X weeks	−0.06 (0.06)	0.07 (0.06)	0.04 (0.05)	−0.03 (0.03)	−0.04 (0.03)
N of campaigns	32	32	32	32	32
N	2,911	2,911	2,911	2,911	2,911

+ $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.

of low and high pride are rarely combined within campaigns and instead offer candidates alternative ways of framing the nation within any given election.

We conclude our discussion of the regression results with the remaining two variables in Table 4: the timing of the campaign speeches and its interaction with incumbency status. Contrary to our expectations, we find little in the way of systematic association between speech timing and the use of the five discursive frames. The one exception involves high pride: challengers—who are generally less likely to use this frame than incumbents—further decrease their emphasis on high pride as Election Day approaches. Given, however, that the interaction effect is only marginally significant and we observe no comparable finding for low pride (or any other frame, for that matter), we hesitate to draw strong conclusions from it. In general, there appears to be scant evidence of temporal shifts in the campaigns’ use of populism, nationalism, and authoritarianism during presidential elections.

Finally, we address Hypothesis 4c, concerning the discursive differences between the 2016 and 2020 Trump campaigns. We argued that Donald Trump’s incumbent status in 2020 may have led him to rely less frequently on populist and low-pride frames than he had in 2016, given that critiques of governing elites and negative evaluations of the nation are less persuasive when the candidate is in power. The incumbency effects observed in the regression analyses further justify these expectations. As illustrated in

Figures 3 and 6, we indeed find that populism and low pride were far less prevalent in Trump's 2020 speeches compared to his 2016 campaign—and even to the campaigns of some mainstream candidates. This underscores the importance of treating populism and national pride—as well as articulations of “thin-centered” ideologies in general (Freeden 1998, 2003)—as attributes of claims rather than political actors. Doing so makes it possible to observe shifts in political strategy in response to contextual factors, including candidates' changing positions in the political field (Bonikowski and Gidron 2016).

Discussion and Conclusion

By leveraging the classification capabilities of large neural language models, we were able to identify subtle and rare political frames in over 70,000 paragraphs of text, and to do so with a degree of accuracy unmatched by traditional text analysis methods. This allowed us to answer our central substantive question: were populism, nationalism, and authoritarianism unique to Donald Trump's radical-right discourse or were these frames also found in the political campaigns of mainstream U.S. presidential candidates between 1952 and 2020?

Our results demonstrate that populism and low levels of national pride were indeed common among mainstream candidates from both parties. Exclusionary nationalism, however, was not—its presence in the corpus was limited to the 2016 and 2020 Trump campaigns. Instead, mainstream candidates relied on two strategies that may have implicitly cued ethnoracially exclusionary beliefs in the electorate: they made use of authoritarian claims that may have served as dog-whistles for racism and xenophobia and they conspicuously refrained from engaging in inclusive nationalist discourse, a mainstay of U.S. electoral politics. Moreover, we show that populism, authoritarianism, and low pride were commonly found among candidates from both parties, whereas high national pride was more prevalent among Republicans and inclusive nationalism was found more frequently in Democratic speeches. These patterns yielded positive correlations between populism and low pride, as well as negative correlations between low and high pride, inclusive nationalism and populism, authoritarianism and populism, and authoritarianism and low pride. Finally, our regression analyses suggest that populism and low pride tend to be employed by challengers, whereas high pride is more typical of candidates representing the incumbent party.

It is important to emphasize that because our corpus is limited to general-election presidential campaigns, we make no claims about how presidential

primaries, state or local elections, media content, or policy debates have used and legitimated populist, nationalist, or authoritarian frames. These scope conditions are particularly relevant for interpreting the infrequent use of explicitly exclusionary claims by candidates other than Donald Trump. This finding should not be seen as a dismissal of the long history of racial appeals in U.S. political culture. On the contrary, it is likely that mainstream campaigns' racist and xenophobic dog-whistles—from Barry Goldwater's support for "states' rights" and Richard Nixon's emphasis on "law and order" through George H. W. Bush's Willie Horton ad to Mitt Romney's flirtation with the birther movement (López 2015; Mendelberg 2017; Youngman 2012)—have drawn their potency from subtly indexing more overt efforts to maintain white racial domination (Bonilla-Silva 2001; Omi and Winant 2014) by actors operating in political domains not covered by our corpus.

Taken together, our study's findings point to several methodological and substantive contributions. First, we demonstrate the utility of neural language models and transfer learning—specifically the BERT family of models—for classifying texts within a supervised framework. The discursive constructs in our analysis share a number of characteristics with a common subset of cultural phenomena that are of interest to social scientists. These object of analysis are not substantive topics in themselves, but frames through which substantive topics are expressed; they tend to occur rarely in texts; they are expressed using polysemic terms; and their meaning is often implicit, generating a wide range of boundary cases that are difficult for human coders—and not just algorithms—to classify. Because of these challenges, many traditional computational methods—including dictionary methods, topic models, and standard supervised machine learning—are ill-equipped to handle the classification of such phenomena. Manual coding, on the other hand, is typically not scalable to large corpora.

In light of these limitations, analysts interested in identifying subtle discursive frames in large corpora have been typically forced to choose from among a number of inadequate methods. The introduction of Transformer-based models, such as BERT and RoBERTa, promises to remedy this situation. While inductive methods like topic modeling or exploratory embeddings-based analysis continue to offer reasonable solutions when the categories of interest are unknown in advance, the high accuracy of neural language models makes them a preferable method for measuring predefined frames. Since scholars often know what phenomena they want to study in light of past theory, the latter methods should be applicable to a wide range of social scientific questions.

In describing the virtues of topic models, DiMaggio, Nag, and Blei (2013) argue that automated text analysis methods should yield results that are interpretable, capture polysemy, and account for the heteroglossia (or multiple voices) in any given text. We believe that the approach we have taken in this study accomplishes these goals, even though it is based on a supervised learning framework. Indeed, part of the reason why Transformer-based models are so powerful is that their logic is consistent with how language actually works: they are able to encode multiple senses of the same word in separate vector representations, thereby directly modeling polysemy. Although the models themselves do not incorporate heteroglossia into the classification process, we did so by choosing paragraphs as the units of analysis and training separate classifiers for the multiple dimensions of radical-right discourse. This allowed us to directly investigate how these frames co-occur within speeches and campaigns. Finally, even though the classification process employed by supervised deep learning models is largely inscrutable due to these models' architecture, the hand-coding of texts for the training set is itself a hermeneutic exercise that allows the analyst to gain a deeper understanding of the frames being studied. By incorporating active learning into our workflow, we not only improved the accuracy of the classifier, but also allowed the algorithm, which directs us to those highly uncertain boundary cases, to further guide our reading of the texts, thereby affording us additional insights into the subtleties of our objects of analysis. In light of these features, we view our application of RoBERTa and active learning as not only an exceptionally accurate mode of text classification, but also one that is in keeping with the central tenets of cultural sociology.

Our methodological contribution is relevant for multiple social scientific domains, but it has the most direct implications for the study of populism—an interdisciplinary area of research well attuned to the importance of discursive frames and their resonance with demand-side political beliefs. Because populism has been increasingly defined in ideational terms (Hawkins et al. 2018), scholars have frequently sought to examine its presence in political texts; in so doing, they have encountered many of the difficulties we have enumerated in this study. The most common solutions have relied on the hand coding of small subsets of large corpora (Hawkins 2009; Hawkins et al. 2019; Jagers and Walgrave 2007; Fahey 2021) and on document-level dictionary-based classification (Rooduijn and Pauwels 2011; Bonikowski and Gidron 2016), methods that have considerable limitations. Our study provides a useful alternative to these approaches for populism scholars interested in extending their investigations to larger document collections without sacrificing measurement accuracy and in relating populist

claims to other discursive aspects of radical politics, a research agenda advocated by recent programmatic interventions into the field (Jenne, Hawkins, and Silva 2021; Rooduijn 2019).

The latter point highlights one of the substantive contributions of our study: rather than focusing on populism alone, we formulate and test hypotheses for a full range of discursive frames typically attributed to radical-right actors. Scholars increasingly agree that the radical right combines populism, nationalism (or nativism), and authoritarianism (Mudde 2007; Rooduijn 2014), but few studies actually measure these three components in political discourse. As a result, we know relatively little about how these frames interact with one another, under what circumstances each is used by candidates, and—most relevant for our purposes—whether their use by radical actors has precedents in mainstream politics. Our approach demonstrates the value of closer alignment between theoretical categories and empirical inquiry in radical-right scholarship. Future research could extend our framework to other country cases and domains of political discourse (e.g., Congressional elections, party manifestos, or parliamentary debates). By doing so, scholars could develop a better understanding of how populism, nationalism, and authoritarianism are bundled together, how incentives to use these frames are shaped by period effects and candidates' positions in political fields, and how campaigns adjust their discursive strategies in response to their competitors.

Finally, by focusing on the three types of frames within a single analytical framework, we were able to demonstrate that, at least in U.S. presidential elections, the sharp analytical boundary between mainstream and radical political discourse may be overstated. While Donald Trump was unusual in the intensity of his rhetoric and in his simultaneous combination of populism, nationalism, and authoritarianism (especially in 2016), he was far from the first U.S. presidential candidate to have relied on these discursive strategies. In fact, populism, low levels of national pride, authoritarianism, and implicit (but not explicit) exclusionary nationalist claims have been commonplace in presidential campaigns over the past six decades.

Should we conclude that such continuities undermine the very distinction between “mainstream” and “radical” politics? In our view, this conceptual boundary is still relevant. Mainstream presidential candidates consistently evoke—even if only as lip service—the normative scripts of liberal democracy and accept—even if reluctantly—the principles of partisan mutual toleration, institutional forbearance, and the peaceful transfer of power (Levitsky and Ziblatt 2018). In a country plagued by persistent racial inequality in access to political, civil, and social rights (Marshall 1950), such acts of compliance with liberal democratic norms enable the continued

pursuit of greater justice—however frustrated and circuitous—via the institutional channels of formal politics. In contrast, radical-right (or far right) parties and politicians seek to foreclose such opportunities and confine political power in perpetuity to the ethnoracial majority (Mudde 2019, 2022; Pirro Forthcoming).

Although the mainstream candidates in our corpus lacked the deeply illiberal and undemocratic predilections of Donald Trump, many of them nonetheless ran campaigns that eagerly encouraged and exploited public antipathies toward elites and minorities and signaled the willingness to turn state institutions against society's most vulnerable members. What our findings suggest is that when mainstream candidates use such incendiary frames to gain electoral support from their more radical constituents, they are engaging in a strategy that may eventually undermine their own political position—and in extreme cases, the very stability of liberal democratic institutions. Even if some mainstream candidates have scant interest in pursuing policies that are in line with their morally charged electoral claims, their campaigns' reliance on populism, declinist and (implicitly) exclusionary nationalism, and authoritarianism risks legitimating cultural tropes that may be subsequently exploited by skillful radical outsiders who not only talk a big game but actually follow through on what they say.

Authors' Note

A replication package including all data and code used in this study is publicly available at <https://bit.ly/3OoVd7K>.

Authors' Contributions

All authors contributed equally to the conceptualization, design, and execution of the study, as well as to the annotation of the data. BB conceived of the project, managed it, and wrote the paper. YL collected and wrangled the data, experimented with a range of analytical methods, and operationalized the hybrid active learning approach. OS introduced the RoBERTa approach to the project, designed and trained the classifiers, executed the analyses, drafted parts of the methods and results sections, and generated the visualizations.

Acknowledgments

The authors are grateful for valuable feedback from Laura Nelson, Carly Knight, Paul DiMaggio, Clayton Childress, Iddo Tavory, Andy Perrin, Jacob Habinek, Adel Daoud, Robert Lieberman, and audience members at the NYU Culture Workshop, the International Roundtable on Computational Social Science at Linköping University's Institute for Analytical Sociology, and the Johns Hopkins University's SNF Agora Institute.

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 iDs

Bart Bonikowski  <https://orcid.org/0000-0003-0107-0644>

Yuchen Luo  <https://orcid.org/0000-0002-0923-7245>

Oscar Stuhler  <https://orcid.org/0000-0001-7391-1743>

Notes

1. Following common convention in the literature (e.g., Gest, Reny, and Mayer 2018; Gidron and Hall 2020; Kivisto 2017), we classify Donald Trump's 2016 and 2020 campaigns as instances of radical-right politics and treat all other Democratic and Republican Party campaigns in the post-War period as "mainstream" (we prefer this label over "moderate" because the latter flattens ideological differences between candidates). If the scope of our study were to include primary campaigns, we could plausibly include several other candidates—for instance, George Wallace or Pat Buchanan—in the "radical-right" category alongside Donald Trump. The classification of specific candidates aside, these binary categories should be understood as ideal types; we rely on them precisely to demonstrate that empirically, there is more discursive continuity between actors conventionally assigned to the two categories than is often acknowledged. Finally, it should be noted that scholars have recently advocated for the use of the descriptor "far right" instead of "radical right," but this has not altered the characterization of Donald Trump as an outlier from the historical mainstream when it comes to U.S. presidential politics (Mudde 2019, 2022; Pirro Forthcoming).
2. Nationalist beliefs and claims also involve judgments about the nation's superiority vis-à-vis other countries, i.e., chauvinism (Huddy, Del Ponte, and Davies 2021; Kosterman and Feshbach 1989). In the interest of streamlining our analyses, however, we do not consider chauvinist frames in the present study.
3. Democratic nominees' use of populist frames could be seen as evidence of their radical-left leanings, but such a conclusion, in our view, would be mistaken. The radical left's political project consists of overturning the capitalist order and dispossessing the existing elites of wealth and power (March 2012). Even though

- radical-left parties (especially in Europe) have gradually shifted from a socialist and communist ideology toward broadly populist appeals (March and Mudde 2005), Democratic presidential nominees have used populism to mobilize support for center-left policies rather than the subversion of existing institutions.
4. Note that the degree to which these methods are unsupervised varies across applications. In many instances, considerable effort goes into the construction of word lists that are then used to explore the embedding space, a process that combines embeddings and dictionary methods.
 5. To be clear, our claim is not that neural language models are suitable only for identifying the kinds of subtle frames covered by our study. Indeed, they tend to outperform other supervised machine learning methods on most classification tasks (Minaei et al. 2021), but the performance differential may be especially marked in difficult cases, where traditional classifiers fare more poorly. Simpler modes of supervised learning may be preferable, on the other hand, for tasks that require local training due to the high specificity of the focal corpus. Alternatively, when the categories to be measured are not known ahead of time, inductive methods, like topic models or interpretive analyses of an embeddings space, may be more useful.
 6. For a more technical introduction to neural language models, we recommend Jurafsky and Martin (2020: Chapters 7 and 9); for a review oriented toward social science applications, see Terechshenko et al. (2020).
 7. Recently, language models have drastically increased in size and have been trained on unprecedented volumes of textual data. The most advanced of these, like GPT-3 (Brown et al. 2020) and OPT (Zhang et al. 2022), contain over one hundred billion parameters and have achieved remarkable performance on complex tasks with minimal input from the end-user. Their ability to generate original text (and images) have led some to view them as potential precursors to generalized artificial intelligence (Johnson and Izhev 2022). These models may provide marginal gains in classification performance over BERT and its derivatives, but the computational and financial costs currently associated with their use make them unattractive for most sociological applications.
 8. RoBERTa, as well as a variety of other language-models, are maintained on huggingface.co, which also provides the Transformers Python library (Wolf et al. 2020).
 9. We merged shorter paragraphs consisting of one or two sentences with adjacent paragraphs.
 10. Because we added the 2020 data to the corpus after the conclusion of our initial analyses, those speeches were not included in the training and test sets. They were labeled by the RoBERTa classifiers fine-tuned with the 1952–2016 data.

11. In presenting all results, we first list the four frames most commonly associated with radical-right discourse (i.e., populism, exclusionary construals of national membership, evocations of low levels of national pride, and authoritarianism), followed by the two remaining types of nationalist appeals (i.e., inclusive national membership and expressions of high pride). The alternative of grouping the four nationalist frames together, while reasonable, would be less consistent with the theoretical assumptions in the literature on radical politics.
12. For comparison, we provide the ROC-AUC metrics in Appendix A. As expected, ROC-AUC is less discriminating in its evaluation of the alternative models.
13. One disadvantage of neural language models pretrained on large contemporary corpora, such as Wikipedia or CommonCrawl (Liu et al. 2019), is that the models may not cover the full range of linguistic variation in target texts drawn from specialized domains or past historical periods. Fine tuning and the supplementation of the training corpus with domain-specific data, however, can help offset this problem. To determine whether differences between the pretraining and target data generate bias in our study, Appendix E reports model performance statistics across subsets of the corpus. Aside from improved recall with speech length, we do not observe meaningful patterns in classification quality across document attributes. Although speeches from 1956-1968 differ significantly from 2008-2016 in terms of recall (not precision), there is no overall monotonic trend in model performance over historical time.
14. Given the high prevalence of populism, exclusionary and declinist nationalism, and authoritarianism in Donald Trump's 2016 campaign, Trump speeches may have had a disproportionate influence on the fine tuning of the RoBERTa classifier. If so, however, this should make our results more conservative, since we may be underestimating the presence of less extreme variants of the frames in the campaigns of mainstream candidates. Validation analyses presented in Appendix E suggest that this may have been the case for several campaigns from the first two decades of our data.
15. Though the McCain finding may seem surprising, it was confirmed via a close reading of the speeches. Whether due to the context of the financial crisis, inspiration from his running mate, Sarah Palin, his self-presentation as a "Maverick," or response to his opponent, Barack Obama, McCain's 2008 campaign appears to have relied heavily on anti-elite language.
16. The patterns in the populism time series differ from those found in the same data by Bonikowski and Gidron (2016). This appears to be primarily a function of different units of analysis. When we binarize populism at the speech rather than paragraph level, our respective results become more closely aligned (see Appendix B).
17. Interestingly, the second- and third-highest mean proportion of inclusive paragraphs among Republican candidates is found in Donald Trump's 2020 and 2016 campaigns. This may be a function of Trump's attempts to offset his

- exclusionary language with vaguely inclusive appeals, as in his patronizing promises to protect African Americans by fixing “inner cities” or the statement “Hispanics love me” (Covert 2016; Saul 2015). Nonetheless, these claims were far outpaced by his explicitly exclusionary discourse.
18. The negative difference between McGovern and Nixon in 1972 is a notable outlier. It is unlikely that this is indicative of Nixon’s turn away from implicit exclusion, since his 1972 campaign directly appealed to a “silent majority” of white voters (Spitzer 2012). More likely, having been accused of racism by the Democrats four years prior, Nixon attempted to soften his authoritarian appeals by couching them in inclusive nationalism. In turn, McGovern did not engage in the same level of inclusive nationalism as most of his Democratic counterparts.
 19. We know from prior research that radical-right actors tend to rely on all three discursive components (i.e., populism, exclusionary and declinist nationalism, and authoritarianism) and this is further confirmed by our trend analysis. Given the high prevalence of these frames in Trump’s speeches, especially in 2016, including them in the PCA and correlation analyses would distort our conclusions about the discursive strategies of mainstream politicians. For results generated using the full corpus, see Appendix F.
 20. For descriptive statistics, see Appendix D. In earlier models, we controlled for average paragraph length (in words) per campaign, but this had no significant effect on any of our dependent variables and did not alter our substantive conclusions.

References

- Akkerman, Agnes, Cas Mudde, and Andrej Zaslove. 2014. “How Populist Are the People? Measuring Populist Attitudes in Voters.” *Comparative political studies* 47:1324-53.
- Annenberg. 2000. *The Annenberg/Pew Archive of Presidential Campaign Discourse*. Philadelphia, PA: The Annenberg School for Communication.
- Artetxe, Mikel, Jingfei Du, Naman Goyal, Luke Zettlemoyer, and Ves Stoyanov. 2022. “On the Role of Bidirectionality in Language Model Pre-Training.” *arXiv* 2205.11726.
- Bonikowski, Bart, Yuval Feinstein, and Sean Bock. 2021. “The Partisan Sorting of ‘America’: How Nationalist Cleavages Shaped the 2016 U.S. Presidential Election.” *American Journal of Sociology* 127:492-561.
- Bonikowski, Bart and Noam Gidron. 2016. “The Populist Style in American Politics: Presidential Campaign Discourse, 1952–1996.” *Social Forces* 94:1593-621.
- Bonikowski, Bart and Oscar Stuhler. 2022. “Reclaiming the Past to Transcend the Present: Nostalgic Appeals in U.S. Presidential Elections.” *Sociological Forum*. Forthcoming.

- Bonikowski, Bart and Daniel Ziblatt. 2020. "Mainstream Conservative Parties Paved the Way for Far-Right Nationalism." *The Washington Post*, Monkey Cage Dec. 2.
- Bonilla-Silva, Eduardo. 2001. *White Supremacy and Racism in the Post-Civil Rights Era*. Boulder, CO: Lynne Rienner Publishers.
- Brown, Peter F, Vincent J Della Pietra, Peter V Desouza, Jennifer C Lai, and Robert L Mercer. 1992. "Class-Based N-Gram Models of Natural Language." *Computational linguistics* 18:467-80.
- Brown, Tom, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. "Language Models Are Few-Shot Learners." *Advances in Neural Information Processing Systems* 33:1877-901.
- Carter, Dan T. 1996. *From George Wallace to Newt Gingrich: Race in the Conservative Counterrevolution, 1963-1994*. Baton Rouge, LA: LSU Press.
- Covert, Bryce. 2016. "Donald Trump's Imaginary Inner Cities." *The Nation*, November 7. (<https://www.thenation.com/article/archive/donald-trumps-imaginary-inner-cities/>)
- Crabtree, Charles, Matt Golder, Thomas Gschwend, and Indridi H Indridason. 2020. "It Is Not Only What You Say, It Is Also How You Say It: The Strategic Use of Campaign Sentiment." *The Journal of Politics* 82:1044-60.
- Dai, Yaoyao and Alexander Kustov. 2022. "When Do Politicians Use Populist Rhetoric? Populism As a Campaign Gamble." *Political Communication* 39:383-404.
- De La Torre, Carlos. 2016. "Left-Wing Populism: Inclusion and Authoritarianism in Venezuela, Bolivia, and Ecuador." *Brown Journal of World Affairs* 23:61-76.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv:1810.04805*.
- DiMaggio, Paul. 2015. "Adapting Computational Text Analysis to Social Science (and Vice Versa)." *Big Data & Society* 2:1-5.
- DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting Affinities Between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding." *Poetics* 41:570-606.
- Topic Models and the Cultural Sciences.
- Dor, Liat Ein, Alon Halfon, Ariel Gera, Eyal Shnarch, Lena Dankin, Leshem Choshen, Marina Danilevsky, Ranit Aharonov, Yoav Katz, and Noam Slonim. 2020. "Active Learning for BERT: An Empirical Study." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 7949-7962.
- Dumais, Susan T. 2004. "Latent Semantic Analysis." *Annual Review of Information Science and Technology* 38:188-230.
- Ethayarajh, Kavin. 2019. "How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings." *arXiv 1909.00512*.

- Evans, James A and Pedro Aceves. 2016. "Machine Translation: Mining Text for Social Theory." *Annual Review of Sociology* 42:21-50.
- Fahey, James J. 2021. "Building Populist Discourse: An Analysis of Populist Communication in American Presidential Elections, 1896–2016." *Social Science Quarterly* 102:1268-88.
- Finkelstein, Federico. 2019. *From Fascism to Populism in History*. Berkeley, CA: University of California Press.
- Freeden, Michael. 1998. "Is Nationalism a Distinct Ideology?" *Political Studies* 46:748-65.
- Freeden, Michael. 2003. *Ideology: A Very Short Introduction*. New York: Oxford University Press.
- Fuhse, Jan, Oscar Stuhler, Jan Riebling, and John Levi Martin. 2020. "Relating Social and Symbolic Relations in Quantitative Text Analysis: A Study of Parliamentary Discourse in the Weimar Republic." *Poetics* 78:101363.
- Gest, Justin, Tyler Reny, and Jeremy Mayer. 2018. "Roots of the Radical Right: Nostalgic Deprivation in the United States and Britain." *Comparative Political Studies* 51:1694-719.
- Gidron, Noam and Peter A Hall. 2020. "Populism As a Problem of Social Integration." *Comparative Political Studies* 53:1027-59.
- Gilens, Martin. 2009. *Why Americans Hate Welfare: Race, Media, and the Politics of Antipoverty Policy*. Chicago, IL: University of Chicago Press.
- Golder, Matt. 2016. "Far Right Parties in Europe." *Annual Review of Political Science* 19:477-97.
- Goodwyn, Lawrence. 1978. *The Populist Moment: A Short History of the Agrarian Revolt in America*. New York: Oxford University Press.
- Hawkins, Kirk A. 2009. "Is Chávez Populist? Measuring Populist Discourse in Comparative Perspective." *Comparative Political Studies* 42:1040-67.
- Hawkins, Kirk A, Rosario Aguilar, Bruno Castanho Silva, ErinK. Jenne, Bojana Kocijan, and Cristóbal Rovira Kaltwasser. 2019. "Measuring Populist Discourse: The Global Populism Database." *Paper presented at the 2019 EPSA Annual Conference, Belfast, UK, June 20–22*.
- Hawkins, Kirk A, Ryan E Carlin, Levente Littvay, and Cristóbal Rovira Kaltwasser. 2018. *The Ideational Approach to Populism: Concept, Theory, and Analysis*. Philadelphia, PA: Routledge.
- Hetherington, Marc J and Jonathan D Weiler. 2009. *Authoritarianism and Polarization in American Politics*. New York: Cambridge University Press.
- Higham, John. 1955. *Strangers in the Land: Patterns of American Nativism, 1860-1925*. New Brunswick, NJ: Rutgers University Press.
- Hoffman, Mark Anthony, Jean-Philippe Cointet, Philipp Brandt, Newton Key, and Peter Bearman. 2018. "The (Protestant) Bible, the (Printed) Sermon, and the

- Word(s): The Semantic Structure of the Conformist and Dissenting Bible, 1660–1780.” *Poetics* 68:89-103.
- Huddy, Leonie, Alessandro Del Ponte, and Caitlin Davies. 2021. “Nationalism, Patriotism, and Support for the European Union.” *Political Psychology* 42:995-1017.
- Huddy, Leonie and Nadia Khatib. 2007. “American Patriotism, National Identity, and Political Involvement.” *American Journal of Political Science* 51:63-77.
- Ivarsflaten, Elisabeth. 2008. “What Unites Right-Wing Populists in Western Europe? Re-Examining Grievance Mobilization Models in Seven Successful Cases.” *Comparative Political Studies* 41:3-23.
- Jagers, Jan and Stefaan Walgrave. 2007. “Populism As Political Communication Style: An Empirical Study of Political Parties’ Discourse in Belgium.” *European Journal of Political Research* 46:319-45.
- Jenne, Erin K, Kirk A Hawkins, and Bruno Castanho Silva. 2021. “Mapping Populism and Nationalism in Leader Rhetoric Across North America and Europe.” *Studies in Comparative International Development* 56:170-96.
- Jensen, Jeffrey L, Daniel Karell, Cole Tanigawa-Lau, Nizar Habash, Mai Oudah, and Dhia Fairus Shofia Fani. 2021. Language Models in Sociological Research: An Application to Classifying Large Administrative Data and Measuring Religiosity. *Sociological Methodology* p. 00811750211053370.
- Johnson, Steven and Nikita Izhev. 2022. A.I. Is Mastering Language. Should We Trust What It Says? *The New York Times*.
- Jurafsky, Dan and James H. Martin 2020. *Speech and Language Processing*. Upper Saddle River, NJ: Pearson Education.
- Karakaya, Yagmur and Penny Edgell. Forthcoming. “From Politics As Vocation to Politics As Business: Populist Social Performance and Narrative Fusion in Trump Rallies.” *Social Forces*.
- Karell, Daniel and Michael Freedman. 2019. “Rhetorics of Radicalism.” *American Sociological Review* 84:726-53.
- Kazerouni, Abbas, Qi Zhao, Jing Xie, Sandeep Tata, and Marc Najork. 2020. Active Learning for Skewed Data Sets. *arXiv preprint 2005.11442*.
- Kazin, Michael. 1998. *The Populist Persuasion: An American History*. Ithaca, NY: Cornell University Press.
- King, Gary, Patrick Lam, and Margaret E Roberts. 2017. “Computer-Assisted Keyword and Document Set Discovery From Unstructured Text.” *American Journal of Political Science* 61:971-88.
- Kivisto, Peter. 2017. *The Trump Phenomenon: How the Politics of Populism Won in 2016*. Bingley, UK: Emerald.
- Kosterman, Rick and Seymour Feshbach. 1989. “Toward a Measure of Patriotic and Nationalistic Attitudes.” *Political Psychology* 10:257-74.

- Kozlowski, Austin C., Matt Taddy, and James A. Evans 2019. "The Geometry of Culture: Analyzing the Meanings of Class Through Word Embeddings." *American Sociological Review* 84:905-49.
- Laclau, Ernesto. 2005. *On Populist Reason*. New York: Verso.
- Lamont, Michèle, Bo Yun Park, and Elena Ayala-Hurtado. 2017. "Trump's Electoral Speeches and His Appeal to the American White Working Class." *The British Journal of Sociology* 68:S156-80.
- Le, Quoc and Tomas Mikolov. 2014. "Distributed Representations of Sentences and Documents." In *Proceedings of the 31st International Conference on International Conference on Machine Learning*, volume 32, pp. 1188-1196.
- Lee, Monica and John Levi Martin. 2015. "Coding, Counting and Cultural Cartography." *American Journal of Cultural Sociology* 3:1-33.
- Levitsky, Steven and James Loxton. 2013. "Populism and Competitive Authoritarianism in the Andes." *Democratization* 20:107-36.
- Levitsky, Steven and Daniel Ziblatt. 2018. *How Democracies Die*. New York: Crown.
- Lieberman, Robert C. 2022. *Shifting the Color Line: Race and the American Welfare State*. Cambridge, MA: Harvard University Press.
- Light, Ryan and Jeanine Cunningham. 2016. "Oracles of Peace: Topic Modeling, Cultural Opportunity, and the Nobel Peace Prize, 1902-2012." *Mobilization: An International Quarterly* 21:43-64.
- Lipset, Seymour Martin. 1990. *Continental Divide: The Values and Institutions of the United States and Canada*. New York: Routledge.
- Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint arXiv:1907.11692*.
- López, Ian Haney. 2015. *Dog Whistle Politics: How Coded Racial Appeals Have Reinvented Racism and Wrecked the Middle Class*. New York: Oxford University Press.
- Manza, Jeff and Ned Crowley. 2018. "Ethnonationalism and the Rise of Donald Trump." *Contexts* 17:28-33.
- March, Luke. 2012. *Radical Left Parties in Europe*. New York: Routledge.
- March, Luke and Cas Mudde. 2005. "What's Left of the Radical Left? The European Radical Left After 1989: Decline and Mutation." *Comparative European Politics* 3:23-49.
- Marshall, Thomas H. 1950. *Citizenship and Social Class*. New York: Cambridge University Press.
- Maxwell, Angie and Todd Shields. 2019. *The Long Southern Strategy: How Chasing White Voters in the South Changed American Politics*. New York: Oxford University Press.
- McVeigh, Rory and Kevin Estep. 2019. *The Politics of Losing: Trump, the Klan, and the Mainstreaming of Resentment*. New York: Columbia University Press.

- Mendelberg, Tali. 1997. "Executing Hortons: Racial Crime in the 1988 Presidential Campaign." *The Public Opinion Quarterly* 61:134-57.
- Mendelberg, Tali. 2017. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton, NJ: Princeton University Press.
- Mercieca, Jennifer. 2020. *Demagogue for President: The Rhetorical Genius of Donald Trump*. College Station, TX: Texas A&M University Press.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
- Minaee, Shervin, Nal Kalchbrenner, Erik Cambria, Narjes Nikzad, Meysam Chenaghlu, and Jianfeng Gao. 2021. "Deep Learning-Based Text Classification: A Comprehensive Review." *ACM Computing Surveys* 54:1-40.
- Molina, Mario and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45:27-45.
- Mudde, Cas. 2004. "The Populist Zeitgeist." *Government and Opposition* 39:541-63.
- Mudde, Cas. 2007. *Populist Radical Right Parties in Europe*. Cambridge, UK: Cambridge University Press.
- Mudde, Cas. 2019. *The Far Right Today*. Medford, MA: John Wiley & Sons.
- Mudde, Cas. 2022. "The Far-Right Threat in the United States: A European Perspective." *The ANNALS of the American Academy of Political and Social Science* 699:101-15.
- Mudde, Cas and Cristóbal Rovira Kaltwasser. 2013. "Exclusionary Vs. Inclusionary Populism: Comparing Contemporary Europe and Latin America." *Government and Opposition* 48:147-74.
- Mudde, Cas and Cristóbal Rovira Kaltwasser. 2017. *Populism: A Very Short Introduction*. New York: Oxford University Press.
- Nelson, Laura K. 2021. Leveraging the Alignment Between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century US South. *Poetics* p. 101539.
- Nelson, Laura K, Derek Burk, Marcel Knudsen, and Leslie McCall. 2021. "The Future of Coding: A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods." *Sociological Methods & Research* 50:202-37.
- Omi, Michael and Howard Winant. 2014. *Racial Formation in the United States*. New York: Routledge.
- Pennington, Jeffrey, Richard Socher, and Christopher D Manning. 2014. "Glove: Global Vectors for Word Representation." In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532-1543.
- Peters, Matthew E, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. *arXiv 1802.05365*.

- Piliawsky, Monte. 1989. "Racial Politics in the 1988 Presidential Election." *The Black Scholar* 20:30-7.
- Pirro, Andrea L. P. Forthcoming. "Far Right: The Significance of An Umbrella Concept." *Nations and Nationalism*.
- Ren, Cheng and Irene Bloemraad. 2022. "New Methods and the Study of Vulnerable Groups: Using Machine Learning to Identify Immigrant-Oriented Nonprofit Organizations." *Socius* 8:23780231221076992.
- Reny, Tyler T, Loren Collingwood, and Ali A Valenzuela. 2019. "Vote Switching in the 2016 Election: How Racial and Immigration Attitudes, Not Economics, Explain Shifts in White Voting." *Public Opinion Quarterly* 83:91-113.
- Roberts, Kenneth M. 1995. "Neoliberalism and the Transformation of Populism in Latin America: The Peruvian Case." *World Politics* 48:82-116.
- Rodriguez, Pedro and Arthur Spirling. 2022. "Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research." *The Journal of Politics* 84:101-115.
- Rooduijn, Matthijs. 2014. "The Nucleus of Populism: In Search of the Lowest Common Denominator." *Government and Opposition* 49:573-99.
- Rooduijn, Matthijs. 2019. "State of the Field: How to Study Populism and Adjacent Topics? A Plea for Both More and Less Focus." *European Journal of Political Research* 58:362-72.
- Rooduijn, Matthijs, Sarah L de Lange, and Wouter van der Brug. 2014. "A Populist Zeitgeist? Programmatic Contagion by Populist Parties in Western Europe." *Party Politics* 20:563-75.
- Rooduijn, Matthijs and Teun Pauwels. 2011. "Measuring Populism: Comparing Two Methods of Content Analysis." *West European Politics* 34:1272-83.
- Rule, Alix, Jean-Philippe Cointet, and Peter S Bearman. 2015. "Lexical Shifts, Substantive Changes, and Continuity in State of the Union Discourse, 1790–2014." *Proceedings of the National Academy of Sciences* 112:10837-44.
- Rydgren, Jens. 2017. "Radical Right-Wing Parties in Europe: What's Populism Got to Do with It." *Journal of Language and Politics* 16:485-96.
- Saito, Takaya and Marc Rehmsmeier. 2015. "The Precision-Recall Plot Is More Informative Than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets." *PLoS one* 10:e0118432.
- Saul, Heather. 2015. "Donald Trump Declares 'The Hispanics Love Me – They Were Chanting for Me' After Being Met by Protesters at US-Mexico Border." *The Independent July*:13–58.
- Scheppele, Kim Lane. 2018. "Autocratic Legalism." *University of Chicago Law Review* 85:545.
- Settles, Burr. 2009. Active Learning Literature Survey. *Computer Sciences Technical Report, University of Wisconsin–Madison* 1648.

- Sides, John and Daniel J Hopkins. 2015. *Political Polarization in American Politics*. New York: Bloomsbury Publishing.
- Sides, John, Michael Tesler, and Lynn Vavreck. 2018. *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America*. Princeton, NJ: Princeton University Press.
- Smith, Rogers M. 1997. *Civic Ideals: Conflicting Visions of Citizenship in US History*. New Haven, CT: Yale University Press.
- Spitzer, Scott J. 2012. "Nixon's New Deal: Welfare Reform for the Silent Majority." *Presidential Studies Quarterly* 42:455-81.
- Stoltz, Dustin S. and Marshall A. Taylor 2019. "Concept Mover's Distance: Measuring Concept Engagement Via Word Embeddings in Texts." *Journal of Computational Social Science* 2:293-313.
- Stoltz, Dustin S and Marshall A Taylor. 2021. "Cultural Cartography with Word Embeddings." *Poetics* 88:101567.
- Sun, Chi, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. How to Fine-Tune BERT for Text Classification?. In *China National Conference on Chinese Computational Linguistics*, pp. 194–206. Springer.
- Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney. 2012. LSTM Neural Networks for Language Modeling. In *Thirteenth Annual Conference of the International Speech Communication Association*.
- Taggart, Paul. 2000. *Populism: Concepts in the Social Sciences*. Philadelphia, PA: Open Press.
- Tavory, Iddo and Stefan Timmermans. 2014. *Abductive Analysis: Theorizing Qualitative Research*. Chicago, IL: University of Chicago Press.
- Terechshenko, Zhanna, Fridolin Linder, Vishakh Padmakumar, Michael Liu, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau. 2020. A Comparison of Methods in Political Science Text Classification: Transfer Learning Language Models for Politics. *SSRN*.
- UCSB. 2021. *The American Presidency Project*. Santa Barbara, CA: UC Santa Barbara.
- Vahudova, Milada A. 2020. "Ethnopolitism and Democratic Backsliding in Central Europe." *East European Politics* 36:318-40.
- Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, AidanN Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *arXiv 1706.03762*.
- Wasow, Omar. 2020. "Agenda Seeding: How 1960s Black Protests Moved Elites, Public Opinion and Voting." *American Political Science Review* 114:638-59.
- Wetts, Rachel and Robb Willer. 2019. "Who is Called by the Dog Whistle? Experimental Evidence That Racial Resentment and Political Ideology Condition Responses to Racially Encoded Messages." *Socius* 5:1-20.

- Weyland, Kurt. 2001. "Clarifying a Contested Concept: Populism in the Study of Latin American Politics." *Comparative Politics* 34:1-22.
- Weyland, Kurt. 2013. "Latin America's Authoritarian Drift: The Threat From the Populist Left." *Journal of Democracy* 24:18-32.
- Whitehead, Andrew L. and Samuel L. Perry. 2020. *Taking America Back for God: Christian Nationalism in the United States*. New York: Oxford University Press.
- Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. "Transformers: State-of-the-Art Natural Language Processing." In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38-45, Online. Association for Computational Linguistics.
- Youngman, Sam. 2012. "Romney Birth Certificate Remark Rekindles Obama Controversy." *Reuters*, August 24 (<https://www.reuters.com/article/usa-campaign-romney/romney-birth-certificate-remark-rekindles-obama-controversy-idINDEE87N0F020120825>).
- Zhang, Susan, Stephen Roller, Naman Goyal, Mikel Artetxe, Moya Chen, Shuohui Chen, Christopher Dewan, Mona Diab, Xian Li, Xi Victoria Lin, et al. 2022. OPT: Open Pre-trained Transformer Language Models. *arXiv 2205.01068*.
- Zhou, Yichu and Vivek Srikumar. 2021. A Closer Look at How Fine-Tuning Changes BERT. *arXiv 2106.14282*.

Author Biographies

Bart Bonikowski is Associate Professor of Sociology and Politics at New York University. Relying on survey methods, computational text analysis, and experimental research, his work applies insights from cultural sociology to the study of politics in the United States and Europe, with a particular focus on nationalism, populism, and the rise of radical-right parties.

Yuchen Luo is a PhD student at New York University's Department of Sociology. She studies political culture in the United States and China with computational, survey-based, and experimental methods.

Oscar Stuhler is a PhD candidate at New York University's Department of Sociology. His research focuses on the development of methods and concepts for the formal analysis of textual representations of social structures.

Appendices

A Model Fit Comparison Based on ROC-AUC

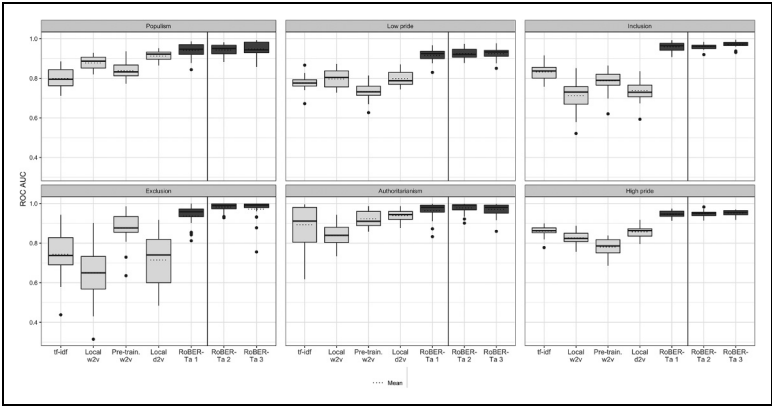


Figure A1. ROC-AUC performance of RoBERTa classifier compared to random forest models.

B Frame Prevalence Binarized at the Speech Level

To supplement our study’s main paragraph-level analyses, Figure B1 shows the campaign prevalence of the six frames binarized at the speech level (i.e., a speech with one or more paragraphs containing a given frame is classified as a positive case). We then compute the share of speeches within each campaign that contain each frame. Besides providing an alternative way of structuring the time series, the speech-level analysis facilitates comparison with Bonikowski and Gidron (2016: p. 1605)’s dictionary-based analysis of populism in the 1952–1996 subset of the same corpus. The overall patterns in the populism time series are strikingly similar.

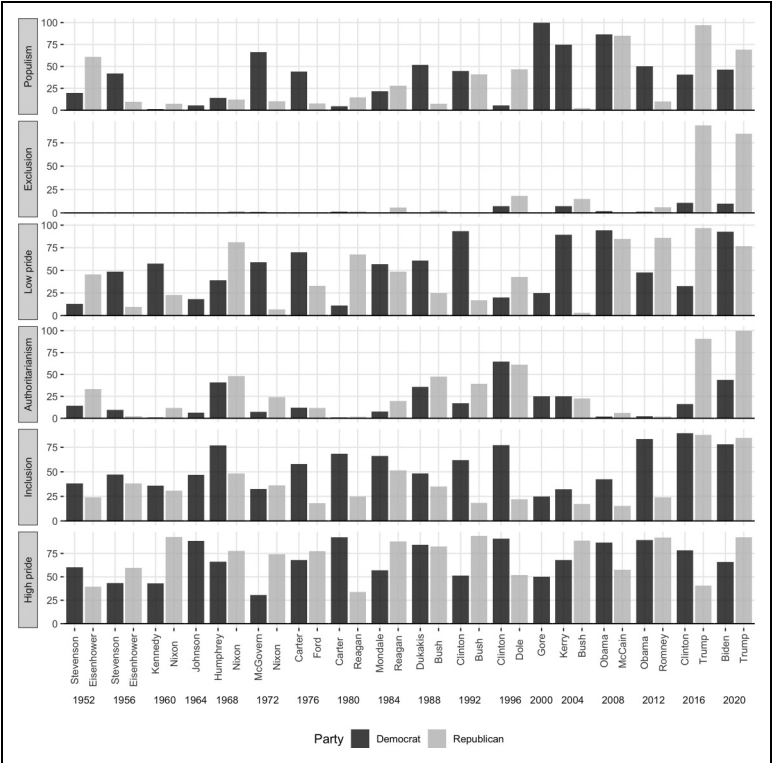


Figure B1. Proportion of speeches containing one or more populist, nationalist, and authoritarian paragraphs by campaign, 1952–2020.

C Principal Component Analysis Model Fit

When generating the principal component analysis visualized in Figure 5, we applied a varimax rotation to the first two components for ease of interpretation. The rotated space differs only minimally from the unrotated space, but leads to a minor redistribution of the variance share accounted for by each of the first two components (less than 1 percentage point); the cumulatively explained variance of the two components remains constant. The variance accounted for by each dimension is shown in Figure C1. Dimensions 2 and 3 account for roughly the same share of variance. Figure C2 replicates the PCA plot contained in the main manuscript but includes dimension 3 instead of dimension 2.

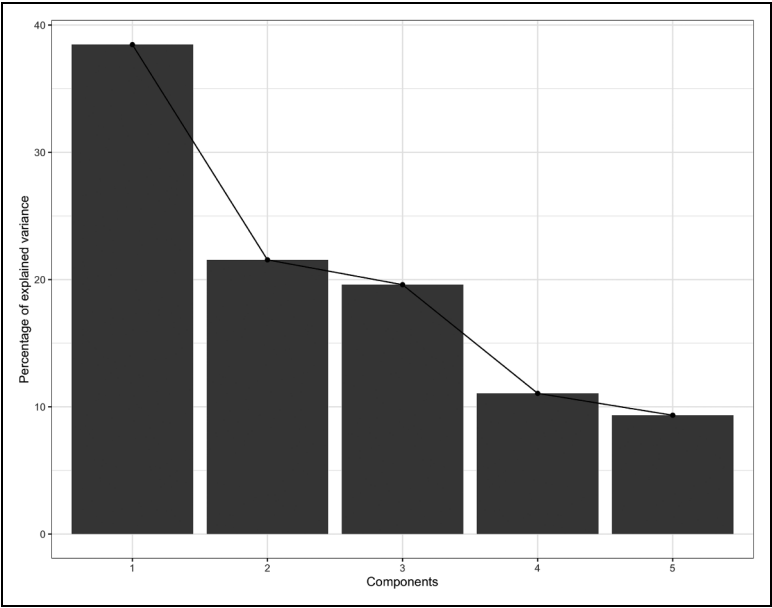


Figure C1. Principal component analysis scree plot.

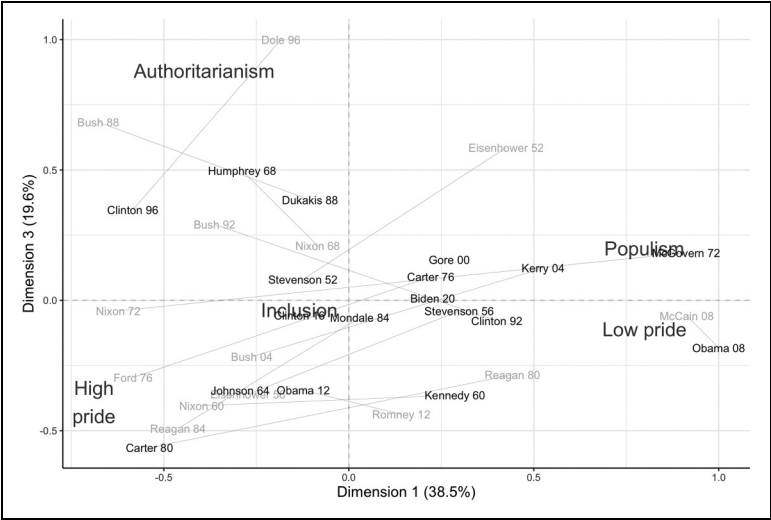


Figure C2. Principal component analysis of discursive frames and campaigns, dimensions 1 and 3.
Note: The figure shows dimensions 1 and 3 of a principal component analysis of campaigns and campaign-level standardized frame prevalence. Axes represent loadings for frames and scores for campaigns. The latter were standardized by the maximum absolute loading across campaigns to fit into the same figure. Edges connect campaigns from the same election. The 2016 Trump campaign and the exclusive nationalism frame were omitted from the analysis.

D Descriptive Statistics for Speech Attributes

Table D1. Descriptive statistics for discursive frames (by paragraph per campaign) and speech attributes (by speech).

	Mean	St. Dev.	Min	Median	Max
<i>DVs: Paragraphs per speech</i>					
Populism	0.53	1.29	0	0	18
High pride	2.14	2.49	0	1	17
Low pride	0.93	1.54	0	0	13

(continued)

Table D1. Continued

	Mean	St. Dev.	Min	Median	Max
Inclusion	1.20	2.41	0	0	34
Authoritarianism	0.45	1.59	0	0	29
<i>IVs: Speech attributes</i>					
Incumbent party	0.54	0.50	0	1	1
Prior president	0.34	0.47	0	0	1
Weeks to election	4.14	2.82	0.14	3.71	16.57
Length (in paragraphhs)	23.69	15.82	1	21	86

E Precision and Recall Across Corpus Subsets

Beyond the aggregate measures shown in the main text, we test whether our classifiers’ performance is associated with any attributes of the corpus, which could lead to potential bias in our substantive conclusions. For instance, given that the RoBERTa model is pretrained on contemporary data, it could perform more poorly on older texts. If so, this could call into question the utility of transfer learning for historical analyses. It is also, possible, however that mismatch between the pretraining and target data is mitigated by the fine-tuning of the model to hand-labeled paragraphs.

In the analyses that follow, we consider all 25 test set predictions for each of our six frames. To test the contextual dependence of *recall*, we first look at the subset of all predictions for paragraphs where the true label was positive (N = 2,691)—that is, all predictions for cases that we manually labeled as containing any of our frames. We then construct a variable indicating whether our models predicted a true positive (1) or a false negative (0). For *precision*, we take the subset of all positive predictions (N = 2,740). This variable again captures whether the prediction was a true positive (1) or a false positive (0). Finally, we average the predictions for the same paragraph and frame, leading to 541 cases for analyzing recall and 738 for precision. We run OLS regressions for both of these dependent variables with standard errors clustered at the paragraph level. Note that the unit of analysis here is the averaged prediction by the RoBERTa models for a given frame occurring in a given paragraph, so some paragraphs feature in the models multiple times.

The regression coefficients in Table E1 suggest that both recall and precision are associated with the number of words in a paragraph (multiplied by 100 for ease of presentation). A one standard deviation increase in paragraph length leads to a 1.88 (recall) and 2.86 (precision) percentage point increase in the

Table E1. Regression of recall and precision on paragraph attributes.

	Recall	Precision
Decade	2.45* (1.11)	0.66 (0.98)
Exclusion (Reference: Authoritarianism)	5.63 (10.55)	13.10 (11.04)
High pride	1.37 (7.12)	−3.78 (6.37)
Inclusion	12.07 (7.50)	6.82 (7.01)
Low pride	−17.76* (7.78)	−4.93 (6.94)
Populism	−9.11 (8.80)	−1.03 (7.80)
Republican	0.61 (4.39)	−2.79 (3.91)
Paragraphs in campaign (standardized)	−1.52 (3.02)	−3.29 (2.43)
Paragraph length in words (standardized)	1.88 (2.04)	2.86 (1.86)
N (unique paragraphs)	485	624
N	541	738
R ²	0.06	0.02

+*p* < .1; **p* < .05; ***p* < .01; ****p* < .001.

probability of a correct prediction. This suggests that our classifiers perform better on longer paragraphs. Neither of these associations is significant, however.

Because we annotated a random sample of paragraphs to fine-tune the RoBERTa models, some campaigns contributed more cases to our training data than others. Therefore, we also test whether having fewer paragraphs from a given campaign negatively impacts our classifiers’ ability to make correct predictions. Surprisingly, both coefficients are negative, the one in the precision model significantly so. This is most likely an artifact, as there is no reason to assume that our classifier will perform better on paragraphs from campaigns that feature less prominently in our training data. In any case, however, this reassures us that our decision to annotate a random sample rather than one stratified by campaign has no major implications for our substantive conclusions.

With regard to party, we find no significant differences in classification performance. Although precision might be marginally (2.79 percentage points)

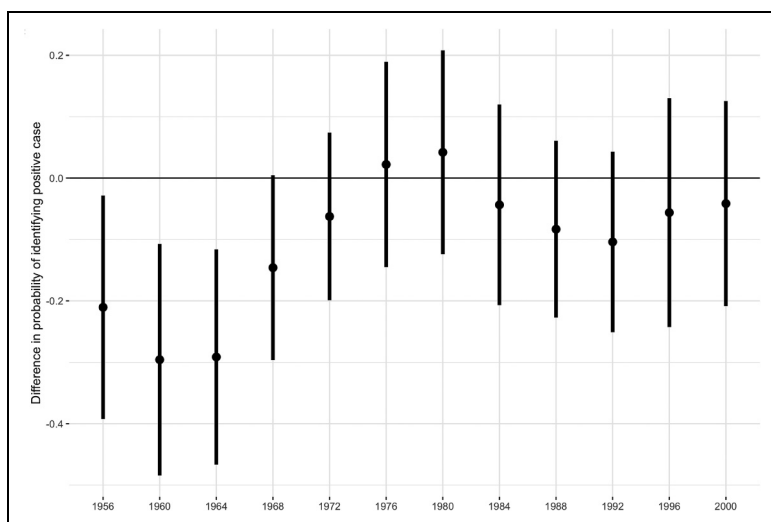


Figure E1. Recall over moving windows of three elections, aggregated across all six frames.

Note: Point estimates represent the predicted difference in recall between a three-election moving window and 2008-2016, along with 95 percent confidence intervals based on standard errors clustered at the paragraph level. For instance, the estimate for the point labeled “2000” indicates that recall was .04 lower in the years 1996-2004 than in 2008-2016 (a difference that is not statistically significant at $p < .05$). The models control for the other variables listed in Table E1.

lower for Republicans than Democrats, given the size of the differences observed in our substantive analyses, this is unlikely to be consequential.

Finally, we turn to the question of temporal bias—that is, whether a language model pretrained on contemporary data and fine-tuned to the target corpus performs more poorly on older historical texts. The results in Table E1 suggest that recall (though not precision) is indeed associated with time. Our classifiers’ ability to generate true positives rather than false negatives is estimated to increase by 2.46 percentage points every 10 years.

To determine whether this pattern is monotonic or whether it is produced by a few outlier campaigns, we run a series of linear models with recall as dependent variable. In each model, we estimate the difference in the probability of correctly identifying positive cases between a window of three elections and the years 2008-2016. The results, illustrated in Figure E1, demonstrate that the pretrained and fine-tuned models do not fare more poorly when

applied to historical data in general, but rather that their recall was lower specifically for campaigns from the late 1950s and early 1960s. Consequently, we should be cautious in interpreting the substantive findings for that time period, as our classifier may underestimate the number of positive cases in this particular subset of the corpus.

The most likely explanation for this finding is that some candidates’ language is inherently difficult to classify. Although sample size limitations preclude robust inferences about the model’s ability to classify paragraphs for each campaign and discursive frame, additional analyses provide suggestive evidence that Humphrey’s 1956, Kennedy’s 1960, and Nixon’s 1968 campaigns were especially challenging for our models—and based on our hand-coding experience, this was the case for human annotators as well.

F Alternative Model Specifications

Figure F1 and Table F1 replicate, respectively, the PCA and correlation analyses and speech-level regression models using the full corpus that includes the 2016 and 2020 Trump campaigns. They should be compared with Figure 5 and Table 4 in the main text.

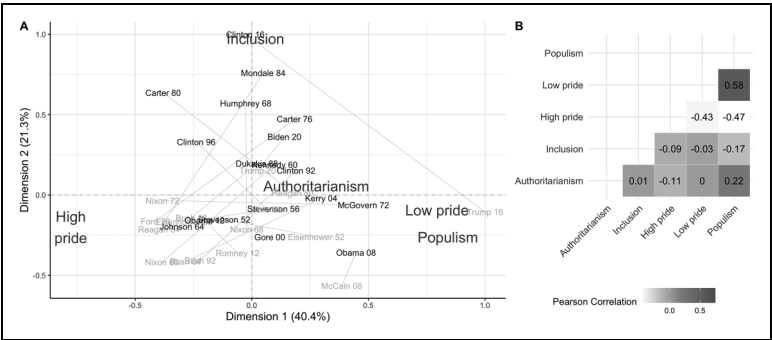


Figure F1. Principal component and correlation analysis of discursive frames and campaigns.

Note: This figure replicates 5 but includes Trump’s 2016 and 2020 campaigns. Panel A shows the first two dimensions of a principal component analysis of campaigns and frame prevalence. The frames were normalized to ensure that each carries equal weight in the analysis. Axes represent loadings for frames and scores for campaigns. The latter were standardized by the maximum absolute loading across campaigns to fit into the same figure. Edges connect campaigns from the same election. Panel B shows the correlations between the five remaining discursive frames.

Table F1. Negative binomial regressions of speech-level frame prevalence on speech attributes (including the 2016 and 2020 Trump campaigns).

	Populism	Authorit.	Inclusion	Low pride	High pride
Republican	0.33 (0.40)	0.53 (0.40)	−0.88*** (0.20)	−0.03 (0.24)	0.29 ⁺ (0.17)
Incumbent party	−1.43*** (0.39)	0.004 (0.39)	0.09 (0.19)	−1.74*** (0.26)	0.84*** (0.18)
Weeks to election (centered)	0.02 (0.03)	−0.03 (0.04)	0.03 (0.04)	−0.001 (0.01)	0.05* (0.02)
Incumbent X weeks	−0.05 (0.06)	0.08 (0.06)	0.04 (0.04)	−0.03 (0.03)	−0.05 ⁺ (0.03)
N of campaigns	34	34	34	34	34
N	2,956	2,956	2,956	2,956	2,956

⁺ $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$.