The Shift to Commitment Politics and Populism: Theory and Evidence

Luca Bellodi

Massimo Morelli

Antonio Nicolò

Paolo Roberti

The Shift to Commitment Politics and Populism: Theory and Evidence[*]

Luca Bellodi[†]   Massimo Morelli[‡]   Antonio Nicolò[§]   Paolo Roberti[¶]

July 25, 2023

Most Recent Version

Abstract

The decline in voters' trust in government and the rise of populism are two concerning features of contemporary politics. In this paper, we present a model of commitment politics that elucidates the interplay between distrust and populism. Candidates supply policy commitments to mitigate voters' distrust in government, shrinking politicians' levels of discretion typical of representative democracies. Alongside commitments, candidates rationally choose the main strategies associated with populism, namely anti-elite and pro-people rhetoric. With novel data on voters' distrust towards the U.S. federal government, which we match with the Twitter activity of more than 2,000 candidates over five congressional elections, we show that distrust is strongly associated with candidates' supply of commitments and populist rhetoric, which are also effective strategies at mobilizing distrustful voters. We also show theoretically that the shift to commitment politics determines greater aversion to checks and balances, and hence even illiberal populism can emerge.

Keywords: Populism, Commitment, Anti-Elite Rhetoric, Trust, Turnout, Agencies of Restraint

JEL codes: D72, D78, P16

1 Introduction

Populism has many faces, and scholars across the social sciences have shown how populist leaders use anti-elite rhetoric to appeal to the grievances of the "pure people" against a "corrupt elite"; how populists capitalize on people's fears and perceived threats, from immigration to automation and more general risk of economic insecurity; how populist leaders have often advocated for illiberal reforms aimed at building an unmediated relationship between the people and their leader, "draining the swamp", or giving the country back to the people. Ultimately, populism has been considered at odds with the core principles of liberal democracy. However, no theory condenses these faces in an integrated account of populism, and the connections between populism – on the one hand – and policy, rhetoric, and democracy – on the other hand – often seem to escape. In this paper we provide novel theory and evidence of populism centred on commitment politics, which shed new light on how these faces are tied to one another, and on how the erosion of citizens' trust in government triggers the populist vicious cycle.

For years, advanced democracies have witnessed a steady decline in voters' trust in government. In the U.S., for instance, the share of individuals who say they trust the government to do what is right dropped from 73% in 1958 to 20% in 2022.[1] Social scientists have identified several sources of such growing distrust, encompassing two broad categories: economic and cultural threats. Algan et al. (2017) find that unemployment increases together with declining trust toward political institutions in Europe. Moreover, the costs imposed on some groups of citizens by globalization (see e.g., Autor et al. 2020, Rodrik 2018, Colantone & Stanig 2018), automation (see e.g., Acemoglu & Restrepo 2020), and the 2008 financial crisis (Algan et al. 2017, Guiso et al. 2021) all increased economic insecurity and distrust towards political institutions. A parallel strand of work has focused instead on the cultural causes of distrust in political institutions. With burgeoning immigration and the erosion of traditional values, trust in political elites' abilities to protect group-based identities plummeted, triggering value change in subsets of the population (see e.g., Laitin 2018, Norris & Inglehart 2019, Berman 2021).[2]

In this landscape of generalized grievances and resentment, populist parties and leaders found fertile ground. In their 70-country-wide study, Funke et al. (2020) show that, in 2018, 16 countries were ruled by a populist (25% of their sample) and similar patterns are reported by the PopuList team for the vote share of populist parties in Europe (Rooduijn et al. 2019). The key trait of these parties and leaders, compared to traditional political actors, is a clear proposed distinction between the "pure" people and the corrupt elites, deemed to be responsible for all

---

[1] See https://www.pewresearch.org/politics/2022/06/06/public-trust-in-government-1958-2022/.

[2] Both the economic and cultural mechanisms are well summarized in a survey article by Guriev & Papaioannou (2022).

society's ills. This people-vs-elites feature is the cornerstone of every definition of populism used in the social sciences (Mudde & Rovira Kaltwasser 2018). Several scholars consider populism a thin-centred ideology which portrays society as divided into the "pure people" and the "corrupt elite" (the so-called "ideational" approach) (Mudde 2004), whereas others view populism as a strategy employed by politicians in an attempt at "winning and exercising power", by creating unmediated ties between the leader and voters (Weyland 2001, 2017). Other definitions used to study populism – especially in Latin America – still focus on candidates' incentives to propose policies that promote the interest of common citizens and signal distance from traditional elites (Dornbusch & Edwards 1991, Acemoglu, Egorov & Sonin 2013). What all these definitions have in common is a clear, moral separation between the people and the elites. As a result, the use of anti-elite and pro-people rhetoric is a standard measure of populism in more recent empirical work too (Hawkins et al. 2019, Gennaro et al. 2021, Di Cocco & Monechi 2021, Dai & Kustov 2022).

In this paper, we show that commitment politics is what links the erosion of voters' trust in government and the rise of populism. Reduced trust in political institutions, together with increased demand for protection, implies greater chances of electoral success for candidates who commit to protection policies, pre-empting the discretion typically granted to politicians in representative government. The demand for protection is to be understood broadly, stemming from the threats discussed above such as immigration, globalization, and economic crises, to protection from various "others", if perceived as a threat to identity and cultural integrity. Once politicians shift from the traditional trustee model of political agency centred on delegated discretion to this commitment strategy, all the other features of populist behaviour follow as rational complementary strategies: anti-elite and pro-people rhetoric, fake-news production, and aversion to checks and balances.

First, we present a simple principal-agent baseline model, where a representative voter (principal) selects a politician (agent) who chooses a policy. The optimal policy is uncertain at the time the politician is appointed, and depends on the realization of a state of the world. The voter selects either a politician who commits to implement a policy that is ex-ante determined (a committed agent) or one who promises to choose the ex-post optimal policy for the voter (a trustee). In the latter case, whether the politician will choose or not the ex-post optimal policy depends on the probability that *the elite* distorts her choice. The choice of commitment is more likely the greater the expected strength and distortion caused by the elite. A lower trust that politicians can resist the elite capture makes the principal (voter) prefer a committed agent who panders to the voter's ex-ante desiderata. While, to keep the model simple and short, we

assume that commitment can be enforced, it is immediate to notice that even in the absence of a commitment technology, the voter can easily monitor and punish with the lack of reelection a politician who does not comply with an electoral commitment. Instead, in the case of full delegation, the voter cannot easily attribute a given policy choice to the state of the world or the elites intervention, and therefore cannot discipline a delegated politician. Pandering to the ex-ante voter's beliefs has been considered an important feature of populism (see e.g., Acemoglu, Egorov & Sonin 2013), but, most important, we characterize the candidate's complementary strategies when introducing electoral competition, showing that they are all consistent with the various faces of populism discussed in the literature.

When we introduce electoral competition between two parties, the two primaries can lead to different choices of candidates to run in the general elections, depending on the different levels of distrust, beliefs, and preferences within the two parties. It is possible that, in the general elections, one party has selected a candidate supplying a greater level of commitment than the other. We show that whenever some difference exists between candidates in terms of their supply of commitment, the most committed candidate has a strong incentive to use anti-elite and pro-people rhetoric. We also show that decreased trust in such cases increases turnout for the more committed candidate and depresses the turnout for a full-delegation/trustee candidate. We then show that the probability of seeing committed populists in equilibrium is the more likely the more homogeneous policy preferences within parties are, and hence the greater is ideological polarization among partisan voters.

We provide novel empirical evidence in support of the predictions of the baseline model on the connection between distrust, shift to commitment politics, and populist rhetoric. In the context of the US congressional elections, we use multilevel regression and post-stratification to create district-level estimates of voters' distrust in the federal government, which we match with the universe of tweets posted by more than 2,000 candidates during the five congressional elections between 2012 and 2020. We use natural language processing and supervised machine learning techniques to detect when a tweet contains a policy commitment or populist rhetoric. We find that voters' distrust is strongly associated with candidates' supply of commitments and populist rhetoric, especially when the candidates tweet about topics where partisan voters have homogeneous preferences. We show that, compared to the 2012 presidential elections, distrustful Republicans were more likely to turn out to vote during the 2016 elections when Trump was the Republican candidate, lending support to the prediction whereby the supply of commitment and populist rhetoric is effective at mobilizing distrustful voters.

After establishing, theoretically and empirically, that the shift to commitment politics is

the logical mechanism that connects the erosion of trust phenomenon with the populism wave, we show in an extension that the shift to commitment politics also determines a demand for reduced checks and balances, hence rationalizing existing evidence on voters' willingness to trade off democratic principles for policy preferences (Graham & Svolik 2020, Alsan et al. n.d.) as well as anecdotal regularities on the anti-democratic steps taken by populist politicians. We show this by introducing in the model an "agency of restraint" – which could be free media, independent bureaucracy, or independent judiciary. In such an augmented model we show that, under general assumptions, voters who choose a committed agent will typically also desire to weaken or eliminate such agency of restraint, while voters who prefer full delegation will typically prefer to have such an agency that produces checks and balances. Similarly, we argue that the committed candidate benefits from making information acquisition costlier for voters, a benefit that can for instance be attained by favouring the spread of fake news. Thus, the shift to commitment politics not only generates a clear shift towards populist rhetoric in electoral campaigns, but it also determines a potential demand for illiberal reforms, which is a pattern that has received sustained attention among political commentators.

Once the connections between distrust, commitment, and populism are drawn, it is easier to derive expectations about the policy consequences of populist government. Our predictions are consistent with the findings in Sasso & Morelli (2021) and Bellodi et al. (2023) – who show that when populists are in power, they replace competent bureaucrats with loyal bureaucrats, to minimize the risk of resistance to the populist commitment agenda by expert bureaucrats. Moreover, there is mounting anecdotal evidence that populist governments from across the political spectrum tend to weaken (besides bureaucracy) also other agencies of restraints, from the media to the judiciary. Evidence from Hungary shows, for instance, that Orbán used state-owned firms' advertising resources to favour media outlets in exchange for positive coverage of the government (Szeidl & Szucs 2021), and checks and balances have been clearly reduced;[3] Mexico's President Obrador recently approved a reform that strips the National Electoral Institute's independence ahead of the 2024 presidential elections;[4] and in Poland, the governing Law and Justice party (PiS) substantially reduced judicial independence and media freedom after a successful electoral campaign dense with policy commitments.[5] The findings in Funke et al. (2020) – who identify a negative long term effect of populist leaders on GDP per capita – are broadly consistent with the expectation that the pursuit of unconditional, ex-ante determined commitments is likely to be very costly for the economy as well. Our paper provides a new,

---

[3]See `https://www.bbc.com/news/world-europe-21740743`). The constitutional court can no longer block executive policy decisions.

[4]See `https://foreignpolicy.com/2023/03/23/lopez-obrador-electoral-reforms-mexico-democracy-ine/`.

[5]See `https://www.bbc.com/news/world-europe-65809525`.

Electronic copy available at: https://ssrn.com/abstract=4520441

integrated framework for the study of populism that sheds light on the reason for all these connected phenomena.

2 Related Literature

This paper contributes to several strands of the literature in political economy and political science. The distinction between the committed agent versus trustee model of political agency has first been introduced by Fox & Shotts (2009) in the context of optimal accountability of incumbents. Along the same dimension, Kartik et al. (2017) show that in electoral competition the equilibrium degree of discretion and ambiguity in candidates' policy platforms depends on the divergence in policy preferences that exist between voters and candidates.[6] Di Tella & Rotemberg (2018) show distrust is also related to betrayal aversion, which strengthens the desire for commitments. While some accounts focus on the credibility of candidates' pledges (Van Weelden 2013), for simplicity our model assumes full credibility of commitments, even though the results would be qualitatively the same when allowing for partial credibility.

Some papers present theories of populism that assume cognitive differences among voters, which we do not invoke in our model. Levy et al. (2022), for instance, depict populist policies as simplistic ones desired by unsophisticated voters, who sometimes win elections because of an intense dislike for the status quo, whereas Crutzen et al. (2020) show that if the people are divided into an informed minority and an uninformed majority, parties tend to cater more to the better-informed or the elite, and hence the common people develop disaffection for the traditional parties, leading to entry incentives for a populist third party. Egorov & Sonin (2021) present a model where an informed minority (the elite) can advise the uninformed majority on candidates' competence, when one candidate is biased towards the elite and the other one is unbiased, and look at the conditions under which the uninformed majority follows the elite's advice. Rather than partitioning citizens along cognitive dimensions, we believe that the critical partition concerns trust, heterogeneous beliefs, and preferences.

A recent set of papers have emphasized changes in social identification, highlighting how the national versus global identity has become the most relevant cleavage, even more relevant than the standard economic left-right ideological cleavage (see e.g., Shayo 2009, Ford & Jennings 2020, Besley & Persson 2019, Bonomi et al. 2021, Gennaioli & Tabellini 2023). These papers are complementary to the distrust and demand of commitments we focus on. Once it becomes

---

[6]On commitment vs flexibility, see also Amador et al. (2006). A similar distinction is present also in Ghosh & Tripathi (2012) and Bueno de Mesquita & Friedenberg (2011), but in their context the committed agent is an "ideologue". In our paper ideology matters for preferences, but it is not a constitutive feature of the populist's strategy.

rational for a number of parties and politicians to choose a commitment strategy, such parties and politicians may become, as a rational consequence, strategic suppliers of messages about the protection of national and communal values. The lower level of trust in political institutions among the audience of such messages implies that the economic left-right dimension drops in salience, since ideological cleavages are of secondary importance compared to the cultural threats that trigger distrust.

Finally, in the extension section on the consequences of the shift to commitment politics for checks and balances, we contribute to the literature on endogenous checks and balances and on the determinants of illiberal reforms. Acemoglu, Robinson & Torvik (2013) show how (poor) voters might prefer weaker checks on the executive when they expect the rich elite to be able to easily bribe the politician, as it is the case in many weakly institutionalized polities. Voters are therefore happy to give up control of the executive as a guarantee that the politician will not be bought by the rich elite. Gratton & Morelli (2022) model checks and balances in a similar way to ours, for checks and balances regimes are treated as a veto player that can block executive decisions. Focusing on populist policies aimed at dismantling expert bureaucracies, Gratton & Lee (2023) show that demand for such illiberal reforms depends on the level of distrust voters have towards public servants.[7]

3   Agency Theory

In this section we present a baseline principal-agent framework, in order to highlight the sources of change on the "demand" side.

A principal G (a party or a voter) delegates at time 1 an agent (a politician) g to choose a policy $q \in \{\ell, h, z\} \equiv Q$ at time 2 in case g is the decision maker. The ex-post optimal policy for G depends on the state of the world at time $t = 2$, which is unknown at time 1 and is observed by the agent g with probability 1 at time 2.[8] The principal chooses $S \subseteq Q$, i.e. the degree of discretion left to the agent. In other words, S is the subset of policies among which the agent is delegated to choose at time 2 after having observed the realization of the state of the world. If $S = Q$, then there is full delegation and the agent can choose any policy, while on the other extreme if $|S| = 1$, the set S is a singleton, and the agent is committed to implement the policy that the principal has chosen ex-ante.

---

[7]The sources of the shift to commitment politics are also entirely consistent with the growing appeal of direct democracy, but we do not connect to such a literature since the direct democracy route is not included in our model.

[8]If we assume that the political agent observes the state of the world with probability related to her competence or ability, we can show that the choice of commitment can also entail lower interest in competence. Here we shut down this channel for the sake of simplicity.

The tension between these strategies, full commitment, partial commitment or full delegation, comes from the fact that the agent can be influenced by an "elite".[9] The elite may distort the agent's choice and has potentially a negative effect on the welfare of the principal. Full delegation allows an unbiased informed agent to choose the ex-post optimal policy for the principal, but it exposes the principal to the risk that the agent is captured by the elite. Commitment, on the contrary, prevents the elite from distorting the agent's choice, but it does not allow to adjust the policy ex-post. We adopt the simplest setting to highlight this tension.

Assume that the principal has a belief $\lambda > 1/3$ that the elite's preferred policy is z; the elite preferred policy is believed to be one of the other two policies with equal probability $\frac{1-\lambda}{2} < \lambda$. In contrast, we assume that the principal's belief that policy z is optimal, γ_z, is lower than the probability she attaches on the event that reform h or l can be optimal, denoted respectively by γ_h and γ_l. This assumption reflects a situation where the principal feels that remaining with the status quo z is unlikely to be optimal, hence there is a desire of change. This implies that z is never ex-ante optimal for the principal in case the principal wants full or partial commitment by the agent. A simple interpretation of the assumptions that $\lambda > \frac{1}{3}$ and z is the least likely policy to be optimal ex-ante is that z could represent a policy that preserves the status quo: the elite is more likely than not to want to preserve the status quo (as the elite often contains the "establishment" that determined the current state of affairs in the first place), while the typical principal (voter) is more likely to desire a reform, especially at time of a crisis.

We denote by q^* the ex-post optimal policy when uncertainty over the realization of the state of the world is resolved. In this section we assume (without loss of generality for the principal agent baseline) that between h and l the ex-ante optimal policy is h.

Payoffs:

– The principal gets a utility normalized to 0 if the implemented policy is q^*. If the implemented policy is a wrong one, then she gets -1.

– The agent's payoff is

$R - \epsilon(\mathbf{1}|_{q \neq q^*})$ – with $\epsilon < R$ – if she respects the contract with the principal;

whereas if she chooses a policy outside S, then the reward R does not accrue.[10]

The fear of elite influence:

The principal assigns a positive probability p that the elite influences the agent in her choice within S, whereas we assume that R is large enough to make sure that the agent never chooses a policy outside the set S chosen by the principal. On the other hand, when S is not a

---

[9]The elite is never an explicit player in our model, and for different situations the elite could mean the establishment, the economic powers, or any combination of external interest groups that could influence policy makers. As it will become clearer, we do not need to be specific about the identity of the elite.

[10]The most obvious interpretation of R is the value of being reconfirmed in the future.

8

singleton, with probability p the agent could receive enough compensation by the elite to choose a suboptimal policy within S.[11]

We can now study which contract a principal offers to the agent. If the principal offers full delegation, $S = Q$, the expected payoff is equal to

$$-p\left[\lambda\left(1-\gamma_z\right) + \frac{1-\lambda}{2}\left(1-\gamma_h\right) + \frac{1-\lambda}{2}\left(1-\gamma_l\right)\right] = -p\left[\lambda\left(1-\gamma_z\right) + \frac{1-\lambda}{2}\left(1+\gamma_z\right)\right].$$

If the principal wants partial commitment, $S = \{h, l\}$, the expected payoff is equal to

$$-p\left[\frac{1-\lambda}{2}\left(1-\gamma_h\right) + \frac{1-\lambda}{2}\left(1-\gamma_l\right)\right] - \gamma_z = -p\left[\frac{1-\lambda}{2}\left(1+\gamma_z\right)\right] - \gamma_z.$$

Thus, partial commitment yields higher expected utility than full delegation if

$$p > \frac{\gamma_z}{\lambda(1-\gamma_z)} \equiv p_1. \tag{1}$$

If the principal offers a full commitment contract, $S = \{h\}$, the expected payoff is equal to $-\gamma_l - \gamma_z$. Thus, full commitment is preferred ex ante to full delegation if

$$p > \frac{(1-\gamma_h)}{\lambda(1-\gamma_z) + (1-\lambda)(1+\gamma_z)/2} \equiv p_2 \tag{2}$$

This implies that the principal chooses full delegation to the agent (trustee) if p is less than the minimum of the two thresholds p_1, p_2 in (1) and (2). Given that p captures *distrust* in the simplest possible way, this already suffices to see that whenever trust goes sufficiently down then political agency choices must shift towards some type of commitment, partial or full.

Proposition 1. *As long as either inequality (1) or (2) (or both) are satisfied, full delegation is never chosen as agency contract.*

Partial commitment is chosen for intermediate levels of distrust and full commitment for higher levels of distrust.

Note that full commitment is preferred to partial commitment when

$$-p\left[\frac{1-\lambda}{2}(1+\gamma_z)\right] - \gamma_z \le -\gamma_l - \gamma_z,$$

$$p \ge \frac{2\gamma_l}{(1-\lambda)(1+\gamma_z)}.$$

---

[11]We do not model the bargaining game between the elite and the politician but we assume that when active (i.e. with probability p) the elite manages to influence the agent to choose the elite's preferred policy if it is within S.

Note also that when the principal is quite sure about what she prefers ex ante, namely γ_h is very high, then the RHS of (2) is very low, and hence the full commitment choice is typically optimal when, on top of low trust, there is also a clear ex ante preference, which we will interpret in the electoral competition model presented below as a proxy for homogeneity of preferences within parties.[12]

The following comparative statics corollary is important:

Corollary 1. *p_1 and p_2 are increasing in γ_z and decreasing in λ, within their restricted domains.*

The importance of this corollary is due to the fact that low γ_z reflects a large perceived difference between the principal's preferences and the (most likely) preferences of the elites that could capture the agent. High λ means that it is most likely that the elite preferred policy is z, therefore it is more likely that commitment is preferred. Thus, to preview the results of the following sections, a committed agent fighting against a fully delegated opponent will have interest in trying to convince voters that λ is high and γ_z is low, which is the model's representation of anti-elite rhetoric.

4 Elections and Mobilization

What happens when we augment the model by allowing for primaries and general elections rather than a simple principal-agent choice, and how do mobilization strategies modify the baseline result? What may explain the existence of an equilibrium in which one party chooses commitment but the other does not? And what role do the fundamentals play in such cases for the determination of the probability of winning for the populist party? Is the ex ante pandering commitment component of populism conducive to the other components of populist politics often emphasized in the political science literature? In this section we offer an answer to all these questions.

4.1 Elections

We extend the model to allow for two-party electoral competition in order to characterize the equilibrium behavior of candidates and to emphasize what determines their probability of electoral success.

Let there be two parties, $G = A, B$. Each of them selects a candidate in a primary election (time $t = 0$) and then the selected candidates $g = a, b$ compete in a general election at time

---

[12]Intuitively, if the principal is a party, γ_h can be interpreted as the percentage of members of the party who hold a strict preference for policy h.

$t = 1$. Each candidate j of party G in the primary proposes an electoral program $e^j = S^j \subseteq Q$. The winner w of the general election chooses a policy $q \in S^w$. For each party G the ex-post optimal policy depends on the state of the world at time $t = 2$, which is unknown at the time of elections. On the other hand, after the general election the elected politician observes the state of the world before making the decision.

A candidate's payoff is the same as the agent's payoff (described in the previous section) if elected, and zero otherwise. The utility of a citizen is the policy utility of the principal in the previous section, minus the cost of voting in case of turnout. Members of the different parties may have different preferences. In particular, members of party G assign a positive probability p^G that the choice of a politician in her electoral platform when in office is influenced by the elite. We denote by γ_q^G the probability that members of G ex ante assign that q will be optimal for them. Finally, $\lambda^G > 1/3$ denotes the probability assigned by members of party G that the policy z will be the one preferred by the elite at time 2. As before we assume that for each party G, $\gamma_z^G < \gamma_t^G$, $t \in \{h, l\}$. We assume that in each party there are a continuum (unit size) of party members. The parameters γ_q^G can be interpreted as the different percentages of voters in each party who believe ex ante that q will be optimal ex post.

Voting in primary elections is costless.[13] Voting is instead costly in the general elections, and the possibility of abstention here matters.

To determine endogenous turnout, we make an "as if pivotal" assumption (Alesina & Rosenthal 1996), i.e., we assume that each citizen turns out if and only if the difference in policy utility for her between the two candidates is higher than her cost of voting.[14] The cost of voting for citizen v of party G is denoted by $c_v^G + \mu^G$, where c_v^G is distributed according to a uniform distribution on $[\underline{\phi}, \overline{\phi}]$, with $0 < \underline{\phi} < \overline{\phi}$ and $\mu^G \in [-\frac{1}{2\psi}, \frac{1}{2\psi}]$ is a party specific voting cost common to all members of G that modifies the distribution of voting costs after the primaries and before the general elections. μ^G is for simplicity distributed uniformly.

In order to limit the attention to interior probabilities of winning, we make the following assumption:

Assumption 1: $\psi \leq \overline{\psi}$, with $\overline{\psi} > 0$ sufficiently low.

This implies that at the time of the primary elections there is sufficient uncertainty over the outcome of the general election due to uncertainty in turnout, irrespective of the electoral program of the candidates who win the primaries.

---

[13]Adding costly voting in the primaries does not change the results presented below if costs are randomly assigned.

[14]The results do not depend on the specific model of turnout chosen (see e.g., Morelli et al. 2021).

4.2 Equilibrium Turnout Effects

We solve the model by backward induction. Consider the general election. Each voter in the general election decides whether to abstain or to vote, and which candidate to vote for. If a citizen does not abstain, then she votes for the candidate who maximizes her expected utility. Let $EU^G(e)$ denote the expected policy utility for a voter of party G if a candidate with platform e is elected. Let e and e' be the electoral programs of the two candidates. Without loss of generality suppose that $EU^G(e) > EU^G(e')$; voter G votes for the former candidate if $EU^G(e) - EU^G(e') \geq c_v^G + \mu^G$, otherwise, she abstains. Candidates' electoral programs are those chosen in the primary elections. Going backwards, citizens are rational, so members of party G in the primaries vote for the candidate with the electoral program that maximizes their expected utility, taking into account the candidates' probability of winning at the general election. We denote by e^{g*} the equilibrium electoral program of candidate g, winner of the primaries of group G. In equilibrium, citizens vote for a candidate of their group, and the turnout of group G depends positively on the difference $EU^G(e^{g*}) - EU^G(e^{-g*})$, hence the probability of winning of candidate g depends positively on $EU^G(e^{g*}) - EU^G(e^{-g*})$ and negatively on $EU^{-G}(e^{-g*}) - EU^{-G}(e^{g*})$.

The following lemma helps characterize the results of the primary elections.

Lemma 1. *Every candidate in the primary election of party G proposes the electoral program e^{g*} that gives to party G members the highest expected policy utility.*

The intuition of the previous result is the following. Suppose that the ex-ante optimal electoral program for party g is different than the ex-ante optimal electoral program for the other party $(-g)$. At the primary election a candidate j of party g can propose the electoral program $e^*(g)$ that maximizes party g members' utility if j wins the general election, or a different electoral program $e'(g)$ that increases the probability that j wins the general election by reducing party $-g$ turnout (that is, an electoral program that members of party $-g$ prefer to $e^*(g)$). Assumption 1 implies that there is a sufficiently large uncertainty on the result of the general election at the time of the primary election, such that party g members prefer a candidate who proposes $e^*(g)$ to any candidate who proposes $e'(g)$. Therefore competition among candidates at the primaries guarantees that the winner proposes $e^*(g)$.[15]

Hence what electoral program each party G chooses in the primary elections is described by the same inequalities presented in the previous section, with the difference that p^G, γ_q^G are party

---

[15]This result can be sustained also with different assumptions: members of a group do not know the preferences for policies of members of the opposing group, or they observe the electoral program of the opposing candidate with a probability lower than 1.

specific. Consistently, in the primary elections, voters may demand a full or partial commitment and may have different willingness to trade the possibility of getting an ex-post optimal policy for the reduction of the probability that the elite can distort politicians' policy ex-post.

The following proposition highlights that citizens distrust towards politicians affects the turnout of party G in the general elections, but in different ways depending on the electoral program offered by the parties.

Proposition 2. • *If the equilibrium electoral program of candidate g is $e^{g*} = Q$ (full delegation), then the higher is p^G, the lower the turnout of party G is, regardless of the opponent's platform.*

• *If the equilibrium electoral programs are $\{S^G = Q, S^{-G} \subset Q\}$ (full delegation vs. commitment), then the higher is p^{-G}, the higher the turnout of party $-G$ is.*

This result, the proof of which follows from Proposition 1 and Lemma 1, establishes that distrust can have opposite effects on turnout of the two competing candidates in general elections when one of them has a partially or fully committed electoral platform. This tells us that not only higher distrust is more likely to induce a commitment choice by parties, but also that higher distrust induces greater turnout for the group whose candidate runs on a commitment platform (full or partial) and a lower turnout for the group whose candidate has full delegation.

4.3 Anti-Elite Rhetoric

In the above analysis of general elections we have assumed for simplicity that the candidates in the general election present themselves with the platform chosen in the primaries and do nothing to affect turnout. It is easy and realistic, however, to assume that in the electoral campaign before the general elections they want to choose campaign rhetoric aimed to make voters update their beliefs about the various parameters, since all of them contribute to the difference in expected utility between the two candidates and hence to the probability of winning. A campaign rhetoric has the general goal of affecting the belief γ_q^G that policy q is optimal for G. Without attempting any microfoundation of how rhetoric can successfully modify such beliefs, we can say that in the model the closest representation of what an anti-elite rhetoric may mean is an attempt (1) to reduce γ_z^G (e.g., "it is not in the interest of people to have the policy likely preferred by the elite"); (2) to increase λ^G (e.g., "the elite will surely push in the direction of the people least preferred policies").[16]

---

[16] Even rhetoric aimed to increase p would go in the same direction, but beliefs about what policy is optimal are easier to affect than more general trust parameters, which tend to move slowly, as all cultural parameters.

The claim of this section is that the standard metric with which parties are called populist or not, namely the use of anti-elite rhetoric, is an epiphenomenon of a commitment platform choice.

Proposition 3. *Whenever a party G chooses commitment, full or partial, and faces a fully delegated opponent, then it is always in the interest of their candidate g to use anti-elite rhetoric in the electoral campaign, pushing γ_z^G down and λ up.*

4.4 Homogeneity and Polarization Effects

Let us consider a situation in which the two parties have ex-ante different preferred policies: $\gamma_h^A > \gamma_l^A$ and $\gamma_l^B > \gamma_h^B$. In this context, an increase in γ_h^A or γ_l^B corresponds to a larger homogeneity of preferences within party, and a larger polarization across parties.[17]

Proposition 4. *A candidate with some degree commitment is more likely if the homogeneity of preferences within her party increases. Moreover, if a party has chosen a fully committed candidate, it is in the interest of the committed candidate in the general election to try to foster even stronger homogeneity, in order to maximize the probability of winning. This implies that if the two parties prefer different policies ex-ante ($\gamma_h^A > \gamma_l^A$ and $\gamma_l^B > \gamma_h^B$), polarization also increases.*

4.5 Discussion on Commitment

The model assumes that R is high enough to make commitment credible by the agent, and that the principal has no cost in "not giving" R (i.e., firing the agent) in case of contract violation. In a nutshell, we assume that a commitment technology exists. Scholars have proposed different mechanisms that endogenously create commitment, which we outline below.

Reelection incentives. In a dynamic framework, voters can punish politicians who deviate from their electoral commitments by not reelecting them again, see the survey by Duggan & Martinelli (2017). This tool could be weakened by the presence of term limits (see also Smart & Sturm 2013).

Cost of lying. There is large evidence (see Abeler et al. 2014, Gneezy et al. 2018) that individuals suffer psychological costs of lying. This applies also to politicians. If lying imposes a cost on politicians (see e.g., Callander & Wilkie 2007, Carrillo & Dewatripont 2008), they

---

[17]Given that the policy space is a finite set, polarization increases when more individuals prefer policies that are different across parties, consistent with the polarization index by Esteban & Ray (1994). The same concept and polarization index has also been used in the political science literature (Posner 2004, Kasara & Suryanarayan 2015).

will not deviate largely from their electoral promises (for experimental evidence, see Corazzini et al. 2014).

Reputational costs. By deviating from their electoral commitments, politicians can damage their reputation both in their future political career, and in the private market (see Alesina 1988, Levy 2007). This can be sufficient to make politicians stick to what they promised.

Citizen candidates/partisan candidates/party selection. Politicians can signal their own preferences for policy by their demographic characteristics and their professional cv. If there is separation of types, politicians will run electoral campaigns with promises that are coherent with their preferences for policies, which will be indeed implemented once elected. Similarly parties can be aggregators of politicians with similar preferences (Caillaud & Tirole 2002, Levy 2004, Snyder & Ting 2002).

There is therefore large evidence that electoral promises are not empty words. However, each of these mechanisms works only under given conditions, and can fail if those conditions are not met. The objective of our theory is to show that distrust can push for a demand and supply of commitment policies, with an asymmetric effect on the turnout of the two candidates, who have also heterogeneous incentives to run anti-elite rhetoric campaigns and to limit the checks and balances of democracy (see extension below). Citizens' incentives to demand commitment may be still present even if they know that the commitment technology may fail.[18] Moreover, the extent to which voters are able to discipline committed politicians is, in many instances (e.g., reelection incentives, reputational costs), inherently linked to the analysis of a dynamic model, which is outside the scope of this paper.

5   Evaluation of the Predictions

In this section we provide evidence in favour of the predictions contained in our propositions, namely that distrust increases the supply of commitment, and that committed agents also engage in populist rhetoric, thus mobilizing distrustful voters. With data on U.S. congressional elections, we show that (i) candidates running in districts with higher distrust use more commitments and populist rhetoric in their tweets, (ii) especially when the tweet pertains to topics with uniform voter preferences, and that (iii) turnout of citizens supporting a commitment (and populist-rhetoric) candidate increases when distrust is large.

The U.S. context lends itself nicely for an empirical test of the predictions. The presence

---

[18]The model can be indeed extended to show that, if with a given probability ρ, the elite distorts the politician's choice outside the commitment set S, the theoretical results are still valid when the probability of elite capture p is sufficiently larger than ρ. Furthermore, the empirical results below on the existence of the link between distrust and commitment politics can be considered even more impressive if people attach only partial credibility to policy commitments.

of single-member districts allows for a precise match between candidates and voters in each district, allowing for sufficient variation in distrust at the district-level. Political campaigns are therefore highly targeted to the constituents of the districts, and the supply of commitments and populist rhetoric have the necessary space to emerge as a result of distrust. In Section *Concluding Remarks*, we discuss how these findings can travel to countries with other institutional configurations.

5.1 Data and methodology

5.1.1 (Dis)Trust in Government

There are two main challenges to estimating how candidates' communication strategy changes with voters' level of distrust. First, there should be a clear partitioning of voters into geographical regions where candidates run for office. Second, there should be data on distrust for said sub-national regions. Looking at Congressional districts/elections in the U.S. allows us to overcome both challenges. Single-member districts create a precise match between a candidate and a district, and opinion surveys with geo-coded responses allow us to use statistical models to derive sub-national estimates from representative samples of respondents at the national level for each of the 435 congressional districts.

We produce novel and time-changing estimates of distrust in the federal government at the congressional district level applying multilevel regression and post-stratification (MRP) techniques on the three most recent waves of the American National Elections Study (ANES 2012, 2016, 2020). Because ANES respondents are representative at the national level, we can obtain valid estimates of distrust for each congressional districts by modelling individual survey responses as a function of individual-level variables and variables at the district- and state-level in a multilevel model and then post-stratify these predictions with census data (Gelman & Little 1997).

We use the three most recent ANES waves (2012, 2016, 2020) and estimate the level of distrust in government of eligible voters across 435 congressional districts in 2012, 2016 and 2020. To measure distrust, we use the question "How often do you trust the government in Washington to do what is right?" and re-code the variables in order to have dichotomous answers. Respondents who "Never" trust the government were coded as 1 (distrust), whereas respondents who answered "Always", "Most of the time", "About half of the time", or "Some of the time" were coded as 0. This question is particularly suitable for the test we aim to perform. First, it explicitly asks respondents about *trust*, without resorting to other proxies or

Table 1: Descriptive Statistics, Distrust.

Year	N. Respondents	% Distrust	SD
2012	5,663	0.06	0.24
2016	4,079	0.13	0.33
2020	7,093	0.09	0.28

Notes: Share of respondents who never trust the government in Washington to do what is right across the three ANES waves.

distinct concepts like confidence of competence.[19] Second, it refers to the federal government in Washington, hence respondents are prompted to think about the same entity when answering the question. In Table 1 below we report the share of respondents who report to never trust the government in Washington across each wave.[20]

We produce the estimates in two steps, that we perform automatically with the `AutoMrP` package in R, which uses machine learning algorithms to improve standard MRP models (Broniecki et al. 2022).[21] First, we fit a multilevel model to the survey data to predict the level of distrust in government for specific categories of respondents (i.e., combining sex-education-age variables into unique categories of "ideal types" distributed across the sub-national units). Once obtained a predicted value for each ideal type, in the second step we calculate the weighted average level of distrust for each district where the weights are determined by how prevalent the ideal types are in the population of each congressional district (i.e., post-stratification step). Data for post-stratification comes from the American Community Survey of the Census Bureau.[22]

In Figure 1, we plot the estimates for 2020. Each same-size hexagon represents a congressional district and darker shades indicate higher distrust. The average share of estimated distrustful voters across the districts is 9.7%, compared to 6% in 2012, and 13% in 2016. Alabama, Louisiana, and Mississippi are the states where average distrust is highest in 2020, all with a share of distrustful individuals above 10%.

---

[19] In the appendix (Table E.8) we show the results are similar when using alternative survey questions: How many of the people running the government are corrupt? (Distrust = 1 for "All"/"Most" responses), Would you say the government is pretty much run by a few big interests looking out for themselves or that it is run for the benefit of all the people? (Distrust = 1 for "Run by a few big interests" responses).

[20] The size is large enough to create valid estimates of district-level responses. Warshaw & Rodden (2012) suggest that national samples of as many as 2,500 people produce reliable estimates for congressional districts. In Section B.1 in the appendix we validate this approach by comparing the estimates obtained from the smaller 2020 ANES sample with the estimates obtained from much larger NationScape sample (Tausanovitch & Vavreck 2021). While in NationScape there are no questions about trust in government, we show that there is a large correlation between the estimated share of respondents who think the economy in the past year worsened and those who say they are mostly interested in politics.

[21] We describe the estimation in greater detail in Section B in the appendix.

[22] With the proposed method we are not able to produce estimates of distrust for Republican and Democratic voters separately, for we do not have census data on party identification of individuals living in congressional districts.

Figure 1: MRP estimates of distrust.
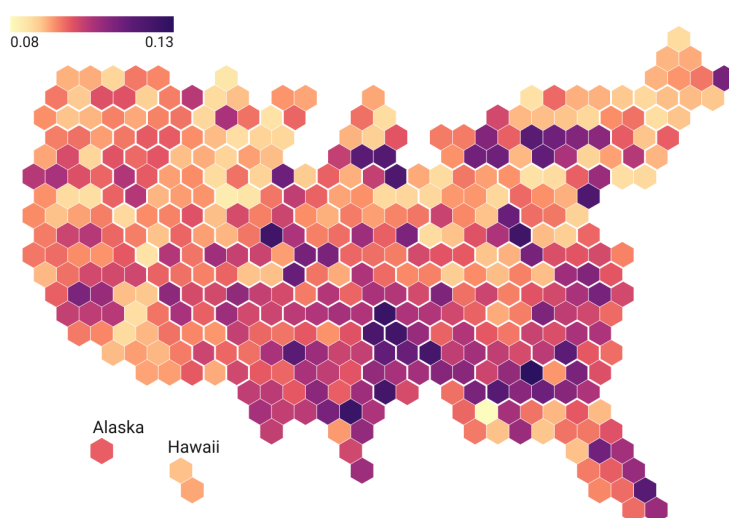
Notes: MRP 2020 estimates of distrust in government (share of voting-age individuals who never trust the government in Washington to do what is right) for each congressional districts.

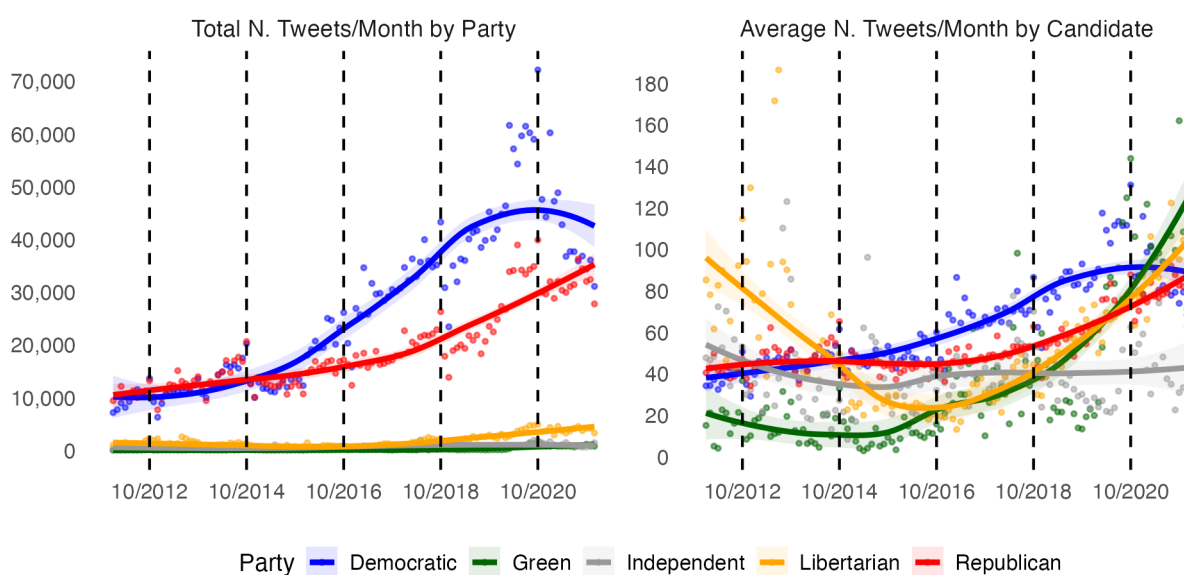### 5.1.2 Supply of Commitment and Populist Rhetoric

To measure the supply of commitments and populist rhetoric we look at the twitter activity of U.S. congressional candidates. We obtain data on the name of candidates and the districts where they ran from the *Candidates in American General Elections* database compiled by Cha et al. (2021). We then scrape the twitter handle of candidates from three sources: Ballotpedia, an online repository on U.S. elections and candidates; the House of Representatives Press Gallery;[23] and Poliwoops, an organisation which tracks deleted tweets by public officials and maintain a list of active twitter accounts.[24] For the candidates not matched in these sources, we programme a scraper that uses Selenium to directly enter a combination of name and surname of the candidates into the search engine of the official Twitter website, and we check manually the candidates who ran more than once.

3,579 unique candidates ran for office between 2012 and 2020, 35% are Democrats and 32% are Republicans. The remaining candidates belong to the Green (4%) and Libertarian (13%) parties, some are Independent (7%) and 9% of candidates' party are classified as "other". We recover at least one account for 64.1% of the candidates. Of the 1,283 candidates for which we do not find an account we are certain that for 11% of them there is no Twitter account. Given the extensive scraping performed, it is highly likely that the remaining 1,145 candidates (who ran only once) do not currently have an active Twitter account.

---

[23]Twitter handle for member of the 117th Congress available at `https://pressgallery.house.gov/member-data/members-official-twitter-handles`.

[24]Dataset available at `https://www.propublica.org/datastore/dataset/politicians-tracked-by-politwoops`.

Figure 2: Twitter activity of candidates.

Notes: Average number of tweets posted per month by Republican and Democratic candidates over time. Vertical dotted lines represent months prior to elections.

We download 5.9 million tweets (in English and excluding re-tweets) from the Twitter API for the period 2012-2021 posted by the candidates with an account. Figure 2 shows the average number of tweets posted by candidates belonging to different parties over time (panel on the right) as well as the total number of monthly tweets posted by candidates belonging to different parties (panel on the left). Despite Republican and Democrats accounting for the largest share in term of tweets' authorship, candidates from the two main parties do not tweet on average more than candidates from minor parties. In the 2020 elections, for instance, Libertarian and Green candidates were as active as Democrats and more active than Republicans on Twitter. Democrats' average number of tweets is larger than that of Republicans by 45%. Democratic candidates post 75 tweets per month on average, whereas Republican candidates 52.

To classify tweets based on whether they use policy commitments and populist rhetoric or not, we build on recent advancements in natural language processing. We train and validate two machine learning classifiers that compute the predicted probability of a tweet containing policy commitments and populist rhetoric, respectively. Our approach consists of the standard steps of any classification exercise. First we produce numerical representation of tweets which will serve as a matrix of predictors. Second, we build and annotate a training dataset used to build the classifier. Third, we assess the performance of the classifier and we select the one with the best performance. Fourth, we use the model to make out-of-sample predictions on the universe of tweets collected.

To produce a numerical representation of tweets we use embeddings, fixed-length vectors capturing the semantic properties of the tweet and allowing us to infer similarity between tweets (i.e., tweets with similar vector representations will have a similar meaning). We compute the vectors with the pre-trained Sentence-BERT language model (SBERT) (Reimers & Gurevych 2019), a modification of the pre-trained BERT network that improves the performance of semantic textual similarity tasks for short texts. Sentence-BERT takes as input the text of the tweet – minimally preprocessed – and produces a fixed-length, dense vector which encodes its meaning and semantic properties in a numerical form.[25]

We then construct two training datasets, one for policy commitments and one for populist rhetoric. Tweets containing commitments and populist rhetoric are relatively rare and two random samples of tweets would yield high class imbalance (i.e., very few instances of commitment and populist rhetoric tweets). To ensure a sufficient balance between classes, which is key for the quality of the classifier, we stratify the data and select the 1,500 tweets with the largest cosine similarity between the vector representation of the tweet and the average vector of a set of 30 commitment (populist rhetoric) tweets generated by ChatGPT and other 1,500 random tweets, for a total of 3,000 tweets for each of the two training datasets.

We annotate the tweets with the assistance of ChatGPT. Among the many capabilities of this large language model, Gilardi et al. (2023) have shown that ChatGPT outperforms crowd-workers for several annotation tasks. Moreover, ChatGPT allows to annotate large training datasets in a highly cost-effective fashion. We pass the $3,000 \times 2$ tweets to the model via the OpenAI API, using the `gpt-3.5-turbo` model. The model takes as input a prompt and returns an answer as in an ordinary chat. In Table 2 we show the prompts we used to ask ChatGPT to annotate the tweets and an example of a response. As a further check on the reliability of ChatGPT as an annotation tool, we compare the labels of ChatGPT to those of decided by a human annotator. From the two training datasets, we extract 100 random tweets and asked an independent coder to decide whether the tweets contained commitments or populist rhetoric. The coder received the same prompt given to ChatGPT. Despite the difficulty of the task, the human annotator and ChatGPT agreed on 82% and 84% of the cases for the commitment and populist-rhetoric tasks, respectively.

Finally, we train several machine learning classifiers on the vector representation of 75% of the tweets in the training dataset, with the goal of achieving good predictions on the 25% of held-out tweets. Among the various models we train, we select a gradient boosting classifier based on its good performance metrics. The classification exercises for commitment and populist

---

[25]We use the `all-MiniLM.L6.v2` model from the `SentenceTrasformers` library in Python.

Table 2: ChatGPT prompt and answer example.

Prompt	Answer
I will show you a tweet posted by a candidate to the U.S. Congress. Candidates often commit to specific policies on social media, outlining the proposals or actions that the they intend to take. Does the tweet contain a commitment to a specific policy or a clear promise of action? Reply with yes or no and give a short justification of your response. This is the tweet: *"It's time we invest in a good education for EVERY student. The Kirwan Commission created a blueprint for how we can do this in MD. This week, @SenatorCardin and I are introducing a bill so the federal gov't can partner with states to make these changes."*	Yes, this tweet contains a commitment to a specific policy. The candidate is proposing to invest in a good education for every student, and is referencing the Kirwan Commission's blueprint for how to do this in Maryland. They are also promising to introduce a bill to the federal government to partner with states to make these changes."
I will show you a tweet posted by a candidate to the U.S. Congress. Candidates sometimes use populist rhetoric in their tweets, appealing to the concerns and grievances of the common people against a perceived elite or establishment. Tweets with populist rhetoric therefore either praise the people or attack the elites. Does the tweet contain populist rhetoric? Reply with yes or no and give a short justification of your response. This is the tweet: *"We need an America that puts working people first! Not parties. Not politicians. Not corporations. It's time for someone who works for all the people. It's time for @Nate_McMurray for Congress – NY27 – Leadership matters! SPECIAL ELECTION June 23".*	Yes, the tweet contains populist rhetoric. The tweet appeals to the concerns and grievances of the common people against the perceived elite or establishment by attacking politicians and corporations and emphasizing the need for someone who works for all the people. The use of the phrase "working people" also suggests a focus on the common people.

Notes: Examples of prompts fed to ChatGPT through the OpenAI API and response given by the model.

rhetoric display very good performance metrics, with accuracy and F1 scores above .8.[26] Tweets with a large predicted probability of belonging to the commitment or populist rhetoric classes (i.e., probability larger than .8) are assigned a score equal to 1, those below the probability threshold to 0.

In Figure 3 we show the share of tweets posted by Democratic and Republican candidates during the three months in the run-up to the elections. While Republicans' share of commitment tweets is much larger than Democrats' in 2016 and at the end of the 2012 campaign, Democrats' doubles from approximately 2% in 2012 and 2016 to an average of 4% in 2020. As for populist rhetoric, Republicans and Democrats display a similar share of tweets classified as populist rhetoric, which substantively increase in the 2020 elections.

5.1.3 Topics

We expect the effect of distrust on the supply of commitment to be stronger when the tweet is about topics on which voters' preferences are highly homogeneous. To identify such topics, we use survey data from the Cooperative (Congressional) Election Study (Kuriwaki 2022) to select in a data-driven way the topic with the largest degree of agreement among individuals identifying with the Republican and Democratic party in any given state and year.[27] The

---

[26]More details on the training of the classifier are reported in Section C.

[27]We group respondents at the state-level to ensure a sufficiently large sample size.

Electronic copy available at: https://ssrn.com/abstract=4520441

Figure 3: Share of commitment and populist rhetoric tweets.
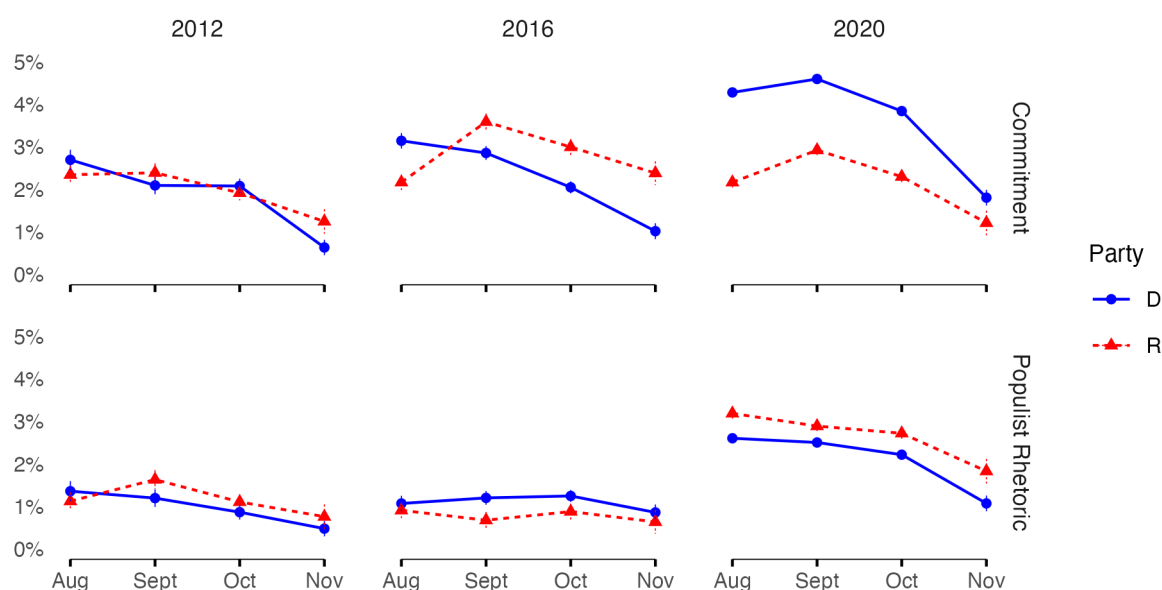
Notes: Share of commitment and populist rhetoric tweets with bootstrapped confidence limits for tweets posted by Republican and Democratic candidates in the September, October, and November (up to election day) of 2012, 2016, and 2020.

CCES study regularly ask a representative sample of U.S. individuals 55 questions about policy preferences (2006-2021). We select the question with the lowest standard deviation in the responses across party-year-state combinations, which we easily map to seven different topics based on the wording of the questions. The seven topics are: immigration, gun regulation, abortion, military issues, environment, health care, and same-sex marriage. For instance, 95.4% of the 813 respondents identifying with the Democratic Party in 2016 in Virginia approve background checks for guns for all sales. The precise wording of the question is *"On the issue of gun regulation, do you support or oppose each of the following proposals? Background checks for all sales, including at gun shows and over the Internet."* The topic we infer from the question is "gun regulation". We then repeat this exercise for both parties over time and across states. In the appendix we report the most homogeneous topics by party-year-state combinations (Table D.6) as well as the mapping of survey questions into topics (Table D.5).

We then build topic-specific dictionaries with the 20 words most similar to the topic label. To learn the semantic relationship between words – and hence select the 20 words semantically most similar to the topic label (e.g., "gun" in the Virginia example) – we train a `word2vec` model (Mikolov et al. 2013) on the 5.9 million tweets we collected, so that the model learns the relationship between words from the language used on Twitter by candidates. The model then produces vector representation of each word and we select the 20 words whose vector representation is closest to the topic label. The words for each of the seven topics/dictionaries

22

are reported in Table D.7 in the appendix.

We consider tweet i to be about topic j, with $J = j_1, ..., j_7$, if tweet i contains at least one of the words in the dictionary of topic j. Finally, if the tweet is about topic j, and topic j is also the most homogeneous topic in the state where the candidate posting the tweet is running, the tweet is considered to be about a homogeneous topic. To continue the example about Democrats in Virginia in 2016, a tweet given by a Democratic candidate on gun regulation in a race for a district in Virginia in 2016 will be considered about a homogeneous topic.

5.2 Evidence on Commitment and Populist Rhetoric

Several candidates ran for office more than once, allowing us to test the relationship between distrust and candidates' rhetoric by exploiting within-candidate variation in distrust and twitter activity. In particular, we estimate the following linear probability model

$$y_{icdt} = \alpha_c + \gamma_d + \eta_{st} + \beta \text{Distrust}_{dt} + \epsilon_{icdt} \tag{3}$$

where y_{icdt} is the probability of a commitment or populist rhetoric tweet, α_c and γ_d are candidate and district fixed effects, and η_{st} are state-by-year fixed effects, to account for time-changing factors at the state level. In line with Corollary **??** and Proposition 3, we expect $\hat{\beta} > 0$. Because β is identified by comparing over-time changes in the twitter activity among candidates who experience a change in the level of distrust in the district where they run, candidates running only once (or twice but not in two of the 2012, 2016, 2020 elections) receive a 0-weight and do not enter the "effective" sample used in the regression (Aronow & Samii 2016). To mitigate the implicit loss of observations that contribute towards the estimation, we also present results without candidate fixed effects and simply conditioning on the party of each candidate. In our preferred specification we also include incumbency status as a covariate. Standard errors are clustered by district.[28]

To precisely match survey-based levels of distrust with Twitter data, we perform the analysis on tweets posted during the electoral campaign, namely during the two months prior to election day (included), in order to precisely match candidates' tweets with the ANES waves.[29]

Besides the positive effect of distrust on the probability of commitment and populist rhetoric tweets, we expect commitments to be more likely when the candidate tweets about topics on which voters' preferences are highly homogeneous (Proposition 4). We therefore expect the average effect of topic homogeneity – i.e., holding distrust fixed – to be positive.

---

[28]The results are almost identical if we cluster the standard errors by candidate.

[29]ANES surveys are conducted in the run-up to the elections, from August to the election day. In Figure E.3, we present results for different windows of time (from 100 to 20 days before the precise election day).

Table 3: Distrust, supply of commitments and populist rhetoric.

	Commitment			Populist Rhetoric		
	(1)	(2)	(3)	(4)	(5)	(6)
Distrust	0.010***	0.012***	0.017***	0.004**	0.005***	0.005**
	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Controls: Party	✓			✓		
Controls: Incumbent	✓		✓	✓		✓
R^2	0.019	0.027	0.029	0.015	0.023	0.023
Observations	274,253	274,253	274,253	274,253	274,253	274,253
District FE	✓	✓	✓	✓	✓	✓
State-Year FE	✓		✓	✓		✓
Candidate FE		✓	✓		✓	✓
Year FE		✓			✓	

Notes: OLS estimates. SE clustered by district. DV is a dummy measure for commitment/populist tweets. Distrust is standardized share of distrustful individuals in each district. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

We test this prediction by adding topic homogeneity to Equation 3, both leaving and removing distrust (to avoid suppression effects (Lenz & Sahn 2021)).

Finally, we estimate a multiplicative interaction model to test whether distrust has a larger effect on commitment when the tweet is about homogeneous topics. Therefore, we estimate the following equation

$$y_{icdt} = \alpha_c + \gamma_d + \eta_{st} + \beta \text{Distrust}_{dt} + \zeta \text{Homogeneous Topic}_i +$$
$$\tau \text{Distrust}_{dt} \times \text{Homogeneous Topic}_i + \epsilon_{icdt} \tag{4}$$

with τ capturing the effect of distrust on commitment for tweets about topics with high homogeneity in the preferences of the constituents of the candidate.

In Table 3 we report the main results on the effect of distrust on commitment and populist rhetoric. The estimates are changes in the probability of commitment and populist tweets resulting from an increase in distrust by +1 standard deviation. A 1 SD increase in the share of voting age individuals who never trust the government in Washington to do what is right is associated with an increase in the probability of a commitment tweet by 1.7 percentage points and of a populist-rhetoric tweet by about half a percentage points when including candidate fixed effects. The effects are small but precisely estimated. Moving from the 10th to 90th percentile in the distribution of distrustful voters (i.e., from 6.4% to 13%) is associated with an increase in the probability of a commitment tweet equal to +5 percentage points, whereas the increase is smaller for the probability of populist-rhetoric tweet, which approximates +1.3

Table 4: Homogeneous topics and supply of commitments.

	Commitment					
	(1)	(2)	(3)	(4)	(5)	(6)
Homog. Topic	0.093***	0.025**	0.089***	0.024**	0.089***	0.024**
	(0.010)	(0.011)	(0.010)	(0.011)	(0.010)	(0.011)
Distrust					0.017***	0.014***
					(0.003)	(0.003)
Controls: Party	✓	✓				
Controls: Incumbent		✓		✓	✓	✓
Topic Dummies		✓		✓		✓
R^2	0.018	0.066	0.029	0.074	0.030	0.074
Observations	255,196	255,196	255,196	255,196	255,196	255,196
District FE	✓	✓	✓	✓	✓	✓
Year FE	✓		✓			
State-Year FE		✓		✓	✓	✓
Candidate FE			✓	✓	✓	✓

Notes: OLS estimates. SE clustered by district. DV is a dummy measure for commitment tweets. Homogeneous topic is a dummy equal to 1 if the tweet's topic matches the most homogeneous topic in any given state-party-year combinations. Distrust is standardized share of distrustful individuals in each district. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

percentage point.[30]

For the supply of commitments, we also detect a positive effect of topic homogeneity (see Table 4). Commitment tweets are about 2.5 percentage points more likely when the tweet is about a topic on which candidates' constituents are highly in agreement with one another, even when holding distrust fixed and including the seven topic dummies.

Finally, distrust also has a much larger effect on the probability of a commitment tweet when the tweet is about a homogeneous topic. In Figure 4 we display the predicted probability of a commitment tweet with varying levels of distrust for homogeneous and non-homogeneous topics. On average, a +10 percentage points in the share of distrustful individuals leads to an increase in the predicted probability of a commitment tweet 18 p.p. larger for tweets about homogeneous topics, compared to the rest of the tweets.[31]

5.3   Evidence on Turnout

When candidates run with commitment platforms, distrustful voters are more likely to turnout (see Proposition 2). We give some empirical support to this prediction looking at the effect of the presidential candidacy of Donald Trump, an uncontroversial example of a presidential candidate

---

[30]In Figure E.3 in the appendix we show how the estimates change when expanding or reducing the window of time analysed.

[31]See Table E.9 in the appendix for full regression table.

Figure 4: Heterogeneous Effects by Homogeneous Topics.
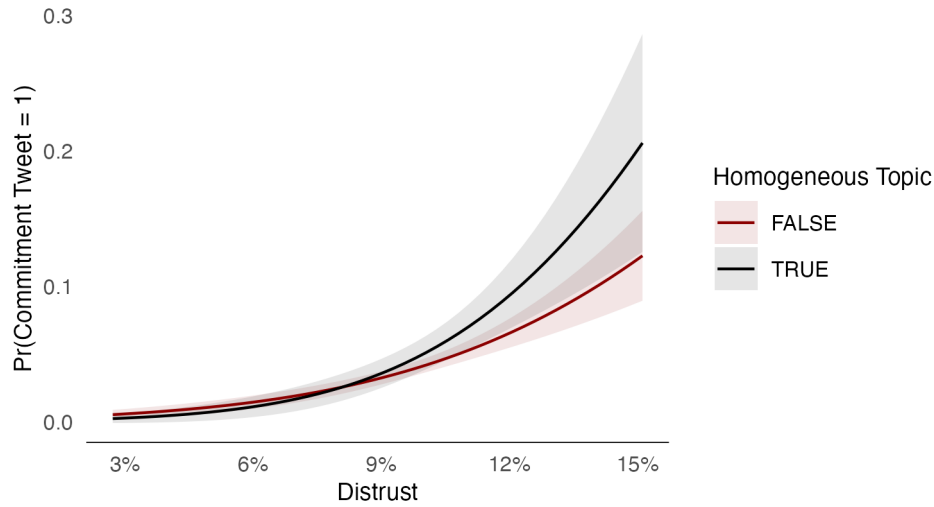
Notes: Predicted probability of a commitment tweet with 95% confidence intervals at different levels of distrust for tweets about homogeneous and non-homogeneous topics. Estimates predicted from a probit model. Difference in predicted conditional probabilities statistically significant at 90% level. Specification includes candidate, state-by-year fixed effects, topic dummies, and incumbency status as a covariate. Standard errors clustered by district.

who employed populist rhetoric and policy commitments (e.g., building a wall, America first, and trade war with China), on the probability of distrustful voters identifying as Republicans to turn out to vote at the presidential elections.

For this test, we use ANES data on self-reported party identification, distrust, and turnout across the three waves/presidential elections used to estimate distrust at the district-level, one without Trump (2012) and two when Trump was running as Republican candidate. ANES self-reported turnout – as it is generally the case for self-reported measures of turnout (Enamorado & Imai 2019) – is markedly higher than actual turnout. In our data it is as high as 86% in 2012, 77% in 2016, and 84% in 2020, against actual figures at 57%, 60%, and 67%.[32]

Our estimand is the effect of the triple interaction between the Trump periods (i.e., election year dummies), self-identifying as a republican, and reporting distrust in the federal government. We estimate the "Trump" effect on the probability of turnout among distrustful Republicans with the following probit model

$$Pr(\text{Turnout}_{idt} = 1) = \Theta(\gamma_d + \delta_t + \gamma_1 \text{Republican}_{idt} + \gamma_2 \text{Distrust}_{idt}$$
$$+ \gamma_3 \text{Republican}_{idt} \times \text{Distrust}_{idt} + \gamma_4 \text{Republican}_{idt} \times \delta_t + \gamma_5 \text{Distrust}_{idt} \times \delta_t$$
$$+ \gamma_6 \text{Republican}_{idt} \times \text{Distrust}_{idt} \times \delta_t + \upsilon \mathbf{X} + \epsilon_{idt})$$

$$(5)$$

---

[32]Data from `https://www.electproject.org/`, accessed on 29 March 2023.

where $Pr(\text{Turnout}_{idt} = 1)$ is self-reported voting of individual i resident in district d in presidential election t. In our preferred specification, we include district fixed effects γ_d, and a set of individual-level covariates $\mathbf{X}$ (i.e., sex, age and race categories, employment status and university education). γ_6 estimates the difference in the probability to turnout in presidential elections for republicans with low trust between the three time periods with and without Trump. We expect this coefficient to be positive for election years 2016 and 2020, when Trump was the republican presidential candidate. Standard errors are clustered by congressional district and we report regression estimates of a linear probability model as a robustness test in the appendix (see Table E.10).

Figure 5 displays the predicted probability of turnout in presidential elections for different categories of respondents. Consistent with an established literature on trust and political participation (Citrin 1974), distrust generally reduces the probability of turnout. In 2012, the predicted probability of turnout for Democrats and Republicans is lower for distrustful respondents. However, in 2016 the relationship flips for republicans with low trust. Although the model does not allow to causally identify the "Trump" effect, the high probability to turnout for distrustful Republicans is strikingly higher compared to that of democrats with similar levels of distrust in the same year, and that of republicans with similar distrust in 2012. Compared to 2012, distrustful republicans are 9 percentage points more likely to turn out to vote at the presidential elections of 2016, during the first Trump candidacy. Interestingly, the effect disappears in the 2020 elections. There could be several reasons why distrustful republicans did not turn out to the same rate in 2020. It is possible that incumbents – who can hardly deliver on all of their previous commitments – do a poorer job of winning distrustful voters, compared to challengers.

6 Important Extensions

In this section we show how the baseline model can be easily generalized to obtain a clear connection with three other important characteristics of populism: aversion to checks and balances, lower competence, and fake-news production.

6.1 Commitment Politics and the Elimination of Checks and Balances

There is ample evidence showing how, once elected, several populist leaders sought to capture (Orbán in Hungary), dismantle (Fujimori in Peru), sabotage (Trump in the US), and reform (Blocher in Switzerland) the state bureaucracy (Bauer & Becker 2020). Sasso & Morelli (2021) and Bellodi et al. (2023) provides theoretical and causal findings in support of populists' pref-

Figure 5: Trump effect on turnout.
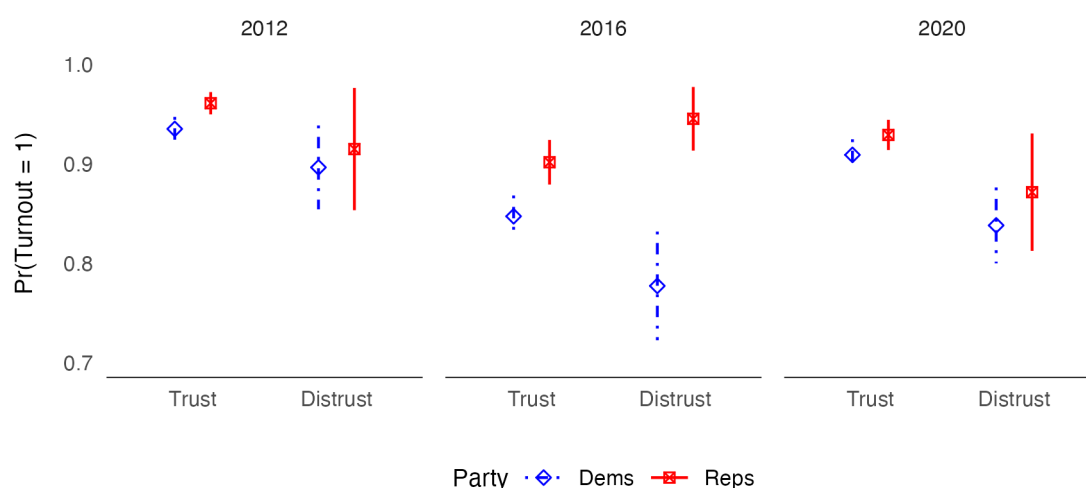
Notes: Predicted probability (with 95% confidence interval) of turnout in the presidential elections for survey respondents self-identifying as Republicans and Democrats with different levels of distrust (Distrust = 1 when respondent some of the time/never trusts the government in Washington to do what is right, and = 0 when respondent always/most of the time/about half of the time trusts the government in Washington to do what is right). Predicted probabilities computed from probit model of respondents' self-reported turnout regressed on a triple interaction Republican × Distrust × Election Year, including a battery of individual-level covariates (age and race categories, university education, employment status, and sex, as well as congressional district fixed effects), with values set to the mean when computing the predicted probability.

erence for less competent bureaucrats, who are less likely to resist the populist agenda. Besides undermining the bureaucracy, populists have often attacked free media and the judiciary, delegitimizing such institutions on the basis of their unelected nature and their interference in government.

To see the consequences of the shift to commitment politics for the survival of free media, independent bureaucracy and/or judiciary, and any other form of checks and balances, consider a simple extension of our baseline principal-agent model (easily extendible further to our consequent electoral competition game). Any "agency of restraint"[33] can be modelled as a player A that can intervene only after the policy-maker has chosen the policy (at time 2 in the original timing). In particular, assuming that z is the status quo and h and l are two possible reforms, we can assume that A can act only as a veto player in case the policy-maker chooses h or l. When the policy is blocked, the status quo z remains. However, whenever the agency wants to block, it succeeds with probability k. What remains to be determined is the criterion determining whether A desires to block or not. The principal assigns a positive probability ρ that the elite manages to influence A.

---

[33]The term "agencies of restraint" can be used to refer to free media, independent public managers and judges, because all of these institutions contribute to create checks for the decision of the executive and balance the executive power in general. In this section we lump them all together in this simple way.

There are two cases in which the elite can distort A's activities:

(i) in case the committed reform is optimal (probability γ_h) and the elite's preferred policy is z (probability λ), with probability ρ the elite manages to influence A and to convince it to intervene to block the reform. The agency is effective in keeping the status quo with probability k.

(ii) in case the status quo is optimal (probability γ_z) and the elite's preferred policy is h (probability $\frac{1-\lambda}{2}$), with probability ρ the lobby manages to influence the agency and to convince it to not intervene.

In case of full commitment and with checks and balances the expected utility for a principal is

$$U_{comm}^{cb} = -\gamma_l - \gamma_z \left(\left(\frac{1-\lambda}{2} \right) \rho + \left(\lambda + \frac{1-\lambda}{2} \right) (1-k) \right) - \gamma_h (\lambda k) \rho. \qquad (6)$$

Let us understand the three terms of the above expected utility under commitment to h when there are checks and balances (cb in short in the superscript). About the first term: when l is optimal the commitment to h always leads to a loss, whether or not the agency intervenes. About the second term: when the status quo is optimal there is a loss when (i) the elite prefers h and manages to influence the agency of restraint to not block the reform and (ii) the elite has no interest to influence the agency of restraint because its preferred policy is either z or l, but the agency does not manage to block the reform (probability $1 - k$). Finally the third term: when h is optimal and the elite prefers z and manages to influence the agency to block the reform.

The derivative of this utility with respect to k tells us if a principal with a committed agent prefers to have more or less checks and balances:

$$\frac{\partial U_{comm}^{cb}}{\partial k} = \gamma_z \left(\lambda + \frac{1-\lambda}{2} \right) - \gamma_h \lambda \rho.$$

Thus, if $\gamma_h > \gamma_z (\frac{1+\lambda}{2\rho\lambda})$, (i.e., the principal believes that the probability that h is optimal is sufficiently larger than the probability that z is optimal), the derivative is negative and the principal prefers to reduce the power of the agency: the higher is ρ and λ, the lower is γ_z the more likely the principal prefers to reduce the agency's power. It follows that populist rhetoric, even if it is limited to affect λ and γ_z, and without explicitly attacking the agency (increase in ρ), induces voters to prefer a reduction of the agency's power.

The expected utility of full delegation is the following. With probability γ_h the optimal policy is the reform h. If the elite's preferred policy is different (probability $\frac{1+\lambda}{2}$), it tries to distort the decision maker' policy (with probability p). A similar reasoning applies when the

29

optimal policy is z and l. The expected utility of the principal under full delegation and without checks and balances is $U_{del} = -p\left((\gamma_h + \gamma_l)\frac{1+\lambda}{2} + \gamma_z(1-\lambda)\right)$.

In presence of the agency, we assume that the elite coordinates its efforts of obtaining the preferred policy trying to influence the agency when it fails to influence the decision maker. In particular, when the optimal policy is either h or l and the elite's preferred policy is z the elite with probability p distorts the agent's decision, but in case it fails (probability $1 - p$), it can still try to influence the agency and block the reform (with probability ρk). In this case the presence of the agency negatively affects the principal's payoff, because it represents a further opportunity for the elite to influence the policy. When, instead, the optimal policy is z and the elite's preferred policy is either l or h, the agency with probability $(1 - \rho)k$ can block the bad reform. In this case the agency limits the bias that the elite determines for the decision maker.

The expected utility of the principal under full delegation and with checks and balances is

$$U_{del}^{cb} = -(\gamma_h + \gamma_l)\left(p\frac{1+\lambda}{2} + \lambda(1-p)\rho k\right) - \gamma_z(1-\lambda)p\left((1-\rho)(1-k) + \rho\right).$$

Similarly to the case of commitment, we can look at the derivative of the utility from full delegation with respect to k:

$$\frac{\partial U_{del}^{cb}}{\partial k} = -(1-\gamma_z)\lambda(1-p)\rho + \gamma_z(1-\lambda)(1-\rho)p,$$

which is positive if

$$\gamma_z \geq \frac{\lambda(1-p)\rho}{\lambda(1-p)\rho + (1-\lambda)(1-\rho)p}.$$

Notice that this threshold goes to zero for p that tends to 1, while the threshold for which more checks and balance increases the utility from full commitment remains positive. Hence, when there is sufficiently large distrust towards politicians, there is a non empty set of parameters $(\rho, \lambda, \gamma_z)$ such that a principal would prefer more checks and balances in case the agent has full delegation, while they would prefer less checks and balances in case of full commitment.

6.2 Incompetence, Fake News, and Information Incentives

In our simple baseline model, the delegated politician sees the realization of the state of the world with probability 1, while the principal never sees the state of the world. Relaxing these two simplifying assumptions yields two more important consequences of the shift to commitment

politics, both consistent with the populism phenomenon in the literature and in our general understanding of it (see also Morelli et al. 2021).

First, the relaxation of perfect understanding of the state of the world by agents yields a clear difference in terms of the desirability of competence: once a principal has chosen a commitment, she doesn't care about competence as much as a principal who chose delegation. Second, if the principal can make a costly effort to figure out the state of the world (to make costlier for the agent to choose the wrong policy), will do so more under delegation, while under commitment it makes no sense. Hence lower information by voters is a consequence rather than a cause of populism.[34]

Moreover, consider a situation in which there is electoral competition between a committed candidate and a fully delegated candidate. Since in case the fully delegated candidate wins the election voters in equilibrium make costly effort to figure out the state of the world, the committed candidate benefits from making costlier for the voters such information acquisition, for instance spreading fake news. In this way, the committed candidate reduces the expected benefits of electing a fully delegated candidate.

7   Concluding Remarks

In this paper we have proposed a novel logical mechanism that links the erosion of trust with the observed populism wave. This mechanism is the rational increase in the demand and supply of policy commitments. More distrust induces voters, and consequently politicians, to shift from the full delegation model of representative democracy to a preference for delegates with tighter commitments, to building walls and closing borders, to anti-immigration or protectionist policies, or other forms of identity protection. In turn, such a shift to commitment politics produces a rational adoption of anti-elite rhetoric, which makes candidates and parties be defined as populists by most recognized definitions and categorizations of populism. We have documented this logical mechanism with a simple principal agent model and with the consequent electoral competition analogue, and we have corroborated our theory with novel evidence on U.S. Congressional candidates and campaigns.

Majoritarian electoral systems might be an important scope condition of the evidence presented in this paper. Clearly an extension of the empirical analysis to different levels of elections in countries with different electoral systems would be harder, especially in the absence of clear geographic representation. However, there is plenty of anecdotal evidence that confirms the importance of the shift to commitment politics even in countries with mixed electoral systems.

---

[34]See Boffa et al. (2023) for another channel linking incompetence and populist support.

Berlusconi's contract with Italians – presented on national television five days before the 2001 general elections – as well as Lega's leader Matteo Salvini's commitments to closed borders are consistent with the idea that commitment politics is not unique to the U.S. or to majoritarian democracies, but it grows where distrust flourishes. Similarly, strategic commitments have been a key feature of the populist strategy of the Five-Star Movement (5SM) in Italy, with the most important being the simple commitment to citizenship income. Evidently the demand of commitments had grown in spite of the fear of partial credibility, and hence the 5SM tried to strengthen credibility by means of other ancillary commitments: to give back to the people part of their salary in case of election; to serve for only two terms (despite no such limit exists in the Italian constitution);[35] to cut substantially the number of parliamentarians; and, most important, a commitment to never renege on the electoral promises made during the campaign, under penalty of expulsion from the party (the so-called *vincolo di mandato*, whose literal translation is "binding mandate").

It is important to note that in this paper we have shown the logic and implications of the growing demand and supply of policy commitments holding constant their credibility, and indeed the empirical evidence represent a conservative test of our theory, given the fact that, in reality, people do believe every commitment and promise to be credible. We leave the dynamics between the credibility and the fulfilment of such commitments, as well as candidates' ex-ante (ex-post) attempts to increase the perception of credibility (fulfilment) of commitments to future research.

Finally, we have shown that the shift to commitment politics is not only a mechanism that determines a rationalization of the rhetorical component of populism, but can also explain the crisis of representative democracy in terms of preferences for reduction of checks and balances and for other illiberal reforms. Future research will have to investigate the dynamic consequences and conditions for the persistence versus reversibility of the shift to commitment politics and its consequences.

---

[35]See the strong emphasis on term limits at `https://beppegrillo.it/circonvenzione-di-elettore/`.

References

Abeler, J., Becker, A. & Falk, A. (2014), 'Representative evidence on lying costs', *Journal of Public Economics* **113**, 96–104.
URL: *http://dx.doi.org/10.1016/j.jpubeco.2014.01.005*

Acemoglu, D., Egorov, G. & Sonin, K. (2013), 'A Political Theory of Populism', *Quarterly Journal of Economics* pp. 771–805.

Acemoglu, D. & Restrepo, P. (2020), 'Robots and jobs: Evidence from us labor markets', *Journal of Political Economy* **128**(6), 2188–2244.

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

Alesina, A. (1988), 'Credibility and Policy Convergence in a Two-Party System with Rational Voters', *American Economic Review* **78**(4), 796–805.
URL: *https://econpapers.repec.org/RePEc:aea:aecrev:v:78:y:1988:i:4:p:796-805*

Alesina, A. & Rosenthal, H. (1996), 'A Theory of Divided Government', *Econometrica* **64**(6), 1311–1341.
URL: *https://www.jstor.org/stable/2171833*

Algan, Y., Guriev, S., Papaioannou, E. & Passari, E. (2017), 'The European Trust Crisis and the Rise of Populism', *Brookings Papers on Economic Activity* (Fall), 309–382.
URL: *https://www.jstor.org/stable/10.2307/90019460*

Alsan, M., Braghieri, L., Eichmeyer, S., Joyce Kim, M., Stantcheva, S. & Yang, D. Y. (n.d.), 'Civil Liberties in Times of Crisis', *American Economic Journal: Applied Economics* .

Amador, M., Werning, I. & Angeletos, G.-M. (2006), 'Commitment vs. Flexibility', *Econometrica* **74**(2), 365–396.

Aronow, P. M. & Samii, C. (2016), 'Does Regression Produce Representative Estimates of Causal Effects?', *American Journal of Political Science* **60**(1), 250–267.

Autor, D., Dorn, D., Hanson, G. & Majlesi, K. (2020), 'Importing political polarization? The electoral consequences of rising trade exposure', *American Economic Review* **110**(10), 3139–3183.

Bauer, M. W. & Becker, S. (2020), 'Democratic Backsliding, Populism, and Public Administration', *Perspectives on Public Management and Governance* **3**(1), 19–31.

Bellodi, L., Morelli, M. & Vannoni, M. (2023), 'A Costly Commitment: Populism, Government Performance, and the Quality of Bureaucracy', *American Journal of Political Science* pp. 1–17.

Berman, S. (2021), 'The Causes of Populism in the West', *Annual Review of Political Science* **24**(1), 1–18.

Besley, T. & Persson, T. (2019), 'Democratic Values and Institutions', *American Economic Review: Insights* **1**(1), 59–76.

Boffa, F., Mollisi, V. & Ponzetto, G. A. (2023), Do Incompetent Politicians Breed Populist Voters? Evidence from Italian Municipalities.
URL: *https://cepr.org/system/files/publication-files/DP17997.pdf*

Bonomi, G., Gennaioli, N. & Tabellini, G. (2021), 'Identity, Beliefs, and Political Conflict', *Quarterly Journal of Economics* pp. 2371–2411.

Broniecki, P., Leemann, L. & Wüest, R. (2022), 'Improved Multilevel Regression with Post-stratification through Machine Learning (autoMrP)', *Journal of Politics* **84**(1), 597–601.

Bueno de Mesquita, E. & Friedenberg, A. (2011), 'Ideologues or Pragmatists?', *Journal of the European Economic Association* **9**(5), 931–951.

Caillaud, B. & Tirole, J. (2002), 'Parties as Political Intermediaries', *Quarterly Journal of Economics* pp. 1453–1489.

Callander, S. & Wilkie, S. (2007), 'Lies, damned lies, and political campaigns', *Games and Economic Behavior* **60**(2), 262–286.

Carrillo, J. D. & Dewatripont, M. (2008), 'Promises, Promises, . . .', *Economic Journal* **118**, 1453–1473.

Cha, J., Kuriwaki, S. & Snyder, J. M. (2021), 'Candidates in American General Elections', *Harvard Dataverse* .

Citrin, J. (1974), 'Comment: The Political Relevance of Trust in Government', *American Political Science Review* **68**(3), 973–988.
URL: *https://www.jstor.org/stable/1959141*

Colantone, I. & Stanig, P. (2018), 'The Trade Origins of Economic Nationalism: Import Competition and Voting Behavior in Western Europe', *American Journal of Political Science* **62**(4), 936–953.

Corazzini, L., Kube, S., Maréchal, M. A. & Nicolò, A. (2014), 'Elections and deceptions: An experimental study on the behavioral effects of democracy', *American Journal of Political Science* **58**(3), 579–592.

Crutzen, B. S. Y., Sisak, D. & Swank, O. H. (2020), Left Behind Voters, Anti-Elitism and Popular Will.

Dai, Y. & Kustov, A. (2022), 'When Do Politicians Use Populist Rhetoric? Populism as a Campaign Gamble', *Political Communication* **39**(3), 383–404.
URL: *https://doi.org/10.1080/10584609.2022.2025505*

Di Cocco, J. & Monechi, B. (2021), 'How Populist are Parties? Measuring Degrees of Populism in Party Manifestos Using Supervised Machine Learning', *Political Analysis* pp. 1–17.

Di Tella, R. & Rotemberg, J. J. (2018), 'Populism and the return of the "Paranoid Style": Some evidence and a simple model of demand for incompetence as insurance against elite betrayal', *Journal of Comparative Economics* **46**(4), 988–1005.
URL: *https://doi.org/10.1016/j.jce.2018.03.001*

Dornbusch, R. & Edwards, S. (1991), The Macroeconomics of Populism in Latin America, *in* R. Dornbusch & S. Edwards, eds, 'The Macroeconomics of Populism in Latin America', University of Chicago Press, Chicago.

Duggan, J. & Martinelli, C. (2017), 'The political economy of dynamic elections: Accountability, commitment, & responsiveness', *Journal of Economic Literature* **55**(3), 916–984.

Egorov, G. & Sonin, K. (2021), 'Elections in Non-Democracies', *Economic Journal* **131**(636), 1682–1716.

Enamorado, T. & Imai, K. (2019), 'Validating Self-Reported Turnout by Linking Public Opinion Surveys with Administrative Records', *Public Opinion Quarterly* **83**(4), 723–748.

Esteban, J.-M. & Ray, D. (1994), 'On the Measurement of Polarization', *Econometrica* **62**(4), 819–851.
URL: *http://www.jstor.org/stable/2951734*

Ford, R. & Jennings, W. (2020), 'The Changing Cleavage Politics of Western Europe', *Annual Review of Political Science* **23**, 295–314.

Fox, J. & Shotts, K. W. (2009), 'Delegates or trustees? A theory of political accountability', *Journal of Politics* **71**(4), 1225–1237.

Funke, M., Schularick, M. & Trebesch, C. (2020), Populist Leaders and the Economy.
URL: *http://hdl.handle.net/10419/226836*

Gelman, A. & Little, T. C. (1997), 'Poststratification into many categories using hierarchical logistic regression', *Survey Methodology* **23**(2), 127–135.

Gennaioli, N. & Tabellini, G. (2023), Identity Politics.
URL: *https://ssrn.com/abstract=4410239*

Gennaro, G., Lecce, G. & Morelli, M. (2021), Mobilization and the Strategy of Populism Theory and Evidence from the United States.
URL: *https://cepr.org/active/publications/discussion_papers/dp.php?dpno=15686*

Ghosh, B. S. & Tripathi, V. (2012), 'Ideologues Beat Idealists', *American Economic Journal: Microeconomics* **4**(2), 27–49.

Gilardi, F., Alizadeh, M. & Kubli, M. (2023), ChatGPT Outperforms Crowd-Workers for Text-Annotation Tasks.
URL: *http://arxiv.org/abs/2303.15056*

Gneezy, U., Kajackaite, A. & Sobel, J. (2018), 'Lying Aversion and the Size of the Lie', *American Economic Review* **108**(2), 419–453.

Graham, M. H. & Svolik, M. W. (2020), 'Democracy in America? Partisanship, Polarization, and the Robustness of Support for Democracy in the United States', *American Political Science Review* **114**(2), 392–409.

Gratton, G. & Lee, B. E. (2023), Drain the Swamp: A Theory of Anti-Elite Populism.

Gratton, G. & Morelli, M. (2022), 'Optimal Checks and Balances Under Policy Uncertainty', *International Economic Review* **63**(2), 549–569.

Guiso, L., Morelli, M., Sonno, T. & Herrera, H. (2021), The Financial Drivers of Populism in Europe.
URL: *http://dx.doi.org/10.2139/ssrn.3922661*

Guriev, S. & Papaioannou, E. (2022), 'The Political Economy of Populism', *Journal of Economic Literature* **60**(3), 753–832.

Hawkins, K. A., Aguilar, R., Castanho Silva, B., Jenne, E. K., Kocijan, B. & Rovira Kaltwasser, C. (2019), Measuring Populist Discourse: The Global Populism Database.
URL: *https://populism.byu.edu/App_Data/Publications/Global Populism Database Paper.pdf*

Kartik, N., Van Weelden, R. & Wolton, S. (2017), 'Electoral Ambiguity and Political Representation', *American Journal of Political Science* **61**(4), 958–970.

Kasara, K. & Suryanarayan, P. (2015), 'When Do the Rich Vote Less Than the Poor and Why? Explaining Turnout Inequality across the World', *American Journal of Political Science* **59**(3), 613–627.

Kuriwaki, S. (2022), 'Cumulative CCES Common Content'.
URL: *https://doi.org/10.7910/DVN/II2DB6*

Laitin, D. D. (2018), 'Populism: the Muslim Connection', *Keynote Address at the CESifo Conference "Globalization and Populism - Past and Present"* .
URL: *http://laitinresearch.stanford.edu/wp-content/uploads/2013/10/PopulismCESkeynoteDL.pdf*

Lenz, G. S. & Sahn, A. (2021), 'Achieving Statistical Significance with Control Variables and Without Transparency', *Political Analysis* **29**(3), 356–369.

Levy, G. (2004), 'A model of political parties', *Journal of Economic Theory* **115**(2), 250–277.
URL: *http://dx.doi.org/10.1016/S0022-0531(03)00254-0*

Levy, G. (2007), 'Decision making in committees: Transparency, reputation, and voting rules', *American Economic Review* **97**(1), 150–168.

Levy, G., Razin, R. & Young, A. (2022), 'Misspecified Politics and the Recurrence of Populism', *American Economic Review* **112**(3), 928–62.
URL: *https://personal.lse.ac.uk/YoungA/Populism_Jan27.pdf*

Mikolov, T., Chen, K., Corrado, G. & Dean, J. (2013), 'Efficient estimation of word representations in vector space', *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings* pp. 1–12.

Morelli, M., Nicolò, A. & Roberti, P. (2021), A Commitment Theory of Populism.
URL: *https://cepr.org/publications/dp16051*

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

Mudde, C. & Rovira Kaltwasser, C. (2018), 'Studying Populism in Comparative Perspective: Reflections on the Contemporary and Future Research Agenda', *Comparative Political Studies* **51**(13), 1667–1693.
URL: *https://doi.org/10.1177/0010414018789490*

Norris, P. & Inglehart, R. (2019), *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*, Cambridge University Press, Cambridge.

Posner, D. N. (2004), 'Measuring ethnic fractionalization in Africa', *American Journal of Political Science* **48**(4), 849–863.

Reimers, N. & Gurevych, I. (2019), 'Sentence-BERT: Sentence embeddings using siamese BERT-networks', *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference* pp. 3982–3992.

Rodriguez, P. L. & Spirling, A. (2022), 'Word embeddings: What works, what doesn't, and how to tell the difference for applied research', *Journal of Politics* **84**(1922658), 101–115.

Rodrik, D. (2018), 'Populism and the economics of globalization', *Journal of International Business Policy* **1**(1-2), 12–33.
URL: *https://doi.org/10.1057/s42214-018-0001-4*

Rooduijn, M., Van Kessel, S., Froio, C., Pirro, A., De Lange, S., Halikiopoulou, D., Lewis, P., Mudde, C. & Taggart, P. (2019), 'The PopuList: An Overview of Populist, Far Right, Far Left and Eurosceptic Parties in Europe'.
URL: *www.popu-list.org*

Sasso, G. & Morelli, M. (2021), 'Bureaucrats under Populism', *Journal of Public Economics* **202**, 104497.
URL: *https://doi.org/10.1016/j.jpubeco.2021.104497*

Shayo, M. (2009), 'A model of social identity with an application to political economy: Nation, class, and redistribution', *American Political Science Review* **103**(2), 147–174.

Smart, M. & Sturm, D. M. (2013), 'Term limits and electoral accountability', *Journal of Public Economics* **107**, 93–102.
URL: *http://dx.doi.org/10.1016/j.jpubeco.2013.08.011*

Snyder, J. M. & Ting, M. M. (2002), 'An Informational Rationale for Political Parties', *American Journal of Political Science* **46**(1), 90.

Szeidl, A. & Szucs, F. (2021), 'Media Capture Through Favor Exchange', *Econometrica* **89**(1), 281–310.

Tausanovitch, C. & Vavreck, L. (2021), 'Democracy Fund + UCLA Nationscape Project, October 10-17, 2019 (version 20211215)'.
URL: *https://www.voterstudygroup.org/data/nationscape*

Van Weelden, R. (2013), 'Candidates, credibility, and re-election incentives', *Review of Economic Studies* **80**(4), 1622–1651.

Warshaw, C. & Rodden, J. (2012), 'How should We measure district-level public opinion on individual issues?', *Journal of Politics* **74**(1), 203–219.

Weyland, K. (2001), 'Clarifying a Contested Concept: Populism in the Study of Latin American Politics', *Comparative Politics* **34**(1), 1–22.
URL: *https://www.jstor.org/stable/422412*

Weyland, K. (2017), Populism : A Political-Strategic Approach, *in* C. R. Kaltwasser, P. Taggart, P. Ochoa Espejo & P. Ostiguy, eds, 'The Oxford Handbook of Populism', Oxford University Press.

Appendix

A Theory Proofs

Proof of Corollary 1. The signs of the derivative of p_1 with respect to γ_z and λ follow from inspection of the expression for p_1. Now consider the derivative of p_2 with respect to γ_z. The derivative of the denominator is $-\lambda + (1-\lambda)/2$, which is negative. The derivative of the numerator crucially depends on additional assumptions of how an increase in γ_z affects γ_l and γ_h. We can assume that γ_h becomes lower or is not affected by an increase in γ_z. In either scenario the derivative of p_2 with respect to γ_z is positive. The derivative of denominator of p_2 with respect to λ is $1 - \gamma_z - 1/2 - \gamma_z/2 = 1/2 - \frac{3}{2}\gamma_z$ which is larger than zero if $\gamma_z \leq \frac{1}{3}$. ∎

Proof of Lemma 1. In the primary elections members of party G vote for the candidate with the electoral program that maximizes their expected utility, taking into account the probability of winning at the general election. We therefore have to compute the probability of winning of a candidate at the general election. Consider a citizen v in group G who must choose between voting or not voting, when the two candidates in the general elections run with electoral programs e and e'. Let us assume, without loss of generality, that program e gives to citizen v in group G a higher utility from policy, among the two electoral programs: $EU^G(e) > EU^G(e')$. Citizens choose their voting action as if pivotal. Then citizen v in group G votes for the candidate running with e if $EU^G(e) - c_v^G - \mu^G > EU^G(e')$, otherwise she abstains. It follows that citizen v with cost of voting below $\underline{c}_v^G := EU^G(e) - EU^G(e') - \mu^G$ votes (for e). If without loss of generality $EU^G(e) > EU^G(e')$ and $EU^{-G}(e) < EU^{-G}(e')$, before the general elections and after discovering μ^G, for group G the probability that candidate running with e wins is

$$
\mathbb{P}\left(\underline{c}_v^G > \underline{c}_v^{-G}\right) = \mathbb{P}\left(\mu^{-G} > EU^{-G}\left(e'\right) - EU^{-G}(e^g) - EU^G(e) + EU^G\left(e'\right) + \mu^G\right) =
$$
$$
\frac{1}{2} + \psi\left(-EU^{-G}\left(e'\right) + EU^{-G}(e^g) + EU^G(e) - EU^G\left(e'\right) - \mu^G\right), \quad (7)
$$

that is, more voters show up for candidate running with program e than for the candidate running with program e' in the general elections. In this case, the expected probability of victory of candidate g at the time of primaries is

$$
\frac{1}{2} + \psi\left(-EU^{-G}\left(e'\right) + EU^{-G}(e^g) + EU^G(e) - EU^G\left(e'\right)\right), \quad (8)
$$

because μ^G has mean 0. In the primaries of party G, depending on the program, there are citizens who will vote in the general elections, and citizens who will abstain. Assume that a candidate wins the primaries with program e. If this candidate also wins the general elections, each citizen v in the first subgroup receives expected utility $EU^G(e) - c_v$, while each citizen v

A1

in the second subgroup receives expected utility $EU^G(e)$. Notice that, if $\psi \to 0$, the probability of winning approaches $\frac{1}{2}$. In this case, when the programs selected in the two primaries are e and e', and $EU^G(e) > EU^G\left(e'\right)$ and $EU^{-G}(e) < EU^{-G}\left(e'\right)$, the expected utility of a voting citizen v in group G is $\frac{1}{2}\left(EU^G(e) - c_v\right) + \frac{1}{2}EU^G\left(e'\right)$. The expected utility of an abstaining citizen v in group G is $\frac{1}{2}EU^G(e) + \frac{1}{2}EU^G\left(e'\right)$. For both subgroups of citizens, the highest expected utility is achieved by the program that maximizes $EU^G(e)$. Given that candidates maximize their probability of winning the primaries, they all propose such electoral program, which is the equilibrium electoral program of the primary. By the discrete nature of the electoral programs, this is true also if ψ is positive and below a given threshold $\underline{\psi}$. ∎

Proof of Proposition 3. Notice that the utility from partial commitment decreases with γ_z^G and increases with λ. Moreover the utility from full commitment decreases with γ_z^G. Finally the utility from delegation increases with γ_z^G because $\lambda \geq \frac{1}{3}$ and decreases with λ because $\gamma_z^G \leq \frac{1}{3}$. Given that that the turnout of a candidate with electoral program e against a candidate with electoral program e' depends positively on the difference $EU^G(e) - EU^G(e')$, the proof follows immediately. ∎

Proof of Proposition 4. Consider party A. Notice that p_1 weakly decreases with γ_h^A, if an increase in γ_h^A negatively affects γ_z^A. Similarly p_2 decreases with γ_h^A, with a primary effect going through the denominator, and a secondary effect through a possible decrease in γ_z^A in the denominator. The same analysis can be done with party B and γ_l^B. Thus by Proposition 1 and Lemma 1 the first of Proposition 4 follows. The second part of the Proposition follows by inspection of the utility from full commitment. ∎

A2

B District-level Estimates of Distrust in Government

To build district-level estimates of distrust we first download survey data from the ANES Data Center. ANES releases surveys every 4 years. We focus on the last three waves (i.e., 2012, 2016, 2020) so that we can match the estimates with the Twitter activity of candidates, which was minimal in the 2008 elections.

To measure distrust we focus on the following survey question: "How often can you trust the federal government in Washington to do what is right?". The possible answers are reported below. We dichotomize the variable so that distrust = 1 if the respondent *never* trusts the government.

- -9. Refused
- -8. Don't know
- 1. Always
- 2. Most of the time
- 3. About half the time
- 4. Some of the time
- **5. Never**

In 2012 some respondents were asked an additional question with a slightly different wording ("How much of the time do you think you can trust the government in Washington to do what is right?) with 4 instead of 5 possible answers ("About half of the time was not included as a possible answer"). Respondents were asked either the 5-scale or the 4-scale question. In both cases, we code distrust = 1 if the answer is *never*. Consistent with common practice, we code as 0 respondents who answered "Don't know" or who refused to answer.

This question is particularly suitable for the test we aim to perform. First, it explicitly asks respondents about *trust*, without resorting to other proxies or other similar but distinct concepts like confidence of competence. Second, it refers to the federal government in Washington, hence respondents are prompted to think about the same entity when answering the question. In fact, if they were asked about trust in politicians or trust in government in general, respondents might have been thinking about their state or local government, thus undermining the comparability of the answers. Clearly, we do not need to assume no interference between respondents' reported trust in federal versus local government. We take the response at face value and we use district and state random effects, as well as other covariates at district level, to account for differences across districts.

As for the choice of individual-level predictors to use both in multilevel model and for post-stratification we select sex, age, and education categories, for they are at the same time

A3

good predictors of distrust and the only variables for which we know the joint distribution at the district level from the U.S. Census. These are also the individual-level predictors commonly used in other MRP applications. At state level researchers are generally able to include a fourth predictor for race, but the U.S. Census reports data by age × sex × education categories only.

To facilitate greater pooling across districts we include in the model several district-level variables that are plausibly correlated with distrust. We include the share of black population, the Gini index, median income, employment rate, and the share of votes for the Republican presidential candidate in the same year at the state level. We include also random effects for the state.

The strength of the individual level predictors is evidenced by the regression estimates reported below. We fit linear probability models following different specifications to show that the individual level variables chosen for the multilevel model are good predictors of distrust and are robust to a series of different specifications.

Table B.1: Individual Predictors of Distrust.

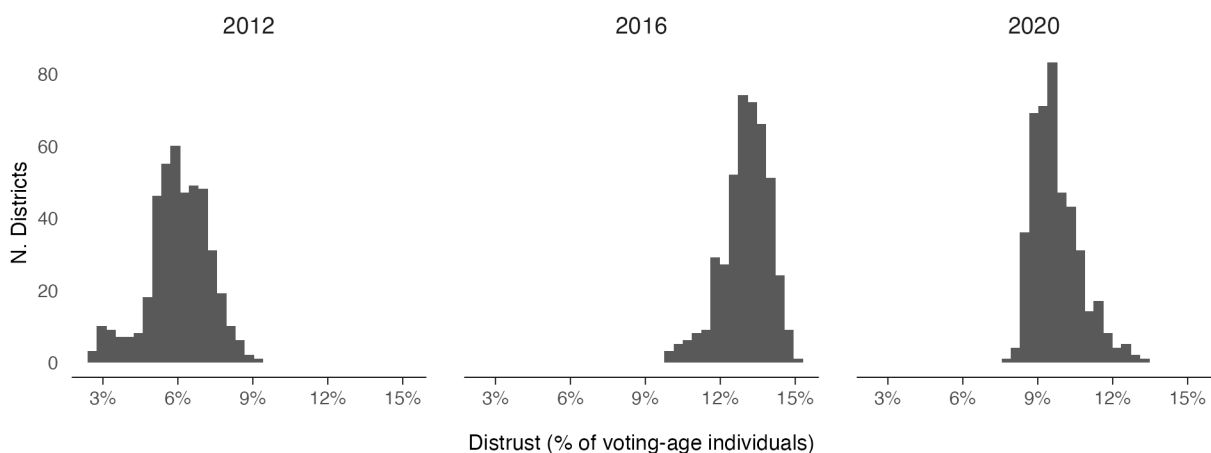|	Distrust			
	(1)	(2)	(3)	(4)
Sex Category [1-2]	-0.018***	-0.020***	-0.021***	-0.021***
	(0.004)	(0.005)	(0.004)	(0.005)
Age [1-5]	-0.014***	-0.014***	-0.014***	-0.016***
	(0.002)	(0.002)	(0.002)	(0.002)
Education [1-5]	-0.009***	-0.008***	-0.011***	-0.011***
	(0.002)	(0.002)	(0.002)	(0.002)
District-level Covariates				
% Black Population			-0.319**	
			(0.125)	
Gini Index			0.402	
			(0.450)	
Median Income			0.013	
			(0.012)	
Employment Rate			0.340	
			(0.301)	
% of Republican Votes (State level)			0.057	
			(0.119)	
R^2	0.006	0.034	0.043	0.095
Observations	16,835	16,835	16,835	16,835
RMSE	0.282	0.278	0.276	0.269
District FE		✓	✓	
Wave FE			✓	
District × Year FE				✓

Notes: OLS estimates. SE clustered by congressional district. DV is a dummy which equals 1 when the respondet said they never trust the government in Washington to do what is right.

We then perform the MRP in two steps. First, we fit a multilevel model to the survey data including a set of random effects, J education categories, K sex categories, M age categories, N districts, and Z states. X is a matrix of predictors that vary at the district-level. We can U.S. citizens as consisting of these different ideal types.

$$
\begin{aligned}
Pr(Distrust_i = 1) = \Phi(\beta_0 + \alpha^{education}_{j[i]} + \alpha^{sex}_{k[i]} + \alpha^{age}_{m[i]} + \alpha^{district}_{n[i]} + \alpha^{state}_{z[i]} + \mathbf{X}^T_{n[i]}\beta) \\
\alpha^{education}_{j[i]} \sim N(0, \sigma^2_{education}), \text{for} j = 1, ...J \\
\alpha^{sex}_{k[i]} \sim N(0, \sigma^2_{sex}), \text{for} k = 1, ...K \\
\alpha^{age}_{m[i]} \sim N(0, \sigma^2_{age}), \text{for} m = 1, ...M \\
\alpha^{district}_{n[i]} \sim N(0, \sigma^2_{district}), \text{for} n = 1, ...N \\
\alpha^{state}_{z[i]} \sim N(0, \sigma^2_{state}), \text{for} z = 1, ...Z
\end{aligned}
\tag{9}
$$

In the second step, we calculate the predicted share of distrustful for each ideal type based on Equation 9, which we denote $\hat{\pi}_{jkmnz}$. We then use post-stratification and calculate the weighted average of $\hat{\pi}_{jkmnz}$ based on the *actual* prevalence of that ideal type in the population of the district. Since the predictions are not linear in the random effects, we need census data on the joint distribution of education, sex, and age in each district. These steps are performed semi-automatically through the `AutoMrP` package in R, which uses an ensemble modelling approach that provides better small-area estimates (Broniecki et al. 2022). We produce estimates of distrust for 436 congressional districts (including the District of Columbia) over three waves, estimated separately for each year. In Figure B.1 we report the distribution of our estimates of distrsuts over time.

Figure B.1: Distribution of distrust estimates over time.

Notes: Distribution of estimates of distrust in government for each congressional districts over the three waves of survey data.

B1 Comparing ANES with NationScape MRP Estimates

To validate our MRP estimates, we show that the district-level estimates that can be obtained from the relatively small ANES sample are comparable to those obtained from much larger samples.

We replicate the MRP method for the year 2020 with survey data from NationScape, an online, weekly survey with approximately 10,000 interviews per week conducted in the period 2019-2021. We keep a large random sample of 50,000 responses recorded before the 2020 election day for which there is information on the congressional district of the respondent, their sex, age, and level education. NationScape does not include questions on trust in government, therefore we cannot directly compare the estimates of distrust used in the analysis. However, we can compare the estimates for two other political questions, one on the perceived state of the economy and one on political interest, which are included both in ANES and NationScape. Table B.2 report the question prompts with the possible answers.

Table B.2: Survey questions, ANES and NationScape.

ANES	NationScape
Now thinking about the economy in the country as a whole, would you say that over the past year the nation's economy has gotten better, stayed about the same, or gotten worse? -9. Refused -8. Don't know 1. Gotten better 2. Stayed about the same **3. Gotten worse**	Would you say that as compared to one year ago, the nation's economy is now better, about the same, or worse? 1. Better 2. About the same **3. Worse** . Respondent Skipped
Sample mean: .61	Sample mean: .48
How often do you pay attention to what's going on in government and politics? -9. Refused **1. Always** **2. Most of the time** 3. About half the time 4. Some of the time 5. Never	Some people follow what's going on in government most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on... **1. Most of the time** 2. Some of the time 3. Only now and then 4. Hardly at all . Respondent Skipped
Sample mean: .63	Sample mean: .43
Respondents: 7,093	Respondents: 414,318

Notes: Question wording and possible answers in ANES and NationScape surveys. Answers re-coded to obtain dichotomous responses, values in bold equal to 1 and 0 otherwise.

As shown in the table, the average responses are quite different in the two surveys. How-

ever, if the size of the ANES survey is large enough, we should observe a strong correlation between the estimates derived from the large NationScape sample and the ANES sample. Figure B.2 shows there is a positive and strong correlation between the estimates. ANES estimates are overall larger than NationScape estimates, consistently with the larger sample means in the ANES sample, but the correlation is very high equal or greater than .8 for both variables. Similarly, the standard error of a linear regression of the NationScape estimates on the ANES estimates is very low. This test suggests that the ANES sample yields estimates comparable to those obtained from a sample size ×7 larger.

Figure B.2: MRP Estimates from ANES and NationScape
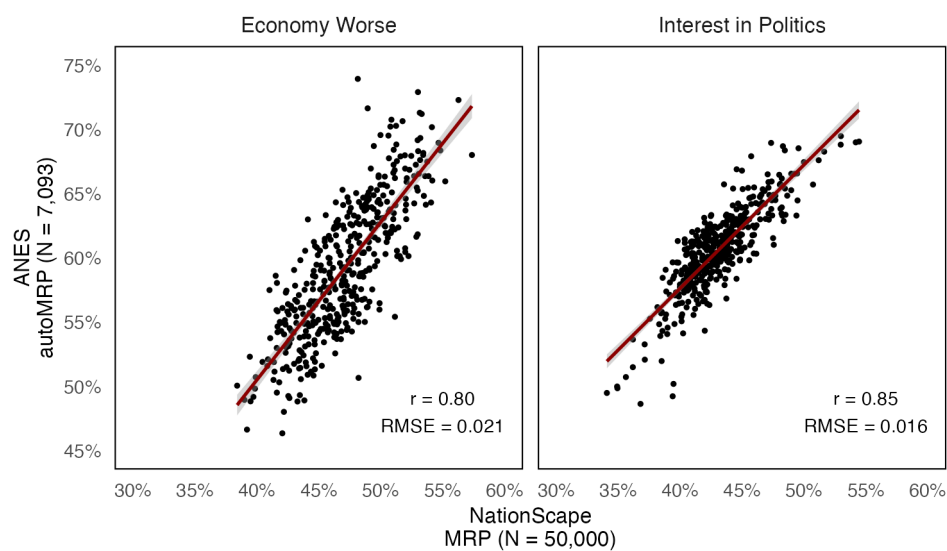
Notes: On the horizontal axis, estimated share of NationScape respondents who believe the economy in the past year has worsened (panel on the left) and who follow what is going on in government most of the times. On the vertical axis, estimated share of ANES respondents for comparable questions.

A7

C   Measuring Commitment and Populist Rhetoric in Tweets

For any natural language processing task, we need a numerical representation of text. Here we build on recent advancements in deep learning and use the pre-trained Sentence-BERT language model (SBERT) (Reimers & Gurevych 2019), a modification of the pre-trained BERT network that improves the performance of semantic textual similarity tasks for short texts. Sentence-BERT takes as input the text of the tweet – minimally preprocessed – and produces a fixed-length, dense vector which encodes its meaning and semantic properties in a numerical form. As a preprocessing step, we simply remove urls and line-break syntax. Producing an embedding for 5.9 million tweets would consume a significant amount of computational resources and may take considerable time and — since we are using a pre-trained model — we produce vector representation of tweets posted in the three elections we studied (i.e., 2012, 2016, 2020). Once we have an embedding for every tweet, we assemble the training dataset on which to train the classifier.

For a classifier to make good predictions, the classes in which documents are labelled by the annotators (i.e., the same categories predicted by the classifier) need to be balanced, namely we should have a similar number of texts in category A and category B (for a 2-class classification task). Classifiers learn better from a balanced distribution of classes in a dataset, hence we ensure balance with the most common technique: oversampling texts in the minority category. Clearly, populist rhetoric and policy commitment tweets are the two minority classes.

To ensure class balance, we first compute the embedding for a "commitment" and "populist rhetoric" poles, given by the the average embedding of 30 examples of tweets containing policy commitment and populist rhetoric generated by ChatGPT. We gave ChatGPT two prompts, one for populist rhetoric and one for policy commitments. For the populist rhetoric prompt, we simply asked "Write me 30 examples of statements given by politicians using anti-elite, populist rhetoric." and we found it to work very well, for populist rhetoric is clear as a concept (although it might be harder to detect it in real text). Conversely, for the commitment prompt we were more specific, for commitment can be understood in multiple ways, not necessarily with regard to a specific policy or action. Hence we prompted ChatGPT with the following command: "Candidates often commit to specific policies on social media, outlining the proposals or actions that they intend to take. Write 30 examples of tweets containing a commitment to a specific policy or a clear promise of action."

The texts produced by ChatGPT are included in the replication scripts. in Table C.3 below we report a subset due to space limitations.

Table C.3: Texts generated by ChatGPT.

Sentences generated by ChatGPT
Policy Commitments
Our veterans deserve the utmost respect and care. I will work tirelessly to improve healthcare services and provide better support for our brave servicemen and women. #SupportOurTroops
I will prioritize criminal justice reform and work towards ending the school-to-prison pipeline. Let's invest in education, mentorship, and community-based programs for at-risk youth. #JusticeForAll
Quality childcare is essential for working families. I will work towards implementing policies that ensure affordable and high-quality childcare options for all parents. #InvestInChildcare
Gun violence must end. I commit to implementing common-sense gun control measures that protect our communities while respecting the rights of responsible gun owners. #EndGunViolence
I believe in the power of arts and culture. If elected, I will support initiatives that promote artistic expression, fund arts programs, and make the arts accessible to all. #ArtsForAll
Populist Rhetoric
The political establishment thinks they know what's best for us, but they're out of touch with the struggles of everyday Americans. We need leaders who will put the people first.
The elite have been using their power to keep us divided and powerless, but we know that solidarity is our greatest strength. Let's come together and build a movement that can't be ignored!
The political elite are more concerned with their own careers than the well-being of the American people. We need leaders who will prioritize the needs of the many over the few.
The elites have been dividing us for far too long, but we know that we're stronger when we stand together. Let's build a movement that empowers all Americans!
The political class thinks they can ignore the will of the people, but we won't be silenced. It's time to demand a government that listens to our voices and represents our needs!

Notes: Subset of the statements generated by ChatGPT.

To ensure the replicability of the annotation, we set the temperature of the model — a parameter that governs the randomness and thus the creativity of the responses — to 0, which means that the responses will be very almost deterministic, yielding the same response over and over again to a given prompt. We leave all other parameters to their default settings.

To compare the validity of ChatGPT's annotation, we extract two random samples of 100 tweets each and ask an independent coder to replicate the annotation task. We give the coder the same prompt given to ChatGPT (see Table 2). The coder and ChatGPT are in agreement 82% (commitment) 84% (populist rhetoric) of the times, which suggests that we can confidently use ChatGPT to train the classifier.

On the annotated dataset, we then train 4 different classifiers, available through the `sklearn` library in Python:

1. Gradient Boosting Classifier (GB)

2. AdaBoost Classifier (AB)

3. Random Forest (RF)

4. Linear Discriminant Analysis (LDA)

We set the model parameters to their default setting. We train the model on the 75% of the

A9

tweets, and we leave the remaining 25% as held-out test set on which to assess the performance of the classifier.

In Table C.4 below, we report the performance metrics of each of the four classifiers. While all classifiers work well on held-out test set, we select the Gradient Boosting classifier, for it is the one performing best. We then train the model on the entire training dataset (this time including the test set too). Finally, we apply the classifier on the universe of the tweets posted in 2012, 2016, and 2020.

Table C.4: Performance metrics of classifiers.

| Classifier | Commitment | | | | Populist Rhetoric | | | |
	F1	Accuracy	Recall	Precision	F1	Accuracy	Recall	Precision
GB	0.84	0.84	0.86	0.83	0.83	0.83	0.86	0.80
AB	0.84	0.84	0.86	0.83	0.83	0.83	0.85	0.81
RF	0.82	0.82	0.84	0.81	0.80	0.81	0.80	0.80
LDA	0.84	0.84	0.86	0.82	0.84	0.84	0.85	0.83

Notes: Performance metrics for each classifier used on two separate 2-classes training datasets: one for policy commitments and one for populist rhetoric. Performance metrics calculated from comparing the model predictions to the annotated labels of held-out test sets (25% of the tweets in a 3,000-tweet training dataset).

One key advantage of using ChatGPT for classifying populist rhetoric and policy commitments is that, while ChatGPT clearly suffers from the same biases of the humans who produced the text on which it has been trained, it is likely to give more stable answer to what are generally difficult questions. Detecting policy commitments and populist rhetoric in tweets is hard, and human annotators might suffer attention decays. Moreover, because tweets can never be fully anonymous, it is likely that coders will adopt a "double-standard" classification when reading tweets (clearly) posted by Democrats or Republicans), displaying a form of confirmation bias. Relying on ChatGPT, besides being a cheap and efficient way of annotating text, also allows to limit human arbitrariness in the classification process.

D Topic Detection

Table D.5: Topics and questions.

Inferred Topic	CCES Question	Years	N. Times
Abortion	abortion_20weeks	2020	1
	abortion_always	2020	1
	abortion_coverage	2020	1
	abortion_prohibition	2016, 2020	23
Environ.	enviro_airwateracts	2016, 2020	2
	enviro_carbon	2016, 2020	21
	enviro_mpg_raise	2020	1
	enviro_renewable	2020	2
Gun Reg.	guns_assaultban	2020	2
	guns_bgchecks	2016	59
Healthcare	healthcare_aca	2012, 2016, 2020	43
	healthcare_medicare	2020	2
Immigration	immig_border	2012, 2020	45
	immig_employer	2012, 2016	6
	immig_legalize	2012, 2020	4
	immig_police	2012	2
	immig_report	2020	1
	immig_services	2012	15
	immig_wall	2020	5
Military	military_democracy	2012, 2016	21
	military_genocide	2020	1
	military_helpun	2012	1
	military_oil	2012, 2016, 2020	12
	military_protectallies	2012, 2016, 2020	33
	military_terroristcamp	2012, 2016	3
Same-sex Marr.	gaymarriage_legalize	2012	13

Notes: Survey questions and inferred topic where partisan respondents expressed highest level of agreement (lowest standard deviation). The table reports the number of times and the years when the question appear as the one displaying lowest standard deviation in any state-party-year groups of respondents.

Table D.6: Most homogeneous topics across state-party-year groups of respondents.

	Democrats			Republicans		
State	2012	2016	2020	2012	2016	2020
AL	Immigration	Gun Reg.	Immigration	Same-sex Marr.	Healthcare	Immigration
AK	Healthcare	Environ.	Abortion, Immigration	Same-sex Marr.	Military	Immigration
AZ	Military	Gun Reg.	Environ.	Military	Healthcare	Immigration
AR	Military	Gun Reg.	Military	Same-sex Marr.	Healthcare	Immigration
CA	Immigration	Gun Reg.	Military	Military	Healthcare	Immigration
CO	Military	Gun Reg.	Immigration	Military	Healthcare	Immigration
CT	Healthcare	Gun Reg.	Abortion	Military	Gun Reg.	Immigration
DE	Military	Gun Reg.	Environ.	Military	Gun Reg.	Abortion
FL	Immigration	Gun Reg.	Environ.	Military	Gun Reg.	Immigration
GA	Immigration	Gun Reg.	Abortion, Immigration	Military	Healthcare	Immigration
HI	Immigration	Gun Reg.	Abortion	Military	Gun Reg.	Military
ID	Healthcare	Gun Reg.	Environ.	Same-sex Marr.	Military	Immigration
IL	Military	Gun Reg.	Environ.	Military	Gun Reg.	Immigration
IN	Military	Gun Reg.	Abortion	Military	Gun Reg.	Immigration
IA	Military	Gun Reg.	Abortion	Military	Healthcare	Immigration
KS	Healthcare	Gun Reg.	Environ.	Military	Healthcare	Immigration
KY	Military	Gun Reg.	Abortion	Same-sex Marr.	Healthcare	Immigration
LA	Immigration	Gun Reg.	Healthcare	Military	Healthcare	Immigration
ME	Same-sex Marr.	Abortion	Immigration	Military	Healthcare	Immigration
MD	Immigration	Gun Reg.	Abortion	Military	Gun Reg.	Immigration
MA	Immigration	Gun Reg.	Environ., Gun Reg.	Military	Military	Immigration
MI	Military	Gun Reg.	Environ.	Military	Healthcare	Immigration
MN	Military	Gun Reg.	Environ.	Military	Healthcare	Immigration
MS	Immigration	Gun Reg.	Healthcare	Same-sex Marr.	Gun Reg.	Immigration
MO	Healthcare	Gun Reg.	Immigration	Military	Healthcare	Immigration
MT	Military	Abortion	Abortion	Military	Healthcare	Immigration
NE	Military	Gun Reg.	Abortion	Military	Healthcare	Immigration
NV	Military	Gun Reg.	Abortion, Environ.	Immigration	Abortion	Immigration
NH	Immigration	Gun Reg.	Environ.	Immigration	Military	Immigration
NJ	Immigration	Gun Reg.	Environ.	Military	Gun Reg.	Immigration
NM	Military	Gun Reg.	Environ.	Military	Healthcare	Immigration
NY	Immigration	Gun Reg.	Environ.	Military	Gun Reg.	Immigration
NC	Immigration	Gun Reg.	Abortion	Same-sex Marr.	Healthcare	Immigration
ND	Military	Gun Reg.	Immigration	Healthcare	Healthcare	Immigration
OH	Immigration	Gun Reg.	Environ.	Immigration	Healthcare	Immigration
OK	Military	Gun Reg.	Abortion	Same-sex Marr.	Healthcare	Immigration
OR	Healthcare	Gun Reg.	Abortion	Military	Military	Immigration
PA	Military	Gun Reg.	Abortion	Military	Gun Reg.	Immigration
RI	Healthcare	Gun Reg.	Abortion	Military	Immigration	Military
SC	Military	Gun Reg.	Abortion	Same-sex Marr.	Gun Reg.	Immigration
SD	Healthcare	Gun Reg.	Environ.	Military	Immigration	Abortion
TN	Immigration	Gun Reg.	Environ.	Same-sex Marr.	Healthcare	Immigration
TX	Immigration	Gun Reg.	Environ.	Military	Healthcare	Immigration
UT	Military	Gun Reg.	Immigration	Same-sex Marr.	Healthcare	Military
VT	Healthcare	Gun Reg.	Environ.	Military	Immigration, Military	Abortion
VA	Immigration	Gun Reg.	Environ.	Military	Healthcare	Immigration
WA	Healthcare	Gun Reg.	Abortion	Military	Immigration	Immigration
WV	Military	Gun Reg.	Abortion	Military	Military	Immigration
WI	Healthcare	Gun Reg.	Abortion	Immigration	Healthcare	Immigration
WY	Healthcare	Military	Environ., Gun Reg., Healthcare	Healthcare	Environ., Healthcare	Immigration

Notes: Inferred topic from quesion displaying lowest standard deviation in responses.

A12

Building Topic-specific Dictionaries

To build topic-specific dictionaries, we extract the 20 words whose vector representation is most similar to the topic label (e.g., immigration) with `word2vec` (Mikolov et al. 2013), an unsupervised algorithm that learns fixed-length feature representations from how often words co-occur with one another, with the assumption the meaning of a word is given by the company it keeps (Rodriguez & Spirling 2022).

First we pre-process the 5.9 million tweets posted by Congressional candidates between 2012 and 2021. We lower case the text, we remove the twitter accounts tagged in the text (e.g., @NAME), we keep ash-tags because they generally convey important information, we remove punctuation, and we convert all nouns to their singular form (e.g., houses → house). We then allow bi-grams to emerge based on how often two words occur next to one another, imposing a minimum count of 200. This step allows for words like "gun_control", or "health_care" to be considered single words when training the model. We finally train the `word2vec` model on the pre-processed corpus of tweets with the `gensim` Python library, estimating 200-dimensional vectors, excluding words appearing less than 10 times, and setting a window size (where to compute word co-occurrences) to 4.

Once we have a word embedding for each word used at least ten times in the corpus, we extract the 20 words with the largest cosine similarity to the topic label. We manually remove false positive to ensure that each word is used almost exclusively in the context of the topic (e.g., we remove the word "government" from the military-specific dictionary, for it can be used in many different contexts without referring to military issues) to and we assemble the eight topic-specific dictionaries reported in Table D.7 below.

Table D.7: Topic specific dictionaries.

Topic	Dictionary
Gun Regulations	gun, firearm, handgun, assault_weapon, weapon_of_war, bump_stock, rifle, weapon, assault_rifle, domestic_abuser, silencer, shotgun, pistol, assaultweapons, weapon_ban, gunfreezones, semiautomatic, concealed_carry
Immigration	immigration, immigrant, asylum_seeker, undocumented_immigrant, refugee, asylumseekers, immigrant_and_refugee, legal_immigration, undocumented, immig, immigration_policy, legal_immigrant, illegal_immigration, migrant, deportation, imm, family_reunification, familybased, assimilation, illegalimmigration, sanctuarycities, chain_migration
Abortion	abortion, lateterm_abortion, infanticide, birth_control, family_planning, abortion_clinic, abortionist, legal_abortion, abortion_provider, unborn_baby, abortion_ban, contraceptive, planned_parenthood, reproductive_health, contraception, unborn_child, title_x, born_alive
Military	military, armed_force, USA_military, military_personnel, servicemembers, service_member, military_readiness, our_troop, uniform, armedforces, dod, warfighters, civilian, troop, navy, coast_guard, department_of_defense, national_defense, military_service, law_enforcement
Environment	environment, our_environment, ecosystem, our_planet, enviro, air_and_water, env, clean_air, airwater, climate, quality_of_life, cleanair, natural_resource, planet, biodiversity, wildlife, energy_sector, waterway, landscape
Health Care	healthcare, health_care, affordable_healthcare, hc, access_to_healthcare, quality_healthcare, heathcare, healthcare_coverage, health_coverage, health, health_insurance, reproductive_healthcare, insurance_coverage, retirement_security, hcare, child_care, universal_healthcare, preventative_care, quality_care, medicare
Same-sex Marriage	gay_marriage, gay, samesex_marriage, polygamy, lgbt, homosexual, lesbian, gaymarriage, marriage_equality, interracial, ssm, conversion_therapy, bisexual, homosexuality, samesex_couple, antigay, doma, samesexmarriage

Notes: Topic specific dictionaries derived from the word2vec model based on the largest cosine similarity between words and the topic label.

A14

E Robustness Tests

Figure E.3: Effect of distrust across different time windows.
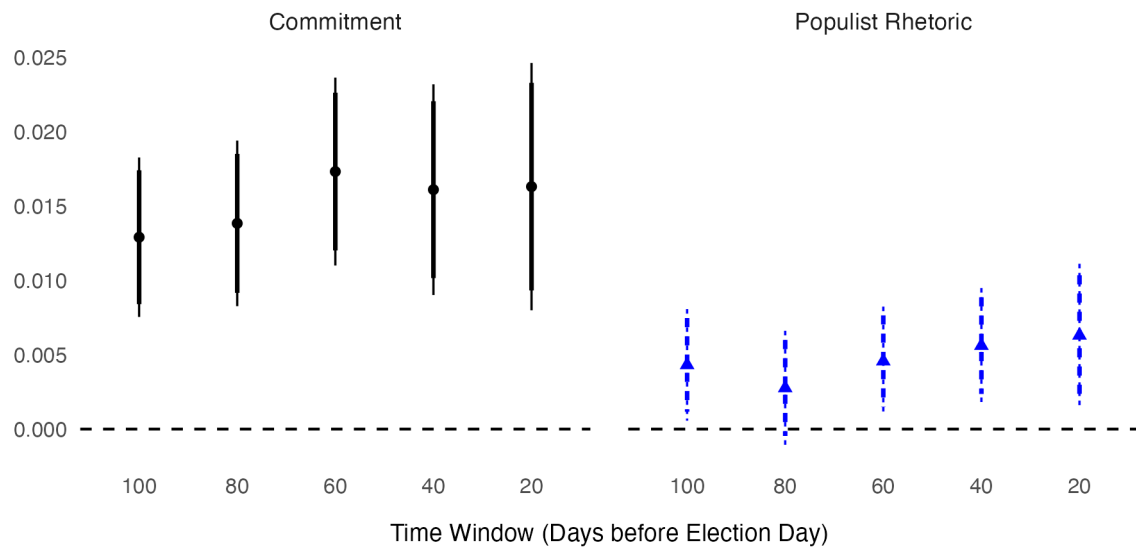
Notes: Estimated effect of distrust with 90 and 95% confidence intervals estimated across different windows of time expressed in days from election day. Specification includes candidate, year, district, state-by-year fixed effects and incumbency status as a covariate. Standard errors clustered by district.

Table E.8: Alternative measures of distrust.

	Commitment			Populist Rhetoric		
	(1)	(2)	(3)	(4)	(5)	(6)
Distrust - Govt. Do Right	0.0173***			0.0046**		
	(0.0032)			(0.0019)		
Distrust - Big Interests		0.0037*			0.0000	
		(0.0020)			(0.0011)	
Distrust - Politians Corrupt			0.0141***			0.0003
			(0.0038)			(0.0020)
Controls: Incumbent	✓	✓	✓	✓	✓	✓
R^2	0.029	0.029	0.029	0.023	0.023	0.023
Observations	274,253	274,253	274,253	274,253	274,253	274,253
Candidate FE	✓	✓	✓	✓	✓	✓
District FE	✓	✓	✓	✓	✓	✓
State-Year FE	✓	✓	✓	✓	✓	✓

Notes: OLS estimates. SE clustered by district. DV is a dummy measure for commitment/populist tweets. Distrust variables are standardized shares of distrustful individuals in each district. Govt. Do Right: never trust the government in Washington to do what is right. Politicians Corrupt: Most/All in government are corrupt. Big Interests: Government run by a few big interests. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

Table E.9: Distrust, supply of commitments and populist rhetoric.

	Commitment			
	(1)	(2)	(3)	(4)
Distrust	0.005**	0.008***	0.010***	0.014***
	(0.002)	(0.002)	(0.003)	(0.003)
Homog. Topic	0.025**	0.025**	0.024**	0.024**
	(0.011)	(0.010)	(0.011)	(0.010)
Distrust × Homog. Topic	0.026***	0.026***	0.026***	0.025**
	(0.010)	(0.010)	(0.010)	(0.010)
Controls: Party	✓	✓		
Controls: Incumbent		✓		✓
Topic Dummies	✓	✓	✓	✓
R^2	0.064	0.067	0.073	0.075
Observations	255,196	255,196	255,196	255,196
District FE	✓	✓	✓	✓
Year FE	✓		✓	
State-Year FE		✓		✓
Candidate FE			✓	✓

Notes: OLS estimates. SE clustered by district. Sample of Democratic and Republican candidates. DV is a dummy measure for commitment/populist tweets. Distrust is standardized share of distrustful individuals in each district. Signif. codes: ***: 0.01, **: 0.05, *: 0.1

Table E.10: Effect of Trump candidacy on Turnout of Distrustful Republicans.

	Pr(Voted in Pres. Elections)		
	(1)	(2)	(3)
Republican	0.053***	0.062***	0.046***
	(0.011)	(0.012)	(0.011)
Distrust	-0.102***	-0.095***	-0.063**
	(0.032)	(0.032)	(0.031)
EY 2016	-0.102***	-0.100***	-0.120***
	(0.012)	(0.012)	(0.012)
EY 2020	-0.005	-0.006	-0.042***
	(0.009)	(0.009)	(0.010)
Republican × Distrust	0.019	0.002	-0.009
	(0.057)	(0.058)	(0.056)
Republican × EY 2016	0.035*	0.024	0.026
	(0.020)	(0.020)	(0.020)
Republican × EY 2020	-0.018	-0.024	-0.018
	(0.015)	(0.016)	(0.015)
Distrust × EY 2016	0.002	-0.002	-0.022
	(0.045)	(0.045)	(0.044)
Distrust × EY 2020	-0.053	-0.052	-0.034
	(0.039)	(0.039)	(0.038)
Republican × Distrust × EY 2016	0.134*	0.146*	0.150**
	(0.074)	(0.075)	(0.072)
Republican × Distrust × EY 2020	0.043	0.051	0.032
	(0.073)	(0.073)	(0.072)
Individual Controls			✓
R^2	0.038	0.073	0.137
Observations	15,086	15,086	14,602
State FE	✓		
Congressional District FE		✓	✓

Notes: OLS estimates (linear probability model). SE clustered by congressional district. Outcome is probability to vote in the presidential elections. Individual-level convariates include: sex, age categories, university education, employment status. Reference election year is 2012. Distrust = 1 if respondent never trusts the government in Washington to do what is right. Signif. codes: ***: 0.01, **: 0.05, *: 0.1