Digital Populism and Electoral Democracy*

Elliot Foote† Weijia Li‡

January 31, 2026

Abstract

We developed a model where a populist produces and distributes anti-"elite" rhetoric, in anticipation of how voters respond to the rhetoric. By doing so, we provide a framework to understand how populism is affected by changes in technology and the institutional environment. The model shows that populism thrives through a combination of digital media and a high cost of voting, under which a populist can win an electoral majority by persuading a small and extreme audience. We then employ the model to study the interaction among digital media, the voting cost, and the degree of voter polarization. We show that a high voting cost may even reverse the capacity of a highly centrist society to deter digital populism, which helps understand the recent rise of populism while most voters largely remained centrist.

*We thank Sascha Becker, Sebastiano Della Lena, Gabriele Gratton, Hongyi Li, Peter Lorentzen, Gerard Roland, Yang Xie, and the participants of the 2022 SIOE conference for valuable discussions. This paper is based on the 2020 thesis of Elliot Foote. All errors are our own.

†Australian Business Growth Fund; elliot.foote@gmail.com.

‡Monash University. weijia.li@monash.edu.

1 Introduction

Since the recent rise of populist movements, there have been calls to strengthen "liberal" institutions to constrain populist leaders, such as the judiciary and independent agencies (Mukand and Rodrik (2020); Guriev and Papaioannou (2022); Gratton and Lee (2024)). Complementing these liberal institutions, we investigate another approach in deterring populism, i.e., a broader participation of ordinary voters in elections. We developed a stylized model to evaluate the relationship between populism and voter participation, especially when a populist can target his rhetoric through digital technology. Populists may claim to represent popular sovereignty, but our analysis suggests that they actually thrive through voter suppression in the digital age. Thus, a democracy should encourage more participation of voters in elections, not less, which contrasts with more elitist solutions to populist threats (e.g., Brennan (2016)).

The model features an election where an establishment incumbent competes against an opposition candidate. The opposition candidate runs either as another establishment candidate or as a populist candidate. A populist opposition can produce a rhetoric to criticize the incumbent. Rather than modeling negative campaigning, we will see that our model attempts to match key features of populism. After producing the rhetoric, the populist distributes his rhetoric through the media. If the populist can only access broad media outlets, he must distribute his rhetoric to the entire national audience. But if the populist can distribute his rhetoric through "targeted media," he can choose a precise subgroup of voters as the audience for his rhetoric (Peitz and Reisinger (2015); Goldfarb and Tucker (2019); Zuboff (2019)). Such capacity to choose the exact audience is achieved through digital media, which allows precise targeting of advertisements (Goldfarb and Tucker (2019)). Lastly, voters decide whether to turn out and, if so, which candidate they support. A voter who turns out pays a voting cost.

By producing and distributing his rhetoric, the populist attempts to win the election by swinging enough voters, while minimizing the persuasion spending. The constrained minimization problem yields the optimal populist rhetoric and the optimal audience for the rhetoric. The optimal rhetoric is inflaming, allowing the populist to win the election by only targeting voters disillusioned with the incumbent. In turn, the optimal targeted audience only includes these disillusioned voters, even though the populist is capable of persuading voters who are currently satisfied with the incumbent. By only targeting disillusioned voters, the populist avoids a discontinuous jump in persuasion spending, thereby generating a "homogeneous" audience for populism.

This endogenous homogeneity of the populist audience determines the comparative statics in a sharp manner. First, the opposition is more likely to run as a populist under a higher voting cost. When the voting cost increases, a populist can replace the most moderate voters in his optimal audience with another group of voters who are more extreme in their ideology, therefore far more amenable to populist rhetoric. Such a replacement reduces persuasion spending. In other words, under a low voting cost, the populist must persuade an audience that is quite moderate and skeptical of the populist rhetoric; but under a high voting cost, the populist only needs to persuade an extreme audience that is especially susceptible to populist rhetoric. In this case, the resulting electoral "majority" that supports populism is actually a small and extreme subset of the broader public.

Second, the opposition is more likely to run as a populist when targeted media is more effective. Effective targeted media allows the populist to send his rhetoric only to a narrow audience, which reduces persuasion spending.

Third, targeted media and a high cost of voting complement each other in producing populism. The populist only benefits from a high voting cost when he can target his rhetoric to the ideal audience with its extreme ideology. The populist does not benefit at all from a high voting cost when he has to distribute his rhetoric to the national audience. Therefore, under more effective targeted media, a higher voting cost produces a larger marginal benefit for the populist. On the flip side of the complementarity, targeted media also produces a larger marginal benefit when the voting cost is high. The populist especially values the capability to target an audience when the populist can win the election by targeting a more extreme audience, a situation that is secured by a higher voting cost.

The complementarity is the first key result of the paper. As the second key result, under a high voting cost, digital populism might be *boosted* by a more centrist society. This result stands in contrast to traditional wisdom that a centrist society is ideal for a democracy to function well (Lipset (1959); Moore (1993)). But the traditional wisdom could be upended by voter suppression in the digital age, which renders a large cluster of centrist voters irrelevant. Our model therefore highlights an insight that has not received sufficient attention: in the digital age, voter suppression might neutralize or even overturn the capacity of a centrist society to deter populism. While traditional wisdom on the centrist society emphasizes the *global* distribution of all voters, what matters in the age of targeted media is the *local* distribution of pivotal voters.

We further shed light on populism by comparing it with traditional negative campaigning (Skaperdas and Grofman (1995)). Unlike a populist, a negative campaigner finds it optimal to target both his own supporters and supporters of the incumbent, an audience that is highly heterogeneous. This heterogeneous audience is too diverse to allow the populist rhetoric of "us" versus "them" (Guriev and Papaioannou (2022). A heterogeneous audience also sharply reduces the appeal of voter suppression to a negative campaigner, who is therefore much less tempted to suppress voters than a populist. In sum, this comparison underscores that the peculiar characteristics of populism are fundamentally shaped by the endogenous homogeneity of its audience.

A substantial body of literature has examined the relationship between populism and liberal democracy. Existing studies primarily focus on how populism collides with liberal institutions that safeguard fundamental rights or mitigate policy failures, such as "constraints on the executive, checks and balances, the rule of law, and independent bureaucratic agencies" (Acemoglu et al. (2013b); Mukand and Rodrik (2020); Sasso and Morelli (2021); Guriev and Papaioannou (2022); Gratton and Lee (2024); Szeidl and Szucs (2025)). To our knowledge, we present the first model demonstrating that digital populism thrives when *electoral* institutions are undermined. Our model shows that digital populism might be a greater threat to democracy than previously recognized.

Many recent papers have uncovered fundamental causes of populism, such as economic grievance, digital technology, social media, and cognitive or psychological biases (e.g., Acemoglu et al. (2013a); Algan et al. (2017); Noury and Roland (2020); Bernhardt et al. (2022); Danieli et al. (2022); Levy et al. (2022); Dal Bo et al. (2023); Della Lena et al. (2025); Gratton and Lee (2025); Szeidl and Szucs (2025)).[1] Our paper contributes to this literature on the conditions of populism. Under a high voting cost, our model uncovers a higher risk of digital populism for a more centrist society, which might explain why populists achieve many electoral victories while the citizenry remain much more centrist than political elites (Gentzkow (2016); Callander and Carbajal (2022)). Meanwhile, the key complementarity result may also help understand why populism shows uneven strength across time and space (Gratton and Lee (2024)), where several liberal democracies with high turnout rates have been more resistant to the recent rise of populism (Taggart (2017); Brett (2019)).

The paper is organized as follows. Section 2 sets up the model. Section 3 solves the model, characterizing the optimal populist message and its optimal audience. Section 4 examines comparative statics, which formalize the two key results of the paper. Section 5 further compares populism and traditional negative campaigning. Section 6 concludes.

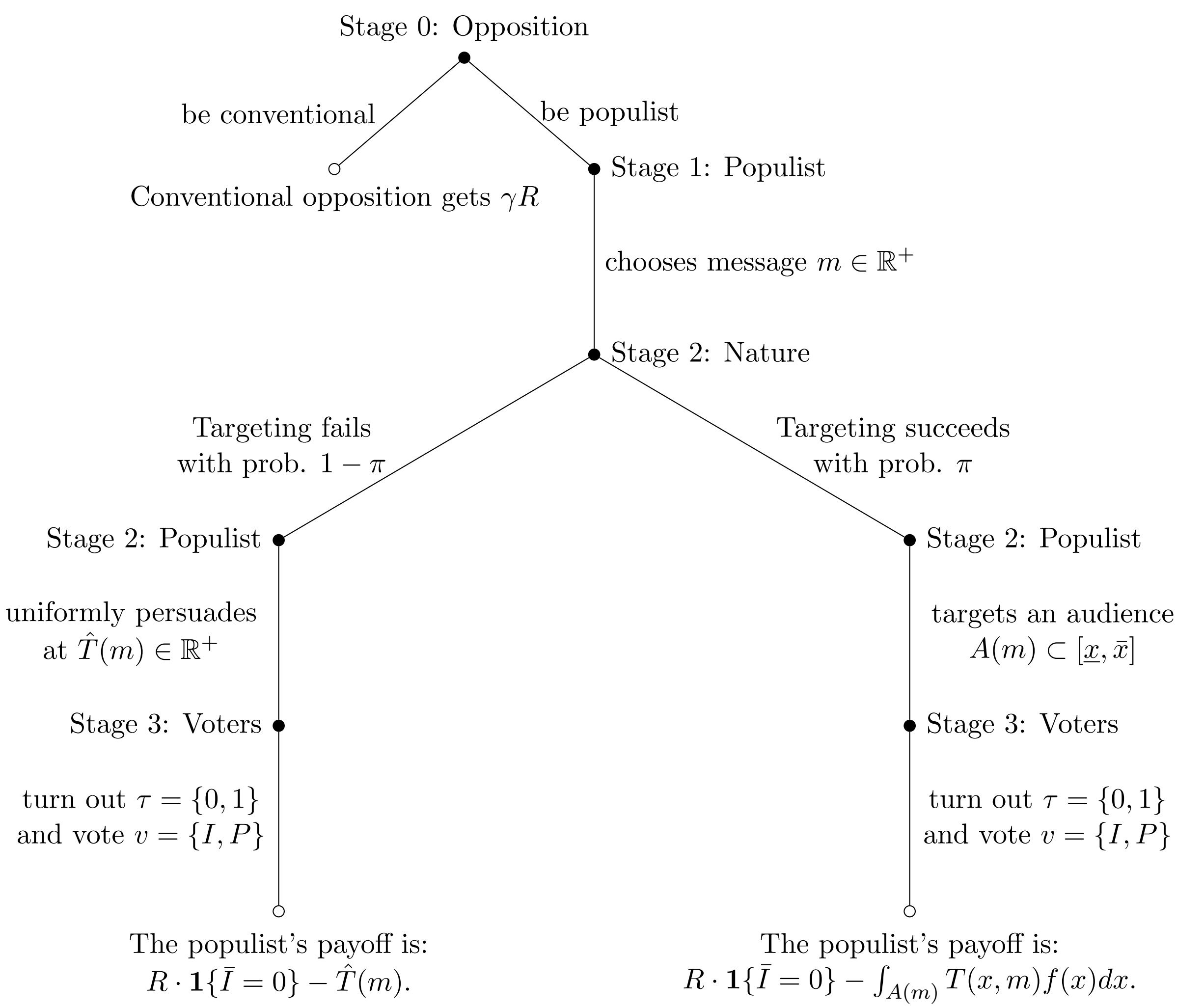
2 Setup of the model

The model is a sequential game. The extensive form is in Figure 1, which will be explained in detail throughout this section. There are three players, an incumbent establishment can-

[1] The development of digital technology is also strongly influenced by political-economic processes outside democracies (Beraja et al. (2023a); Beraja et al. (2023b)).

didate, an opposition candidate, and a unit mass of voters. A voter's satisfaction with the incumbent is $x \in [\underline{x}, \bar{x}] \subset \mathbb{R}$, with $-\infty \leq \underline{x} < 0 < \bar{x} \leq \infty$. Voter satisfaction follows the cumulative distribution function $F(\cdot)$, and the probability density function is $f(\cdot)$. The density function $f(\cdot)$ is bounded, has full support on its domain, is uni-modal, and is symmetric around the mean/median μ. This setup focuses our attention on a "centrist" distribution of voter preferences. But even in such a centrist environment, we will show that digital populism can be menacing.

Figure 1: Extensive form of the game



The average satisfaction with the incumbent is $\mu = E[x]$. We assume that $\mu > 0$, while a voter's satisfaction with a populist is normalized to 0. The assumption $\mu > 0$ embeds the notion that the incumbent has spent a lot in persuasion while he was in office before the game in Figure 1 starts. So average voters begin with a higher satisfaction with the incumbent than a potential populist. Meanwhile, because of his extensive and exhaustive persuasion during his previous term, the incumbent cannot further shift voters' opinions. A passive candidate like this serves as the alternative option for voters in many election models. For instance, this setting is prevalent in canonical models of electoral accountability, where voters can replace the incumbent with a passive candidate operating in the background (Persson and Tabellini (2002); Fearon (2011); Gehlbach (2021)). This type of setting allows researchers to focus on the more novel aspects of the environment.

At Stage 0, the opposition runs as a conventional or populist candidate. If the opposition runs as a conventional candidate, the game ends. The conventional opposition wins the election with an exogenous probability $\gamma \in (0,1)$.[2] An electoral victory confers an exogenous utility of $R \in \mathbb{R}^{++}$ to the opposition. If the opposition runs as a populist, the game continues.

At Stage 1, the populist chooses a message $m \in \mathbb{R}^{+}$.

At Stage 2, the populist tries to convince voters of the message m. If a voter is convinced of the message m, his satisfaction with the incumbent is reduced to

$$x - m. \tag{1}$$

To convince a voter whose *ex ante* satisfaction with the incumbent is x, the populist needs to spend

$$T(x, m) \in \mathbb{R}^{++} \tag{2}$$

on the voter. The derivative T_x is positive: it is more difficult to persuade a voter who is more satisfied with the incumbent. The other derivative T_m is also positive: it is more difficult to convince a voter of a more inflaming message.

The persuasion protocol depends on $\pi \in (0,1)$, the effectiveness of a media targeting technology. With probability $1-\pi$, the targeting technology fails, and the populist can only access a broad media outlet (e.g., national media). As such, the populist can only implement uniform persuasion: the populist spends the same amount

$$\hat{T}(m) \in \mathbb{R}^{+} \tag{3}$$

on all voters. If $\hat{T} < T(x,m)$, a voter x is unconvinced of the message m and his satisfaction with the incumbent remains at x. If $\hat{T} \geq T(x,m)$, a voter x is convinced of the message m, and his satisfaction with the incumbent is reduced to $x - m$.

With probability π, the targeting technology succeeds. The populist can choose a targeted

[2] In a conventional electoral competition, the opposition competes with the incumbent on a traditional policy space. Following the logic of Dowsian electoral competition, the opposition chooses the same policy platform as the incumbent, and the winner is determined randomly (Gehlbach (2021)).

audience for the message m:

$$A(m) \subset [\underline{x}, \bar{x}]. \tag{4}$$

To each voter with $x \in A(m)$, the populist spends $T(x, m)$. The set $A(m)$ can be any Lebesgue measurable subset of $[\underline{x}, \bar{x}]$.

At Stage 3, voters decide whether to turn out and, if so, which candidate to support. Denote $2c > 0$ as the cost of voting (2 is multiplied to simplify algebra). Voters follow a simple turnout rule: if a voter is convinced of a message m, she turns out if and only if

$$|x - m| \geq 2c; \tag{5}$$

if a voter is unconvinced of any message, the voter turns out if and only if

$$|x| \geq 2c. \tag{6}$$

The turnout decision is shaped by a standard trade-off between the cost and the benefit of voting. The right-hand side of Equation 5 is the cost of voting. The left-hand side is the difference in utility from electing the incumbent $(x - m)$ versus the populist (0), the "benefit" for a voter who is convinced of a message m. For a voter who is unconvinced of any message, her satisfaction with the incumbent remains at x. Her turnout decision is described by Equation 5 by setting $m = 0$, yielding Equation 6. Assume that $2c > \mu$, under which the turnout rate in equilibrium is strictly less than 100%.[3]

The decision rule leads to a plausible pattern of voter turnout. Voters only turn out if they have relatively strong preferences over candidates. A voting citizen either strongly prefers the incumbent over the populist or the other way around. A citizen does not vote if she finds not much of a difference between the incumbent and the populist.

The turnout decision also implies which candidate a voter supports. For the population with $x \geq m + 2c$, they vote for the incumbent because the incumbent in power yields a higher payoff than the populist $(x - m \geq 2c > 0)$. Similarly, the population with $x \leq m - 2c$ vote for the populist.

The solution concept is the subgame perfect equilibrium. A player maximizes his expected payoff. Specifically, if targeting fails, the populist's payoff is:

$$R \cdot \mathbf{1}\{\bar{I} = 0\} - \hat{T}(m). \tag{7}$$

The function $\mathbf{1}\{\cdot\}$ is the indicator function. The parameter $\bar{I} = 1$ if more than 50% of the

[3]Even under compulsory voting that is well enforced (e.g., since 1924 in Australia), there are still around 4%–10% of the population who do not turn out to vote (Australian Electoral Commission (2023)).

voting citizens support the incumbent; otherwise, $\bar{I} = 0$.

Similarly, if targeting succeeds, the populist's payoff is:

$$R \cdot \mathbf{1}\{\bar{I} = 0\} - \int_{A(m)} T(x, m) f(x) dx. \tag{8}$$

We focus our attention on a reasonable parametric space. First, $R < T(-2c + 2\mu, 2\mu)$, meaning that it is expensive to broadcast a populist message uniformly to the entire national audience. Also, $R > \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x) dx$, so that it is relatively cheap to indoctrinate a populist message when the message can be targeted to a precise and "small" audience (specifically, the audience with $x \in [-2c, -2c + 2\mu]$).

The setup so far might still be a bit too general, potentially including other political phenomena such as negative campaigning (Section 5). To focus our attention on populism, we impose the following assumption.

Assumption 1. *1. There exists* $\underline{T} > 0$ *and* $m_1 \in (0, \mu)$*, for all* $x \in [\underline{x}, \bar{x}]$ *and for all* $m \in [m_1, \infty)$*,*

$$T_x(x, m) > \underline{T}. \tag{9}$$

2. There exists $\tilde{t} > 0$ *sufficiently small and* $m_2 \in (0, \mu)$*, for all* $x \in [\underline{x}, \bar{x}]$ *and for all* $m \in [m_2, \infty)$*,*

$$T_m(x, m) < \tilde{t}. \tag{10}$$

The assumption has two parts. The first part says that for a message strongly against the incumbent (m large enough), it is strictly more expensive to persuade a voter who is more satisfied with the incumbent. So individuals differ a lot from each other in their opinions of an inflaming message against the incumbent: while voters already angry with the incumbent are highly receptive to the message, supporters of the incumbent are much more resistant. The second part says that for a message strongly against the incumbent, another message that is even more critical only raises the cost to persuade an individual by a negligible amount. In other words, beyond a threshold of rhetorical extremity, additional attacks on the incumbent are largely redundant presumably because they convey little incremental information and elicit little additional affect.

As we will see, Assumption 1 drives the populist to target his message to an audience that is homogeneous in their negative evaluation of the establishment incumbent. Without Assumption 1, the audience can be quite heterogeneous in their opinions of the establishment incumbent, which is inconsistent with a key feature of populism as uniformly anti-elite (Gratton and Lee (2025); Szeidl and Szucs (2025)). In Section 5, we will further clarify

Assumption 1 by comparing populism with traditional negative campaigning, where the audience can indeed be heterogeneous. We also want to note that our model does not attempt to capture the "essence" of populism. Our model is more suitable as a device to understand the consequences of important features of populism.

3 Analysis of the Model

3.1 Stage 2: the persuadable audience

The model is solved by backward induction. Prior to Stage 2, the populist has chosen a message m in Stage 1. Before we derive the optimal persuasion strategies of the populist, we characterize the persuadable audience that can be swung by a message m. Denote the persuadable audience as

$$\bar{A}(m) \subset [\underline{x}, \bar{x}]. \tag{11}$$

It is easy to verify Lemma 1.

Lemma 1. *Suppose that the populist spends enough to convince all voters of a message m. The message m can only change the voting behavior of voters with:*

$$x \in \bar{A}(m) \equiv [-2c, -2c+m] \cup [2c, 2c+m]. \tag{12}$$

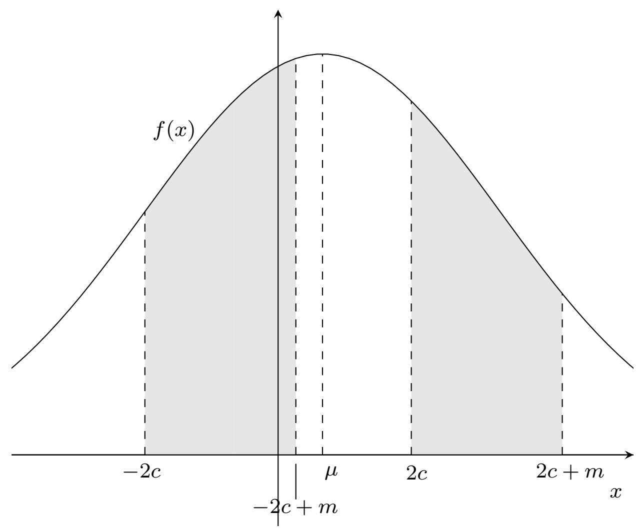
All proofs are in Appendix A. The persuadable audience $\bar{A}(m)$ is illustrated in Figure 2. The message m can only swing two groups of voters. Voters with $x \in [-2c, -2c+m]$ would not vote for the populist without being convinced of the message m. But if convinced, these voters cast ballots for the populist. We call this group "marginal populist voters."

In the other component of $\bar{A}(m)$, voters with $x \in [2c, 2c+m]$ would vote for the incumbent without being convinced of the message. But if convinced, they would not vote for the incumbent. We call this group "marginal incumbent voters."[4]

The populist message cannot influence the other three groups of voters. Voters with $x < -2c$ always vote for the populist, and voters with $x > 2c+m$ always vote for the incumbent. Voters with $x \in (-2c+m, 2c)$, if they exist, always abstain.

[4]If the first group $[-2c, -2c+m]$ intersect with the group $[2c, 2c+m]$, then the marginal incumbent voters are $[2c, 2c+m] \backslash [-2c, -2c+m]$.

Figure 2: the maximal audience that can be swung by a message m



Notes: the horizontal axis is the voter satisfaction x, the vertical axis is the probability density $f(x)$, and the lightly shaded area includes all persuadable voters.

3.2 Stage 2: the optimal audience when targeting succeeds

With probability π, the media targeting technology allows the populist to choose a precise audience for the populist message. The following proposition characterizes the optimal audience for any populist message.

Proposition 1. *There exists a unique $\bar{m} \in [2\mu, \infty]$ and a unique $\underline{m} \in [\mu, 2\mu)$, such that the optimal audience as a function of the message m is:*

$$A^*(m) = \begin{cases} [-2c, -2c + 2\mu] & \text{for } m \in [2\mu, \bar{m}) \\ [-2c, -2c + m] \cup [2c, 2\mu - (m - 2c)] & \text{for } m \in [\mu, \underline{m}) \\ \emptyset & \text{for } m \in [0, \underline{m}) \cup [\bar{m}, \infty) \end{cases} . \tag{13}$$

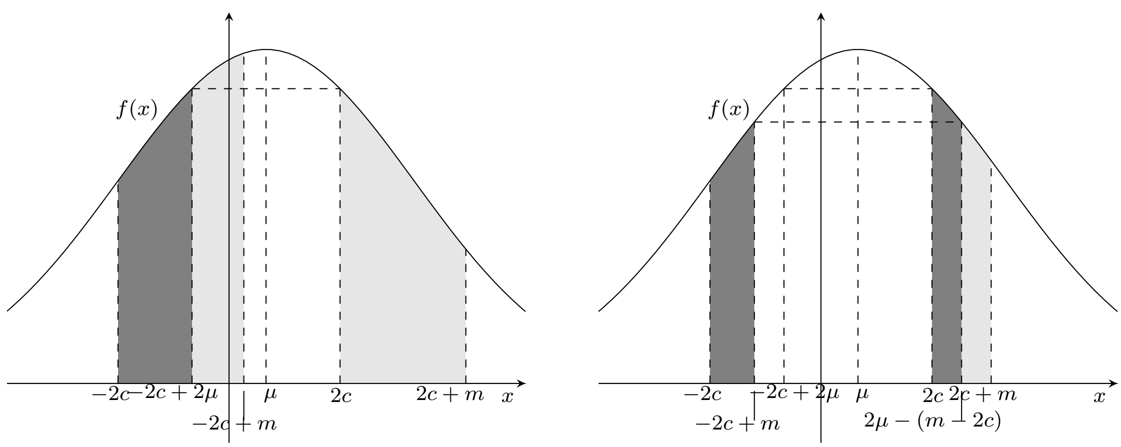
What is the substantive message of Proposition 1? There are three cases to discuss.

A homogeneous audience for an inflaming message. The first line of the optimal audience function 13 says that, if the message m is inflaming enough ($m \in [2\mu, \bar{m})$), the populist only sends the message to the audience $[-2c, -2c+2\mu]$, which is quite homogeneous. This result is illustrated in the left panel of Figure 3. The darkly shaded area is the optimal audience $A^*(m)$ and the lightly shaded area includes all other voters in the persuadable audience $\bar{A}(m)$. The persuadable audience consists of two groups, marginal populist voters

$[-2c, -2c+m]$ and marginal incumbent voters $[2c, 2c+m]$. But the optimal audience, $[-2c, -2c+2\mu]$, only includes marginal populist voters.

Why is the group $[-2c, -2c+2\mu]$ the optimal audience? This audience secures just enough votes so that the populist wins the election. Meanwhile, in the persuadable audience $\bar{A}(m)$, the group $[-2c, -2c+2\mu]$ is the least satisfied with the incumbent. Targeting this group minimizes the persuasion spending.

Figure 3: The optimal audience for a message m



Notes: the left panel shows the optimal audience for an "inflaming" message (the precise definition in Proposition 1). The right panel shows the optimal audience for a "moderate" message (the precise definition in Proposition 1). In both panels, the darkly shaded area is the optimal audience for a message m, and the lightly shaded area includes all other voters that could be persuaded but did not receive a populist message. The horizontal axis is voter satisfaction x, and the vertical axis is the probability density $f(x)$.

A heterogeneous audience for a moderate message. In the optimal audience function 13, the second line shows the case of a relatively moderate message ($m \in [\mu, \underline{m})$). The persuasion strategy is illustrated in the right panel of Figure 3. Again the darkly shaded area is the optimal audience $A^*(m)$ and the lightly shaded area includes all other voters that are in the persuadable audience $\bar{A}(m)$. The optimal audience includes all marginal populist voters *and, crucially, also some marginal incumbent voters.*

This audience of two separate groups is therefore heterogeneous. Under a moderate message, had the populist only persuaded marginal populist voters, the populist's vote share would be smaller than the incumbent's. To win the election, the populist must also convince some marginal incumbent voters to abstain from the election.

An empty audience for a message that is extremely inflaming or weak. Finally, the third line in the optimal audience function 13 shows the case of a message that is extremely inflaming or weak ($m \in [0, \underline{m}) \cup [\bar{m}, \infty)$). The populist does not send such a message to any voter. When the message is too weak, the populist cannot win half of the total votes even if he convinces every persuadable voter. When the message is too inflaming, it is too costly to convince voters of such an inflaming message. In either case, it is better for the populist to lose the election.

3.3 Stage 2 when targeting fails

With probability $1 - \pi$, the targeted media fails to work. The populist can still access the national media. But on the national media, the populist can only spend the *same* amount on *all* voters. The populist's optimal strategy is characterized by the following proposition.

Proposition 2. *Suppose that targeting fails. Then the populist spends nothing on uniform persuasion:*

$$\text{for any } m \in \mathbb{R}^+,\ \hat{T}^*(m) = 0. \tag{14}$$

With only access to the national media, the populist spends nothing to persuade voters. So the entire population is unconvinced of the populist message, and the voting population elects the incumbent. Proposition 2 captures the idea that uniform persuasion of a populist message is too expensive. The populist would rather lose the election than spend a tremendous amount, most of which would be wasted on irrelevant voters.

3.4 Stage 1: the optimal message and its optimal audience

Given the optimal audience function 13, the populist chooses the message m in Stage 1. The optimal message and the equilibrium path are as follows.

Proposition 3. *The optimal message is:*

$$m^* = 2\mu. \tag{15}$$

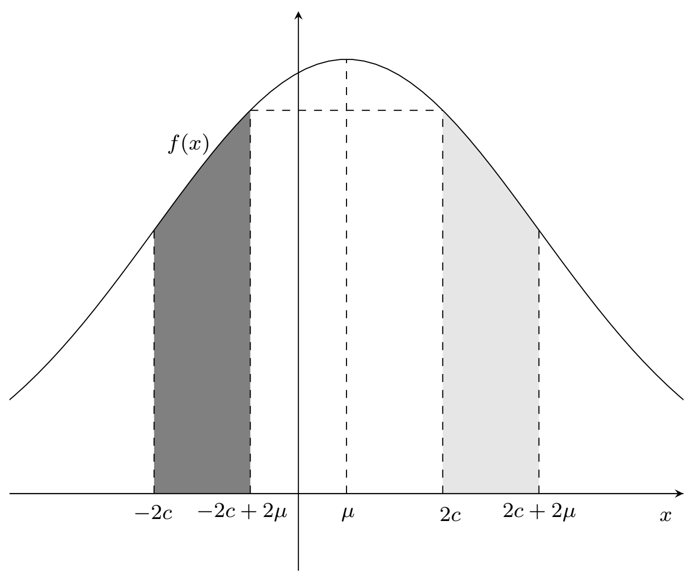
With probability π, the message $m^* = 2\mu$ convinces the optimal audience

$$A^*(m^*) = [-2c, -2c + 2\mu]. \tag{16}$$

With probability $1 - \pi$, the message $m^* = 2\mu$ convinces no voters.

The Proposition is illustrated in Figure 4. The darkly shaded area is the optimal audience for the optimal message $m^* = 2\mu$. The lightly shaded area includes all other voters in the persuadable audience. The optimal message $m^* = 2\mu$ is the least inflaming message that allows a winning populist to only persuade marginal populist voters, ignoring all marginal incumbent voters.

Figure 4: The optimal audience for the optimal message $m^* = 2\mu$



Notes: the darkly shaded area is the optimal audience for the optimal message $m^* = 2\mu$; the lightly shaded area includes all other voters that could be persuaded, but the populist does not send the message to them. The horizontal axis is voter satisfaction x, and the vertical axis is the probability density $f(x)$.

Proposition 3 is the central proposition that prepares the subsequent comparative statics. Thus, we informally discuss the proof of the proposition in more detail. First, $m^* = 2\mu$ yields a higher payoff than any message that is more inflaming. Note that any sufficiently inflaming message targets the same audience, i.e., marginal populist voters who are the most dissatisfied with the incumbent. Therefore, the message 2μ saves on the persuasion spending than any message that is more inflaming.

Second, $m^* = 2\mu$ yields a higher payoff than any message that is less inflaming. There are only two effects to consider.

- The persuasion spending effect: the same as before, a more inflaming message is more costly to indoctrinate on any voter, reducing the populist's payoff.
- The replacement effect: but a more inflaming message generates another effect. A more inflaming message expands the group of marginal populist voters, while shrinking the

group of marginal incumbent voters who must be persuaded. In other words, a more inflaming message allows the populist to replace a group of marginal incumbent voters with a group of marginal populist voters, the latter being much more amenable to the populist message. Thus, the replacement effect reduces the persuasion spending.

The key observation is that the replacement effect dominates the persuasion spending effect. The replacement effect must be qualitatively large because the marginal populist voters are *discontinuously* easier to persuade than marginal incumbent voters. The persuasion spending drops *discontinuously* because the two groups of voters are separated by a non-zero measure of citizens who abstain from the election. Meanwhile, the persuasion spending effect is qualitatively "small" because each individual voter is largely fixed in their opinion of messages strongly critical of the incumbent. Thus, $m^* = 2\mu$ yields a higher payoff than any message that is less inflaming.

In equilibrium, the populist chooses $m^* = 2\mu$, a message inflaming enough so that the populist does not need to convince any marginal incumbent voters. The optimal audience is a connected set, henceforth quite homogeneous. This homogeneous audience is consistent with the defining feature of populism as a sharp demarcation between "us" versus "them."

3.5 Stage 0: populism or conventional politics

Proposition 3 yields the maximal payoff for the opposition to be a populist. The populist chooses the optimal message $m^* = 2\mu$ and obtains:

$$\pi[R - \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x) dx]. \tag{17}$$

In the payoff 17, the media targeting technology succeeds with probability π. The populist obtains $[R - \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x) dx]$, the payoff as a populist who wins the election by only persuading marginal populist voters. With probability $1 - \pi$, the media targeting technology fails. The populist does not attempt to convince anyone of the message $m^* = 2\mu$ and loses the election, receiving 0.

If the opposition runs as a conventional candidate, his payoff is γR, $\gamma \sim U[0, 1]$.[5] Thus, the opposition runs as a populist if and only if

$$\pi[R - \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x) dx] \geq \gamma R. \tag{18}$$

[5] One can easily incorporate more incumbent advantage by assigning $\gamma \sim U[0, \bar{\gamma}]$, with small $\bar{\gamma}$ indicating a large incumbent advantage when the opposition runs as a conventional candidate. The parameter $\bar{\gamma}$ can be a function of μ, the incumbent advantage if the opposition runs as a populist. All main results stay the same when $\bar{\gamma}$ is a function of μ.

We obtain the key object of the model, the probability that the opposition runs as a populist:

$$\Pi \equiv \frac{\pi[R - \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x) dx]}{R}. \tag{19}$$

4 Comparative Statics

4.1 Targeted media, the voting cost, and populism

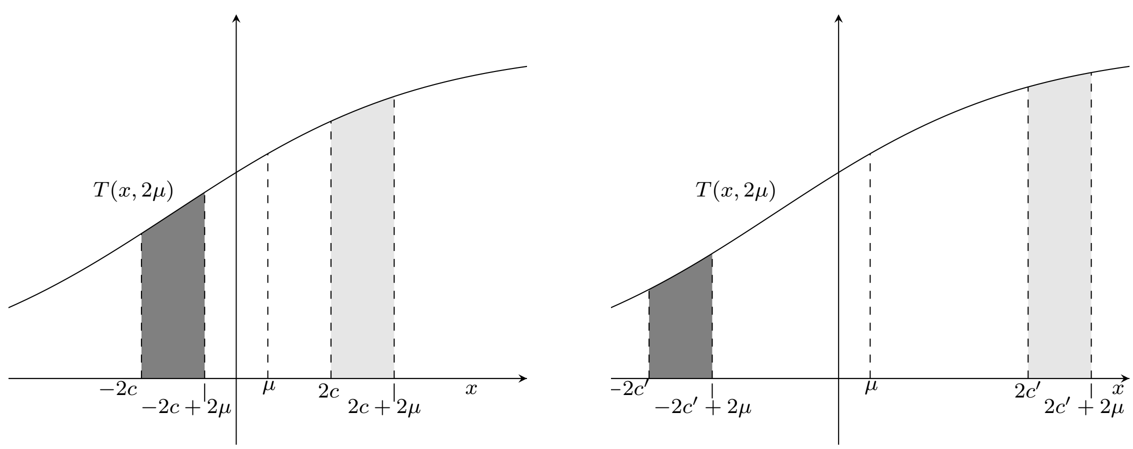
The next proposition is the first key result of the paper. It states comparative statics on Π, the probability of populism, with respect to targeted media and the voting cost.

Proposition 4. *1.* $\Pi_c > 0$*: the opposition is more likely to run as a populist if the voting cost is higher (c increases).*

2. $\Pi_\pi > 0$*: the opposition is more likely to run as a populist if the targeting technology improves (π increases).*

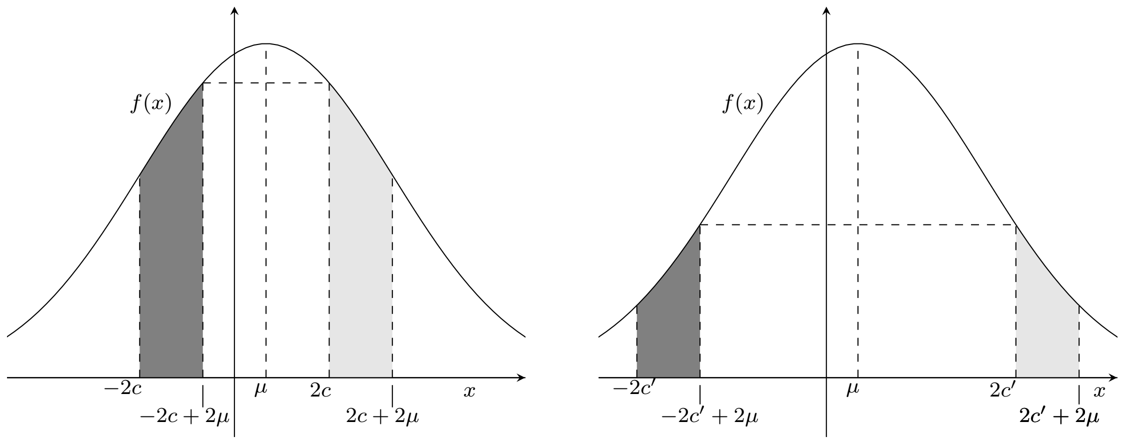
3. The cross derivative $\Pi_{\pi c} > 0$*: in a democracy with a higher voting cost, the marginal effect of targeted media in producing populism is larger.*

Figure 5: A higher voting cost replaces an "angry" audience with an even "angrier" one



Notes: the two panels show how the cost to persuade targeted voters shifts when the voting cost increases from c (the left panel) to $c' > c$ (the right panel). In both panels, the darkly shaded area is the cost to persuade the optimal audience under a uniform distribution, while the lightly shaded area includes all other voters that could be persuaded but do not receive a populist message. The horizontal axis is voter satisfaction x, and the vertical axis is the persuasion spending $T(x, 2\mu)$.

Figure 6: A higher voting cost narrows the optimal audience



Notes: the two panels show that the optimal audience shrinks when the voting cost increases from c (the left panel) to $c' > c$ (the right panel). In both panels, the darkly shaded area is the optimal audience, while the lightly shaded area includes all other voters that could be persuaded but do not receive a populist message. The horizontal axis is voter satisfaction x, and the vertical axis is the probability density $f(x)$.

A high voting cost promotes populism. The first result shows that a high voting cost produces strong populism. There are two effects that work in the same direction. First, a higher voting cost replaces an already angry audience with another audience that is even angrier with the incumbent. It becomes cheaper for the populist to convince the angrier audience, further reducing the persuasion spending. This effect is illustrated in Figure 5. As before, the darkly shaded area is the optimal audience, and the lightly shaded area includes other voters in the persuadable audience. Notice that the first effect is driven by the key result of Proposition 3, the result that a populist only targets his message to marginal populist voters, never attempting to persuade marginal incumbent voters. Had the optimal audience included marginal incumbent voters, a higher voting cost would dictate the populist to also persuade a group of voters who are even more enthusiastic supporters of the incumbent, a competing effect that increases the persuasion spending (see detailed analysis in Section 5). The first effect only unambiguously benefits a populist because the optimal audience is highly homogeneous.

The second effect is illustrated in Figure 6: a higher voting cost ($c' > c$) also narrows the optimal audience for the populist message. When it is more costly to vote, the populist only needs to target an even smaller audience to secure the election, further reducing the persuasion spending. This effect is driven by the "centrist" distribution of voter preference, with most voters centering around the average voter. In a twist, with a higher voting cost, the society's intrinsic centrism actually contributes to populism. We discuss more extensively on the paradoxical role of a centrist society in the next section (Section 4.2).

These two effects work together that exacerbate each other, since the populist always chooses a homogeneous audience (Proposition 3). Under a higher voting cost, the optimal audience becomes smaller, and the smaller audience is more disillusioned with the incumbent. Therefore, we have uncovered the mechanism behind our key claim, the claim that digital populism is hostile towards a broad participation of voters.

Targeted media promotes populism. The second result of Proposition 4 shows that targeted media might also promote populism, in line with existing literature (e.g., Della Lena et al. (2025); Szeidl and Szucs (2025)). When a populist can only access the national media, the populist would have wasted most of the persuasion spending on irrelevant voters. But when the targeting technology is effective, the populist targets a narrow and receptive audience, minimizing the persuasion spending (Proposition 3). Through targeted media, populist persuasion can secure the election in a cost-efficient manner.

The complementarity between targeted media and the voting cost. In the third result of Proposition 4, the inequality $\Pi_{\pi c} > 0$ shows that more effective targeting technology increases the marginal effect of a higher voting cost in producing populism. A higher voting cost reduces the persuasion spending only when the populist can precisely target his message. Thus, when targeted media is highly effective, populists are especially enthusiastic about voter suppression.

For the other side of the inequality $\Pi_{\pi c} > 0$, a higher voting cost increases the marginal effect of the targeted media in producing populism. With a higher voting cost, more effective targeting helps the populist save more on persuasion spending than the case with a lower voting cost. Thus, in democracies with a high voting cost, populists are also especially enthusiastic about targeted media. The complementarity explains the notable resistance of a few democracies against digital populism (Kaltwasser et al. (2017); Moffitt (2017); Brett (2019)). In these democracies, a low cost of voting induces a large and moderate audience that a populist must persuade. Even armed with targeted media, it is costly to persuade such a large and moderate audience. Note that because the complementarity relies on $\Pi_c > 0$ (a high voting cost promotes populism), the complementarity is also driven by the high homogeneity of the populist audience.

Remark: the complementarity and the "Anglophone divergence." The complementarity may also help understand another empirical puzzle: the qualitative literature has highlighted an "Anglophone divergence" between Australia and the United States in the strength of populism (Brett (2019)). The literature has underscored compulsory voting as a major institution in Australia that deters populism, but this cannot easily explain why populism was also weak in the United States until the 2010s. Our model, however, highlights that the benefit of compulsory voting was only activated in the age of digital media, therefore incorporating both the temporal and spatial variations in the Anglophone divergence. It was not compulsory voting *per se*, but its interaction with digital media, that is decisive.

Remark: endogenous voter suppression. The voting cost is an exogenous parameter in the main model. It is straightforward to endogenize it by extending the main model. After winning an election, a populist leader can choose to increase the voting cost from $2c$ to $2(c+e)$ by paying a cost of $Q(e)$. Such a populist leader will indeed choose to raise the voting cost, which benefits his re-election campaign. This captures the key insight that a digital populist in power is a significant threat to *electoral* democracy.

Because this insight is a direct application of Proposition 4, we leave the technical and detailed exposition to Appendix B.

4.2 Voter polarization and digital populism

All our previous results are obtained under a symmetric, unimodal distribution of voter preference. Such a distribution is centrist, with "most" voters clustering around the mean (also the mode) and relatively few voters with extreme preference. Thus, the model has already predicted that even a centrist society is not immune to populist forces. The prediction engages with the large literature on voter polarization (Fiorina and Abrams (2008); Barber et al. (2015); McCarty (2019)), which finds that even though American political elites are becoming highly polarized, American citizens have long remained centrist and moderate. Thus, our model might help think about why digital populism can arise in a centrist society.

To dive deeper into how a centrist society affects the emergence of populism, this section analyzes how populism responds to a *change* in voter polarization. We obtain another key result of the paper, which might seem more counter-intuitive: a more centrist distribution of voters may actually *encourage* digital populism. We characterize the condition for this paradoxical result, showing that the key driver is again the interaction between high voting cost and targeting technology.

To describe how centrist a society is in a general way, we look at a space of symmetric, unimodal probability density functions with full support on the real line ($x \in \mathbb{R}$). The space is indexed by α:

$$\{f(x;\alpha)\}_{\alpha \in \mathbb{R}^+}. \tag{20}$$

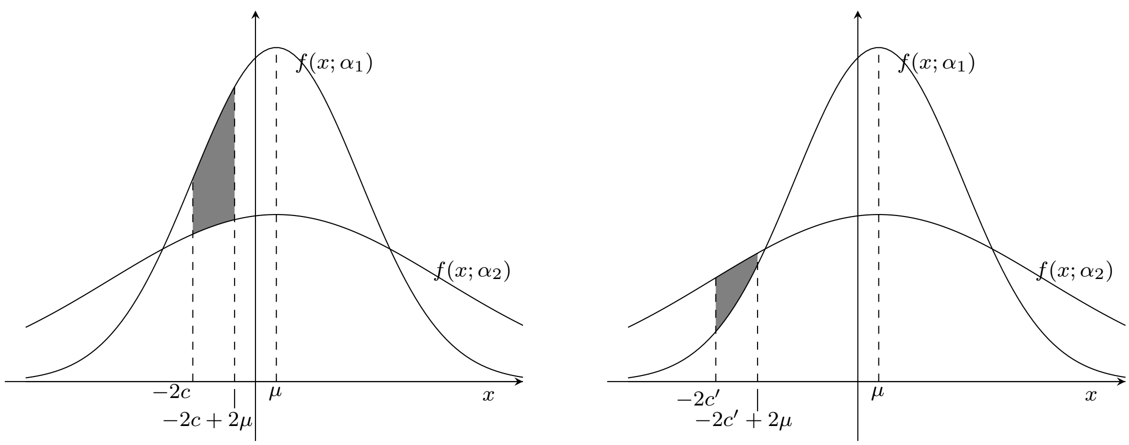
The density functions are ordered: for any $\alpha_1 > \alpha_2 > 0$,

$$|f'(x;\alpha_1)| > |f'(x;\alpha_2)| \text{ for all } x \in \mathbb{R}\backslash\{\mu\}. \tag{21}$$

Such an ordering implies that $\{f(x;\alpha)\}$ is ranked by second-order stochastic dominance. A larger α means that voters are more clustered around the mean μ, therefore *more centrist*.[6]

With respect to α, we will derive the comparative statics of Π, the probability that the opposition runs as a populist. To do this, notice that the domain $x \in (-\infty, \mu]$, the probability density functions $\{f(x;\alpha)\}$ satisfy a version of the single crossing property. And because these functions are symmetric probability density functions, any two of them must cross exactly once on $x \in (-\infty, \mu]$. This intuitive result is proved in the following proposition, allowing us to sharply characterize the diametrically different effects of a centrist society on populism under different voting costs.

Figure 7: Under low/high voting cost, a more centrist distribution of voters deters/encourages populist forces.



Notes: the two panels show how a more centrist distribution of voters may produce opposite effects on digital populism, depending on the voting cost. In both panels, the darkly shaded area is the change in voting cost when the population becomes more centrist (i.e., the voter distribution moves from $f(x;\alpha_2)$ to $f(x;\alpha_1)$, where $\alpha_1 > \alpha_2$). The left panel shows the change under a low voting cost c; the right panel shows the change under a high voting cost $c' > c$. The horizontal axis in either panel is voter satisfaction x, and the vertical axis is the probability density $f(x;\alpha)$.

Proposition 5. *For any $\alpha_1 > \alpha_2$, there exists a "cutoff" $\mu^* \in (-\infty, \frac{\mu}{2}]$ so that the followings are true.*

[6] As an example, if we restrict the space to normal distributions $\{f(x;\mu,\sigma^2)\}$ with mean μ and variance σ^2, we can assign α to be $1/\sigma^2$, the inverse of the variance. But for more general distributions, the parameter α may capture extra information than variance or second-order stochastic dominance.

1. *If* $c < -\mu^*$, $\Pi(\alpha_1) < \Pi(\alpha_2)$. *In other words, when the voting cost is sufficiently low, a more centrist society further deters digital populism.*

2. *If* $c > \mu - \mu^*$, $\Pi(\alpha_1) > \Pi(\alpha_2)$. *In other words, when the voting cost is sufficiently high, a more centrist society further encourages digital populism.*

The results are illustrated in Figure 7. The shaded area is the *change* of optimal targeted audience when the probability density function shifts from $f(x; \alpha_2)$ to $f(x; \alpha_1)$. We focus on the domain $x \in (-\infty, \mu]$ where all interesting actions happen. As voters become more clustered around the "average voter," there must be a lower probability density of voters somewhere else. Under the ordering 5, a more centrist distribution must increase the probability density above a threshold, while simultaneously shrinking the probability density below a threshold. In other words, as the society becomes more centrist, the local effect on probability density is exactly the opposite for preferences below and above the threshold. When the populist can target his message, what matters is the *local* probability density of his ideal audience. The populist pays no attention to probability density elsewhere. Persuasion is more expensive only when the local probability density of the populist audience increases.

How local probability density behaves is determined by the voting cost. When the voting cost is sufficiently low, all voters in the targeted audience are more moderate than the threshold voter. A more centrist distribution raises the size of the targeted audience, making populist persuasion more expensive. But when the voting cost is sufficiently high, all voters in the targeted audience are more extreme than the threshold voter. A more centrist distribution shrinks the size of the targeted audience, making populist persuasion even cheaper. In sum, a more centrist society encourages populism under a high voting cost, but only when the populist can target his message.

Proposition 5 uncovers a *possibility* result: it does not assert that a centrist society always supports digital populism, as made clear by the first part of Proposition 5. Nevertheless, the second part of Proposition 5 does question the traditional wisdom that a strong consensus among voters unconditionally induces democracy to produce desirable outcomes (Tocqueville (2000); Lipset (1959); Moore (1993)). This notion is appealing, but our model shows that the notion may not work in the age of targeted media, particularly so in democracies with low voter turnouts. While the traditional wisdom focuses more on the *global* distribution of voters, under targeted media, only the *local* distribution matters.

Proposition 5 also uncovers a nuanced interaction between a low voting cost and a centrist society in deterring populism. Interpreted through Proposition 5, compulsory voting in Australia might have amplified the effect of the strong centrist tendency among Australian voters, therefore producing an exceptionally successful resistance against populism in Australia so far (Moffitt (2017); Brett (2019)). Our analysis also uncovers structural reasons why such successes are not guaranteed to continue: media targeting is becoming more effective, and, less obviously, a rise in voter polarization may especially benefit populists *under compulsory voting* (Part 1 of Proposition 5).

For completeness, we also analyze how the average voter satisfaction (μ) affects populism. The result is in Appendix C. With some caveats, the model confirms that strong digital populism can be supported by a population that is already highly unsatisfied with the incumbent, the case with economic recession or stagnation. Because the insight is more conventional, we leave it to an appendix.

5 Populism and traditional negative campaigns

In our analysis of populism, we have assumed that each voter is essentially saturated in their repulsion or enthusiasm over populist messages above a threshold (Assumption 1). This assumption drives a homogeneous audience as optimal for a populist, supporting the central populist rhetoric of "us" versus "them" (Guriev and Papaioannou (2022)) and the appeal of voter suppression to a digital populist (Proposition 4).

What if we discard Assumption 1? The optimal audience can indeed be highly heterogeneous in their evaluation of the establishment incumbent, therefore inconsistent with the homogeneous populist attack on the establishment (Gratton and Lee (2025); Szeidl and Szucs (2025)) and the populist rhetoric of "us" versus "them" (Guriev and Papaioannou (2022)). So this alternative setup does not capture populism. The setup, however, might be a good description of traditional negative campaigning, as our analysis will show.

Specifically, we keep the same setup as the main model, with the only exception that Assumption 1 is replaced with the following one:

Assumption 2.

$$\int_{-2c}^{-2c+2\mu} T_m(x, 2\mu) f(x) dx > f(2c)[T(2c, 2\mu) - T(-2c + 2\mu, 2\mu)]. \tag{22}$$

Under Assumption 2, voters become a lot more suspicious of a message more critical of the incumbent (the derivative T_m is large). This contrasts with the main model, where voters find little difference among populist messages above a threshold. We can prove the following result, which applies better to a traditional negative campaigner.

Proposition 6. *Suppose that Assumption 2 holds.*

1. *The optimal message m^* satisfies $m^* < 2\mu$, and the optimal audience is*

$$[-2c, -2c + m^*] \cup [2c, 2\mu - (m^* - 2c)]. \tag{23}$$

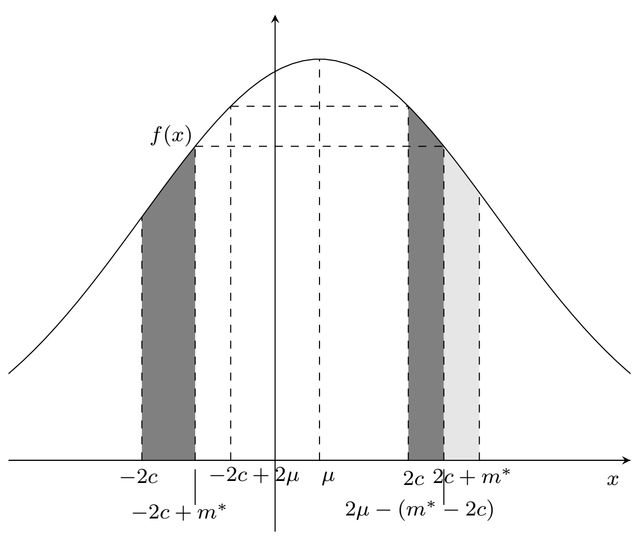
   In other words, the negative campaigner needs to target the optimal message m^* to both his own base and the base of the incumbent.

2. *Denote Π as the probability that the opposition runs as a negative campaigner.*

   (a) The sign of Π_c is ambiguous because, under a higher voting cost (c increases), the negative campaigner also needs to target more enthusiastic supporters of the incumbent.

   (b) $\Pi_\pi > 0$: negative messaging is more likely under a better targeting technology.

Figure 8: For traditional negative campaigns, the optimal audience is heterogeneous.



Notes: the figure shows that, under traditional negative campaigns, the optimal audience for the optimal message m^* is heterogeneous. The darkly shaded area is the optimal audience; the lightly shaded area includes all other voters that could be persuaded, but the negative campaigner does not send the message m to them. The horizontal axis is voter satisfaction x, and the vertical axis is the probability density $f(x)$.

A heterogeneous audience. In Proposition 6, the optimal message m^* balances between two effects, therefore less inflaming than the optimal populist message at 2μ. A message more critical of the incumbent allows the negative campaigner to replace marginal incumbent voters with voters in his own base, the latter far more receptive to his criticism of the

Table 1: Populism and negative campaigns in the digital era

	Populism	Traditional negativism
The targeted audience	homogeneous	heterogeneous
Opponent candidates attacked?	yes	yes
Opponent voters attacked?	yes	no
Thrives in a flawed democracy?	yes	much less so

incumbent. Yet a more critical message is also met with *significantly* stronger resistance from all targeted voters, raising the cost of persuasion (Assumption 2). Balancing the two effects, the optimal message makes it necessary for the negative campaigner to energize marginal voters in his base and also to persuade marginal incumbent voters to abstain. This audience of two distinct groups is highly heterogeneous (the darkly shaded area in Figure 8), in contrast to the homogeneous audience for a populist message.

A high voting cost produces an ambiguous effect on negative campaigning. Precisely because the audience is heterogeneous, a high voting cost produces an ambiguous effect on negative campaigning. A higher voting cost also necessitates the persuasion of more enthusiastic voters for the incumbent, raising the persuasion spending for a negative campaigner. This significantly reduces the benefit of voter suppression to a negative campaigner, different from the unambiguous appeal to a populist.

Yet a negative campaigner still unambiguously benefits from targeting technology, just like a populist. Both types of political actors save on their persuasion spending when they can target pivotal voters.

The heterogeneous audience and the avoidance of "us"-versus-"them" rhetoric. As an informal remark, a negative campaigner should find it challenging to spread a rhetoric of "us" versus "them," precisely because his message must be sent to two very different groups of voters. This contrasts with the main model, where the highly homogeneous audience allows the populist to demarcate the voters into "us" versus "them," without worrying about backlashes from voters in the targeted audience.

Summary: a comparison between negative campaigning and populism. Table 1 summarizes the comparison between populism and traditional negative campaigning. Because populism targets a highly homogeneous audience, a populist can attack both his establishment opponent and voters for the opponent, therefore supporting a divisive rhetoric of "us" versus "them" *among the voters*. This divisive rhetoric is less feasible for a traditional negative campaigner because his optimal audience is highly heterogeneous.

Equally important, it is the homogeneous audience that causes digital populism to thrive in a flawed democracy that limits access to ballots. Limits to ballot access is much less helpful to a negative campaigner, if any. This is because the heterogeneous audience of a negative campaigner includes supporters of his opponent. The last row of Table 1 might explain why digital populism seems to be far more pervasive in democracies already under stress, while negative campaigning is common in all types of democracies.

We can also allow a negative campaigner in power to endogenously suppress voters, as a populist leader does at the end of Section 4. In this case, a negative campaigner in power has a much weaker incentive to suppress voters, again in sharp contrast to a populist leader. To summarize, a comparative analysis of digital populism and negative campaigning further shed light on their distinct features, especially in their sharply different interactions with electoral democracy.

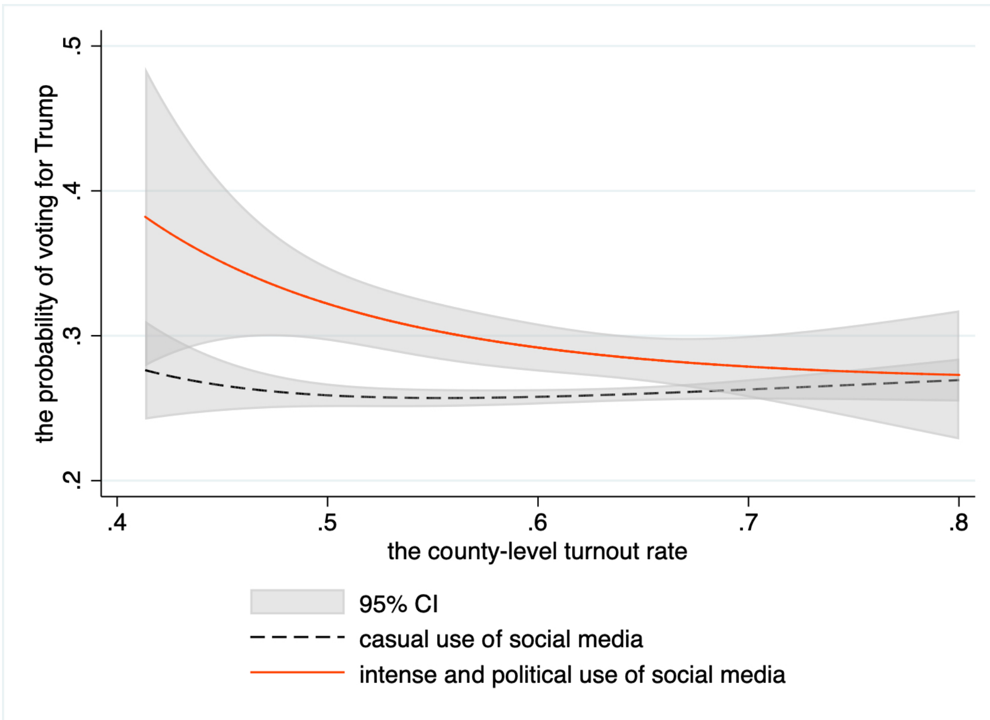
6 Conclusion

Different from traditional negative campaigners, populists in the digital area thrive through voter suppression, which allows them to win an electoral "majority" that is, in fact, a small and extreme subset of the broader public. Therefore, a particularly powerful weapon to deter digital populism is high voter participation – we need more democracy, not less, to activate the centrist tendency of ordinary voters (McCarty (2019)).

Our theoretical model might also have broad empirical relevance. Though systematic empirical analysis is beyond the scope of this paper, Figure 9 shows that, during the first ascendance of Trump presidency in 2016, "intense and political" use of social media was correlated with a higher propensity to vote for Trump, but only so in counties with low turnouts. Such correlation is indeed much more attenuated in counties with high turnouts. These empirical patterns suggest that a high voting cost might indeed interact with digital media in a significant manner, potentially inducing populists to run for office. Regressions and more details are presented in Appendix D, and we leave systematic empirical work for future research.

Our tractable model can potentially be employed to study other questions about populism. For instance, a recent literature begins to pay significant attention to the interaction between political institutions and cultural factors (Cantoni et al. (2018); Besley (2020); Gorodnichenko and Roland (2021); Bisin and Verdier (2024)). Specific to our paper, it might be fruitful to formalize how the turnout decision is affected by the culture and value of "civic duty" in voting, which has long been recognized as first order in the turnout decision (Feddersen (2004); Herrera et al. (2016)). By integrating "civic duty" into our framework through, for example, the cultural transmission model (Bisin and Verdier (2001); Della Lena and Panebianco (2021)), such a model would be able to analyze how civic culture interact with electoral institutions and media environment in affecting populism.

Figure 9: The estimated probability of a voter supporting Donald Trump in the 2016 American presidential election



Votes in 2016 presidential election and use of social media are from the 2016 Congressional Cooperative Election Survey. The county-level turnout rate in 2016 is the ratio of the total votes in a county (from MIT Election Lab) to the estimate of Citizen Voting Age Population of the county (from the 2012-2016 American Community Survey). Data are fit to fractional polynomials, and turnout rates are winsorized at the 1% level. More details and regression analysis are in Appendix D.

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2013a. A political theory of populism. *The Quarterly Journal of Economics* **128**: 771–805.

Acemoglu, Daron, James A Robinson, and Ragnar Torvik. 2013b. Why do voters dismantle checks and balances? *Review of Economic Studies* **80**: 845–875.

Algan, Yann, Sergei Guriev, Elias Papaioannou, and Evgenia Passari. 2017. The European trust crisis and the rise of populism. *Brookings papers on economic activity* **2017**: 309–400.

Australian Electoral Commission. 2023. Voter turnout – previous events. `https://www.aec.gov.au/elections/federal_elections/voter-turnout.htm`. Updated: 7 November 2023.

Barber, Michael, Nolan McCarty, Jane Mansbridge, and Cathie Jo Martin. 2015. Causes and consequences of polarization. *Political Negotiation: A Handbook* **37**: 39–43.

Beraja, Martin, Andrew Kao, David Y Yang, and Noam Yuchtman. 2023a. AI-tocracy. *The Quarterly Journal of Economics* **138**: 1349–1402.

Beraja, Martin, David Y Yang, and Noam Yuchtman. 2023b. Data-intensive innovation and the state: Evidence from AI firms in China. *The Review of Economic Studies* **90**: 1701–1723.

Bernhardt, Dan, Stefan Krasa, and Mehdi Shadmehr. 2022. Demagogues and the economic fragility of democracies. *American Economic Review* **112**: 3331–3366.

Besley, Timothy. 2020. State capacity, reciprocity, and the social contract. *Econometrica* **88**: 1307–1335.

Bisin, Alberto, and Thierry Verdier. 2001. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic theory* **97**: 298–319.

Bisin, Alberto, and Thierry Verdier. 2024. On the joint evolution of culture and political institutions: Elites and civil society. *Journal of Political Economy* **132**: 1485–1564.

Brennan, Jason. 2016. *Against Democracy*. Princeton, NJ: Princeton University Press.

Brett, Judith. 2019. *From Secret Ballot to Democracy Sausage: How Australia Got Compulsory Voting*. Text Publishing.

Callander, Steven, and Juan Carlos Carbajal. 2022. Cause and effect in political polarization: A dynamic analysis. *Journal of Political Economy* **130**: 825–880.

Cantoni, Davide, Jeremiah Dittmar, and Noam Yuchtman. 2018. Religious competition and reallocation: The political economy of secularization in the Protestant Reformation. *The Quarterly Journal of Economics* **133**: 2037–2096.

Dal Bo, Ernesto, Frederico Finan, Olle Folke, Torsten Persson, and Johanna Rickne. 2023. Economic and social outsiders but political insiders: Sweden's populist radical right. *The Review of Economic Studies* **90**: 675–706.

Danieli, Oren, Noam Gidron, Shinnosuke Kikuchi, and Ro'ee Levy. 2022. Decomposing the Rise of the Populist Radical Right .
URL `https://ssrn.com/abstract=4255937`

Della Lena, Sebastiano, Luca Paolo Merlino, and Yves Zenou. 2025. Affective polarization, media outlets, and opinion dynamics. Revise and resubmit, *Review of Economic Studies*.

Della Lena, Sebastiano, and Fabrizio Panebianco. 2021. Cultural transmission with incomplete information. *Journal of Economic Theory* **198**: 105373.

Fearon, James D. 2011. Self-enforcing democracy. *The Quarterly Journal of Economics* **126**: 1661–1708.

Feddersen, Timothy J. 2004. Rational choice theory and the paradox of not voting. *Journal of Economic perspectives* **18**: 99–112.

Fiorina, Morris P, and Samuel J Abrams. 2008. Political polarization in the American public. *Annual Review of Political Science* **11**: 563–588.

Gehlbach, Scott. 2021. *Formal models of domestic politics*. Cambridge University Press.

Gentzkow, Matthew. 2016. Polarization in 2016. *Toulouse Network for Information Technology Whitepaper* : 1–23.

Goldfarb, Avi, and Catherine Tucker. 2019. Digital economics. *Journal of Economic Literature* **57**: 3–43.

Gorodnichenko, Yuriy, and Gerard Roland. 2021. Culture, institutions and democratization. *Public Choice* **187**: 165–195.

Gratton, Gabriele, and Barton E Lee. 2024. Liberty, security, and accountability: The rise and fall of illiberal democracies. *Review of Economic Studies* **91**: 340–371.

Gratton, Gabriele, and Barton E Lee. 2025. Drain the swamp: A theory of anti-elite populism. *BAFFI Centre Research Paper* .

Guriev, Sergei, and Elias Papaioannou. 2022. The political economy of populism. *Journal of Economic Literature* **60**: 753–832.

Herrera, Helios, Massimo Morelli, and Salvatore Nunnari. 2016. Turnout across democracies. *American Journal of Political Science* **60**: 607–624.

Kaltwasser, Cristóbal Rovira, Paul A Taggart, Paulina Ochoa Espejo, and Pierre Ostiguy (eds.) . 2017. *The Oxford Handbook of Populism*. Oxford University Press.

Levy, Gilat, Ronny Razin, and Alwyn Young. 2022. Misspecified politics and the recurrence of populism. *American Economic Review* **112**: 928–962.

Lipset, Seymour Martin. 1959. Some social requisites of democracy: Economic development and political legitimacy. *American Political Science Review* **53**: 69–105.

McCarty, Nolan. 2019. *Polarization: What Everyone Needs to Know®*. Oxford University Press.

Moffitt, Benjamin. 2017. Populism in Australia and New Zealand. In Kaltwasser, Cristóbal Rovira, Paul A Taggart, Paulina Ochoa Espejo, and Pierre Ostiguy (eds.) *The Oxford Handbook of Populism*, chapter 6. Oxford: Oxford University Press.

Moore, Barrington. 1993. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Beacon Press.

Mukand, Sharun W, and Dani Rodrik. 2020. The political economy of liberal democracy. *The Economic Journal* **130**: 765–792.

Noury, Abdul, and Gerard Roland. 2020. Identity politics and populism in Europe. *Annual Review of Political Science* **23**: 421–439.

Peitz, Martin, and Markus Reisinger. 2015. The economics of Internet media. In *Handbook of Media Economics*, volume 1. Elsevier, 445–530.

Persson, Torsten, and Guido Tabellini. 2002. *Political economics: explaining economic policy*. MIT press.

Sasso, Greg, and Massimo Morelli. 2021. Bureaucrats under populism. *Journal of Public Economics* **202**: 104497.

Skaperdas, Stergios, and Bernard Grofman. 1995. Modeling negative campaigning. *American Political Science Review* **89**: 49–61.

Szeidl, Adam, and Ferenc Szucs. 2025. A model of populism as a conspiracy theory. *American Economic Review* **115**: 3214–3247.

Taggart, Paul. 2017. Populism in Western Europe. In Kaltwasser, Cristóbal Rovira, Paul A Taggart, Paulina Ochoa Espejo, and Pierre Ostiguy (eds.) *The Oxford Handbook of Populism*, chapter 13. Oxford: Oxford University Press.

Tocqueville, Alexis de. 2000. *Democracy in America*. The University of Chicago Press. Harvey C. Mansfield and Delba Winthrop (trans.).

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Profile books.

A Proofs

Lemma 1. *Suppose that the populist spends enough to convince all voters of a message m. The message m can only change the voting behavior of voters with:*

$$x \in \bar{A}(m) \equiv [-2c, -2c + m] \cup [2c, 2c + m]. \tag{12}$$

Proof. Before the population is convinced of the message m, the voting behavior is:

$$\begin{cases} \text{vote for the populist} & \text{if } x \in [\underline{x}, -2c) \\ \text{abstain} & \text{if } x \in [-2c, 2c) \\ \text{vote for the incumbent} & \text{if } x \in [2c, \bar{x}] \end{cases}$$

After the population is convinced of the message m, the voting behavior is:

$$\begin{cases} \text{vote for the populist} & \text{if } x \in [\underline{x}, -2c+m] \\ \text{abstain} & \text{if } x \in (-2c+m, 2c+m] \\ \text{vote for the incumbent} & \text{if } x \in (2c+m, \bar{x}] \end{cases}$$

Suppose that $-2c+m < 2c$, or $m < 4c$. We have five groups of voters:

$$\begin{cases} \text{always vote for the populist} & \text{if } x \in [\underline{x}, -2c) \\ \text{abstain} \to \text{populist vote} & \text{if } x \in [-2c, -2c+m] \\ \text{always abstain} & \text{if } x \in (-2c+m, 2c) \\ \text{incumbent vote} \to \text{abstain} & \text{if } x \in [2c, 2c+m] \\ \text{always vote for the incumbent} & \text{if } x \in (2c+m, \bar{x}] \end{cases}$$

The group that changes its voting behavior is:

$$\bar{A}(m) = [-2c, -2c+m] \cup [2c, 2c+m].$$

Suppose that $-2c+m \leq 2c$, or $m > 4c$. We also have five groups of voters:

$$\begin{cases} \text{always vote for the populist} & \text{if } x \in [\underline{x}, -2c) \\ \text{abstain} \to \text{populist vote} & \text{if } x \in [-2c, 2c] \\ \text{incumbent vote} \to \text{populist vote} & \text{if } x \in (2c, -2c+m) \\ \text{incumbent vote} \to \text{abstain} & \text{if } x \in [-2c+m, 2c+m] \\ \text{always vote for the incumbent} & \text{if } x \in (2c+m, \bar{x}] \end{cases}$$

The group that changes its voting behavior is:

$$\bar{A}(m) = [-2c, 2c+m] = [-2c, -2c+m] \cup [2c, 2c+m].$$

□

Proposition 1. *There exists a unique* $\bar{m} \in [2\mu, \infty]$ *and a unique* $\underline{m} \in [\mu, 2\mu)$*, such that the optimal audience as a function of the message* m *is:*

$$A^*(m) = \begin{cases} [-2c, -2c + 2\mu] & \text{for } m \in [2\mu, \bar{m}) \\ [-2c, -2c + m] \cup [2c, 2\mu - (m - 2c)] & \text{for } m \in [\mu, \underline{m}) \\ \emptyset & \text{for } m \in [0, \underline{m}) \cup [\bar{m}, \infty) \end{cases} . \tag{13}$$

Proof. We discuss three different cases.

1. Consider the case where $m \in [2\mu, \infty)$**.**

For any $m \in [2\mu, \infty)$, we first derive the minimal persuasion spending that secures the election. This gives us the payoff the populist obtains if he wins the election through targeted persuasion. We then compare the payoff from winning the election with deliberately losing the election by targeting no voters.

For any $m \in [2\mu, \infty)$, the minimal cost of targeted persuasion is:

$$T^*(m) \equiv \min_{A(m) \in \bar{A}(m)} \int_{A(m)} T(x, m) f(x) dx \tag{24}$$

such that

$$\int_{\left(A(m) \cap [-2c, -2c+m]\right) \cup [\underline{x}, -2c]} f(x) dx \geq \int_{[2c, \bar{x}] \setminus A(m)} f(x) dx. \tag{25}$$

The constraint 25 ensures that the populist will be elected. The left hand side is the vote share received by the populist. Voters with $x \in [\underline{x}, -2c]$ always support the populist. Voters with $x \in [-2c, -2c + m]$ also support the populist upon receiving the message m at an intensity $T(x, m)$. Thus, by sending the message to $A(m)$, the populist gains additional voters with $x \in [-2c, -2c + m] \cap A(m)$. The populist receives support from all voters with:

$$x \in \left(A(m) \cap [-2c, -2c + m]\right) \cup [\underline{x}, -2c].$$

The right hand side of constraint 25 is the vote share received by the incumbent. Voters with $x \in [2c, \bar{x}]$ support the incumbent had they not received the message m. Upon receiving the message, voters with $x \in A(m)$ withdraw their support of the incumbent. Thus, the incumbent receives support from all voters with:

$$x \in [2c, \bar{x}] \setminus A(m).$$

The optimality of $A^*(m) = [-2c, -2c + 2\mu]$ is intuitive, since these are voters who are the least satisfied with the incumbent in the persuadable audience $\bar{A}(m)$, while securing exactly 50% of total votes for the populist. Notice that by symmetry,

$$F(-2c + 2\mu) = 1 - F(2c).$$

$F(-2c+2\mu)$ is the vote share of the populist who chooses the audience $A^*(m)$, and $F(2c)$ is the vote share of the incumbent.

But the full proof is tedious, because we need to show that

$$A^*(m) \text{ produces the lowest persuasion cost among } \textit{all} \text{ Lebesgue measurable subsets of } \bar{A}(m). \tag{26}$$

We will show the case where $[-2c, -2c+m] \cap [2c, 2c+m] = \emptyset$. For the other case where $[-2c, -2c+m] \cap [2c, 2c+m] \neq \emptyset$, the proof is almost identical and omitted. We will also omit another almost identical step for $m < 2\mu$ later.

To show the claim 26, We proceed in two steps.

Step 1. *Any audience A' with $\int_{\left(A' \cap [-2c, -2c+m]\right) \cup [\underline{x}, -2c]} f(x)dx > \int_{[2c, \bar{x}] \setminus A'} f(x)dx$ is strictly suboptimal.*

Denote the Lebesgue measure of a subset of $\mathbb{R}$ as

$$\nu(\cdot).$$

Notice that $\nu(A') > 0$. There are two cases to consider: $\nu(A' \cap [-2c, -2c+m]) > 0$ and $\nu(A' \cap [2c, 2c+m]) > 0$. We focus on the first case. The proof for the second case is, again, almost identical.

Denote that $D = A' \cap [-2c, -2c+m]$. We have just assumed that $\nu(D) > 0$. We first prove the following lemma.

Lemma *Fix any $\epsilon > 0$. For any measurable set $E \subset [-2c, -2c+m]$ with $\nu(E) > 0$, there exists a subset $\Delta \subset E$ such that $\nu(\Delta) > 0$ and $\nu(\Delta) < \epsilon$.*

We first partition the set $[-2c, -2c+m)$. Note that for a fixed integer $h \geq \frac{2}{m}$, we can partition the set $[-2c, -2c+m)$ as follows:

$$[-2c, -2c+m) = \cup_{i=1}^{mh} [-2c + \frac{i-1}{h}, -2c + \frac{i}{h}).$$

For each interval $[-2c + \frac{i-1}{h}, -2c + \frac{i}{h})$, the measure is $\frac{2}{h} > 0$. All the partitions are disjoint. Thus, we can apply the additivity of Lebesgue measure, and note that the intersection of two measurable sets is measurable:

$$\nu(E) = \nu\Bigg(E \cap [-2c, -2c+m)\Bigg)$$

$$= \nu\Bigg(E \cap \Big\{ \cup_{i=1}^{mh} [-2c + \frac{i-1}{h}, -2c + \frac{i}{h})\Big\}\Bigg)$$

$$= \nu\Bigg(\cup_{i=1}^{mh} \Big\{E \cap [-2c + \frac{i-1}{h}, -2c + \frac{i}{h})\Big\}\Bigg)$$

$$= \Sigma_{i=1}^{mh}\nu\Bigg(E \cap [-2c + \frac{i-1}{h}, -2c + \frac{i}{h})\Bigg) > 0.$$

Denote $\Delta_i = E \cap [-2c + \frac{i-1}{h}, -2c + \frac{i}{h})$. Therefore there exists an $i' \in \{1, ..., mh\}$, such that $\nu(\Delta_{i'}) > 0$. Notice that $\nu(\Delta_{i'}) \leq \nu([-2c + \frac{i-1}{h}, -2c + \frac{i}{h})) = \frac{2}{h}$. Thus, Take $h^* = \max\{[\frac{4}{\epsilon}+1], [\frac{2}{m}+1]\}$, $\nu(\Delta_{i'}) \leq \frac{\epsilon}{2}$. Gather the facts that $\nu(\Delta_{i'}) > 0$, $\nu(\Delta_{i'}) < \epsilon$, and $\Delta_{i'} \subset E$, the lemma is proved.

Next, pick

$$0 < \epsilon' < \frac{\int_{(D\cup[\underline{x}, -2c])} f(x)dx - \int_{[2c,\bar{x}]\setminus A'} f(x)dx}{\sup_x f(x)}.$$

Apply the above lemma, we can find a subset of $D = A' \cap [-2c, -2c+m]$, denote as $\Delta_{\epsilon'}$, such that $\nu(\Delta_{\epsilon'}) > 0$ and $\nu(\Delta_{\epsilon'}) < \epsilon'$. Suppose the populist does not send the message to the set $\Delta_{\epsilon'} \subset D$. The votes that the populist receives is:

$$\int_{(D\cup[\underline{x}, -2c])-\Delta_{\epsilon'}} f(x)dx$$

$$= \int_{(D\cup[\underline{x}, -2c])} f(x)dx - \int_{\Delta_{\epsilon'}} f(x)dx$$

$$\geq \int_{(D\cup[\underline{x}, -2c])} f(x)dx - \nu(\Delta_{\epsilon'}) \sup_x f(x)$$

$$> \int_{(D\cup[\underline{x}, -2c])} f(x)dx - \epsilon' \cdot \sup_x f(x)$$

The populist still wins the election:

$$\int_{(D\cup[\underline{x}, -2c])} f(x)dx - \epsilon' \cdot \sup_x f(x),$$

$$> \int_{(D\cup[\underline{x},-2c])} f(x)dx - \frac{\int_{(D\cup[\underline{x},-2c])} f(x)dx - \int_{[2c,\bar{x}]\setminus A'} f(x)dx}{\sup_x f(x)} \cdot \sup_x f(x)$$

$$= \int_{[2c,\bar{x}]\setminus A'} f(x)dx.$$

And the populist generates a strictly positive saving on persuasion cost

$$\int_{(D\cup[\underline{x},-2c])-\Delta_{\epsilon'}} T(x,m)f(x)dx$$

$$= \int_{(D\cup[\underline{x},-2c])} T(x,m)f(x)dx - \int_{\Delta_{\epsilon'}} T(x,m)f(x)dx$$

$$\leq \int_{(D\cup[\underline{x},-2c])} T(x,m)f(x)dx - \overbrace{\nu(\Delta_{\epsilon'}) \cdot T(\inf \Delta_{\epsilon'}, m) f(\inf \Delta_{\epsilon'})}^{>0}$$

$$< \int_{(D\cup[\underline{x},-2c])} T(x,m)f(x)dx.$$

Step 2 *we show that the audience* $A^* = [-2c, -2c+2\mu]$ *reaches the lowest persuasion spending among any measurable set* A *such that* $\int_{\left(A\cap[-2c,-2c+m]\right)\cup[\underline{x},-2c]} f(x)dx = \int_{[2c,\bar{x}]\setminus A} f(x)dx.$

By contradiction, suppose that there exists a measurable set $A' \subset \bar{A}$ such that

1. $\nu(A' - A^*) > 0$,
2. $\int_{A'} T(x,m)f(x)dx < \int_{A^*} T(x,m)f(x)dx$,
3. $\int_{\left(A'\cap[-2c,-2c+m]\right)\cup[\underline{x},-2c]} f(x)dx = \int_{[2c,\bar{x}]-A'} f(x)dx.$

The last condition is obtained from Step 1. The last condition implies that:

$$\int_{A'} f(x)dx = \int_{A^*} f(x)dx. \tag{27}$$

Condition 27 implies that:

$$\int_{A^*-A'} f(x)dx = \int_{A'-A^*} f(x)dx \tag{28}$$

To prove 28, by definition,

$$\int_{A^*} f(x)dx = \int_{(A^*\cap A')\cup(A^*-A')} f(x)dx$$

$$= \int_{(A^*\cap A')} f(x)dx + \int_{(A^*-A')} f(x)dx.$$

Similarly, we have:

$$\int_{A'} f(x)dx = \int_{(A^* \cap A')} f(x)dx + \int_{(A' - A^*)} f(x)dx.$$

Recall that $\int_{A'} f(x)dx = \int_{A^*} f(x)dx$ (Equation 27), we obtain Equation 28.

Also, notice that for any $x' \in (A' - A^*)$, $x' \in \bar{A} = [-2c, -2c + m] \cup [2c, 2c + m]$ and $x' \notin [-2c, -2c + 2\mu]$. Thus,

$$\text{for any } x' \in (A' - A^*),\ x' > -2c + 2\mu. \tag{29}$$

We now compute a lower bound on the cost to persuade the audience A':

$$\begin{aligned}
\int_{A'} T(x,m)f(x)dx &= \int_{(A' \cap A^*) \cup (A' - A^*)} T(x,m)f(x)dx \\
&= \int_{A' \cap A^*} T(x,m)f(x)dx + \int_{A' - A^*} T(x,m)f(x)dx. \qquad (30) \\
&> \int_{A' \cap A^*} T(x,m)f(x)dx + T(-2c + 2\mu, m) \int_{A' - A^*} f(x)dx \\
&= \int_{A' \cap A^*} T(x,m)f(x)dx + T(-2c + 2\mu, m) \int_{A^* - A'} f(x)dx \\
&> \int_{A' \cap A^*} T(x,m)f(x)dx + \int_{A^* - A'} T(x,m)f(x)dx \\
&= \int_{A^*} T(x,m)f(x)dx.
\end{aligned}$$

A contradiction. This finishes the proof that an optimal audience is $A^*(m) = [-2c, -2c+2\mu]$. The proof also establishes that any other optimal audience can only differ from $A^*(m) = [-2c, -2c + 2\mu]$ by a measure of zero.

With $A^*(m) = [-2c, -2c + 2\mu]$, we obtain the maximal payoff to a populist who sends out a message $m \geq 2\mu$ to the audience:

$$U(m) \equiv R - \int_{-2c}^{-2c+2\mu} T(x,m)f(x)dx,$$

$U(m)$ is monotonically decreasing in m:

$$U'(m) = -\int_{-2c}^{-2c+2\mu} T_m(x,m)f(x)dx < 0.$$

We have assumed that:

$$U(2\mu) = R - \int_{-2c}^{-2c+2\mu} T(x,2\mu)f(x)dx > 0.$$

Thus, there exists a unique $\bar{m} \in (2\mu, \infty]$ such that:

$$\begin{cases} U(m) < 0 & \text{if } m \in (\bar{m}, \infty] \\ U(m) \geq 0 & \text{if } m \in (2\mu, \bar{m}] \end{cases}.$$

If $m \in (\bar{m}, \infty]$, the populist refuses to target his message m to any voters, obtaining a payoff of zero:

$$A^*(m) = \emptyset.$$

If $m \in [2\mu, \bar{m}]$, the populist targets the message m to:

$$A^*(m) = [-2c, -2c + 2\mu].$$

2. Consider the case where $\mu \in [\mu, 2\mu]$.

The populist solves a problem that is identical to the one in Equation 24 and Equation 25, with the only difference that $m \in [\mu, 2\mu]$.

The minimal persuasion cost that secures the election is:

$$T^*(m) \equiv \min_{A(m) \in \bar{A}(m)} \int_{A(m)} T(x,m)f(x)dx \tag{31}$$

such that

$$\int_{\left(A(m)\cap[-2c,-2c+m]\right)\cup[\underline{x},-2c]} f(x)dx \geq \int_{[2c,\bar{x}]\setminus A(m)} f(x)dx. \tag{32}$$

We now characterize the optimal audience for the problem 31-32. If $m = 2\mu$, we have shown that $A^*(m) = [-2c, -2c + 2\mu]$.

If $m < 2\mu$, then just targeting $[-2c, -2c + m]$ is not enough:

$$\text{the populist's vote share} = F(-2c + m)$$

$$< F(-2c + 2\mu) = 1 - F(2c) = \text{the incumbent's vote share}.$$

The populist also needs to target some voters with $x \in [2c, 2c+m]$.[7] Denote the targeted voter with the maximal satisfaction as $\hat{x}$. $\hat{x}$ should ensure that:

$$\text{the incumbent's vote share} = \int_{\hat{x}}^{\bar{x}} f(x)dx$$

$$\leq \int_{\underline{x}}^{m-2c} f(x)dx = \text{the populist's vote share},$$

$$1 - F(\hat{x}) \leq F(m-2c),$$

$$F(2\mu - \hat{x}) \leq F(m-2c),$$

$$\hat{x} \geq 2\mu - (m-2c).$$

Thus, the optimal $\hat{x}$ is:

$$\hat{x}^* = 2\mu - (m-2c).$$

Notice that $\hat{x^*} = 2\mu - (m-2c) \in [2c, 2c+m] \Leftrightarrow m \in [\mu, 2\mu]$. So $\hat{x}^*$ is feasible.

Thus, for $m \in [\mu, 2\mu]$, the optimal audience for the populist is:

$$A^*(m) = [-2c, -2c+m] \cup [2c, 2\mu - (m-2c)].$$

The minimal persuasion spending is:

$$T^*(m) = \int_{-2c}^{-2c+m} T(x,m)f(x)dx + \int_{2c}^{2\mu-(m-2c)} T(x,m)f(x)dx.$$

The minimal persuasion spending decreases with m:

$$T^*_m(m) = f(-2c+m)T(-2c+m,m) + \int_{-2c}^{-2c+m} T_m(x,m)f(x)dx$$

$$-f(2\mu-(m-2c))T(2\mu-(m-2c),m) + \int_{2c}^{2\mu-(m-2c)} T_m(x,m)f(x)dx$$

$$< f(-2c+m)T(-2c+m,m) - f(2\mu-(m-2c))T(2\mu-(m-2c),m) + \tilde{t}$$

$$= f(-2c+m)T(-2c+m,m) - f(-2c+m)T(2\mu-(m-2c),m) + \tilde{t}$$

$$= f(-2c+m)\overbrace{[T(-2c+m,m) - T(2\mu-(m-2c),m)]}^{<-2\underline{T}(\mu-m+2c)<0} + \tilde{t} < 0$$

[7]Note that when $m \in [\mu, 2\mu]$, the two intervals of the persuadable audience $\bar{A}(m) = [-2c, -2c+m] \cup [2c, 2c+m]$ do not overlap: $m \leq 2\mu$ and $\mu < 2c \Rightarrow m < 4c \Rightarrow -2c+m < 2c$.

for $\tilde{t}$ sufficiently small.

So the maximal payoff by sending out the populist message, $U(m) = R - T^*(m)$, increases with m. Also,

$$U(2\mu) = R - \int_{-2c}^{-2c+2\mu} T(x,m)f(x)dx - \int_{2c}^{2c} T(x,m)f(x)dx$$

$$= R - \int_{-2c}^{-2c+2\mu} T(x,m)f(x)dx > 0.$$

Therefore, there exists a unique $\underline{m} \in [\mu, 2\mu)$ such that:

$$\begin{cases} R - T^*(m) \geq 0 & \text{if } m \in [\underline{m}, 2\mu] \\ R - T^*(m) < 0 & \text{if } m \in [\mu, \underline{m}) \end{cases}.$$

When $m \in [\underline{m}, 2\mu]$, the populist targets the message m to $A^*(m) = [-2c, -2c+m] \cup [2c, 2\mu - (m-2c)$. When $m \in [\mu, \underline{m})$, the populist does not send the message to any voter, $A^*(m) = \emptyset$.

3. Consider the case where $m \in [0, \mu)$.

The populist cannot win even if he convinces all voters in the persuadable audience $\bar{A}(m) = [-2c, -2c+m] \cup [2c, 2c+m]$. The populist's vote share is

$$F(-2c+m) < F(-2c+\mu),$$

while the incumbent's vote share is:

$$1 - F(2c+m) > 1 - F(2c+\mu) = F(-2c+\mu).$$

Thus,

$$F(-2c+m) < 1 - F(2c+m).$$

Therefore, the populist should send the message to no voters: $A^*(m) = \emptyset$. □

Proposition 2. *Suppose that targeting fails. Then the populist spends nothing on uniform persuasion:*

$$\textit{for any } m \in \mathbb{R}^+,\ \hat{T}^*(m) = 0. \tag{14}$$

Proof. Denote

$$\underline{T}(m)$$

as the minimum spending that secures an election for the populist when she broadcast a message $m \in \mathbb{R}^+$ to all voters $[\underline{x}, \bar{x}]$.

The optimal uniform spending for $m \in [0, \mu)$ is zero. Suppose that the message $m \in [0, \mu)$. Even if the populist spends $T(\bar{x}, m)$ so that he convinced everyone of the message m, the populist cannot win the election:

$$\text{the populist's vote share} = F(-2c + m) < F(-2c + \mu)$$

$$= 1 - F(\mu + 2c) < 1 - F(m + 2c) = \text{the incumbent's vote share.} \tag{33}$$

$F(-2c + \mu) = 1 - F(\mu + 2c)$ because $f(\cdot)$ is symmetric around μ. Equation 33 show that $\underline{T}(m) = \infty$. Thus, for any $m \in [0, \mu)$, the populist sets $\hat{T}^*(m) = 0$ and loses the election:

$$\text{the populist's vote share} = F(-2c) = 1 - F(2\mu + 2c)$$

$$< F(2c) = \text{the incumbent's vote share.}$$

The populist's payoff is 0.

The optimal uniform spending for $m \in [2\mu, \infty)$ is zero. Suppose that the message $m \in [\mu, \infty)$. Then a populist could win if he spends $T(\bar{x}, m)$ and convince everyone of the message m. The populist could swing voters with

$$x \in \bar{A}(m) = [-2c, -2c + m] \cup [2c, 2c + m].$$

Further suppose that $m \in [2\mu, \infty)$. To win the election with minimal spending, the populist spends

$$\underline{T}(m) = T(-2c + 2\mu, m)$$

With such a spending, the populist can swing voters with

$$x \in [-2c, -2c + 2\mu].$$

The persuasion spending for an $m \in [2\mu, \infty)$ is:

$$\int_{\underline{x}}^{\bar{x}} T(-2c + 2\mu, m) f(x) dx = T(-2c + 2\mu, m).$$

Such a populist obtains a payoff lower than losing the election:

$$R - T(-2c + 2\mu, m) < R - T(-2c + 2\mu, 2\mu) < 0.$$

Thus, for any $m \in [2\mu, \infty)$, the populist sets $\hat{T}^*(m) = 0$.

The optimal uniform spending for $m \in [\mu, 2\mu]$ is zero. Suppose that the message $m \in [\mu, 2\mu)$. To win the election with minimal spending, the populist sets

$$\underline{T}(m) = T(2\mu - (m - 2c), m)$$

and swings voters with

$$x \in [-2c, -2c + m] \cup [2c, 2\mu - (m - 2c)].$$

$T(2\mu - (m - 2c), m)$ decreases with m:

$$\frac{\partial}{\partial m} T(2\mu - (m - 2c), m) = -T_x + T_m < -T_x + \tilde{t} < 0,$$

with $T_m < \tilde{t}$ by Assumption 1. Thus, with a message $m \in [\mu, 2\mu)$, the populist obtains a payoff:

$$R - T(2\mu - (m - 2c), m) \tag{34}$$

$$< R - T(2\mu - (2\mu - 2c), 2\mu) \tag{35}$$

$$= R - T(2c, 2\mu) \tag{36}$$

$$< R - T(-2c + 2\mu, 2\mu) < 0. \tag{37}$$

Equation 34 to Equation 35 because $T(2\mu - (m - 2c), m)$ decreases with m. Equation 36 to Equation 37 because $2c > -2c + 2\mu \Leftrightarrow c > 2\mu$.

Thus, for any $m \in [\mu, 2\mu)$, the populist sets $\hat{T}^*(m) = 0$. □

Proposition 3. *The optimal message is:*

$$m^* = 2\mu. \tag{15}$$

With probability π, the message $m^* = 2\mu$ convinces the optimal audience

$$A^*(m^*) = [-2c, -2c + 2\mu]. \tag{16}$$

With probability $1 - \pi$, the message $m^* = 2\mu$ convinces no voters.

Proof. As a preliminary step, notice that formulating a message $m \in [0, \underline{m}) \cup (\bar{m}, \infty)$ yields a payoff of 0, lower than formulating the message 2μ:

$$0 < \pi[R - \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x) dx].$$

Therefore,

$$m^* = \arg\max_{m \in [\underline{m}, \bar{m}]} \bar{V}(m),$$

where

$$\bar{V}(m) = \begin{cases} \pi[R - \int_{-2c}^{-2c+2\mu} T(x,m) f(x) dx] & \text{if } m \in [2\mu, \bar{m}) \\ \pi[R - \int_{-2c}^{-2c+m} T(x,m) f(x) dx - \int_{2c}^{2\mu-(m-2c)} T(x,m) f(x) dx]. & \text{if } m \in [\underline{m}, 2\mu] \\ 0 & \text{if } m \in [0, \underline{m}) \cup [\bar{m}, \infty]. \end{cases}.$$

First, $\bar{V}'(m) < 0$ for $m \in [2\mu, \bar{m})$:

$$\bar{V}'(m) = -\pi[\int_{-2c}^{-2c+2\mu} T_m(x,m) f(x) dx] < 0.$$

Second, $\bar{V}'(m) > 0$ for $m \in [\underline{m}, 2\mu]$:

$$\begin{aligned} \bar{V}'(m) = -\pi\Big\{ & T(-2c+m, m) f(-2c+m) + \int_{-2c}^{-2c+m} T_m(x,m) f(x) dx \\ & -T(2\mu-(m-2c), m) f(2\mu-(m-2c)) + \int_{2c}^{2\mu-(m-2c)} T_m(x,m) f(x) dx \Big\} \\ < -\pi\Big\{ & T(-2c+m, m) f(-2c+m) - T(2\mu-(m-2c), m) f(2\mu-(m-2c)) + \tilde{t} \Big\} \\ = -\pi\Big\{ & T(-2c+m, m) f(-2c+m) - T(2\mu-(m-2c), m) f(-2c+m) + \tilde{t} \Big\} \\ = -\pi\Big\{ & f(-2c+m) \underbrace{[T(-2c+m, m) - T(2\mu-(m-2c), m)]}_{<-2\underline{T}(\mu-m+2c)<0} + \tilde{t} \Big\} > 0. \end{aligned}$$

Thus, the opposition's payoff is maximized at

$$m^* = 2\mu.$$

Following Proposition 1, the optimal convinced audience is:

$$A^*(m^*) = [-2c, -2c+2\mu].$$

Following Proposition 2, when targeting fails, the opposition spends $\hat{T}(2\mu) = 0$ on the national media and loses the election. $\square$

Proposition 4. *1. $\Pi_c > 0$: the opposition is more likely to run as a populist if the voting cost is higher (c increases).*

2. $\Pi_\pi > 0$: the opposition is more likely to run as a populist if the targeting technology improves (π increases).

3. The cross derivative $\Pi_{\pi c} > 0$: in a democracy with a higher voting cost, the marginal effect of targeted media in producing populism is larger.

Proof.

$$\Pi_\pi(\pi, c) = \frac{[R - \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x) dx]}{R} > 0.$$

$$\Pi_c(\pi, c) = -\frac{\pi}{R} \cdot \frac{\partial}{\partial c} \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x) dx$$

$$= \frac{2\pi}{R} \cdot [T(-2c + 2\mu, 2\mu) f(-2c + 2\mu) - T(-2c, 2\mu) f(-2c)] > 0$$

as $T(-2c+2\mu, 2\mu) > T(-2c, 2\mu)$ and $f(-2c+2\mu) > f(-2c)$; $f(-2c+2\mu) > f(-2c)$ because $f(\cdot)$ is unimodal and is symmetric around μ, with $-2c + 2\mu < \mu$, or $2c > \mu$.

$$\Pi_{\pi c}(\pi, c) = \frac{2}{R} \cdot [T(-2c + 2\mu, 2\mu) f(-2c + 2\mu) - T(-2c, 2\mu) f(-2c)] > 0.$$

$\square$

Proposition 5. *For any $\alpha_1 > \alpha_2$, there exists a "cutoff" $\mu^* \in (-\infty, \frac{\mu}{2}]$ so that the followings are true.*

1. *If $c < -\mu^*$, $\Pi(\alpha_1) < \Pi(\alpha_2)$. In other words, when the voting cost is sufficiently low, a more centrist society further deters digital populism.*

2. *If $c > \mu - \mu^*$, $\Pi(\alpha_1) > \Pi(\alpha_2)$. In other words, when the voting cost is sufficiently high, a more centrist society further encourages digital populism.*

Proof. **Lemma 1** There exists $\underline{\mu} < \mu$, such that for all $x \in (\underline{\mu}, \mu]$, and $\alpha_1 > \alpha_2$, $f(x; \alpha_1) > f(x; \alpha_2)$.

By symmetry,

$$\int_{-\infty}^{\mu} f(x; \alpha_1) dx = \int_{-\infty}^{\mu} f(x; \alpha_2) dx = \frac{1}{2}.$$

Claim 1 There exists a set E of measure $\nu(E) > 0$, such that $f(x;\alpha_1) > f(x;\alpha_2)$.

By contradiction, suppose Claim 1 is not true. Further suppose $f(x;\alpha_1) = f(x;\alpha_2)$ for all $x \in (-\infty,\mu]\backslash\Delta$,with $\nu(\Delta) = 0$. Then $f'(x;\alpha_1) = f'(x;\alpha_2)$ for all $x \in (-\infty,\mu]/\Delta$, a contradiction. Then for all $x \in (-\infty,\mu]\backslash\Delta$, with $0 \leq \nu(\Delta) < \infty$, and $\nu((-\infty,\mu]\backslash\Delta) > 0$, $f(x;\alpha_1) < f(x;\alpha_2)$. For $x \in \Delta$, $f(x;\alpha_1) = f(x;\alpha_2)$. Therefore $\int_{-\infty}^{\mu} f(x;\alpha_1)dx > \int_{-\infty}^{\mu} f(x;\alpha_1)dx$. A contradiction.

Claim 2 There exists a connected set E' of measure $\nu(E') > 0$, such that $f(x;\alpha_1) > f(x;\alpha_2)$.

Pick any two $x_1, x_2 \in E$, $x_1 < x_2$, we want to show that for any $x \in (x_1,x_2)$, $g(x) \equiv f(x;\alpha_1) - f(x;\alpha_2) > 0$. By contradiction, suppose that there exists $x_1, x_2 \in E$, such that for an $x' \in (x_1,x_2)$, $g(x') = f(x;\alpha_1) - f(x;\alpha_2) \leq 0$. By continuity, there exists an $\tilde{x} \in (x_1,x')$ such that $g'(\tilde{x}) = \frac{g(x')-g(x_1)}{x'-x_1} < 0$, or $g'(\tilde{x}) = f'(\tilde{x};\alpha_1) - f'(\tilde{x};\alpha_2) < 0$, a contradiction. We establish Claim 2 by setting $E' = E \cup \{x \in \mathbb{R} : x = x_1 + x_2, x_1 \in E, x_2 \in E\}$. Notice that because E' is connected with $\nu(E') > 0$, it is an interval. Also, $\inf E' < \mu$.

Claim 3 For all $x \in (\inf E', \mu]$, $f(x;\alpha_1) > f(x;\alpha_2)$.

Note that $g(\inf E') \geq 0$ and $g'(x) > 0$ for $x \in (-\infty,\mu]$. Claim 3 is immediately established.

Let $\underline{\mu} = \inf E$. Lemma 1 is proved.

Lemma 2 There exists $\hat{\mu} < \mu$, such that for all $x \in (-\infty,\hat{\mu})$, and $\alpha_1 > \alpha_2$, $f(x;\alpha_1) < f(x;\alpha_2)$.

Claim 1 There exists a set D of measure $\nu(D) > 0$, such that $f(x;\alpha_1) < f(x;\alpha_2)$.

Because $\int_{\underline{\mu}}^{\mu} f(x;\alpha_1)dx > \int_{\underline{\mu}}^{\mu} f(x;\alpha_2)dx$ and $\int_{-\infty}^{\mu} f(x;\alpha_1)dx = \int_{-\infty}^{\mu} f(x;\alpha_2)dx$, there must be a measurable set D with $\nu(D) > 0$, such that $f(x;\alpha_1) < f(x;\alpha_2)$.

Claim 2 There exists a connected set D' of measure $\nu(D') > 0$, such that $f(x;\alpha_1) < f(x;\alpha_2)$.

Pick any two $x_1, x_2 \in D$, $x_1 < x_2$, we want to show that for any $x \in (x_1,x_2)$, $g(x) = f(x;\alpha_1) - f(x;\alpha_2) < 0$. By contradiction, suppose there exists an $x' \in (x_1,x_2)$, such that $g(x') \geq 0$. By continuity, there exists an $\tilde{x} \in (x',x_2)$, such that $g'(\tilde{x}) = \frac{g(x_2)-g(x')}{x_2-x'} < 0$, or $g'(\tilde{x}) = f'(\tilde{x};\alpha_1) - f'(\tilde{x};\alpha_2) < 0$, a contradiction. Define $D' = D \cup \{x \in \mathbb{R} : x = x_1 + x_2, x_1 \in D, x_2 \in D\}$, we have proved Claim 2. Note that D' is an interval with $\nu(D') > 0$.

Claim 3 For all $x \in (-\infty, \sup D']$, $f(x;\alpha_1) > f(x;\alpha_2)$.

This is simply established by the fact that $g(\sup D') \leq 0$ and $g'(x) > 0$ for $x \in (-\infty,\mu]$. Let $\hat{\mu} = \sup D'$, the lemma is established.

Lemma 3 There exists a unique $\mu^* \in (\infty, \frac{\mu}{2})$, such that $g(x) < 0$ for $x \in (-\infty, 2\mu^*)$ and $g(x) > 0$ for $x \in (2\mu^*, \mu]$.

Notice that for the sets D' and E' we have constructed, $\sup D' \leq \inf E'$. If $\sup D' = \inf E'$, the claim is established by setting $2\mu^* = \sup D' = \inf E'$. If $\sup D' < \inf E'$, by continuity and monotonicity of $g(x)$, there exists a unique $2\mu^* \in (\sup D', \inf E')$, such that $g(x) < 0$ for $x \in (-\infty, 2\mu^*)$ and $g(x) > 0$ for $x \in (2\mu^*, \mu]$. This proves Lemma 3.

Now recall that

$$\Pi(\alpha) \equiv \frac{\pi[R - \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x; \alpha) dx]}{R}.$$

If $-2c > 2\mu^*$ or $c < -\mu^*$, for all $x \in (-2c, -2c + 2\mu)$, $f(x; \alpha_1) > f(x; \alpha_2)$. Thus, $\Pi(\alpha_1) < \Pi(\alpha_2)$.

If $-2c + 2\mu < 2\mu^*$ or $c > \mu - \mu^*$, for all $x \in (-2c, -2c + 2\mu)$, $f(x; \alpha_1) < f(x; \alpha_2)$. Thus, $\Pi(\alpha_1) > \Pi(\alpha_2)$. □

Proposition 6. *Suppose that Assumption 2 holds.*

1. *The optimal message m^* satisfies $m^* < 2\mu$, and the optimal audience is*

$$[-2c, -2c + m^*] \cup [2c, 2\mu - (m^* - 2c)]. \tag{23}$$

   In other words, the negative campaigner needs to target the optimal message m^* to both his own base and the base of the incumbent.

2. *Denote Π as the probability that the opposition runs as a negative campaigner.*

   (a) *The sign of Π_c is ambiguous because, under a higher voting cost (c increases), the negative campaigner also needs to target more enthusiastic supporters of the incumbent.*

   (b) *$\Pi_\pi > 0$: negative messaging is more likely under a better targeting technology.*

Proof. Consider the payoff to the negative campaigner as a function of the message m:

$$\bar{V}(m) = \begin{cases} \pi[R - \int_{-2c}^{-2c+2\mu} T(x, m) f(x) dx] & \text{if } m \in [2\mu, \bar{m}) \\ \pi[R - \int_{-2c}^{-2c+m} T(x, m) f(x) dx + \int_{2c}^{2\mu-(m-2c)} T(x, m) f(x) dx]. & \text{if } m \in [\underline{m}, 2\mu] \\ 0 & \text{if } m \in [0, \underline{m}) \cup [\bar{m}, \infty] \end{cases}.$$

For $m \in [\underline{m}, 2\mu)$, the derivative $\bar{V}'(m)$ is:

$$\bar{V}'(m) = -\pi \Big\{ T(-2c + m) f(-2c + m) - T(2\mu - (m - 2c), m) f(2\mu - (m - 2c)) +$$

$$+\int_{-2c}^{-2c+m} T_m(x,m)f(x)dx + \int_{2c}^{2\mu-(m-2c)} T_m(x,m)f(x)dx\Bigg\}$$

Due to symmetry of $f(\cdot)$ around μ, we have $f(-2c+m) = f(2\mu-(m-2c))$. Therefore,

$$\bar{V}'(m) = -\pi\Bigg\{[T(-2c+m) - T(2\mu-(m-2c),m)]f(2\mu-(m-2c))+$$

$$+\int_{-2c}^{-2c+m} T_m(x,m)f(x)dx + \int_{2c}^{2\mu-(m-2c)} T_m(x,m)f(x)dx\Bigg\}.$$

We have:

$$\lim_{m\to(2\mu)^-} \bar{V}'(m) = -\pi\Bigg\{[T(-2c+2\mu) - T(2c,m)]f(2c) + \int_{-2c}^{-2c+2\mu} T_m(x,m)f(x)dx\Bigg\} < 0$$

because:

$$\int_{-2c}^{-2c+2\mu} T_m(x,m)f(x)dx > [T(2c,m) - T(-2c+2\mu)]f(2c).$$

In this case, there exists an $d > 0$, such that for all $\delta < d$,

$$\bar{V}(2\mu+\delta) > \bar{V}(2\mu) = \pi[R - \int_{-2c}^{-2c+2\mu} T(x,m)f(x)dx] > 0.$$

Because $V(\underline{m}) = 0$ by construction, we also have for all $\delta < d$,

$$\bar{V}(2\mu+\delta) > V(\underline{m}).$$

Therefore, the optimal m^* on the interval $[\underline{m}, 2\mu]$ is interior: $m^* \in (\underline{m}, 2\mu)$, and it satisfies:

$$\bar{V}'(m^*) = -\pi\Bigg\{[T(-2c+m^*) - T(2\mu-(m^*-2c),m^*)]f(-2c+m^*)+$$

$$+\int_{-2c}^{-2c+m^*} T_m(x,m^*)f(x)dx + \int_{2c}^{2\mu-(m^*-2c)} T_m(x,m^*)f(x)dx\Bigg\} = 0.$$

Notice that m^* is the global optimum because $\bar{V}'(m) < 0$ for all $m \in (2\mu, \bar{m})$, with $\bar{V}(m)$ remaining continuous at $m = 2\mu$ (though generically not differentiable). Therefore, the optimal audience is:

$$[-2c, -2c+m^*] \cup [2c, 2\mu-(m^*-2c)],$$

with both intervals non-empty. The probability that the opposition candidate runs as a negative campaigner is:

$$\Pi(\pi, c) = \frac{\pi}{R}[R - \int_{-2c}^{-2c+m^*} T(x, m^*) f(x) dx - \int_{2c}^{2\mu-(m^*-2c)} T(x, m^*) f(x) dx].$$

$$\frac{\partial}{\partial \pi}\Pi(\pi, c) = \frac{1}{R}[R - \int_{-2c}^{-2c+m^*} T(x, m^*) f(x) dx - \int_{2c}^{2\mu-(m^*-2c)} T(x, m^*) f(x) dx]$$
$$> R - \int_{-2c}^{-2c+2\mu} T(x, m) f(x) dx > 0.$$

By the envelope theorem

$$\frac{\partial}{\partial c}\Pi(\pi, c) = \frac{1}{R}\frac{\partial}{\partial c}\max_m \pi\Big[R - \int_{-2c}^{-2c+m} T(x, m) f(x) dx - \int_{2c}^{2\mu-(m-2c)} T(x, m) f(x) dx\Big]$$

$$= \frac{\pi}{R}\Bigg\{ - \Big[T(-2c + m^*) f(-2c + m^*) \cdot (-2) - T(-2c, m^*) f(-2c) \cdot (-2)\Big]$$

$$-[T(-2\mu - (m^* - 2c), m^*) f(2\mu - (m^* - 2c)) * 2 - T(2c, m^*) f(2c) * 2]\Bigg\}$$

$$= \frac{2\pi}{R}\Bigg\{[T(-2c + m^*) f(-2c + m^*) - T(-2c, m^*) f(-2c)]$$

$$+[T(2c, m^*) f(2c) - T(-2\mu - (m^* - 2c), m^*) f(2\mu - (m^* - 2c))]\Bigg\}.$$

The term $[T(-2c+m^*, m^*)f(-2c+m^*) - T(-2c, m^*)f(-2c)] > 0$ captures a similar mechanism as the main model, where a higher voting cost allows the negative campaigner to focus on a smaller group ($f(-2c) < f(-2c + m^*)$) of more receptive voters ($T(-2c, m^*) < T(-2c + m^*, m^*)$).

But the term $[T(-2\mu - (m^* - 2c), m^*)f(2\mu - (m^* - 2c)) - T(2c, m^*)f(2c)]$ is ambiguous because a higher voting cost induces the negative campaigner to focus on more enthusiastic voters of the incumbent $T(-2\mu - (m^* - 2c), m^*) > T(2c, m^*)$, albeit a smaller group of them. □

B Voter suppression by a populist leader

This short section extends the logic of the main model to study the behavior of a populist in power. We formalize a simple and important insight: because a higher voting cost reduces persuasion spending (Proposition 4 in the paper), a digital populist in power faces a strong temptation to raise the voting cost through various measures of voter suppression. Populism indeed may produce some of the most detrimental consequences on voting rights.

Specifically, we endogenize investment in voter suppression as a choice for a populist politician. The model now has two periods. The first period is the same as the main model, with the difference that the utility of the office, now denoted as $\tilde{R}$, includes the populist's expected payoff in the second period. The second period remains similar, with only two differences. First, the populist politician can only run the re-election campaign as a populist; he cannot run as an establishment candidate. Second, after it is revealed whether media targeting is successful for the re-election campaign, the populist politician may increase the voting cost to $c+e$ by spending $Q(e)$, with $Q'(e) > 0$ and $Q''(e) > 0$. Then the populist chooses the optimal message and the optimal audience. We assume that $e \leq \bar{e}$, $\bar{e}$ being the maximal increase in voting cost that does not wipe out a facade of democracy. The populist runs for re-election for the exogenous utility R and then retires. As in the main model, it is also too expensive to persuade the entire audience of a populist message ($T(-2(c+\bar{e})+2\mu, 2\mu) > R$).

The optimal message and the optimal audience are the same as the main model. The populist incumbent's payoff is:

$$\max_{e} \left\{ R + \pi[-Q(e) - \int_{-2(c+e)}^{-2(c+e)+2\mu} T(x, 2\mu) f(x) dx + R] \right\}. \tag{38}$$

This yield a simple proposition:

Proposition 7. *1. $e^* > 0$: a populist makes a positive investment in voter suppression.*

2. As $Q'(e^*) \to 0$, $e^* = \bar{e}$: if it is costless to suppress voters, the populist excludes as many voters as possible, including his own base.

Proof. Examine the derivative of Equation 38 with respect to e:

$$-Q'(e) + 2\Big\{T[-2(c+e)+2\mu, 2\mu] f[-2(c+e)+2\mu] - T[-2(c+e), 2\mu] f[-2(c+e)]\Big\}.$$

Notice that $T[-2(c+e)+2\mu, 2\mu] > T[-2(c+e), 2\mu]$ and $f[-2(c+e)+2\mu] \geq f[-2(c+e)]$. The second inequality is true because $-2(c+e)+2\mu \leq \mu$; otherwise, the populist would have won the votes from more than one half of the population, a contradiction. The proposition is established immediately. □

The proposition formalizes the notion that in the digital age (when targeting succeeds), populism has an intrinsic impulse to exclude as many voters as possible. Existing literature focuses on how populism threatens liberal institutions in advanced democracies, such as the media and separation of powers. It turns out that the problem is far worse: digital populism can be a lethal threat to the very foundation of an *electoral* democracy. Indeed, extreme digital populism is indistinguishable from dictatorship.

In the first period, the opposition solves the same game as the baseline model, with the utility of winning the office as:

$$\tilde{R} = \{R + \pi[R - \int_{-2(c+e^*)}^{-2(c+e^*)+2\mu} T(x, 2\mu) f(x) dx - Q(e^*)]\},$$

Thus, the probability that the opposition runs as a populist is:

$$\tilde{\Pi} \equiv \frac{\pi}{R}[\tilde{R} - \int_{-2c}^{-2c+2\mu} T(x, \mu) f(x) dx].$$

The extended model affirms the key comparative statics in the main model, as summarized in the following proposition.

Proposition 8. *1.* $\tilde{\Pi}_c > 0$*: the opposition is more likely to run as a populist under a higher voting cost.*

2. $\tilde{\Pi}_\pi > 0$*: the opposition is more likely to run as a populist if the targeted media is more effective.*

3. $\tilde{\Pi}_{\pi c} > 0$ *:the opposition's probability of running as a populist features a complementarity between targeted media and the voting cost.*

Proof.

$$\tilde{\Pi} = \frac{\pi}{R}\Big[-\int_{-2c}^{-2c+2\mu} T(x, \mu) f(x) dx + R\Big] + \frac{\pi^2}{R}[-Q(e^*) - \int_{-2(c+e^*)}^{-2(c+e^*)+2\mu} T(x, 2\mu) f(x) dx + R].$$

$$\tilde{\Pi}_c = \frac{\pi}{R}\Big\{\frac{\partial}{\partial c}\big[-\int_{-2c}^{-2c+2\mu} T(x, \mu) f(x) dx\big] + \frac{\pi^2}{R}\overbrace{\big[\frac{\partial}{\partial c}\big[-\int_{-2(c+e^*)}^{-2(c+e^*)+2\mu} T(x, 2\mu) f(x) dx\big]}^{\text{Envelope Theorem}}\Big\} > 0.$$

$$\tilde{\Pi}_{\pi} = \frac{1}{R}\Big[-\int_{-2c}^{-2c+2\mu} T(x,\mu)f(x)dx + R\Big] + \frac{2\pi}{R}[-Q(e^*) - \int_{-2(c+e^*)}^{-2(c+e^*)+2\mu} T(x,2\mu)f(x)dx + R] > 0,$$

noticing that

$$\begin{aligned} &-Q(e^*) - \int_{-2(c+e^*)}^{-2(c+e^*)+2\mu} T(x,2\mu)f(x)dx + R \\ &> Q(0) - \int_{-2c}^{-2c+2\mu} T(x,2\mu)f(x)dx + R \\ &= -\int_{-2c}^{-2c+2\mu} T(x,2\mu)f(x)dx + R > 0. \end{aligned}$$

$$\tilde{\Pi}_{\pi c} = \frac{1}{R}\Big\{\frac{\partial}{\partial c}\big[-\int_{-2c}^{-2c+2\mu} T(x,\mu)f(x)dx\big] + \frac{2\pi}{R}[\overbrace{\frac{\partial}{\partial c}\big[-\int_{-2(c+e^*)}^{-2(c+e^*)+2\mu} T(x,2\mu)f(x)dx\big]}^{\text{Envelope Theorem}}\Big\} > 0.$$

□

C Voter satisfaction and populism

This section analyzes how average voter satisfaction with the incumbent (μ) affects populism. Voter dis-satisfaction, manifested as economic and sociocultural grievances, is usually cited as the major cause of populism. Although the explanation commands strong intuitions, the empirical findings are mixed (Guriev and Papaioannou (2022)). Our analysis replicates an intuitive mechanism that it is easier to convince angrier voters of a populist message. But we also uncover an additional mechanism in a centrist society, a mechanism that countervails the intuitive mechanism. The next proposition characterizes the populist strength when voters become more satisfied with the incumbent.

Proposition 9. $\Pi_\mu(\pi, c, \mu)$ *can be decomposed into two competing effects: in the first effect,* μ *reduces* Π*; in the other effect,* μ *raises* Π*.*

Proof. Recall that

$$\Pi(\pi, c, \mu) = \frac{\pi[R - \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f(x; \mu) dx]}{R}. \tag{39}$$

Notice that we denote the probability density function as $f(x; \mu)$ to highlight that μ is the key parameter.

$$\Pi_\mu = -\frac{2\pi}{R}[T(-2c+2\mu, 2\mu)f(-2c+2\mu; \mu) + \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f_\mu(x; \mu) dx + \int_{-2c}^{-2c+2\mu} T_m(x, 2\mu) f(x; \mu) dx]. \tag{40}$$

$-\frac{2\pi}{R} T(-2c + 2\mu, 2\mu) f(-2c + 2\mu; \mu) < 0$ is the effect of the additional marginal audience.

$-\frac{2\pi}{R} \int_{-2c}^{-2c+2\mu} T(x, 2\mu) f_\mu(x; \mu) dx > 0$ is the effect of the shrinking existing audience. $f_\mu(x; \mu) < 0$ because for all $x \in [-2c, -2c + 2\mu]$, $x < \mu$ and $f(\cdot)$ is symmetric at μ and uni-modal.

$-\frac{2\pi}{R} \int_{-2c}^{-2c+2\mu} T_m(x, 2\mu) f(x; \mu) dx < 0$ is the effect of a more inflaming message, and it is second order by Assumption 1 in the paper. □

A more satisfied citizenry produces two effects.

1. The addition of a marginal audience: the populist needs to persuade a new marginal group of voters, making it more costly to be a populist.

2. The shrinking of the existing audience: the existing audience of the populist message shrinks, making it cheaper to be a populist.

The first effect is straightforward. As the average voter becomes more satisfied with the incumbent (μ increases), the populist needs to expand the optimal audience and persuade all voters with

$$x \in [-2c, -2c + 2\mu]. \tag{41}$$

The persuasion of the additional audience at $x = -2c + 2\mu$ increases the cost of being a populist.

At the same time, we have a second effect. When the average voter becomes more satisfied with the incumbent, the entire distribution of voter satisfaction shifts rightward. Because the optimal audience is to the left of the average voter, the probability mass of the optimal audience shrinks everywhere. The effect of a shrinking audience is driven by the setup that, generically, the society is predominantly centrist. Per the logic of this setup, the higher satisfaction of the average voter reflects the higher satisfaction of the centrist society in general, as well as the shrinking mass of dissatisfied voters. The effect disappears if the voter satisfaction follows a uniform distribution, the most polarized distribution under our setup. In this case, the prediction is un-ambiguous. The model only features the intuitive effect that higher voter satisfaction adds a marginal audience for the populist message. Notice that, under a uniform distribution, all results in the key Proposition 4 hold. Under a higher voting cost, even though the size of the optimal targeted audience remains the same, the audience becomes more angry with the incumbent, driving all the results in Proposition 4.

D Data appendix

In this section, we document how Figure 9 is constructed, as well as extra regression analysis that further confirms its patterns. We use the following data-sets:

- The Cooperative Congressional Election Survey (CCES) from https://cces.gov.harvard.edu;
- County Presidential Election Returns 2016 from https://electionlab.mit.edu/data;
- the Citizen Voting Age Population (CVAP) Special Tabulation From the 2012-2016 American Community Survey (https://www.census.gov/programs-surveys/decennial-census/about/voting-rights/cvap.2016.html).

From these datasets, we construct the variables in Figure 9 as follows.

Trump votes For each respondent in the CCES, we know whether they voted for Trump in the 2016 General Election.

Turnout rates in 2016 We estimate turnout rates as the ratio of total votes in the 2016 Presidential Election (the MIT Election Lab) to the estimate of Citizen Voting Age Population (the 2012-2016 American Community Survey).

Intense and political use of social media CCES asks whether the respondent uses social media in general and, in addition, whether the respondent "politically" uses social media for each of the following five activities:

- posted a story, photo, video, or link about politics;
- posted a comment about politics;
- read a story or watched a video about politics;
- followed a political event;
- forwarded a story, photo, video, or link about politics to friends.

We define "intense political use of social media" as a dummy, which equals to one if and only if the respondent uses social media for all the above five activities in CCES.

To further document the complementarity between targeted media and the turnout rate, we run the following regression:

$$Trump_{ics} = \beta_1 turnout_c + \beta_2 media_{ics} + \gamma turnout_c * media_{ics} + X_{ics} + \alpha_s + u_{ics}. \qquad (42)$$

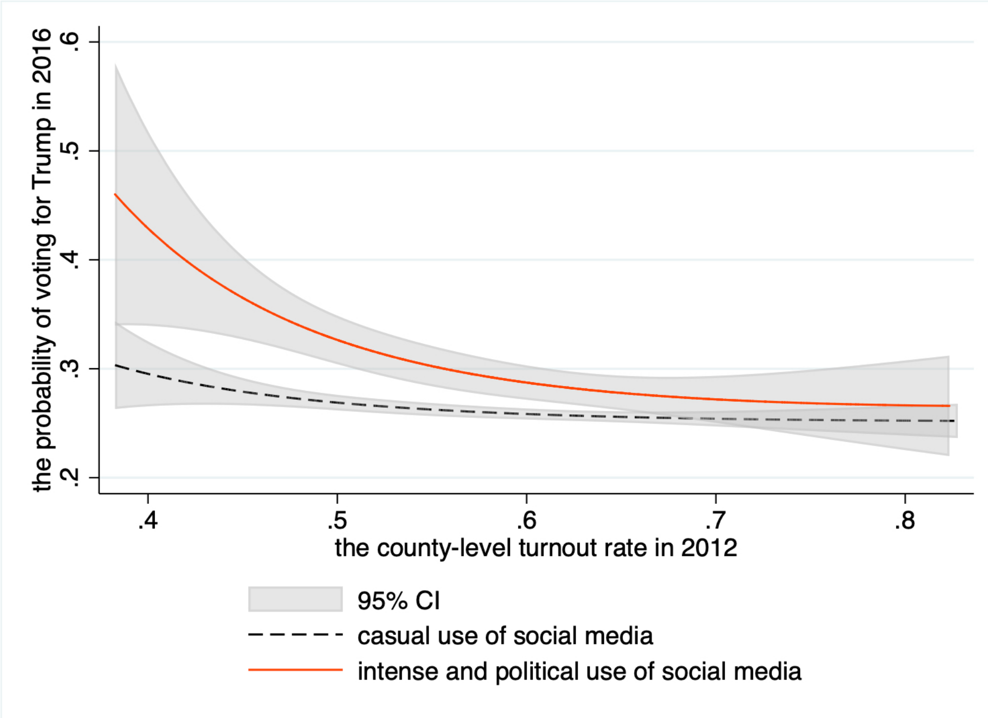
The variable $Trump_{ics}$ is a dummy whether the respondent i living in the county c and in the state s votes for Trump in the 2016 Presidential Election. The variable $turnout_c$ is the turnout rate in the county c. The variable $media_{ics}$ is a dummy for social media use. We will look at the political and intense use of social media, as well as the general use of social media. X_{ics} is a vector of control variables at the individual level. We discuss later more details about the control variables. α_s is the state fixed effects. Standard errors are clustered at the county level.

Table 4 shows the results with the sample restricted to social media users, which is 70.1% of the sample. $media_{ics}$ equals one if the respondent uses social media intensely and politically; $media_{ics}$ equals zero if the respondent uses social media but not in such an intense or political manner. Column (1) to Column (3) include no controls, while Column (4) to Column (6) add control variables successively. Non-political controls include gender, race, family income, education, age, and employment status. Political controls include the ideology of the respondent, which party the respondent is identified with, and whether the respondent approves of the jobs of President Obama, Congress, and the Supreme Court. Column (3) to Column (6) show that the interaction term is statistically significant, regardless of control variables. Given the context, it is not surprising that these political controls are the most powerful predictors for Trump votes. But even conditional on political controls, the interaction term remains significant. To interpret the point estimates, in Column (6), a one standard deviation decrease in turnout rate (0.083) predicts a $0.083 * 0.131 = 1.0873\%$ increases in a Trump vote, but only for those voters who use social media intensely and politically. Also, notice that when the respondent does not use social media in such a manner, a lower county turnout is much less correlated with Trump votes. This suggests that a high voting cost might not be important without the rise of social media.

Table 5 shows that the general use of social media does not matter. The variable $media_{ics}$ equals one when the respondent uses social media; $media_{ics}$ equals zero when the respondent does not use social media. We can see that the interaction term is small and statistically insignificant. Thus, even for respondents in counties with an extremely low turnout rate, having a social media account does not produce support for populism. What matters is an intense and political use of social media.

Furthermore, we can use estimates of 2012 turnout rates instead to proxy the voting cost. They are constructed in a similar manner: we estimate turnout rates as the ratio of total votes in the 2012 Presidential Election (the MIT Election Lab) to the estimate of Citizen Voting Age Population (the 2012-2016 American Community Survey). Table 4 and Table 5 run the same regressions as Table 2 and Table 3, only replacing the turnout rates with the ones from 2012 Presidential Election. Table 4 and Table 5 confirm similar patterns as Table

Figure 10: The estimated probability of a voter supporting Donald Trump in the 2016 American presidential election



Votes in 2016 presidential election and use of social media are from the 2016 Congressional Cooperative Election Survey. The turn out rate in 2012 presidential election is the ratio of the total votes in the 2012 Presidential Election (from MIT Election Lab) to the estimate of Citizen Voting Age Population (from the 2008-2012 American Community Survey). Data are fit to fractional polynomials, and turnout rates are winsorized at the 1% level.

2 and Table 3. Figure 10 also shows a similar pattern with Figure 9 in the paper. Potentially inducing more support for populism, better targeted media and a higher voting cost might encourage populists to run for office. Yet all empirical patterns documented here are merely suggestive; systematic empirical work is left for future research.

Table 2: Turnout, political use of social media, and Trump votes

	(1)	(2)	(3)	(4)	(5)	(6)
			Trump votes			
intense and political use of social media	0.0339***	0.0339***	0.168***	0.166***	0.106**	0.0901***
	(0.00706)	(0.00703)	(0.0523)	(0.0523)	(0.0505)	(0.0327)
the county turnout rate		-0.00587	0.0160	-0.0274	-0.189***	-0.0284
		(0.0566)	(0.0569)	(0.0674)	(0.0530)	(0.0252)
interaction			-0.218***	-0.214**	-0.168**	-0.131**
			(0.0841)	(0.0837)	(0.0808)	(0.0523)
state fixed effects	No	No	No	Yes	Yes	Yes
Non-political individual controls	No	No	No	No	Yes	Yes
Political individual controls	No	No	No	No	No	Yes
N	45215	45215	45215	45215	45199	44945
R^2	0.001	0.001	0.001	0.011	0.146	0.519
adj. R^2	0.001	0.000	0.001	0.010	0.143	0.517

Standard errors are in parentheses and are clustered at the county level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is whether the individual voted for Trump in 2016. "Intense and political use of social media"= 1 is defined in the text. The county turnout rate is from 2016 Presidential Election. The interaction term is the product of the county turnout rate and the intense and political usage of social media. Non-political controls include gender, race, family income, education, age, and employment status. Political controls include the ideology of the respondent, which party the respondent is identified with, and whether the respondent approves of the jobs of President Obama, Congress, and the Supreme Court.

Table 3: Just having a social media account does not matter much

	(1)	(2)	(3)	(4)	(5)	(6)
			Trump votes			
social media	-0.0915***	-0.0915***	-0.0432	-0.0525*	0.00435	0.00181
	(0.00415)	(0.00417)	(0.0309)	(0.0308)	(0.0274)	(0.0213)
county turnout		0.0184	0.0741	0.0450	-0.159***	-0.0315
		(0.0573)	(0.0721)	(0.0801)	(0.0609)	(0.0333)
interaction			-0.0787	-0.0657	-0.0461	-0.000622
			(0.0498)	(0.0496)	(0.0447)	(0.0345)
state FEs	No	No	No	Yes	Yes	Yes
Non-political individual controls	No	No	No	No	Yes	Yes
Political individual controls	No	No	No	No	No	Yes
N	64455	64455	64455	64455	64434	64059
R^2	0.009	0.009	0.009	0.018	0.154	0.540
adj. R^2	0.008	0.008	0.009	0.018	0.152	0.539

Standard errors are in parentheses and are clustered at the county level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is whether the individual voted for Trump in 2016. "Social media"= 1 if and only the individual uses any social media. The county turnout rate is from 2016 Presidential Election. The interaction term is the product of the county turnout rate and usage of social media. Non-political controls include gender, race, family income, education, age, and employment status. Political controls include the ideology of the respondent, which party the respondent is identified with, and whether the respondent approves of the jobs of President Obama, Congress, and the Supreme Court.

Table 4: Turnout in 2012, political use of social media, and Trump votes

	(1)	(2)	(3)	(4)	(5)	(6)
	Trump votes					
intense and political use of social media	0.0339***	0.0340***	0.167***	0.166***	0.110**	0.0866***
	(0.00706)	(0.00704)	(0.0472)	(0.0471)	(0.0450)	(0.0284)
the turnout rate in 2012		-0.0925*	-0.0703	-0.225***	-0.279***	-0.0467*
		(0.0543)	(0.0546)	(0.0658)	(0.0507)	(0.0242)
the interaction			-0.220***	-0.217***	-0.176**	-0.127***
			(0.0767)	(0.0763)	(0.0729)	(0.0460)
state FEs	No	No	No	Yes	Yes	Yes
Non-political individual controls	No	No	No	No	Yes	Yes
Political individual controls	No	No	No	No	No	Yes
N	45215	45215	45215	45215	45199	44945
R^2	0.001	0.001	0.001	0.012	0.147	0.519
adj. R^2	0.001	0.001	0.001	0.011	0.144	0.517

Standard errors are in parentheses and are clustered at the county level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is whether the individual voted for Trump in 2016. "Social media"= 1 if and only the individual uses any social media. The county turnout rate is from 2012 Presidential Election. The interaction term is the product of the county turnout rate and usage of social media. Non-political controls include gender, race, family income, education, age, and employment status. Political controls include the ideology of the respondent, which party the respondent is identified with, and whether the respondent approves of the jobs of President Obama, Congress, and the Supreme Court.

Table 5: Just having a social media account does not matter much

	(1)	(2)	(3)	(4)	(5)	(6)
			Trump votes			
intense and political use of social media	-0.0915*** (0.00415)	-0.0917*** (0.00416)	-0.0566** (0.0284)	-0.0662** (0.0283)	-0.00441 (0.0252)	-0.00544 (0.0193)
the turnout rate in 2012		-0.0749 (0.0556)	-0.0338 (0.0704)	-0.174** (0.0752)	-0.255*** (0.0571)	-0.0605** (0.0308)
interaction			-0.0580 (0.0463)	-0.0446 (0.0462)	-0.0326 (0.0415)	0.0113 (0.0316)
state FEs	No	No	No	Yes	Yes	Yes
Non-political individual controls	No	No	No	No	Yes	Yes
Political individual controls	No	No	No	No	No	Yes
N	64455	64455	64455	64455	64434	64059
R^2	0.009	0.009	0.009	0.019	0.155	0.540
adj. R^2	0.008	0.009	0.009	0.018	0.152	0.539

Standard errors are in parentheses and are clustered at the county level. * p<0.1, ** p<0.05, *** p<0.01. The dependent variable is whether the individual voted for Trump in 2016. "Social media"= 1 if and only the individual uses any social media. The county turnout rate is from 2012 Presidential Election. The interaction term is the product of the county turnout rate and usage of social media. Non-political controls include gender, race, family income, education, age, and employment status. Political controls include the ideology of the respondent, which party the respondent is identified with, and whether the respondent approves of the jobs of President Obama, Congress, and the Supreme Court.