When Do Politicians Use Populist Rhetoric?

Populism as a Campaign Gamble[*]

Yaoyao Dai[†]   Alexander Kustov[‡]

Draft: December 21, 2021

Forthcoming in *Political Communication*

Abstract

Why do some politicians employ populist rhetoric more than others within the same elections, and why do the same politicians employ more of it in some elections? Building on a simple formal theoretical model of two-candidate elections informed by the ideational approach to populist communication, we argue that the initially less popular political actors are more likely to use populist rhetoric in a gamble to have at least some chance of winning. To test the empirical implications of our argument, we construct the most comprehensive corpus of U.S. presidential campaign speeches (1952-2016) and estimate the prevalence of populist rhetoric across these speeches with a novel automated text analysis method utilizing active learning and word embedding. Overall, we show the robustly greater use of populism among the presidential candidates with the lower polling numbers regardless of their partisanship or incumbency status.

Keywords: Populism, Elections, Campaign Rhetoric, Automated Text Analysis

Introduction

Over the past two decades, the world has been witnessing the emergence of a "populist zeitgeist" (Mudde, 2004) with populist parties and candidates gaining increasing electoral support across countries. To understand this change, scholars examined various economic and cultural factors that could have potentially increased the *demand* for populism among voters, such as the rising dissatisfaction with democracy in Latin America (e.g., Remmer, 2012; Singer, 2018), xenophobic reactions to migration crises in Europe and the U.S. (e.g., Rydgren, 2008; Inglehart and Norris, 2017), or popular backlash to globalization around the world (e.g., Swank and Betz, 2003; Schmuck and Matthes, 2017; Van Hauwaert and Van Kessel, 2018; Rooduijn and Burgoon, 2018). Regardless of their preferred explanation of this change, however, in studying the *supply* of populism many scholars view it as an opportunistic communication strategy employed by politicians to exploit those widespread grievances (e.g., Bonikowski and Gidron, 2016; Moffitt, 2016; Engesser et al., 2017; Heiss and Matthes, 2020).

While the increasing demand for populism can explain its electoral success in general, it is still unclear why some politicians are more likely to employ populist rhetoric than others, especially within the same electoral context when the popular demand is arguably fixed. After all, the upsurge of populist rhetoric across regions is evidently concentrated among particular political actors rather than being spread throughout the political system. For instance, one can easily observe the recent populist spike in the 2016 U.S. presidential election, led by Donald Trump and Bernie Sanders (Oliver and Rahn, 2016). However, many candidates also ran non-populist campaigns within the same election (Hawkins and Kaltwasser, 2018a). At the same time, the same politicians often vary significantly in their use of populism across various elections. Some candidates that relied heavily on populist rhetoric in one election, such as Eisenhower in 1952 and Nixon in 1968, used less populist rhetoric in their other campaigns (Bonikowski and Gidron, 2016). So, when do–and don't–politicians *strategically* decide to use populism?

To address this question, we provide a simple game-theoretic model of two-candidate elections and argue that the candidates' decision to use populist rhetoric may be fruitfully viewed as a campaign *gamble*. Since the effect of populism is polarizing, and its appeal may depends on various volatile structural factors, there is always some uncertainty whether it would be on net beneficial for candidates in a particular election. In line with this intuition, we then formally show that, despite its possible risks, the political actors with lower popular support are more likely to use populist rhetoric to have at least some chance of winning.

To test the empirical implications of our theoretical model, we construct the most comprehensive corpus of U.S. presidential campaign speeches from 1952 to 2016 and estimate the prevalence of populist rhetoric across these speeches. In doing so, we develop a novel automated text analysis method that utilizes active learning, an interactive and iterative supervised machine learning method, and Doc2vec, an advanced natural language processing model. Compared to the dictionary method commonly used for measuring populist rhetoric, our method is better at capturing complex language features and the underlying concept. In line with our theoretical expectations, we show the greater use of populist rhetoric among the presidential candidates with the initially lower polling numbers (of either party and regardless of their establishment status). These results are further robust to a number of alternative specifications.

Our contribution is thus two-fold. Theoretically, we bridge the previously disconnected ideational and game-theoretic approaches to the study of populism by providing a formal model of populist rhetoric as a risky campaign gamble, elucidating when political actors may strategically decide to be populist. Empirically, we provide the most comprehensive estimates of populism across U.S. presidential campaigns, corroborating the intuition of the unprecedented use of populist rhetoric by Donald Trump in the 2016 elections. In doing so, we introduce a new efficient measurement method to capture the necessary and sufficient conditions of populist rhetoric, which can be used across different elections and political contexts.

Populist Communication as Strategic Campaigning

Its popularity notwithstanding, populism has been a highly contested term. To improve its conceptual clarity and ensure generalizability across contexts, over the last decade a growing number of scholars from multiple disciplines have gradually adopted the "ideational" approach to populism. The ideational approach views populism as a set of ideas depicting society as divided into two homogeneous and antagonistic groups–the "good" people and the "corrupt" elites–and emphasizing that politics should reflect the general will of the people (Mudde, 2004; Mudde and Kaltwasser, 2018; Hawkins et al., 2018). To that end, this approach constructs a minimalist definition highlighting the shared core of populist ideas and separating the concept of populism from its possible causes and consequences. According to such ideational definition, populism is not a complete ideology–it has little policy content and thus also has to be combined with other major ideologies. Accordingly, it has been utilized by politicians and parties from across the political spectrum (Gidron and Bonikowski, 2013; Akkerman et al., 2013). Within the ideational approach, scholars have studied populism as either such thin-centered ideology (e.g., Mudde, 2004), a political discourse (Bonikowski and Gidron, 2016), a political style (Moffitt, 2016), or a set of public attitudes (Wuttke et al., 2020).

Building on this literature, in this paper, we study populism in political discourse as a claim-making or communication which can be used by diverse political actors to gain an electoral advantage during their campaigns. In doing so, as in the other research on populist communication, we do not treat populism as a fixed feature of politicians or parties (see Moffitt, 2016). Instead, the same political actors can be more or less populist across different campaigns (such as Eisenhower's significantly more populist 1956 campaign compared to his 1952 campaign) or, sometimes, even within the same campaign (Bonikowski and Gidron, 2016).

Although the demand for populist rhetoric certainly varies across space and time, many scholars who study the supply side of populism usually take it as given and

explore how this demand is then exploited by opportunistic politicians. In line with this reasoning, it has been increasingly shown that some of the major sources of populism, such as related to anti-immigration attitudes, are rather stable and robust to various economic and demographic shocks (Kustov et al., 2021; Dennison and Geddes, 2019).[1] Furthermore, not only these underlying attitudes but the individual propensity for populist voting itself may withstand major economic and political crises (Gidron and Mijs, 2019; Kustov, 2021). Although these (non)findings may seem counter-intuitive, they dovetail well with the evidence that populist attitudes are largely a function of stable personality traits such as openness to experience and agreeableness (Bakker et al., 2016), as well as authoritarianism and ethnocentrism (Assche et al., 2019).

But while the literature on the supply side of populist communication rightly assumes that opportunistic politicians can strategically exploit stable and widespread popular anti-immigration and anti-elitist attitudes, there is very little elaboration on why populist rhetoric is used by some actors more than others even within the same electoral context.[2] Unfortunately, scholars of populism rarely specify what games politicians are actually playing, what other campaign strategies are available, and why populist rhetoric may or may not constitute someone's best response in equilibrium. One of the biggest omissions in the literature is perhaps the relative neglect of the potential costs incurred by politicians who employ populist rhetoric.

What are the potential electoral benefits and costs of employing populist rhetoric? The often cited "benefit" of populism is the increased turnout of politically disaffected citizens who may find such rhetoric appealing and vote for populist parties (Huber and Ruth, 2017). Populist rhetoric is thus conventionally viewed as one of the top-

---

[1] For some evidence of the stability of anti-elitist attitudes, see Motta (2018).

[2] Although various candidate attributes might lead to different levels of credibility in making populist anti-elite appeals, most populist candidates who claim to truly represent the people are themselves elites (Mudde and Kaltwasser, 2018). In line with the idea that much of the identity of the people and the elite is constructed, Castanho Silva (2019) shows that when the populist candidates come to power, their supporters still view them as being a part of the people. In other words, the establishment status does not prevent candidates from making credible populist claims.

down strategies for voter mobilization (Weyland, 2001). These purely mobilizational effects of populism, however, have been recently questioned both theoretically and empirically (e.g., Ardag et al., 2019). Most prominently, populist rhetoric can also be deliberately used to *de*mobilize voters for mainstream parties by amplifying negativity in politics and potentially triggering popular distrust in democracy (see Ansolabehere and Iyengar, 1995). Consistent with this idea, a comprehensive empirical test of the populism effects on turnout finds that the emergence of successful populist parties may indeed demobilize a substantial share of new voters (Immerzeel and Pickup, 2015).

Immerzeel and Pickup (2015), however, also find that populism may mobilize those who vote for mainstream, non-populist parties. In other words, besides its potential electoral benefits (for a particular politician or party), populist rhetoric can also impose significant electoral costs by repelling those voters who reject the associated moralizing and anti-pluralist views (Hameleers et al., 2019). All in all, similar to findings on negative campaigning (Lau and Rovner, 2009; Krupnikov, 2011), populist rhetoric does not appear to be always effective at (de)mobilizing voters in an electorally advantageous way, especially when potential alternatives are considered (Bornschier, 2017). Accordingly, while many scholars try to explain the relative success of radical parties by emphasizing the electoral effectiveness of populism, this raises the follow-up question of why not all politicians would use populist rhetoric even more often.

Meanwhile, the existing game theoretic treatment of populism has so far been disconnected from these conceptual and empirical debates by treating supply-side populism as a "distorted" left-wing ideology (Acemoglu et al., 2013) or anti-institutionalism (Acemoglu et al., 2013). While insightful, these formal conceptualizations of populism do not speak to the latest theoretical understanding of populism as a strategic campaign rhetoric as defined by the ideational approach (Mudde and Kaltwasser, 2018).[3] These models are thus not well-suited for explaining why political actors of similar

---

[3]For a notable exception, see Serra (2018).

ideology may differ in their use of populist rhetoric while facing the same demand (or similarly "populist-friendly" electorate) within the same electoral context. In the same vein, the models that treat populism–similar to ideology–as a stable characteristic of political actors also fail to explain why the same politicians may differ in their use of populist rhetoric under different electoral circumstances.

A Précis of Populism as a Campaign Gamble Model

To bridge the divide between these two sets of literature, our theoretical model of populist rhetoric adopts the ideational conceptualization of populism while also building on the earlier game theoretic literature on political campaigning. We model populism as a particular type of communication that creates a moralized divide between *us* and *them*. In addition to claiming to represent *us*–the virtuous people–populist rhetoric portrays *them*–the political opponents and the (out)group associated with these opponents–as morally corrupt (Mudde and Kaltwasser, 2018; Hawkins et al., 2018). In line with the idea of a moralized combat, empirical literature demonstrates the pervasive negativity of populist campaigning (Nai, 2018). When it comes to the game theoretical literature, especially relevant for our purposes are thus the models of negative campaigning, which can help elucidate the strategic logic of gloom-ridden populist rhetoric as a campaign gamble within a certain electoral context (e.g., Skaperdas and Grofman, 1995).

Building on Skaperdas and Grofman (1995), our model of populism as a campaign gamble assumes a standard electoral race with the two as-if identical political candidates who decide whether to allocate their effort to "conventional" or "populist" campaigning in an attempt to improve their electoral chances. While the use of conventional campaigning is assumed to primarily mobilize additional support by focusing on one's own policy platform, populist campaigning is assumed to have a polarizing effect–appealing to some and appalling others.[4]

---

[4]This assumption is consistent with the recent empirical findings indicating that exposure to pop-

Specifically, populist rhetoric can demobilize the existing support of the opponent but can also backfire such that there is some chance that it would demobilize the candidate's own support or mobilize support for the opponent, depending on the underlying (dis)agreement with populism among the supporters for the candidate and opponent. While populist rhetoric may be appealing to more people under some structural background conditions (e.g., in a downturn economy with increasing unemployment), there is always some uncertainty whether it would be on net benefit for the candidate in a particular election. Consequently, we argue that the decision to use populism may be fruitfully viewed as a campaign gamble.

Given those minimal assumptions, we then formally prove the following proposition: *"the candidate with a lower pre-existing support is expected to use more populist campaign rhetoric relative to his opponent."* (for the detailed description of the model, its assumptions and results, see Appendix A). Intuitively, despite its possible risks, the candidates with the lower level of prior popular support are more likely to use populism to have at least some chance of winning. In turn, this can help explain why some politicians employ populist rhetoric more than others within the same elections, or why the same politicians decide to use it more in some particular elections.

Empirical Strategy and Design

There can be multiple ways to test the empirical implications of our model. While in principle we should be able to test our main proposition in any two-candidate elections, the well-documented universe of U.S. presidential campaigns with (normally) two major contestants provides a particularly great case for this purpose. So far, populism has been largely explored as a prominent feature of Latin American and European politics, but there has been growing attention to populism in the United States (Hawkins and

---

ulist messages enhances both prior agreement and disagreement with populism (Müller et al., 2017; Hameleers et al., 2018).

Littvay, 2019). This has been especially true since the recent rise of the Tea Party and, subsequently, Donald Trump as a part of the Republican party, as well as Bernie Sanders as a part of the Democratic party. While there are important sociological and institutional differences between the U.S. and other industrialized democracies (Taylor et al., 2014), few factors besides the number of parties should systematically impact our general theoretical argument regarding the greater strategic use of populist rhetoric among the initially loosing candidates.[5] In fact, when it comes to the demand side of populism, the U.S. patterns are arguably similar to other high-income countries, including the rising dissatisfaction with mainstream parties, immigration salience, and economic grievances (Inglehart and Norris, 2017). When it comes to the supply side, while the U.S. political system offers much less opportunity for organized populist parties compared to countries with proportional representation, it still provides ample opportunities for populist candidacies (Lee, 2019). In line with this, Bonikowski and Gidron (2016) confirm that populist rhetoric has historically been a rather common feature of U.S. presidential campaigning across all parties.

Data

Despite the growing interest in studying populism in U.S. presidential elections, existing research has either focused on the recent 2016 election (see Hawkins and Littvay, 2019; Lacatus, 2019) or more historical cases (Bonikowski and Gidron, 2016). We expand the scope of this research by building a comprehensive U.S. presidential campaign corpus of 4, 314 speeches from 1952-2016. The speeches are collected from two data sources: The Annenberg/Pew Archive of Presidential Campaign Discourse (Annenberg, 2000) and The American Presidency Project (Woolley and Peters, 2008) hosted at the University of California, Santa Barbara. The Annenberg/Pew Archive of Presidential Campaign Discourse includes transcripts of campaign speeches delivered by the Democratic and

---

[5]For the potential role of multi-party competition for our argument, see Discussion.

Republican presidential nominees between September 1st and the election day, as well as their nomination acceptance speeches. Overall, it covers 12 elections and 21 presidential campaigns from 1952 to 1996 with 2, 406 speeches, which have been previously used to examine populist rhetoric by Bonikowski and Gidron (2016).[6] The American Presidency Project is an online database of presidential documents hosted at the University of California, Santa Barbara. We use the American Presidency Project to expand on the Annenberg/Pew Archive data by adding five most recent elections from 2000 to 2016 and incorporating all speeches delivered during presidential campaigns. The average speech length is 2,167 words, and 90% of the speeches are between 500 words to 5000 words.[7]

We create a measure of populist rhetoric in these campaign speeches using our original machine learning method (see below). While our speech data and populism measurement include speeches and populism score by all candidates and span from the day of candidacy announcement to the election day, we only include the speeches delivered by the final candidates from the two parties from January to the election day to test our theoretical model, which results in 3, 436 speeches. We use monthly polling results for the candidates throughout their respective campaigns to measure their electoral advantage: whether and what extent each candidate was leading in the polls across campaign months (Gallup, 2013).[8] In addition to the level of populism rhetoric and electoral advantage, we also include several candidate and speech characteristics in our data, such as party membership, party incumbency, and speech length. For summary statistics of all major variables, see Table B1 in Appendix.

---

[6]For the distribution of speeches across presidential campaigns, see B1.

[7]It is important to note that, unlike the more recent American Presidency Project which covers all campaign speeches starting January 1 (for 1952-2016), The Annenberg/Pew Archive only spans the last three campaign months starting September 1 (for 1952-1996). While we decided to utilize all of the available information from the combined unbalanced panel dataset in our main analysis, restricting the whole sample to the last three months to harmonize the monthly coverage of the panel yields substantively similar results (see Table B3 in Appendix).

[8]The polling results include the earliest possible candidate polling for each month. For some campaign months with no available polling, we impute an estimate based on the closest available polls.

Measurement of Populist Rhetoric

Populist rhetoric in politician speeches has previously been measured using either human-coded content analysis (e.g., Jagers and Walgrave, 2007; Hawkins, 2009; Hawkins and Kaltwasser, 2018b) or dictionary-based automated text analysis (e.g., Rooduijn and Pauwels, 2011; Bonikowski and Gidron, 2016; Heiss and Matthes, 2020). Scholars have recently begun to evaluate supervised machine learning methods in measuring populism in texts (Hawkins and Silva, 2018; Dai, 2019). Although human-coded methods generally have high content validity, they are also costly and time-consuming. Given that we have over 4,000 lengthy speeches, we use automated text analysis to code populist rhetoric. Instead of using the common dictionary-based method, however, we use a novel approach that utilizes active learning, an interactive and iterative machine learning method, random forests, a supervised machine learning model, and Doc2vec, a word embedding model. As we demonstrate below, this new method is potentially better at capturing complex language features, as well as the necessary and sufficient conditions of the underlying theoretical concept.

Our measurement strategy follows the same definition as in our theoretical model, which defines populism as a particular style of communication. This style portrays society in moral terms as being divided into two homogeneous and antagonistic groups, the good people versus the corrupt elite, while emphasizing that politics should reflect the general will of the people (Mudde, 2004; Mudde and Kaltwasser, 2018). "The people" and "the elite" are both constructed, which allows for much flexibility in using populist rhetoric. Indeed, those politicians who use populist rhetoric essentially get to define who counts as the elite, so they are not necessarily against all conventional elites based on their socioeconomic or political status (in many cases, populist politicians are actually a part of such conventional elites). Although the specific identity of the people

and the elite varies across contexts, the elite is always portrayed as a morally corrupt actor who is ignoring or subverting the interests of the people by favoring some special interests domestically and/or abroad.

Populism is thus a multi-dimensional concept which follows a certain necessary and sufficient structure (Wuttke et al., 2020; Goertz, 2006). Following such structure, in our measurement a text is considered populist *if* and *only if* it (1) recognizes the people instead of the elite as the only legitimate source of power (*people-centric*); (2) creates separation between us and them (*anti-pluralist*); and, in doing so, (3) stipulates the separation of us and them on moral grounds (*good versus evil*) (Hawkins, 2009; Dai, 2019). None of the necessary components of populism alone can clearly distinguish it from other related concepts. For example, people-centrism is a shared feature between populism and liberal democracy. What separates populism from liberal democracy is moralized anti-pluralism (Müller, 2017). In other words, while the people are the only legitimate source of power in both populism and liberal democracy, their understanding of what constitutes the people differ. Liberal democracy takes a pluralist understanding of the people, recognizes the possible conflicts between different groups, and thus emphasizes compromises. In contrast, the people in populism are one romanticized homogeneous group and are victimized by the corrupt elite. Therefore, simply praising the people or criticizing the elite is not enough to be counted as being populist. A populist rhetoric also needs to be moralizing and anti-pluralist. Similarly, although necessary, anti-pluralism alone cannot distinguish populism from other anti-pluralist ideology such as elitism and nationalism (Müller, 2017; Mudde and Kaltwasser, 2017).

Given its complex structure, populist rhetoric can hardly be contained or measured at a level of single words or even sentences. In fact, recent studies on populist rhetoric on social media find that, while a collection of social media posts can contain all the components of populism, any single post usually contains only one or two of its core components (Engesser et al., 2017). At the same time, any sufficiently long text can

potentially have all of the components across at least some of its parts.

For the purposes of our dataset on presidential campaign rhetoric, we thus divide all speeches (with the average speech length of $2,167$ words) into sub-speeches of 10 paragraphs, resulting in $16,729$ sub-speeches.[9] Dividing the speech data into sub-speeches also improves the performance of our classification model, because there is less noise and irrelevant information on the sub-speech level relative to the much longer speech level. Therefore, we treat each sub-speech as a document for our supervised classification task. We classify sub-speech documents as populist (as opposed to non-populist) if and only they contain all of the necessary and sufficient conditions described above. After applying the supervised learning algorithm, each sub-speech document has a score of either 1 (populist) or 0 (non-populist). We then aggregate this sub-speech level classification into a speech level measurement of populism by simply calculating the proportion of populist sub-speeches in a given speech.

As in any supervised approach, the first step is to obtain or to produce labeled data for a classification algorithm to learn (D'Orazio et al., 2014). Given that populist rhetoric is not a common feature in most campaign speeches, any random sample of documents can only result in a handful of positive cases, which is inefficient and can lead to sub-optimal model performance. Therefore, we adopt active learning to assist the labeling process, which is an interactive and iterative supervised machine learning method that is able to query the most informative cases, such as the ones with most uncertainties, for the human coders to code and thus can achieve higher model performance with much fewer labeled instances. Active learning has shown success in many machine learning applications (see e.g., Tong and Koller, 2001; Ertekin et al., 2007; Qi et al., 2008; Settles et al., 2008; Settles, 2009), and political scientists

---

[9]To ensure that each document contains complete information, we divide the speeches at the paragraph level instead of an equal number of words and sentences. We choose 10 paragraphs so that we have enough content to identify all of the necessary and sufficient conditions of populism. This decision is informed by our hand-coding experiments when setting up the coding rule. However, we don't have empirical evaluation on different sampling windows. Future research might be needed to identify a range of acceptable window sizes.

have recently begun to realize the potential of active learning (Miller, Linder, and Mebane Jr, Miller et al.). The process is as follows: we initially randomly selected 73 out of $4,314$ speeches, which contains 407 sub-speeches, for the human coder to code. To account for the possible variation of populist language over time, the random sample is stratified so that we have at least two speeches from every decade in the sample. 48 out of 407 (11.79%) sub-speeches were categorized as populist. As a part of this sample, a smaller set of 69 sub-speeches were then coded by the second coder to evaluate inter-coder reliability.[10] 88% of the time, two coders agreed with each other: 12 sub-speeches were coded as populist and 49 sub-speeches were coded as non-populist by both coders. Accordingly, both hand-coding highly correlated with each other: the Pearson correlation coefficient between the two coders' hand-coding was 0.72.[11] The discrepancies between the two coders were resolved using majority rule with a third coder.

Second, we train a random forest classifier that learns the rules from the initial hand-coded sample in the first step to predict the (non-)populist document class as close as possible to the human coder. In vectorizing the documents and words, we use Doc2vec, a neural network-based word embedding model from Natural Language Processing (Le and Mikolov, 2014). Unlike the common "bag-of-words" approach, which represents documents using the simple counts of as-if independent words, Doc2vec learns to maintain the semantic and syntactic relationships by vectorizing words and documents in a dense vector space. As a result, words similar in their meaning and documents similar in their contexts are positioned close to each other in the vector space.[12] Furthermore, Doc2vec reduces the high dimensionality of the raw text data

---

[10]Because populist rhetoric is a rare event in our data, we up-sampled populist documents for the second coder to avoid having all non-populist documents. The second coder did not know the proportion of potentially populist documents.

[11]The *Krippendorff's alpha* is 0.68, which is above the minimum reliability recommended by Krippendorff (2004). For comparison, it is also in between the *Krippendorff's alpha* values reported by other hand-coding methods in measuring populism (Hawkins, 2009; Hawkins and Silva, 2018).

[12]For more details on the algorithm and how we trained our Doc2vec model, see Online Appendix C.

and significantly improves the model performance relative to the same classification algorithm that uses a document-term matrix constructed by the "bag-of-words" approach. After vectorizing all documents, we train a random forest classifier on the word/document vectors to separate populist and non-populist documents.[13] To evaluate the classifier and to avoid over-fitting, we use cross validation and only train the model on 80% of the hand-coded data (training set) and reserve the rest of the sample (test set) for testing.[14]

Third, we apply the trained classifier in the second step to the full corpus, which predicts the probability of each document being populist (or not). Fourth, we then apply a query function (an active learner) to obtain the most informative documents for labeling. There are many querying strategies. In our case, we use uncertainty sampling that queries those documents that the classifier is most uncertain about (Settles, 2009; Danka and Horvath, 2018).[15] Fifth, we code the queried documents and add the newly coded instances to the training set, re-train the classifier, and query new documents to code (repeat the second to fifth steps). We repeated the process 9 times and labeled an additional 180 documents (sub-speeches). As expected, the new documents queried by the active learner contain much more positive (populist) instances than a ransom sample: 74 out of the 180 (41.11%) documents are labeled as being populist.[16] In the end, we labeled 587 documents total and identified 122 (20.78%) populist documents.

In the end, our model achieves an 91% accuracy in the test set (i.e., making the same prediction as the human coder). Because populist rhetoric is a rare event in our data (only 11.79% of sub-speeches are coded as populist in the initial random sample),

---

[13]Given that the positive class only accounts for 20% of the training data, we used stratified sampling when training the random forest classifier.

[14]During the training phase, we also use probability calibration with 5-fold cross-validation, i.e., the training set is further split into five folds, and probabilities for each fold are then averaged for prediction.

[15]Specifically, we use classification entropy as the uncertainty measure, a built-in uncertainty measure in modAL a Python3 library for active learning (Danka and Horvath, 2018).

[16]For the model performance before and after the active learning, see Appendix Figure B2. The precision recall AUC of the model before the active learning is 0.67. The PR AUC is increased to 0.74 after the active learning.

we provide two additional performance metrics to evaluate the out-of-sample model performance in Figure 1. We first report the receiver operating characteristic (ROC) curve and examine the corresponding area under the curve (AUC) in Figure 1a. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) in out-of-sample prediction for all thresholds. As can be seen from the plot, our model achieves a high AUC of 0.93, comparing to a 0.50 AUC for an unskilled or random classification model, while suffering only a minor false positive rate to obtain a high true positive rate in out-of-sample prediction. For example, to correctly identify 80% of populist documents, our model makes only about 10% of false positive predictions. In other words, less than 10% of the documents that are predicted as populist are not populist.

Since we have a highly-skewed dataset with 80% cases being the negative class (non-populist), we also report the precision-recall curve (PRC) in Figure 1b, which is more sensitive when evaluating binary classifier for the minority class (Saito and Rehmsmeier, 2015). A precision-recall curve plots precision against recall for all thresholds. Precision is the fraction of true positive cases among true positive and false positive cases, while recall is the fraction of the true positive cases among the true positive and false negative cases. As can be seen from the plot, our model achieves a high Precision Recall AUC of 0.74, while an unskilled or random classifier only has a Precision Recall AUC of 0.14.[17] The PRC curve of our model is also consistently above the baseline PRC.

When applying the classifier to the full corpus, a total of 427 sub-speeches (2.6%) are predicted as being populist. To better illustrate the model performance or what a populist sub-speech may look like, below, we discuss two randomly drawn sub-speeches from either party, which were classified as populist by the algorithm (outside of the

---

[17]Ulinkskaitė and Pukelis (2021) adopt a similar model pipeline and evaluate several classification models on classifying populist manifestos. Our model performance is similar to their best model, which is an ensemble of two classifiers, with a precision of 0.73.

Figure 1: Out-of-sample Performance of Classification Model

(a) ROC Curve   (b) PRC Curve
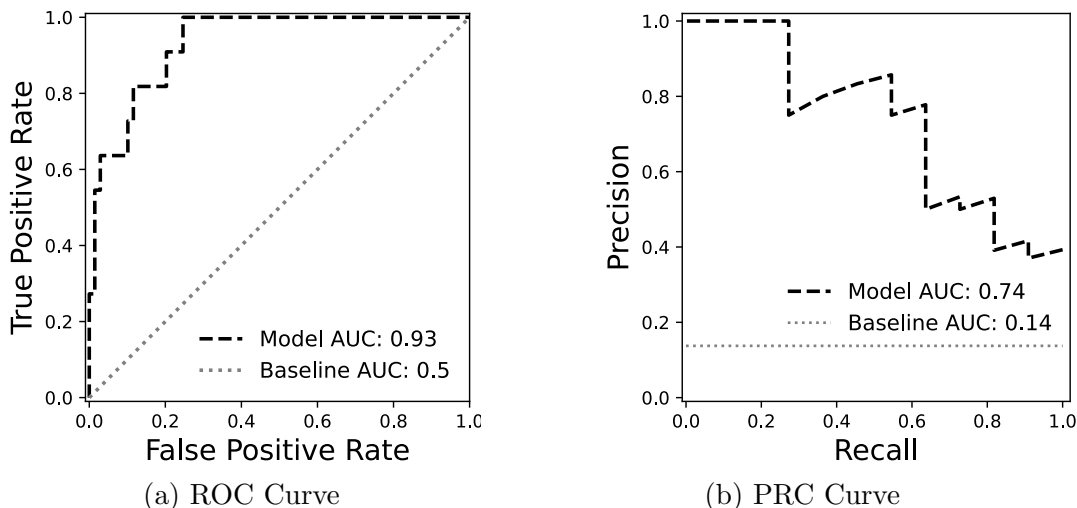
Note: The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) in out-of-sample prediction for all thresholds. The TPR captures the proportion of populist documents correctly classified as populist, while the FPR captures the proportion of out-of-sample documents that were incorrectly classified as populist. A ROC curve for a classifier that perfectly predicts class membership (populist and non-populist) would form a 90-degree angle in the upper left corner of the plot–the TPR would be 1, and the FPR would be 0. A classifier with no predictive power, one that performs randomly, would have a ROC curve that follows the diagonal dashed gray line in Figure 1a. The PRC curve plots the precision against the recall. Precision is the fraction of true positive cases among true positive and false positive cases. Recall is the fraction of true positive cases among true positive and false negative cases. A PRC curve for a classifier that perfectly predicts class membership (populist and non-populist) would form a 90-degree angle in the upper right corner of the plot. The baseline AUC for an unskilled or random classifier is the proportion of the minority (populist) class among all cases in the test set, which is 0.14 in our case.

hand-coded sample). Because the sub-speeches are still quite long, we only include the highlighted parts of these sub-speeches for demonstration. In the first example of Barack Obama's campaign speech in 2008, he creates a separation between Main Street (us) and Wall Street (them). While Main Street is innocent, Wall Street is greedy, irresponsible, and corrupted. Furthermore, Wall Street is the reason behind the economic suffering of Main Street. To that end, Obama also claims that, while he represents the millions of innocent people, the political establishment in Washington represents special interests.

> ***Barack Obama 2008***: Third, I said that we cannot and will not simply bailout Wall Street without helping the millions of innocent homeowners who are struggling to stay in their homes. ... I said that I would not

17

allow this plan to become a welfare program for the Wall Street executives whose greed and irresponsibility got us into this mess. ... We don't just need a plan for bankers and investors, we need a plan for autoworkers and teachers and small business owners. ... That means taking on the lobbyists and special interests in Washington. That means taking on the greed and corruption on Wall Street ... It is time to reform Washington. *(Remarks in Detroit, Michigan. September 28th, 2008.)*

Similarly, in Donald Trump's 2016 campaign speech, he creates a moralized separation between us (the American people) and them, while claiming to represent all Americans. In his narrative, the American people have been failed by the corrupt status quo. While the exact identity of the corrupt few here is somewhat ambiguous, it includes all the people who disagree with his campaign and policies because of their vested interests.

Donald Trump 2016: Change is coming. All the people who've rigged the system for their own personal benefit are trying to stop our change campaign because they know that their gravy train has reached its last stop. It's your turn now. This is your time ... We are fighting for all Americans ... who've been failed by this corrupt system. We're fighting for everyone who doesn't have a voice. Hillary Clinton is the candidate of the past. Ours is the campaign of the future. In this future, we are going to pursue new trade policies that put American workers first – and that keep jobs in our country ... The era of economic surrender is over. *(Remarks at a Rally at the Pensacola Bay Center in Pensacola, Florida. September 9th, 2016.)*

So far, our measurement of populism has been at the sub-speech level: a sub-speech is either populist with a score of 1 or not populist with a score of 0. To measure populist rhetoric at the speech level, we then calculate the proportion of populist sub-speeches in a given speech. Similarly, we calculate the proportion of populist sub-speeches among candidate speeches to create the level of populist rhetoric at the candidate level.

Compared to the dictionary-based approach commonly used to measure populist rhetoric (e.g., Bonikowski and Gidron, 2016; Rooduijn and Pauwels, 2011), our measurement has several advantages. The dictionary-based method measures populist

rhetoric by counting the words associated with populism in each document using a pre-defined dictionary. First, this essentially word-level measurement is based on the "bag-of-words" assumption and cannot take the words' meaning and contexts into account. A populist speech with similar but not identical phrases as in the populism dictionary would be categorized as non-populist. In fact, our previous example from Trump's 2016 speech would not be considered populist by the dictionary approach such as in Bonikowski and Gidron (2016), because it does not contain any words from their populism dictionary. It can be easily seen, however, that many phrases in that Trump's speech are synonymous with the words in their dictionary. Instead of "special interest", Trump uses "personal benefit"; and instead of "forgotten Americans", Trump claims to fight for "everyone who doesn't have a voice." With the Doc2vec component, our method is able to learn the semantic relationships between such words and phrases. Therefore, our model is still able to make an accurate prediction when a document uses similar but not identical words as in the populist documents used to train the model.

Second, the dictionary-based method assumes that all dictionary words have equal importance in measuring the level of populist rhetoric, which cannot capture the idea of necessary and sufficient conditions in defining populism (Wuttke et al., 2020). As a result, a non-populist speech with many references to "the people" without criticizing the corrupt few would have a high populist score. In contrast, our decision tree based random forest algorithm can learn complex non-linear relationships between the document features and its class, which resembles the necessary and sufficient conditions more closely.[18]

---

[18]For example, when a document mentions "American people" a lot, the algorithm will not immediately give it a high probability of being populist. Instead, it will trigger the next decision rule, such as whether the document also uses phrases like "special interests" and "betrayed."

Analysis and Results

We start by presenting our measure descriptively and verifying some of the previous stylized findings on populist rhetoric in U.S. presidential elections (Bonikowski and Gidron, 2016; Hawkins and Kaltwasser, 2018b; Lacatus, 2019) using our expanded data and new method. In Figure 2, we show the more detailed estimates of the use of populist rhetoric by candidate, campaign, and party. Similar to Bonikowski and Gidron (2016), we find that populist rhetoric is not a stable commitment by a particular candidate or party (for a comparison of our measures, see Figure B3). For those candidates who run in multiple elections, for instance, the prevalence of populist rhetoric often varies across campaigns. It is important to note the extremely high usage of populist rhetoric in 2016 elections by the Republican candidate Donald Trump. While this is in line with the intuition of many political observers, our research provides the first comparative quantitative assessment of his outlier populist status in U.S. general elections. Of course, our theoretical model could not predict the sheer extent of Trump's populism, but it is in line with our general expectations based on his lower support initially and throughout the campaign.

Populist Rhetoric as a Function of Electoral Advantage

The main proposition derived from our model predicts that candidates are more likely to use populist rhetoric when they are confronted with lower pre-existing support (or electoral disadvantage) relative to their opponent. While most of the variation in electoral advantage is across candidate-years, there have been significant monthly fluctuations in support within each election. Therefore, as an initial test of our theory, we visualize the average share of populist rhetoric in speeches depending on whether the speech was given by a candidate who was leading in the most recent poll (relative to that speech) in Figure 3. As can be seen, campaign speeches are indeed almost twice

20

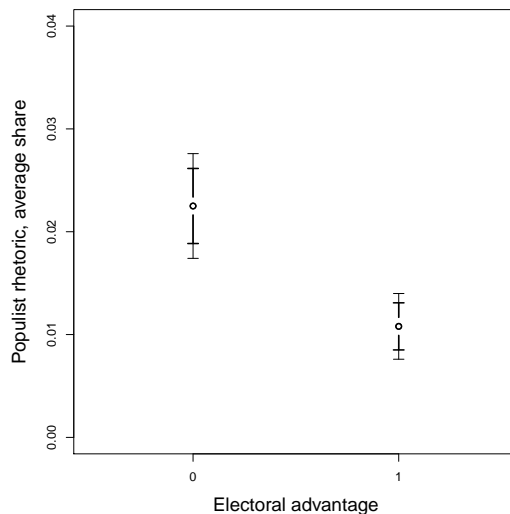Figure 2: Average Share of Populist Rhetoric across Campaigns
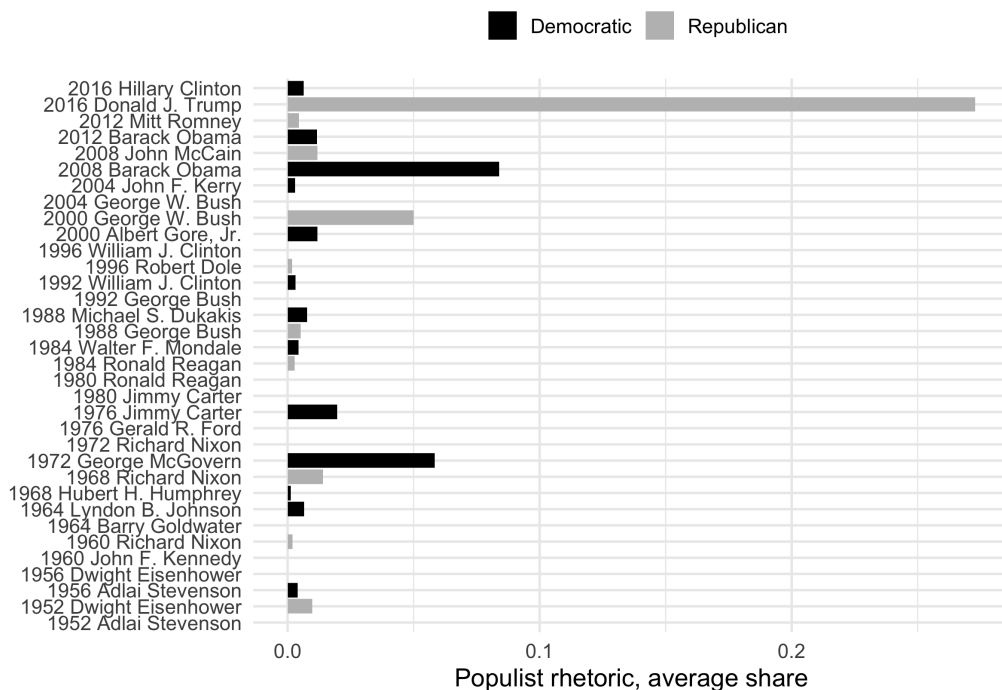
Note: The plot indicates the average share of populist sub-speeches across U.S. presidential campaigns based on our measure.

more likely to contain populist rhetoric under an electoral disadvantage.

To further test our theory, below we estimate and report the results from OLS and nonlinear regressions with different specifications. All models control for variables that are likely to influence the use of populist rhetoric based on previous studies. Most important, we include a binary indicator of incumbency status in all models to account for the fact that the incumbents tend to use populist rhetoric less (see Taggart, 2000; Bonikowski and Gidron, 2016; Herkman, 2017). In addition, we also control for partisanship. Since longer speeches are more likely to contain populist (or any other) rhetoric, we also account for speech length (standardized to be between 0 and 1) in most specifications. Given that some campaign cycles (and their parts) appear to be much more populist than others as seen in the descriptive analysis, we also include month and year fixed effects in some models.

Figure 3: Electoral Advantage and Populist Rhetoric in U.S. Presidential Speeches

Note: The figure indicates the average share of populist rhetoric in speeches depending on whether a certain presidential candidate is experience an electoral advantage in the most recent polls. The bars indicate 95% (84%) confidence intervals.

First, we fit a series of linear models with basic candidate controls and fixed effects reported in Table 1. Overall, these baseline OLS results provide strong support for our theory: the speeches of winning candidates generally contain less populist rhetoric, which also translates into a substantial difference of nearly two standard deviations.

Next, since presidential speeches are nested within months and campaigns (or candidate and years), we also consider multilevel or mixed models (see Table 2). Compared to OLS, these models arguably allow more flexibility in handling the variation in populist rhetoric due to clustering. As a first step, we fit a baseline multilevel model with random intercepts for months, years, and campaigns to check how much the use of populist rhetoric in presidential speeches varies across different levels. As we can see from the ICC estimates in Table 2 (Model 1), populism varies almost as much between as within campaigns (with candidates and years accounting for 37% and 9% of all variation respectively). We then investigate the extent to which this variation can be accounted for by candidates' electoral advantage along with other characteris-

22

Table 1: Populist Rhetoric as a Function of Electoral Advantage (Speech Level)

	(1)	(2)	(3)
Electoral Advantage	−0.012***	−0.012***	−0.012***
	(0.003)	(0.003)	(0.003)
Party Incumbency	−0.023***	−0.025***	−0.030***
	(0.003)	(0.003)	(0.003)
Partisanship (GOP)	0.003	0.003	0.001
	(0.003)	(0.003)	(0.003)
Speech Length	*No*	*Yes*	*Yes*
Month FE	*No*	*Yes*	*Yes*
Year FE	*No*	*No*	*Yes*
Observations	3,436	3,436	3,436
Adjusted R^2	0.021	0.023	0.205

Note: All models are OLS regressions of the average share of populist rhetoric in U.S. presidential speeches on electoral advantage in the recent monthly polls and other candidate or speech characteristics. Robust standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.

tics. Overall, similar to as before, the results from multilevel modeling indicate that speeches given by losing candidates contain more populist rhetoric. Importantly, we also see that the electoral advantage coefficient does not change much regardless of a particular model specification (also see Appendix).[19]

Robustness Checks

Our findings are robust to a number of additional tests with no change in their substantive interpretation. First, we replicate our analysis on the original sub-speech level with 11,839 valid observations used for coding the average share of populist rhetoric across speeches (see Table B2, 1). Second, we check whether the electoral advantage can explain the variation of populist rhetoric *within* campaigns after including candi-

---

[19]While neither our theory nor our empirical tests are aimed to compare the relative strength of various predictors, our results are in line with the previous research showing the relevance of incumbency and the irrelevance of partisanship for predicting populist rhetoric (Bonikowski and Gidron, 2016).

Table 2: Populist Rhetoric as a Function of Electoral Advantage (Mixed Models)

	Random Effects	Mixed Effects
	(1)	(2)
Electoral Advantage		-0.014^{**}
		(0.005)
Party Incumbency		-0.010
		(0.007)
Partisanship (GOP)		0.030
		(0.025)
Pseudo-R^2($marginal$)	0	0.033
Pseudo-R^2($conditional$)	0.494	0.496
Months SD/ICC (N = 11)	0.019/3.3%	0.018/3.3%
Years SD/ICC (N = 17)	0.031/9.4%	0.031/9.2%
Candidates SD/ICC (N = 24)	0.062/36.7%	0.060/35.3%
Residual SD/ICC	0.073/50.6%	0.073/52.1%
Observations	3,436	3,436
Log Likelihood	4,051	4,047

All models are (random intercept) multilevel regressions of the average share of populist rhetoric among U.S. presidential candidates clustered within months, years, and candidates.
The standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.

date fixed effects (see Table B2, 2). Third, we test whether the results hold even after excluding the unusually populist 2016 campaign by Donald Trump (see Table B2, 3) or any other particular election or candidate (not shown). Fourth, we constrain our analysis to three last months of the campaign, which excludes all primary campaign speeches with potentially different electoral incentives (Table B3).

Fifth, we consider that the variation in our main independent variable of electoral advantage primarily comes from candidates (and, to a less extent, changes within candidates across years and months). Consequently, we re-run our analysis at the much smaller candidate-year-month level as opposed to the speech level as in previous analyses (Table B4). While these results based on 189 observations are statistically much weaker, they similarly–in line with our theoretical model–indicate that those who have an advantage use less populist rhetoric.

Finally, and perhaps most important, we consider a number of alternative model specifications which also take into account that the relative rarity and overdispersion of populist rhetoric in our data (see Table B5). In particular, we fit several negative binomial regressions with the count of populist sub-speeches per speech as the dependent variable. Given a large number of zero values in our dependent variable, we also include (two-part) zero-inflated negative binomial regressions. Altogether, consistent with our previous results, these models suggest that losing candidates use more populist rhetoric in their speeches (and they are also more likely to use it in the first place).

Discussion

Many people appear to believe that politics is about the righteous Manichean fight between "the good people" and "the corrupt few." Similarly, various anti-elitist and anti-pluralist attitudes are rather widespread in the electorate. As a result, scholars often use this popular demand of populism to explain the rise and fall of populism:

politicians can exploit these attitudes in an opportunistic fashion to win elections when the demand is particularly high. However, the demand-side explanations cannot explain the puzzle of why populist rhetoric is not omnipresent when the demand is rather fixed within the same elections. Populism might be more credible for some than others. It is also only one possible rhetorical device and one set of popularly appealing ideas among many (Hawkins and Rovira Kaltwasser, 2017). Nonetheless, as of now, the study on supply side of populism can hardly explain the likely strategic rather than just principled *non-use* of populist rhetoric among the majority of political actors most of the time.

To remedy this omission, we introduce a new formal model of populism as a campaign gamble and argue that politicians are more likely to employ populism under the condition of (as-if) exogenous electoral disadvantage. We then corroborate the empirical implications of this proposition by measuring populist rhetoric in U.S. presidential campaign speeches using an original supervised machine learning algorithm and modeling the use of populism as a function of the initial polling results. In support of our hypothesis, we find that candidates with an electoral disadvantage, not necessarily the "challengers," are more likely to use populist rhetoric, even after accounting for various confounding factors such as incumbency status or partisanship.

By formally defining the potential costs and benefits of populist rhetoric in terms of voter (de)mobilization, our model of populism provides a fruitful way to explain when political actors may decide (not) to be populist as a part of the empirically testable equilibrium candidate strategies. By following a minimal conceptualization of populism that is independent of politicians' attributes and policy positions, the model bridges the previously disconnected ideational and game theoretic approaches to the study of populism. At the same time, our further empirical examination of populism in U.S. presidential rhetoric allows testing this and other related theories by deriving precise speech-level estimates of populist rhetoric. In doing so, we improve upon

the dictionary-based methods from previous research by introducing a new measurement method that better captures the necessary and sufficient conditions of populism. In turn, this gives us comprehensive, comparable measures of populism across U.S. presidential campaigns, corroborating the intuition of the likely unprecedented use of populist rhetoric by Donald Trump in 2016 election.

Our research is not without limitations. First, our model can only speak to a limited set of considerations regarding the strategic use of populism. For instance, we do not address why there is a high popular demand for populist rhetoric or, relatedly, why it can be sometimes effective in the first place. In that sense, we cannot explain variation in populist rhetoric that is unrelated to electoral support such as based on candidates' personality or electoral institutions. Second, we cannot explain why candidates may choose to use other non-ideological types of political rhetoric such as related to clientelism (Hawkins and Rovira Kaltwasser, 2017). Third, despite a large number of analyzed speeches, the available variation in candidates' polling results across campaign months is rather limited. Fourth, while we provide general models of strategic populism use and its measurement, our empirical test has been solely based on the U.S. presidential campaign data. However, none of these limitations likely challenge our main result regarding the greater strategic use of populist rhetoric among the initially losing candidates.

While further theoretical and empirical additions are beyond the scope of this paper, one can easily expand on our results in the future. When it comes to the theory, the model can be generalized to multiple actors and time periods, which would allow deriving hypotheses about the use of populism rhetoric across a variety of election types throughout the entire campaign. While a multi-party competition per se likely does not change the logic of our theoretical argument regarding the lesser incentive to use of populist rhetoric for the winning candidate (especially when in majority), it would be informative to consider the potential role of electoral thresholds and coalition-building

considerations in terms of shaping the incentives for weaker parties in a proportional representation system. It would be also useful to examine the role of mobilization as opposed to persuasion, loss aversion, as well as the uncertainty of populism effectiveness in more detail. Finally, one can complement our account by considering alternative strategic, formal conceptualizations of populism that are still in line with the ideational approach. For instance, it may be fruitful to model populism as a special symbolic type of turnout and vote buying (Nichter, 2008).

When it comes to further empirical implications, one can examine campaign rhetoric in U.S. congressional races or other non-U.S. two-candidate elections, which might provide much larger samples within the same electoral context. Combined with a multi-player model extension, one can also consider the use of populist rhetoric in primary elections, as well as multi-party elections outside of the United States. We believe it may be especially fruitful to devise empirical comparisons of the strategic (non-)use of populist campaign rhetoric across countries and different election types. In doing so, one can build on some of the recent research expanding text analysis methods to non-English languages and multi-lingual corpora (Dai, 2019; Dai and Radford, 2018; Halterman et al., 2018).

References

Acemoglu, D., G. Egorov, and K. Sonin (2013). A Political Theory of Populism. *Quartely Journal of Economics*, 771–805.

Acemoglu, D., J. A. Robinson, and R. Torvik (2013). Why do voters dismantle checks and balances? *Review of Economic Studies 80*(3), 845–875.

Akkerman, A., C. Mudde, and A. Zaslove (2013). How Populist Are the People? Measuring Populist Attitudes in Voters. *Comparative Political Studies 47*(9), 1324–1353.

Annenberg, P. P. C. (2000). Annenberg/Pew Archive of Presidential Campaign Discourse CD-ROM.

Ansolabehere, S. and S. Iyengar (1995). *Going negative: How political advertisements shrink and polarize the electorate.* Stanford University Press.

Ardag, M. M., B. C. Silva, J. P. Thomeczek, F. B.-R. Steffen, and L. Littvay (2019). Populist Attitudes and Political Engagement: Ugly, Bad, and Sometimes Good? *Journal of Representative Democracy*, 1–24.

Assche, J. V., A. Roets, A. V. Hiel, and K. Dhont (2019). Diverse Reactions to Ethnic Diversity: The Role of Individual Differences in Authoritarianism. *Current Directions in Psychological Science*.

Bakker, B. N., M. Rooduijn, and G. Schumacher (2016). The psychological roots of populist voting: Evidence from the United States, the Netherlands and Germany. *European Journal of Political Research 55*(2), 302–320.

Bonikowski, B. and N. Gidron (2016). The Populist Style in American Politics: Presidential Campaign Discourse, 1952–1996. *Social Forces 94*(4), 1593–1621.

Bornschier, S. (2017). Populist Mobilization Across Time and Space: An Introduction. *Swiss Political Science Review 23*(4), 301–312.

Castanho Silva, B. (2019). "he's not one of them!" antiestablishment supporters of populist governments in bolivia and ecuador. *The Journal of Politics 81*(3), 1085–1089.

Dai, Y. (2019). Measuring Populism in Context: A Supervised Approach with Word Embedding Models. In *Manifesto Corpus Conference, Berlin*.

Dai, Y. and B. Radford (2018). Multilingual word embedding for zero-shot text classification. In *Annual Meeting of the American Political Science Association, Boston*.

Danka, T. and P. Horvath (2018). modAL: A modular active learning framework for Python. Available on arXiv at `https://arxiv.org/abs/1805.00979`.

Dennison, J. and A. Geddes (2019). A Rising Tide? The Salience of Immigration and the Rise of Anti-Immigration Political Parties in Western Europe. *The Political Quarterly 90*(1), 107–116.

D'Orazio, V., S. T. Landis, G. Palmer, and P. Schrodt (2014). Separating the wheat from the chaff: Applications of automated document classification using support vector machines. *Political Analysis 22*(2), 224–242.

Engesser, S., N. Ernst, F. Esser, and F. Büchel (2017). Populism and social media: How politicians spread a fragmented ideology. *Information, communication & society 20*(8), 1109–1126.

Ertekin, S., J. Huang, L. Bottou, and L. Giles (2007). Learning on the border: active learning in imbalanced data classification. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, pp. 127–136.

Gallup (2013). Gallup Presidential Election Trial-Heat Trends, 1936–2008. Technical report.

Gidron, N. and B. Bonikowski (2013). Varieties of Populism: Literature Review and Research Agenda.

Gidron, N. and J. J. B. Mijs (2019). Do Changes in Material Circumstances Drive Support for Populist Radical Parties? Panel Data Evidence from the Netherlands during the Great Recession, 2007–2015. *European Sociological Review 35*(5), 637–650.

Goertz, G. (2006). *Social science concepts: A user's guide.* Princeton University Press.

Goldberg, Y. and O. Levy (2014). Word2vec explained: Deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*.

Halterman, A., N. Halterman, and J. Irvine (2018). How right wing is right wing populism? evidence from the manifesto corpus.

Hameleers, M., L. Bos, and C. H. de Vreese (2018). Selective exposure to populist communication: How attitudinal congruence drives the effects of populist attributions of blame. *Journal of Communication 68*(1), 51–74.

Hameleers, M., C. Reinemann, D. Schmuck, and N. Fawzi (2019). The Persuasiveness of Populist Communication. Conceptualizing the Effects and Political Consequences of Populist Communication from a Social Identity Perspective. In *Communicating populism. Comparing interactions between politicians, media, and citizens across Europe*, pp. 1–38. Routledge.

Hawkins, K. and L. Littvay (2019). *Contemporary US Populism in Comparative Perspective*. Elements in American Politics. Cambridge University Press.

Hawkins, K. A. (2009). Is Chávez populist? Measuring populist discourse in comparative perspective. *Comparative Political Studies 42*(8), 1040–1067.

Hawkins, K. A., R. E. Carlin, L. Littvay, and C. R. Kaltwasser (2018). *The ideational approach to populism: Concept, theory, and analysis*. Routledge.

Hawkins, K. A. and C. R. Kaltwasser (2018a). Measuring populist discourse in the united states and beyond. *Nature human behaviour 2*(4), 241–242.

Hawkins, K. A. and C. R. Kaltwasser (2018b). Measuring populist discourse in the United States and beyond. *Nature Human Behaviour 2*(4), 241–242.

Hawkins, K. A. and C. Rovira Kaltwasser (2017). What the (Ideational) Study of

Populism Can Teach Us, and What It Can't. *Swiss Political Science Review 23*(4), 526–542.

Hawkins, K. A. and B. C. Silva (2018). Textual analysis: Big data approaches. In *The Ideational Approach to Populism*, pp. 27–48. Routledge.

Heiss, R. and J. Matthes (2020). Stuck in a nativist spiral: Content, selection, and effects of right-wing populists' communication on facebook. *Political Communication 37*(3), 303–328.

Herkman, J. (2017). The life cycle model and press coverage of nordic populist parties. *Journalism studies 18*(4), 430–448.

Huber, R. A. and S. P. Ruth (2017). Mind the Gap! Populism, Participation and Representation in Europe. *Swiss Political Science Review 23*(4), 462–484.

Immerzeel, T. and M. Pickup (2015). Populist radical right parties mobilizing 'the people'? The role of populist radical right success in voter turnout. *Electoral Studies 40*, 347–360.

Inglehart, R. and P. Norris (2017). Trump and the populist authoritarian parties: the silent revolution in reverse. *Perspectives on Politics 15*(2), 443–454.

Jagers, J. and S. Walgrave (2007). Populism as political communication style: An empirical study of political parties' discourse in Belgium. *European Journal of Political Research 46*(3), 319–345.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research 30*(3), 411–433.

Krupnikov, Y. (2011). When Does Negativity Demobilize? Tracing the Conditional Effect of Negative Campaigning on Voter Turnout. *American Journal of Political Science 55*(4), 796–812.

Kustov, A. (2021). Testing the Backlash Argument: Voter Responses to Pro-immigration Reforms. *Working Paper (https://alexanderkustov.org/files/Kustov_Backlash.pdf)*.

Kustov, A., D. Laaker, and C. Reller (2021). The Stability of Immigration Attitudes: Evidence and Implications. *The Journal of Politics 83*(4), 1478–1494.

Lacatus, C. (2019). Populism and the 2016 American election: Evidence from official press releases and twitter. *PS - Political Science and Politics 52*(2), 223–228.

Lau, R. R. and I. B. Rovner (2009). Negative Campaigning. *Annual Review of Political Science 12*(1), 285–306.

Le, Q. and T. Mikolov (2014). Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 1188–1196.

Lee, F. E. (2019). Populism and the American Party System: Opportunities and Constraints. *Perspectives on Politics*, 1–45.

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean (2013). Distributed representations of words and phrases and their compositionality. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pp. 3111–3119.

Miller, B., F. Linder, and W. R. Mebane Jr. Active learning approaches for labeling text: Review and assessment of the performance of active learning approaches. *Political Analysis 28*(4), 532–551.

Moffitt, B. (2016). *The global rise of populism: Performance, political style, and representation.* Stanford University Press.

Motta, M. (2018). The Dynamics and Political Implications of Anti-Intellectualism in the United States. *American Politics Research 46*(3), 465 –498.

Mudde, C. (2004). The Populist Zeitgeist. *Government and Opposition 39*(4), 541–563.

Mudde, C. and C. R. Kaltwasser (2017). *Populism: A very short introduction.* Oxford University Press.

Mudde, C. and C. R. Kaltwasser (2018). Studying Populism in Comparative Perspective: Reflections on the Contemporary and Future Research Agenda. *Comparative Political Studies 51*(13), 1667–1693.

Müller, J.-W. (2017). *What is populism?* Penguin UK.

Müller, P., C. Schemer, M. Wettstein, A. Schulz, D. S. Wirz, S. Engesser, and W. Wirth (2017). The polarizing impact of news coverage on populist attitudes in the public: Evidence from a panel study in four european democracies. *Journal of Communication 67*(6), 968–992.

Nai, A. (2018). Fear and Loathing in Populist Campaigns? Comparing the Communication Style of Populists and Non-populists in Elections Worldwide. *Journal of Political Marketing.*

Nichter, S. (2008). Vote Buying or Turnout Buying? Machine Politics and the Secret Ballot. *American Political Science Review 102*(1), 19–31.

Oliver, J. E. and W. M. Rahn (2016). Rise of the trumpenvolk: Populism in the 2016 election. *The ANNALS of the American Academy of Political and Social Science 667*(1), 189–206.

Qi, G.-J., X.-S. Hua, Y. Rui, J. Tang, and H.-J. Zhang (2008). Two-dimensional active learning for image classification. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE.

Remmer, K. L. (2012). The rise of leftist–populist governance in latin america: the roots of electoral change. *Comparative Political Studies 45*(8), 947–972.

Rooduijn, M. and B. Burgoon (2018). The paradox of well-being: do unfavorable socioeconomic and sociocultural contexts deepen or dampen radical left and right voting among the less well-off? *Comparative Political Studies 51*(13), 1720–1753.

Rooduijn, M. and T. Pauwels (2011). Measuring populism: Comparing two methods of content analysis. *West European Politics 34*(6), 1272–1283.

Rydgren, J. (2008). Immigration sceptics, xenophobes or racists? radical right-wing voting in six west european countries. *European Journal of Political Research 47*(6), 737–765.

Saito, T. and M. Rehmsmeier (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one 10*(3), e0118432.

Schmuck, D. and J. Matthes (2017). Effects of economic and symbolic threat appeals in right-wing populist advertising on anti-immigrant attitudes: The impact of textual and visual appeals. *Political Communication 34*(4), 607–626.

Serra, G. (2018). The electoral strategies of a populist candidate: Does charisma discourage experience and encourage extremism? *Journal of Theoretical Politics 30*(1), 45–73.

Settles, B. (2009). Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences.

Settles, B., M. Craven, and L. Friedland (2008). Active learning with real annotation costs. In *Proceedings of the NIPS workshop on cost-sensitive learning*, pp. 1–10. Vancouver, CA:.

Singer, M. (2018). Delegating away democracy: how good representation and policy successes can undermine democratic legitimacy. *Comparative Political Studies 51*(13), 1754–1788.

Skaperdas, S. and B. Grofman (1995). Modeling Negative Campaigning. *American Political Science Review 89*(1), 49–61.

Swank, D. and H.-G. Betz (2003). Globalization, the welfare state and right-wing populism in western europe. *Socio-Economic Review 1*(2), 215–245.

Taggart, P. (2000). Populism: Concepts in the social sciences. *Philadelphia: Open*.

Taylor, S. L., M. S. Shugart, Lijphart Arend, and Grofman Bernard (2014). *A Different Democracy: American Government in a 31-Country Perspective*. New Haven: Yale University Press.

Tong, S. and D. Koller (2001). Support vector machine active learning with applications to text classification. *Journal of machine learning research 2*(Nov), 45–66.

Ulinkskaitė, J. and L. Pukelis (2021). Identifying populist paragraphs in text: A machine-learning approach. *arXiv preprint arXiv:2106.03161*.

Van Hauwaert, S. M. and S. Van Kessel (2018). Beyond protest and discontent: A cross-national analysis of the effect of populist attitudes and issue positions on populist party support. *European Journal of Political Research 57*(1), 68–92.

Weyland, K. (2001). Clarifying a Contested Concept: Populism in the Study of Latin American Politics. *Comparative Politics 34*(1), 1–22.

Woolley, J. T. and G. Peters (2008). The American presidency project. *Santa Barbara, CA. Available from World Wide Web: http://www. presidency. ucsb. edu/ws*.

Wuttke, A., C. Schimpf, and H. Schoen (2020). When the whole is greater than the sum of its parts: On the conceptualization and measurement of populist attitudes and other multidimensional constructs. *American Political Science Review*, 1–19.

Supplementary Material

When Do Politicians Use Populist Rhetoric?
Populism as a Campaign Gamble

Online Appendix A:

Model of Populism as a Campaign Gamble

Building on Skaperdas and Grofman (1995), our model assumes a standard office-seeking two-candidate race where each candidate has a certain level of pre-existing electorate support among the decided (or henceforth "mobilized") voters, while the rest of the (potential) electorate is undecided (or henceforth "unmobilized"). We further stipulate that candidates can use a combination of "conventional" and "populist" campaign rhetoric to improve their electoral chances. While conventional rhetoric is assumed to help mobilize additional support among the unmobilized, populist rhetoric is assumed to demobilize the opponent's pre-existing support. The use of populism, however, can backfire such that there is some chance that it can demobilize the candidate's own pre-existing support (or, equivalently, mobilize more votes for the other candidate). Overall, we show that, despite the potential risk, the *ex-ante* losing candidate is more likely to use populism to have at least some chance of winning.

Our basic model of populist rhetoric as a campaign gamble is of imperfect information. There are two candidates (or parties), A and B. Both candidates observe each others' level of pre-existing support α_i and the share of unmobilized electorate $\overline{\alpha} = 1 - \alpha_A + \alpha_B$. Then, each candidate simultaneously decides to allocate its effort to populist (p_i) or non-populist, conventional (c_i) rhetoric so that $p_i + c_i = 1$.

While political campaigning can have a number of aims including changing voter's preferences over candidates, we assume that the primary function of conventional campaign rhetoric is mobilizing electoral support among the (currently) unmobilized. Put formally, let $m^A(c_A, c_B)$ and $m^B(c_A, c_B)$ indicate the share of unmobilized electorate ultimately attracted by candidates A and B such that, for any given combination pair of conventional campaign strategies pursued by both candidates, $m^i(c_i, c_j)$ is increasing in c_i and decreasing in c_j. To that end, we also assume that all of the unmobilized

are equally and ultimately susceptible to mobilization by either candidate such that $m^A(c_A, c_B) = m^B(c_A, c_B)$ and $m^A(c_A, c_B) + m^B(c_A, c_B) = 1$. As a result, the candidates attract the same share of the unmobilized electorate when they decide to allocate the same amount of effort to conventional campaigning. Since the function m is symmetric ($m^i(c_i, c_j) = 1 - m^j(c_j, c_i)$), we can simply denote $m^A(c_A, c_B)$ by m and $m^B(c_A, c_B)$ by $1 - m$. Finally, we assume that conventional campaigning has diminishing returns (so that m is a concave function: $d^2m/dc_i^2 \leq 0$).

Unlike conventional campaigning to attract the unmobilized, we assume that the primary function of populist rhetoric is demobilizing the opponent's pre-existing support. In line with some of the empirical literature described above, however, we also assume that populist campaigning can backfire by demobilizing the candidate's own current supporters. Put formally, for any given combination pair of populist campaign strategies p_i pursued by both candidates, let $\alpha_i(p_i + Ep_j)$ be the resulting decrease in pre-existing support share for candidate i, where E indicates the relative effectiveness of populist campaigning (or to what extent the opponent is hurt more than the candidate). In our base model, we assume that $E > 1$. Importantly, at least in terms of the relative electoral advantage, the backfire effect of populist campaigning that demobilizes one's own support is equivalent to the one that mobilizes the electorate to vote for the opponent. In other words, although the model focuses on populism as primarily a tool for demobilization, the main distinctive feature of populist rhetoric is ultimately assumed to be its greater riskiness (relative to non-populist rhetoric).

We can now summarize the final overall support that each candidate gets after deciding on their use of conventional and populist rhetoric. To simplify, given that $p_i + c_i = 1$, we can represent the resulting support (α_i') as just a function of c_i:

$$\alpha_i' = \alpha_i + \overline{\alpha}m - \alpha_i(1 - c_i + E(1 - c_j)) \tag{1}$$

Similar to other campaign strategy literature, we assume that candidates ultimately care about maximizing their winning margin or electoral advantage. Consequently, given equation 1, we can define the utility function for each candidate as follows:

$$u_i(c_i, c_j) = \alpha'_i - \alpha'_j = \overline{\alpha}(2m(c_i, c_j) - 1) - \alpha_i(1 - c_i + E(1 - c_j)) + \alpha_j(1 - c_j + E(1 - c_i)) + \alpha_i - \alpha_j \quad (2)$$

After formulating the strategic form of our game, we can now proceed with determining the possible Nash equilibria. We can say that a campaign strategy pair (c_A^*, c_B^*) is an equilibrium if $u_i(c_i^*, c_j^*) \geq u_i(c_i, c_j^*)$ for all $c_i, i \neq j$. Let $u'_i = du_i(c_i, c_j)/dc_i$, $m'_i = dm/dc_i$, and assume that $m''_i = d^2m/dc_i^2 \leq 0$. Then, we can find the first derivative and characterize the marginal benefits of putting extra effort into conventional and populist campaigning as follows:

$$u'_i(c_i, c_j) = \overline{\alpha}2m'_i - (E\alpha_j - \alpha_i). \quad (3)$$

We can then similarly derive $u''_i = \overline{\alpha}2m''_i \leq 0$. Conditional on the assumptions above being satisfied, we can now show that candidates' utility function u_i is concave in their own strategy c_i and thus that there exists a pure-strategy Nash equilibrium.

Given equation 3, both candidates would only devote a non-zero effort to both campaign strategies ($0 < c_A^* < 1$ and $0 < c_B^* < 1$) if and only if their marginal benefits and costs are equalized. Quite naturally, this implies that $E\alpha_A > \alpha_B$ and $E\alpha_B > \alpha_A$ simultaneously, which necessarily requires that the derivatives in equation 3 are equal to zero for both candidates:

$$
\begin{aligned}
\overline{\alpha}2m'_A(c_A^*, c_B^*) - (E\alpha_B - \alpha_A) = 0 &\implies m'_A(c_A^*, c_B^*) = (E\alpha_B - \alpha_A)/(2\overline{\alpha}) \\
-\overline{\alpha}2m'_B(c_A^*, c_B^*) - (E\alpha_A - \alpha_B) = 0 &\implies -m'_B(c_A^*, c_B^*) = (E\alpha_A - \alpha_B)/(2\overline{\alpha})
\end{aligned}
\quad (4)
$$

Now suppose that one of the candidates has more pre-existing support $\alpha_A > \alpha_B$.

We can then show that $E\alpha_A - \alpha_B > E\alpha_B - \alpha_A$ and, by equation above, $m'_A(c^*_A, c^*_B) < -m'_B(c^*_A, c^*_B)$ so that $m(c^*_A, c^*_B) > 1/2$ under our assumptions. In turn, this is equivalent to $c^*_A > c^*_B$ or $p^*_A < p^*_B$, which gives us our main result: *"the candidate with a lower pre-existing support is expected to use more populist campaign rhetoric relative to his opponent"* (**Proposition 1**).

In addition to this general result it may also be instructive to examine two special cases where one of the candidate allocates all effort to either conventional or populist campaigning (c^*_i is equal to 0 or 1). First, suppose that the pre-existing support is lower for one of the candidates ($\alpha_i > \alpha_j$) and that the effectiveness of populist rhetoric is relatively low ($E\alpha_j \leq \alpha_i$). Then, in line with equation 3, $u'_i(c_i, c_j) > 0$ given that $\alpha_i \geq E\alpha_j$. Consequently, candidate i would only do conventional campaigning in equilibrium ($c^*_i = 1$ is the optimal choice regardless of c_j). Second, consider a function m with a finite derivative $m'_i(0, c_j)$ and sufficiently low α_i (or sufficiently high E). Then, regardless of c_j, it must be true that $u_i(0, c_j) = \overline{\alpha}2m'_i(0, c_j) - (E\alpha_j - \alpha_i) \leq 0$. In other words, candidate i would only do populist campaigning in equilibrium ($c^*_i = 0$ is the optimal choice regardless of c_j). In sum, although this is less realistic than the general proposition 1, *if the candidate's pre-existing support is sufficiently low (high) or populist rhetoric is sufficiently (in)effective, then the candidate is expected to fully engage in populist (conventional) campaigning.* Importantly, the results hold even if we introduce some uncertainty about the (in)effectiveness of populism and relax the assumption that $E > 1$, i.e., that populist rhetoric is necessarily hurting the opponent more than the candidate instigator of such rhetoric (not shown).

Online Appendix B:

Tables and Figures

Figure B1: Distribution of Speeches across Presidential Campaigns
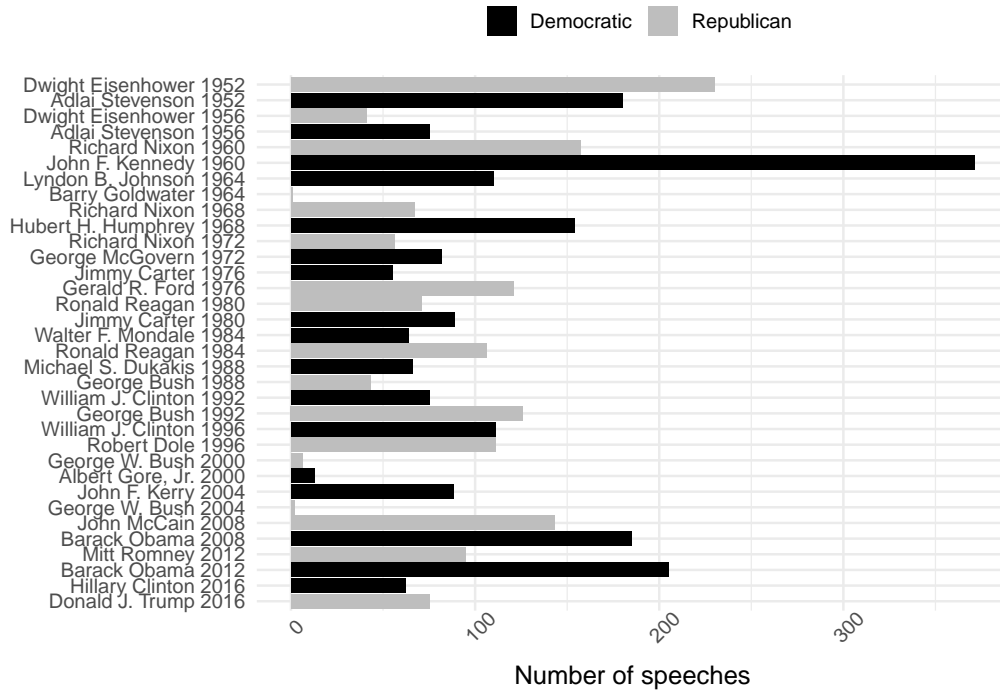

Note: The plot visualizes the distribution of speeches across campaigns in our combined dataset.

Table B1: Summary Statistics

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Populism Rhetoric (Average Share)	0.02	0.09	0	0	0	1
Electoral Advantage (Binary)	0.53	0.50	0	0	1	1
Electoral Advantage (Percent)	0.44	6.40	−29.50	−4.00	5.00	18.00
Party Incumbency	0.46	0.50	0	0	1	1
Party Membership (Republican)	0.42	0.49	0	0	1	1
Speech Length (Standardized)	0.19	0.12	0.00	0.10	0.26	1.00

Full Speech-level Data (n = 3,436). The mean speech length is 2,167 words.

Figure B2: Out-of-sample Model Performance before and after Active Learning
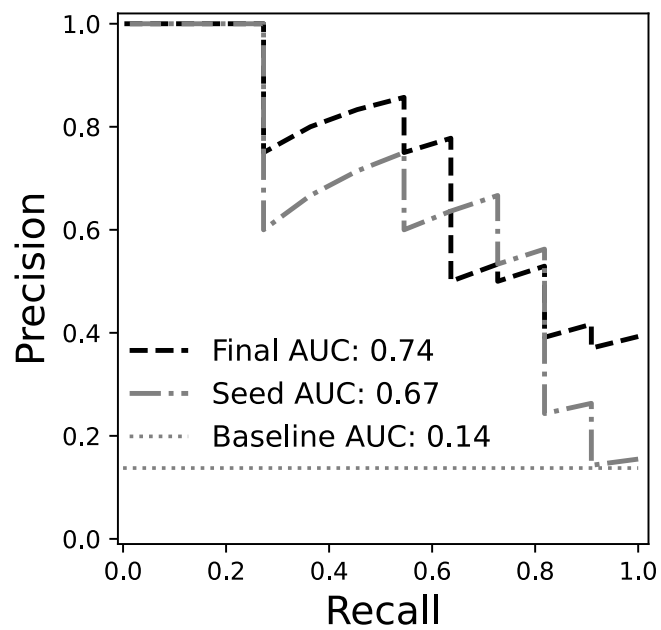
Note: The plot indicates the precision recall AUCs for the random forest classifier using the seed training data and the final model after populating the training data with active learning.

Table B2: Populist Rhetoric as a Function of Electoral Advantage (Robustness Checks)

	Sub-speech Level	Speech Level	
	(1)	(2)	(3)
Electoral Advantage	−0.013***	−0.015**	−0.015**
	(0.003)	(0.005)	(0.005)
Party Incumbency	−0.039***	−0.003	−0.003
	(0.003)	(0.008)	(0.007)
Partisanship (GOP)	0.005	0.015	0.014
	(0.003)	(0.015)	(0.013)
Speech Length	*N/A*	*Yes*	*Yes*
Month FE	*Yes*	*Yes*	*Yes*
Year FE	*Yes*	*Yes*	*Yes*
Candidate FE	*No*	*Yes*	*Yes*
Excluding D. Trump	*No*	*No*	*Yes*
Observations	11,839	3,436	3,361
Adjusted R^2	0.131	0.322	0.136

Note: All models are OLS regressions of populist rhetoric in U.S. presidential speeches on electoral advantage in the recent monthly polls and other candidate or speech characteristics. Standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.

Table B3: Populist Rhetoric as a Function of Electoral Advantage (Sep.-Nov. Only)

	(1)	(2)	(3)	(4)	(5)
Electoral Advantage	−0.013***	−0.013***	−0.012***	−0.021	−0.020*
	(0.003)	(0.003)	(0.003)	(0.011)	(0.009)
Party Incumbency	−0.021***	−0.023***	−0.027***	−0.004	−0.003
	(0.003)	(0.003)	(0.003)	(0.007)	(0.006)
Partisanship (GOP)	0.002	0.002	0.00004	0.023	0.021
	(0.003)	(0.003)	(0.003)	(0.017)	(0.014)
Speech Length	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Month FE	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Year FE	*No*	*No*	*Yes*	*Yes*	*Yes*
Candidate FE	*No*	*No*	*No*	*Yes*	*Yes*
Excluding D. Trump	*No*	*No*	*No*	*No*	*Yes*
Observations	2,778	2,778	2,778	2,778	2,725
Adjusted R^2	0.021	0.024	0.274	0.433	0.246

All models are OLS regressions of the average share of populist rhetoric in U.S. presidential speeches on electoral advantage in the three last monthly polls and other characteristics.
The standard errors are given in parentheses, *p<0.05; **p<0.01; ***p<0.001.

Table B4: Populist Rhetoric as a Function of Electoral Advantage (Candidate-Year-Month Level)

|	(1)	(2)
Electoral Advantage	−0.009	−0.011
	(0.009)	(0.009)
Month FE	*No*	*Yes*
Year FE	*No*	*Yes*
Observations	189	189

All models are OLS regressions of the average share of populist rhetoric among U.S. presidential candidates on electoral advantage in the most recent monthly polls aggregated at the candidate-year-month level. The standard errors are given in parentheses.

Table B5: Populist Rhetoric as a Function of Electoral Advantage (Rare Event Models)

	Negative Binomial				Zero-inflated Negative Binomial			
	(1)	(2)	(3)	(4)	(5a)	(5b)	(6a)	(6b)
Electoral Advantage	-0.860^{***}	-0.397^{*}	-0.371	-0.544	-1.865^{***}	-0.610	-1.216^{***}	-4.114^{***}
	(0.197)	(0.202)	(0.231)	(0.290)	(0.296)	(0.341)	(0.298)	(1.042)
Candidate Controls	Yes	Yes	Yes	Yes	No	No	Yes	Yes
Sub-Speech Count	No	Yes	Yes	Yes	No	No	Yes	No
Month FE	No	Yes	Yes	Yes	No	No	Yes	Yes
Year FE	No	No	Yes	Yes	No	No	Yes	Yes
Candidate FE	No	No	No	Yes	No	No	No	No
Observations	3,436	3,436	3,436	3,436	3,436	3,436	3,436	3,436
Log Likelihood	-824	-734	-569	-539	-850	-850	-502	-502

All models are negative binomial regressions of the populist sub-speech count in the speeches of U.S. presidential candidates. The last four columns show the two-part zero-inflated negative binomial regressions (5 and 6) with the first (a) column indicating the count model coefficients and the second (b) column indicating the zero-inflation model coefficients. The standard errors are given in parentheses. *p<0.05; **p<0.01; ***p<0.001.

Figure B3: Average Share of Populist Rhetoric across Years
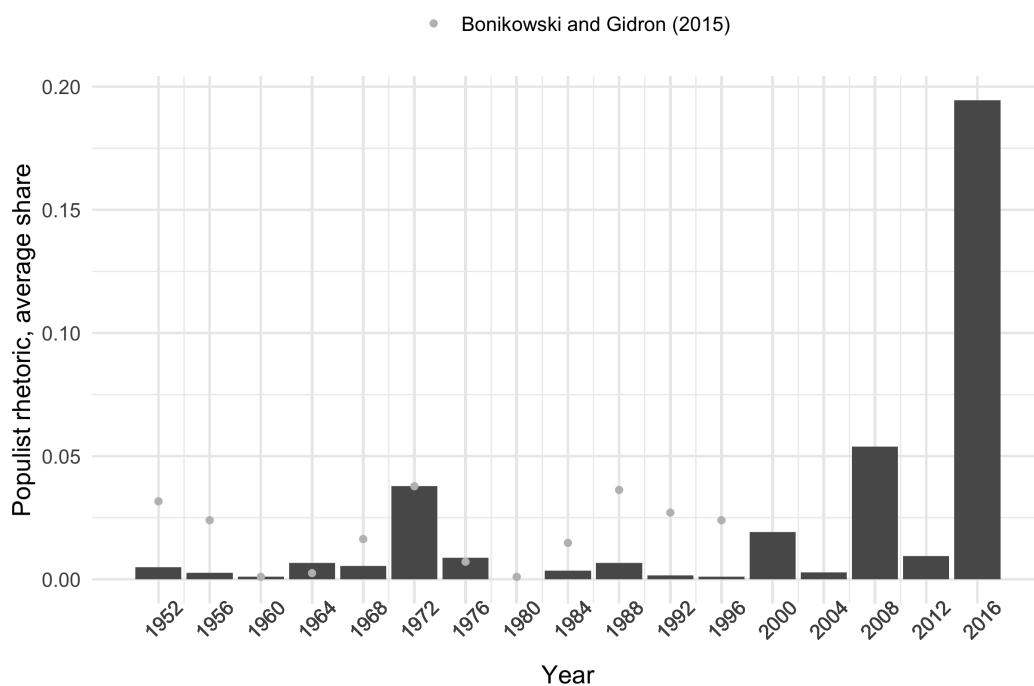
Note: The bar plot indicates the average share of populist sub-speeches in U.S. presidential campaigns across years based on our measure and Bonikowski and Gidron (2016). Since we have a sub-speech level measure and Bonikowski and Gidron (2016) measure is at word level, the two measures are on different scales. To make it more comparable, their measure is re-scaled to have the same lower and upper limit as our measure between 1952 to 1996. To match and to compare with Bonikowski and Gidron (2016), we create the annual average level of populist rhetoric by calculating the proportion of populist sub-speeches in a year. As can be seen, both measures have similar trends between 1952 and 1996 with 1972 and 1988 as the most populist elections while 1960 and 1980 as the least populist elections.

Online Appendix C:

Word Embedding Models

Word embedding is a type of language model that maps words or sentences and documents into vectors of real numbers. Unlike the common 'bag-of-words' method of vectorization, in which one unique word is one dimension, word embedding represents words and documents in a dense continuous vector space with many fewer dimensions and positions semantically and syntactically similar words close to each other in this vector space. The method of word embedding is based on a distributional hypothesis in linguistics theory, which states that the meaning of a word is a function of its contexts or surrounding words. Unlike the 'bag-of-words' assumption, which treats words as independent atomic units, the distributional hypothesis aims to model the meaning of a word and assumes that the meaning of a word is given, and can be approximated, by the sets of contexts in which the word appears. In effect, the underlying idea is that words that frequently appear in the same contexts are likely to have a similar meaning.

There are several different ways to train word embedding. In this paper, we use a Doc2vec model (Le and Mikolov, 2014), which is based on the more foundational Word2vec model (Mikolov et al., 2013). We begin by describing the Word2vec model. The Word2vec model is a neural network based model that takes each unique word in the vocabulary of a corpus as an input. The input word, represented as a one-hot vector, is then multiplied by a dense, real-valued weights matrix of size $V \times d$, where V is the length of the vocabulary in the corpus and d is the chosen size of the hidden layer or 'embedding'.[1] By multiplying the $1 \times V$ input vector for a word with the $V \times d$ weights matrix, a $1 \times d$ vector is generated; this is the word's vector representation, v_{word}. The model then uses this vector representation of the input target word as the input to a softmax classifier to predict which of the V words in the vocabulary

---

[1] We choose $d = 150$, in keeping with standard practice.

are likely to be the context words of the input word. Context words are those that appear in a certain range of words before and after the current/target word. The model learns the embedding or the parameters in the hidden layer by finding the parameters that maximize the predicted probability of true context words. In other words, the Word2vec model seeks to set parameters θ to maximize the conditional probability of contexts C when observing the target word T: $p(C|T; \theta)$ for all words in the vocabulary (Mikolov et al., 2013; Goldberg and Levy, 2014).[2] Therefore, mathematically, the model assigns similar parameters to words that are used interchangeably in the same contexts.

Because maximizing $p(C|T; \theta)$ for all target and possible contexts is expensive to compute and there are more words that do not appear together than words that often appear together, we adopt negative sampling skip-gram in training the model. In negative sampling skip-gram, the input layer contains target-context word pairs. The target-context pairs are generated by taking the target word at index i and pairing it with all context words from $i - k$ to $i + k$ given a window size k.[3] For every true target-context word pair, we generate s negative samples; these are target-context word pairs that are not observed in the actual text corpus.[4] The output layer contains dummy values 1 and 0 indicating whether the input pair is a true target-context pair that co-locates in the texts (1) or a negative/fake pair that does not appear together in the texts (0). The predicted value given an input pair is computed by taking the dot product of the target word vector (target embedding) and the context word vector (context embedding) and then applying the logistic function, $\sigma(\cdot)$. The model uses small non-zero random values as the initial parameters in the hidden layer to produce

---

[2] Word2vec encompasses two different, related neural-network based models, including the continuous bag-of-words (CBOW) and skip-gram (SG) models (Mikolov et al., 2013). The SG model, which is used and explained here, inputs a target word from a text and attempts to predict the target word's likely context words. The CBOW model does the reverse. Given a set of context words, CBOW attempts to predict the context's target word.

[3] In our model we use $k = 10$.

[4] In our model, we use $s = 10$.

the embedding/word vector. Stochastic gradient descent is then used to optimize the parameters through back-propagation to minimize the logarithmic loss between $\sigma(v_{\text{target word}} \cdot v_{\text{context word}})$ and the true value $[0, 1]$.

Expanding the Word2vec model to the document level is simple; each document is labeled with an ID and treated as one unit (like a word). This document ID is positioned within the text in the document. For example, suppose we have a one-sentence document labeled as Doc1: "We are fighting for the forgotten Americans." The document ID is treated as one unit and positioned within its text: "We are fighting for *Doc1* the forgotten Americans."[5] The negative sampling algorithm can now be applied to both the target word and the document, which is treated as a target word. In this way, the documents sharing similar texts or content are positioned close to each other in the vector space (Le and Mikolov, 2014).

---

[5]In practice, the model is adjusted so that the *Doc1* token occurs in all of document 1's words' contexts and all of the document 1's words appear in the *Doc1* token's context.