A Model of Populism as a Conspiracy Theory

By Adam Szeidl and Ferenc Szucs*

We model populism as the dissemination of a false "alternative reality", according to which the intellectual elite conspires against the populist for purely ideological reasons. If enough voters are receptive to it, this alternative reality—by discrediting the elite's truthful message—reduces political accountability. Elite criticism, because it is more consistent with the alternative reality, strengthens receptive voters' support for the populist. Alternative realities are endogenously conspiratorial to resist evidence better. Populists, to leverage or strengthen beliefs in the alternative reality, enact harmful policies that may disproportionately harm the non-elite. These results explain previously unexplained facts about populism.

JEL: D03, D72, D82, D83
Keywords: populism, conspiracy theory, alternative reality, propaganda, misbeliefs, distrust

Populist leaders paint a grim picture of the world. The populist ideology is often centered around a false narrative, an *alternative reality*, in which a conspiracy of the elite shapes major events to the detriment of the people. In this narrative, the elite is not only "corrupt", as described in leading accounts of populism (Guriev and Papaioannou, 2022); it is conspiratorial and all-powerful. For example, a central part of Donald Trump's political narrative is that the 2020 US election was stolen by a conspiracy of the deep state. The populist narrative matters because it shapes supporters' beliefs: the majority of Republicans believe that Trump did not lose the 2020 election legitimately.[1] These sorts of misbeliefs are potentially highly consequential.

Populism is also associated with a profound decline in political accountability. Funke, Schularick and Trebesch (2023) show that populist leaders, despite substantially reducing GDP per capita, stay in power for twice as long as non-populists. Moreover, populists seem to achieve electoral success despite widely-publicized acts that would normally be extremely damaging: for example, Donald Trump, a convicted felon, won the 2024 election.

Existing economic models do not explain the use of conspiratorial narratives

* Emails: szeidla@ceu.edu, ferenc.szucs@su.se. An earlier draft of this paper was titled "The Political Economy of Alternative Realities." We thank Nageeb Ali, Ruben Durente, Ruben Enikolopov, Matthew Gentzkow, Sergei Guriev, Helios Herrera, Botond Koszegi, Kristof Madarasz, Maria Petrova, Giacomo Ponzetto, Jesse Shapiro, David Stromberg, Ekaterina Zhuravskaya, and conference and seminar audiences for comments and discussions. We thank the European Research Council (ERC), under the European Union's Horizon 2020 research and innovation programme grant agreement number 724501, and under the Horizon Europe grant agreement number 101163526; and the Austrian Science Fund (FWF), under grant F100600, for funding.

[1] See Figure 1 below. People not only claim to hold these beliefs, they act on them, as illustrated by the 2021 January 6 attack on the United States Capitol.

in politics and their association with reduced accountability. Models of political narratives do not study conspiratorial narratives (Eliaz and Spiegler, 2020; Eliaz, Galperti and Spiegler, 2022). Models of reduced accountability work through the silencing of the media, or through repression (Guriev and Treisman, 2020; Egorov and Sonin, 2024), hence do not explain reduced accountability in populist democracies that have an independent media. And in models of populism, populist policies are a positive signal, so that populism is associated with *increased* accountability (Acemoglu, Egorov and Sonin, 2013; Bellodi et al., 2023).

We present a new theory of populism that helps us understand both conspiratorial narratives and reduced accountability. In our theory, the goal of populism is to provide a false *alternative reality* that discredits the intellectual elite's message about the politician. Specifically, we assume that populist propaganda can (partially) persuade voters that the elite conspires to criticize the politician's competence purely because they disagree with his ideology. This alternative reality discredits elite criticism that would normally reveal the politician's type. "Bad" politicians, expecting elite criticism, propagate the alternative reality to remain in power. Thus, our theory predicts both the use of conspiratorial propaganda and its association with reduced accountability.

We formalize these ideas in a model which explicitly incorporates the false alternative reality. Beyond explaining our motivating facts, this model makes several new predictions. It predicts that truthful elite criticism can backfire and strengthen some voters' support for the politician; that alternative realities are endogenously conspiratorial to better resist evidence; and that populists, despite their "pro-people" rhetoric, may set policies that disproportionately harm the non-elite. These results offer a new understanding of populism.

In our model, presented in Section 2, an incumbent politician is characterized by a type dimension, e.g., competence, along which he can be good or bad. Voters do not directly observe the politician's type but form beliefs over it, and political accountability is measured with the accuracy of their beliefs. The politician and the intellectual elite send messages which affect these beliefs. First, the politician chooses whether to send conspiratorial propaganda. Then the elite (including the news media), having received an informative signal about the politician's type, sends a message that reports on that signal. We assume that the elite consists of a continuum of small members, who individually cannot influence voters and thus report about the signal truthfully.

A share α of voters are receptive to propaganda, so that propaganda exogenously and counterfactually increases their prior belief in the alternative reality (AR). The AR is a state of the world with zero objective probability, which differs from the objective reality in precisely one way: In the AR the continuum of small elite members can coordinate—effectively conspire—and thus can collectively choose their message to influence voters. It follows that if elite members sufficiently dislike the politician, perhaps because they disagree with his ideology, then in the AR they will always report that politician bad. Intuitively, in the AR the "fake news media" criticize Trump's competence not because he is incompetent, but because he is "anti-woke." In turn, a voter persuaded by propaganda partially believes this alternative reality and distrusts elite criticism.

We analyze the model in Section 3. We show that an equilibrium of the following form emerges. (i) In the objective reality only the bad politician sends propaganda, and the elite always reports truthfully. (ii) In the alternative reality both the good and the bad politician sends propaganda, and the elite always criticizes the politician. Intuitively, in reality the good politician has no reason to send propaganda as he expects praise from the elite. The bad politician, who expects criticism, has an incentive to send propaganda if doing so discredits elite criticism. Discrediting only works if the narrative of the alternative reality is plausible: if it is incentive compatible for the conspiring elite to criticize even a good politician. This holds provided that elite members sufficiently dislike the politician (sufficiently disagree with his ideology). Under this assumption, the equilibrium admits the above form; otherwise it does not feature propaganda. These results immediately predict the equilibrium use of conspiratorial propaganda, and its association with reduced accountability. Thus, our model helps explain our motivating facts.

The model also yields new theoretical implications. First, it predicts that propaganda *inverts* the effect of the elite's message on receptive voters, so that elite criticism increases their beliefs that the politician is good. The key intuition is that for a receptive voter who experienced propaganda, the elite's message is primarily informative about the nature of reality. In particular, he knows that in the alternative reality the (conspiring) elite always criticizes, while in the objective reality the (honest) elite only sometimes criticizes. Thus, observing elite criticism is more consistent with the alternative reality, increases his posterior of the alternative reality, and with it, his posterior that the politician is good. It follows that truthful critical information can increase receptive voters' support for the politician, a result that overturns standard intuitions about the impact of information in political economics.

Second, the politician's choice of when to send propaganda *amplifies* receptive voters' misbelief about the alternative reality. Intuitively, the politician supplies the alternative reality precisely when events, such as elite criticism, are expected to be consistent with it. The receptive voter neglects this correlation, implying that his misbelief is (on average) strengthened by events. As a result, even propaganda that plants a small initial misbelief can have large societal effects.

These implications help explain previously unexplained facts. The inversion result explains a key fact in contemporary US politics: that the four criminal indictments against Trump in 2023 were accompanied by an *increase* in his support among Republican voters (Swan et al., 2023). This reaction by supporters of the presumptive party of law and order is puzzling, especially when compared to the case of Nixon, who lost Republican support after Watergate. Our inversion result explains the increased support for Trump by predicting that it was the *causal* *effect* of the indictments. This prediction is in line with survey evidence that Republicans claimed to increase support for Trump due to the indictments. It is also in line with new evidence we present that scandals of Republican politicians caused an increase in the donations they received from Trump supporters. Moreover, the mechanism for inversion, increased beliefs in the alternative reality, is consistent with the fact that following the indictments, Republicans sharply increased their beliefs in the conspiracy theory that the 2020 election was stolen. Finally, our model explains the contrast between Trump and Nixon through the logic that only Trump had a sufficiently large ideological cleavage with the elite to make the alternative reality plausible. Nixon, representing the more educated party (Republicans around 1970), could not credibly argue that the intellectual elite conspired to remove him.

The amplification result may help explain why beliefs in the alternative reality are so prevalent (e.g., held by most Republicans), a fact that seems difficult to attribute purely to the persuasive effect of propaganda. Consistent with the logic of amplification, we argue that populists in multiple countries supplied the alternative reality precisely in situations where it matched headline facts.

In Section 4 we develop two applications of the model. First, we investigate the reason that alternative realities are often conspiracy theories. In our basic model, the conspiracy was purely by assumption. We now allow the politician to choose between two types of alternative realities: one in which elite members have a low lying cost but cannot conspire, and another in which they can also conspire. Sending propaganda about the latter is more expensive. We show that the conspiracy theory often dominates, because it solves a collective action problem of the elite. Intuitively, each elite member's lie about the politician benefits all other elite members, resulting in a within-elite externality which the conspiracy internalizes. Thus, the ability to conspire makes the elite more powerful. As a result, the conspiracy alternative reality is more attractive to the politician, because the more powerful elite can explain away even more credible critical evidence. We believe that this is the first formal explanation for the prevalence of political conspiracy theories.

This analysis predicts that alternate realities are often resistent to evidence, because in response to more credible criticism the politician can "upgrade" the narrative from a lying elite to a conspiring elite. Upgrading is socially harmful, because it increases distrust in the elite beyond politics. Once the voter contemplates an elite conspiracy, he fears that the conspiracy's interests may be driving elite messages even in other domains. This logic helps explain why misbeliefs under populism extend beyond politics, including Republicans' general distrust in science.

In our second application, we investigate the effect of conspiratorial populism on government policy. This is an important topic since populism is associated with large economic and non-economic costs (Guriev and Papaioannou, 2022). We find that populists introduce harmful policies for two distinct reasons. First, there is a direct effect of reduced accountability: populism enables "bad" politicians to maintain power, who then enact "bad" policies. Second, our model predicts that populists will choose harmful polcies *purely to trigger the elite.* The intuition follows from the inversion result: since elite criticism increases the support of receptive voters, the politician chooses harmful policies to invite elite criticism.

Harmful policies also emerge in the Acemoglu, Egorov and Sonin (2013) model, where populists signal their independence from the elite using policies that disproportionately harm the elite. The key difference is that our model can also account for harmful policies that *do not* disproportionately harm the elite. As a result, our model helps explain the previously unexplained fact that populists, despite their pro-people rhetoric, are not actually siding with the "people": their policies seem to hurt the non-elite as much as they hurt the elite. Indeed, Funke, Schularick and Trebesch (2023) show that populists reduce GDP per capita without meaningfully reducing inequality, i.e., that they seem to cause equal economic harm to the elite and the non-elite. Populists also favor specific policies that especially harm the "people." They tend to be massively corrupt (Zhang, 2024), thereby reducing the quality of government services; they implement tariffs that harm their core supporters (Fajgelbaum et al., 2019); and they oppose environmental policies that would help the non-elite (Friedman, Plumer and Stevens, 2025). We conclude that the current wave of populism will likely create substantial harm to both the elite and the non-elite.

Our paper builds on overlapping literatures in political, behavioral, and information economics. We build on theories of populism (Acemoglu, Egorov and Sonin, 2013; Bellodi et al., 2023; Agranov, Eilat and Sonin, 2023) and identity politics (Besley and Persson, 2021; Bonomi, Gennaioli and Tabellini, 2021). Our main contribution to this work is a conspiracy-theory-based model of populism. Our model makes a number of new predictions about populism, including reduced accountability, the inverted effect of elite criticism, the emergence of conspiracy theories, and broadly harmful policies.

We also build on work studying the supply of misinformation in politics, including the supply of hatred (Glaeser, 2005), media capture (Besley and Prat, 2006), censorship and positive propaganda (Guriev and Treisman, 2020; Egorov and Sonin, 2024), and worldview politics (Ash, Mukand and Rodrik, 2021). Much of this work assumes that voters update in a Bayesian fashion, as in models of Bayesian persuasion (Kamenica and Gentzkow, 2011). We contribute to this work by modeling misinformation as a strategic alternative reality, an approach potentially portable to other settings; and with the aforementioned new implications.

Our modelling approach builds on theories of distorted belief formation in political economics. Theories of motivated beliefs in an alternative state of the world have been used to study beliefs in a just world (Bénabou and Tirole, 2006), groupthink (Bénabou, 2013), inefficient policy-making (Levy, 2014) and partisan disagreement (Le Yaouanq, 2023). Theories of model misspecification have been used to study persuasion (Bénabou, Falk and Tirole, 2018; Galperti, 2019;

Schwartzstein and Sunderam, 2021; Aina, 2023) and political narratives (Eliaz and Spiegler, 2020; Eliaz, Galperti and Spiegler, 2022). We contribute to this work with a model-misspecification-based approach to model conspiratorial populism, and with its new implications.

Finally, we build on a multidisciplinary literature studying misinformation and conspiracy theories, especially in political science and psychology, reviewed for example by Nyhan (2020) and Douglas et al. (2019). Our main contribution to this work is a formal model of conspiracy theories.

I. Model

A. Motivation

Our model is motivated by two observations. First, in a number of democracies, the populist ideology is centered around a conspiratorial alternative reality, according to which the elite conspires to attack the competence of the populist purely because they disagree with his ideology. This alternative reality goes beyond the Mudde (2004) and Guriev and Papaioannou (2022) descriptions of populist ideology, which emphasize the antagonism between the "pure people" and the "corrupt elite", in that here the elite is conspiratorial and all-powerful. The following examples illustrate.

- In the United States, Trump claims that the deep state and the media conspire to criticize him (e.g., claim him a criminal) because they find his cultural values too conservative. Trump talks about the conspiracy explicitly, "Either the deep state destroys America, or we destroy the deep state" (Allen, 2023); ties the incentives of the conspiracy to cultural values, "they won't hesitate to ramp up their persecution of Christians, pro-life activists"; and suggests that the goal of the conspiracy is to limit conservative values "they want to silence me because I will never let them silence you" (Corasaniti and Gabriel, 2023).
- In Hungary, Orban claims that the members of the "Soros network"—including Brussels and the media—conspire to attack him for dismantling checks and balances because they find him too anti-immigration. Orban is explicit about this narrative. "And we understand what is happening. George Soros has bought people, he has bought organisations, he is feeding them out of the palm of his hand, Brussels is under his influence, and it is his plan that the Brussels machine is implementing in the case of immigration. They want to remove the fence, they want to let in millions of immigrants and they want to divide them up on a compulsory basis. And they want to punish those who do not obey." (Kocsis, 2017).
- In Israel, Netanyahu claims that the judiciary and the media conspire to attack him on charges of corruption because they find him too anti-Palestinian. As Horovitz (2020) explains, Netanyahu's thesis is that "a

> strong, pro-annexation, right-wing prime minister is facing an illicit attempt — perpetrated by a vast, leftist alliance of politicians, media, cops and state prosecutors — to oust him because of his ideology and policies".

Our second observation is that in the same democracies, supporters of the populist leader tend to hold misbeliefs consistent with these alternative realities.

- In the US, the majority of Republicans believe that Biden did not win the 2020 election legitimately (see Figure 1 below). Since large-scale election fraud requires a conspiracy, these beliefs reflect beliefs in the deep state conspiracy.
- In Hungary, the majority of Orban's supporters believe in the existence of a Soros-plan (hvg.hu, 2017), i.e., the conspiracy that the Soros network is bringing migrants into Europe.
- In Israel, a large fraction of Netanyahu's supporters doubt the corruption charges against him (Navot, 2022), suggesting beliefs in a conspiracy of the justice system.

The fact voters believe in the specific and often elaborate alternative reality supplied by the politician (e.g., the "Soros-plan") suggests that these beliefs are at least partly driven by the supply of populist propaganda.[2] This motivates our model in which a politician can supply a conspiratorial alternative reality to change voter beliefs.

B. *Setup*

Players, types, and actions. Our model is an information game in which a politician and the intellectual elite send messages to influence voters' beliefs about the politician's type. Both the intellectual elite, which represents the news media, and the voters consist of a unit mass of members. Each elite member has limited influence: each sends its message to an audience of voters who have measure zero. In turn, each voter has access to the message of exactly one elite member, i.e., consumes exactly one news media.[3]

At the beginning of the game the politician's type $\theta_c \in \{0, 1\}$ is realized, where $\theta_c = 1$ with probability q_c. Here $\theta_c = 1$ means that the politician is "good" and $\theta_c = 0$ means that the politician is "bad." Good politicians are valued by both elite members and voters. We refer to θ_c as competence, but it could represent some other broadly valued attribute such as being honest (as opposed to corrupt), lawful (as opposed to criminal), or democratic (as opposed to authoritarian). We assume that θ_c is observed only by the politician.

[2] There is much evidence that the political supply of misinformation changes beliefs (Yanagizawa-Drott, 2014; Adena et al., 2015; Blouin and Mukand, 2019; Barrera et al., 2020; Ajzenman, Cavalcanti and Da Mata, 2023) and that populists value the supply sufficiently to capture media (Mcmillan and Zoido, 2004; Szeidl and Szucs, 2021).

[3] We present a formal construction of the elite and the voters in online Appendix A.1.

Table 1—Timing and allocation of information

Stage:	0	1	2
Politician	$\theta_c,\ \theta_r$	$\hat{p}$	$\hat{s}$
Elite		$\hat{p},\ \hat{\theta}_c,\ \theta_r$	$\hat{s}$
Receptive voter		$\hat{p}$	$\hat{s}$
Unreceptive voter			$\hat{s}$

After observing his type, the politician has an opportunity with probability $1-\beta$ (where $0 < \beta < 1$) to send propaganda. Only the politician knows whether he has the opportunity to send propaganda. We let $p = 1$ denote that the politician sends propaganda, and $p = 0$ that he does not, either because he did not have the opportunity or because he chose not to. The role of β is to ensure that the absence of propaganda does not fully reveal the politician's type.

Members of the elite observe the propaganda realization and receive a signal $\hat{\theta}_c \in \{0, 1\}$ about the politician's type θ_c. This signal is correct ($\hat{\theta}_c = \theta_c$) with probability $\pi \in (0.5, 1]$. All elite members receive the same signal; voters do not receive a signal. We think of π as relatively high. Then, each elite member j sends a message $s_j \in \{0, 1\}$ about the signal to its zero measure of voters, where $s_j = 1$ means that the signal is good. We sometimes refer to $s_j = 1$ as praise and $s_j = 0$ as criticism.

There are two kinds of voters. A share α are receptive to propaganda, and observe both the elite's message and propaganda. The remaining share $1 - \alpha$ are unreceptive, and only observe the elite's message, but not propaganda. This is a stylized representation of the idea that unreceptive voters primarily follow the news media and are less exposed to propaganda. Each elite member j has an audience consisting of a share α of receptive and a share $1 - \alpha$ of unreceptive voters.

We assume that the elite's message s_j and propaganda p are subject to vanishing noise. This ensures that beliefs are well-defined off the equilibrium path. With probability ε_e, perfectly correlated across elite members, every elite member's realized message $\hat{s}_j$ is the opposite of the message s_j sent; and with independent probability ε_p, realized propaganda $\hat{p}$ is the opposite of the propaganda p sent. We let ε_e and ε_p go to zero and characterize the equilibrium in the limit.

Alternative reality. To model the alternative reality, we assume that there is a state of the world $\theta_r \in \Theta_r = \{R, AR\}$, where R represents the objective reality and AR the alternative reality. We assume that the true prior probability of $\theta_r = AR$ is zero.[4] The difference between the two realities is that in R the elite cannot, but in AR the elite can coordinate. Thus, if $\theta_r = R$, then each elite member j chooses her message s_j individually to maximize her own utility, but if $\theta_r = AR$, then the elite collectively chooses an identical message $s_j = s$ for all of

[4] In principle we could allow this prior to be positive, but to make the results stark we set it to zero.

its members to maximize the sum of their utilities.[5]

At the beginning of the game all voters hold the correct prior about θ_r, but propaganda exogenously increases receptive voters' prior that $\theta_r = AR$ to $q_{ar} > 0$. We let $q_r = 1 - q_{ar}$. We encode the change in the prior by assuming that each receptive voter i has a mind type $\theta_{mi} \in \{N, P\}$ (for normal and persuaded), that i becomes persuaded if and only if he encounters propaganda, and that the prior of receptive voter i as a function of his mind type is $\mu^0_{rec,i}(\theta_r = AR|\theta_{mi}) = 1_{\{\theta_{mi}=P\}} \cdot q_{ar}$. Each receptive voter then updates from his prior in a Bayesian fashion. Since either all or none of the receptive voters encounter propaganda, their mind types are identical and denoted by θ_m.

Motives. We assume that the payoffs of the politician and the elite are determined by voters' average beliefs about θ_c. Assuming that payoffs depend on beliefs simplifies presentation because it allows us to abstract from voters' preferences and actions. In online Appendix A.2 we show that a probabilistic voting model with a common preference shock provides microfoundations for this assumption, through the logic that voters' beliefs govern the probability that the politician is reelected, which in turn determines payoffs.

We define voters' average posterior belief about θ_c as

$$\bar{\mu}(\theta_c = 1|\hat{p}, \hat{\mathbf{s}}) = \alpha \cdot \overline{\mu_{rec,i}}(\theta_c = 1|\hat{p}, \hat{s}_{j(i)}, \theta_m) + (1-\alpha) \cdot \overline{\mu_{un,i}}(\theta_c|\hat{s}_{j(i)}).$$

On the left-hand-side, the conditioning shows that the average posterior depends both on realized propaganda $\hat{p}$ and the full collection of realized elite messages $\hat{\mathbf{s}} = (\hat{s}_j)_{j\in\text{elite}}$. In the first term on the right-hand-side, $\mu_{rec,i}(\theta_c = 1|\hat{p}, \hat{s}_{j(i)}, \theta_m)$ stands for the belief of receptive voter i who observes realized propaganda $\hat{p}$, belongs to the audience of elite member $j(i)$ and thus observes elite message $\hat{s}_{j(i)}$, and has mind type θ_m (which in turn is pinned down by $\hat{p}$). The bar means that this belief is averaged across all receptive voters i. In the second term, $\overline{\mu_{un,i}}(\theta_c|\hat{s}_{j(i)})$ is the average belief over unreceptive voters i, who only observe the elite's message $\hat{s}_{j(i)}$, not propaganda.[6]

Using these voter beliefs, we define the preferences of elite member j as

$$U_{ej} = (\theta_c - \kappa) \cdot \bar{\mu}(\theta_c = 1|\hat{p}, \hat{\mathbf{s}}). \tag{1}$$

Here $\theta_c - \kappa$ reflects that the elite likes competence θ_c but dislikes the incumbent politician by κ, where κ measures the ideological disagreement between the incumbent and the elite. These terms are multiplied by voters' average posterior belief, which, intuitively, governs the probability that the incumbent stays in power. We further assume that each elite member j has a small preference for sending a truthful message s_j, thus if otherwise indifferent tells the truth.

Our assumptions imply that in state R, because each elite member acts indepen-

[5] In our microfoundation in online Appendix A.2 we show that sending an identical message is the optimal strategy for a coordinating elite.

[6] We use i to denote both receptive and unreceptive individual voters.

dently and influences a zero measure of voters, each sends her message truthfully. In contrast, in state AR, because elite members act as a single decision maker, they choose their message to maximize (1). In both states, all elite members send an identical message, denoted s. Thus, for the purposes of characterizing behavior, we can represent the elite as a single player which maximizes

$$U_e = 1_{\{\theta_r = AR\}} \cdot (\theta_c - \kappa)\bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) + 1_{\{\theta_r = R\}} 1_{\{s = \hat{\theta}_c\}}. \tag{2}$$

The first term, active in the AR, represents the collective interests of the elite. The second term, active in R, represents that in isolation, each elite member chooses to tell the truth.

Since all elite members send the same message, all receptive voters, and all unreceptive voters, form the same beliefs. Thus, we can represent them with a representative receptive voter and a representative unreceptive voter, respectively.

The preferences of the politician are given by

$$U_p = \bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) - f \cdot p. \tag{3}$$

The first term captures the politician's preference to get reelected, which is governed by voters' belief about his type. The second term captures the cost $f > 0$ of sending propaganda.

Timing. In summary, the model consists of the following stages.

0) The politician's type θ_c and the reality state θ_r are realized and observed by the politician.

1) With probability β the politician cannot send propaganda; with probability $1 - \beta$ he can, and he decides on propaganda $p \in \{0, 1\}$. Propaganda is subject to trembles. The elite observes the reality state θ_r, the realized propaganda $\hat{p}$, and receives a signal $\hat{\theta}_c$ on the politician's type (correct with probability π).

2) The elite sends message $s \in \{0, 1\}$, which is subject to trembles. Voters observe the realized message $\hat{s}$. Receptive voters (share α) also observe the realized propaganda message $\hat{p}$. If $\hat{p} = 1$ then receptive voters' become persuaded ($\theta_m = P$), implying that their prior that $\theta_r = AR$ changes to $q_{ar} > 0$. Voters form posterior beliefs.

C. *Equilibrium*

Our equilibrium concept is a version of perfect Bayesian equilibrium that recognizes our framework's departure from common priors and rationality. We assume that actors correctly anticipate each others' strategies, compute expected utilities using their subjective beliefs, and choose strategies to maximize these expected utilities. We also assume that actors update in a Bayesian fashion. The trembles ensure that these updates are well defined.

The key novelty in this equilibrium is the Bayesian updating of the receptive voter. We assume that in stage 2, the posterior of the receptive voter is computed from the prior associated with his mind type θ_m. This definition allows the persuaded voter to make Bayesian inference from the elite's message and propaganda; but the order of updating is that first propaganda changes his prior, and then he makes the inference. Because aside from this novelty our equilibrium concept is standard, we relegate the formal definition to the online Appendix.

Equilibrium selection. Given the complexity of our game we expect multiple equilibria, and we introduce the following criteria for selection. First, we focus on equilibria which are *politician-pure*: in which all politician types in all states use pure strategies. Second, among these equilibria, we focus on *politician-optimal* equilibria, which maximize the ex ante expected utility of the incumbent politician in state R. We refer to equilibria satisfying these conditions as *PPO* equilibria.

D. *Discussion of model assumptions*

Modeling alternative realities. Central to our model is to explicitly incorporate the false alternative reality that voters may believe in. Importantly, this alternative reality contains optimizing agents, who impose constraints on real-world outcomes that parallel the out-of equilibrium constraints of perfect Bayesian equilibrium. Indeed, perfection requires that agents, even at information sets never reached, behave optimally; whereas we require that agents, even in imaginary states, behave optimally. This approach of explicitly modeling a strategic alternative reality—also used by Bénabou (2013) in a different setting—may be portable to other systems of misbeliefs in economics.

Elite conspiracy. In our model, the elite conspiracy emerges by assumption. Moreover, as we show in Section III.A, our qualitative results would also obtain in a framework without a conspiracy, in which the key difference between the R and the AR is that in the latter the elite has a lower lying cost. We chose to incorporate the conspiracy into our basic model both because it is realistic (Douglas et al., 2019) and because it highlights that allowing for coordination fundamentally alters the equilibrium. We endogenize the conspiracy in Section III.A by showing that when the politician can choose between a lying cost and a conspiracy narrative, he will often prefer the latter because it makes the elite appear more powerful.

Propaganda is only observed by part of the electorate. In our model unreceptive voters do not observe propaganda, implying that they are neither manipulated by it nor learn from it about the politician's type. We think of unreceptive voters as those who primarily consume the traditional media, while receptive voters as those who primarily consume social media and propaganda news. Unreceptive voters focus on the elite's message and do not internalize the politician's narrative. In contrast, receptive voters consume propaganda and fully internalize its narrative, but still encounter headline news, consistent with evidence that even strong partisans are aware of the headlines (Angelucci and Prat, 2024).

The assumption that unreceptive voters do not observe propaganda at all is for tractability. The key part of this assumption is that unreceptive voters do not fully update from propaganda, so that they may still find the elite's message informative. To demonstrate this, in online Appendix A.5 we develop two realistic ways of allowing unreceptive voters to learn from propaganda. In the first, a bounded share of unreceptive voters update from propaganda, while the rest do not. In the second, all unreceptive voters update from propaganda, but the alternative reality falsely claims that they do not, i.e., that they remain malleable to the elite's lies. In both cases, our main results continue to hold.

Receptive voters are a minority. In the main text below we assume that $\alpha < 0.5$, i.e., receptive voters are a minority. But in online Appendix A.4 we show that our main results hold for $\alpha > 0.5$ as well, albeit may require mixed strategies. We further note that for the pure strategy equilibrium we only need that receptive voters *believe* $\alpha < 0.5$, i.e., that only a minority see through the conspiracy, even if the true α is larger. Real-world conspiracy theories often assume that believers are a minority (Douglas et al., 2019).

Belief changes. We assume that propaganda can exogenously change the prior beliefs of receptive voters. This assumption is consistent with the descriptive evidence in Section I.A that supporters tend believe in the specific alternative reality disseminated by the politician. However, it is reasonable to assume that misbeliefs are also shaped by voters' demand. To address this issue, in online Appendix A.7 we develop a simple model of the demand for misbeliefs based on motivated beliefs. This model provides microfoundations for the reduced-form framework presented here.

II. Results

A. Equilibrium

We will characterize the equilibrium for $\pi < 1$ large. Empirically, this is the right parameter range, as the signal of the elite is plausibly fairly informative but imperfect. From the perspective of the analysis, assuming that π is large means that we can simplify some derivations by working them out for $\pi = 1$ and then using arguments based on continuity.

ASSUMPTION 1: *The elite wants to remove the politician irrespective of his type:*

$$\kappa > 1.$$

This assumption captures that the ideological disagreement between the politician and the elite is large. Recalling from (1) that the utility of the elite is $(\theta_c - \kappa) \cdot \bar{\mu}(\theta_c = 1|\hat{p}, \hat{s})$, since the assumption implies that $\theta_c - \kappa < 0$ for any value of θ_c, it implies that the elite—if it can influence voters—wants to minimize voters' average belief that the politician is good. This assumption ensures the incentive compatibility of elite criticism in the AR.

ASSUMPTION 2: *For π approaching 1, the cost of propaganda is smaller then its gain:*

$$f < \alpha \hat{q}_c.$$

Here, as we will explain in detail below, $\hat{q}_c = q_{ar} q_c / (q_{ar} + q_r(1 - q_c))$ is the persuaded voter's limiting posterior belief (as $\pi \to 1$), after observing propaganda and criticism, that the politician is good. Assumption 2 captures that propaganda has the potential to improve outcomes for the politician. The left-hand side is the cost of propaganda, while the right-hand side is the limit of the gain from propaganda as π approaches one. This gain derives from increasing the beliefs about the politician for the share α of receptive voters, from (approximately) zero to (approximately) $\hat{q}_c$. As we explain after stating the result, this assumption ensures the incentive compatibility of the politician's equilibrium strategy.[7]

PROPOSITION 1: *If Assumptions 1 and 2 hold, and $\alpha < 0.5$, then there exists $\bar{\pi} < 1$ such that for $\pi > \bar{\pi}$ in the unique PPO equilibrium*

1) In the reality (R):

- *The elite reports truthfully,*
- *The politician sends propaganda if he can and is bad.*

2) In the alternative reality (AR):

- *The elite always reports that the politician is bad,*
- *The politician sends propaganda if he can.*

All proofs are in the online Appendix. At a high level, the intuition for the result is as follows. In reality (part 1 of the result), the good politician has no reason to send propaganda as he will most likely be praised by the elite. The bad politician, who will likely be criticized by the elite, does have an incentive, and will do so by Assumption 2 if propaganda succeeds in discrediting criticism. But discrediting elite criticism requires a persuasive alternative explanation for that criticism: here an elite conspiracy (part 2 of the result). For this conspiracy theory to be persuasive, it is necessary that members of the elite, if they could, would in fact conspire to act against the politician. This is ensured by Assumption 1 which states that members of the elite sufficiently dislike the politician. The narrative then is that conspiring elite members always criticize, leaving the politician no choice but to spread propaganda to counter the elite's lies.

To see more precisely how discrediting works, note that in equilibrium, the receptive voter's posterior about θ_c after propaganda and elite criticism equals

$$\mu_{rec}(\theta_c = 1 | \hat{p} = 1, \hat{s} = 0, \theta_m = P) = \frac{q_c q_{ar}}{q_{ar} + q_r \pi (1 - q_c)}. \tag{4}$$

[7] Assumption 2 implies that when the share α of voters persuadable by propaganda is higher, even a lower q_{ar}, i.e., less persuasive propaganda, is sufficient to incentivize the politician.

To understand this expression, recall that since $\hat{p} = 1$, the receptive voter is persuaded ($\theta_m = P$) and assigns prior q_{ar} that reality is AR. His posterior belief that the politician is good conditional on propaganda and criticism is then formed by Bayes' rule. The numerator measures the joint probability that (i) the politician is good (q_c), (ii) propaganda, which as the politician is good only happens in the AR (q_{ar}), and (iii) criticism, which always happens in the AR.[8] The denominator measures the probability of propaganda and criticism. In the AR (q_{ar}) the politician always sends propaganda (if he can) and the elite always criticizes, explaining the first term. In the R (q_r), the bad politician sends propaganda ($1-q_c$) and the elite criticizes when it receives a correct signal (π), explaining the second term.

The key is that as $\pi \to 1$, these beliefs converge to

$$\hat{q}_c = \frac{q_c q_{ar}}{q_{ar} + q_r(1-q_c)} > 0 \tag{5}$$

so that even as the elite's message becomes arbitrarily precise, the persuaded voter's beliefs after elite criticism remain bounded away from zero. This is because the persuaded voter entertains the possibility of the AR, and in the AR the elite sends criticism even when the politician is good. This limits the perceived informativeness of the elite's message for the persuaded voter. It follows that $\hat{q}_c$ measures (for π large) the effectiveness of discrediting, explaining why it appears in Assumption 2.

B. *Detailed logic of equilibrium*

To fully flesh out the equilibrium logic, we derive voter beliefs and explain how these beliefs ensure incentive compatibility for the elite and the politician.

Voter beliefs. In the proposed equilibrium, the receptive voter's posterior, *absent propaganda* ($\hat{p} = 0$), as a function of the elite's message $\hat{s}$ is

$$\begin{aligned} \mu_{rec}(\theta_c = 1 | \hat{p} = 0, \hat{s}, \theta_m = N) =& \hat{s}\frac{\pi q_c}{\pi q_c + (1-\pi)(1-q_c)\beta} \\ &+ (1-\hat{s})\frac{(1-\pi)q_c}{(1-\pi)q_c + \pi(1-q_c)\beta}. \end{aligned} \tag{6}$$

On the left-hand side, note the receptive voter's mind type θ_m: because $\hat{p} = 0$, he is normal (N) and retains his prior that reality is R. On the right-hand side, the first term is active when the voter receives a good message from the elite ($\hat{s} = 1$). Such a message typically comes when the politician is good, but may also come when the politician is bad if the elite's signal is incorrect. However, in the latter case, the politician must not be able to send propaganda, otherwise in the proposed profile we would observe $\hat{p} = 1$. The formula then follows via Bayes' rule. The numerator is the probability that the politician is good (q_c) and

[8] These terms should also be multiplied by $1-\beta$ to reflect that the politician can send propaganda, but all terms in the denominator should also be multiplied by $1-\beta$ so we divided through with it.

the signal is correct (π); while the denominator also includes the probability that the politician is bad ($1 - q_c$), the signal is incorrect ($1 - \pi$) and the politician cannot send propaganda (β). The second term, active when the elite sends a bad message ($\hat{s} = 0$), follows analogous logic. Observe that as π approaches one, these beliefs converge to $\hat{s}$: for large π the elite's report almost fully reveals the politician's type.[9]

The receptive voter's posterior, *after propaganda*, as a function of the elite's message $\hat{s}$ is

$$\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s}, \theta_m = P) = (1 - \hat{s})\frac{q_{ar}q_c}{q_{ar} + q_r\pi(1 - q_c)}. \tag{7}$$

Equation (4) already derived these beliefs after elite criticism ($\hat{s} = 0$). In the case of elite praise ($\hat{s} = 1$) the expression is zero: the AR elite never sends praise, and in R only the bad politician sends propaganda.

Finally, the beliefs of the unreceptive voter are similar to those of the receptive voter absent propaganda (6), except that the unreceptive voter does not observe propaganda and hence does not infer from its absence, so that we do not have the β factors in the denominator.

Incentive compatibility of the elite. Having characterized beliefs, we explain why the elite follows the proposed equilibrium. In R, the behavior of the elite is straightforward: because its members are atomistic and cannot influence voter beliefs, they report truthfully. In the AR, since the elite acts as a single actor and wants to lower voter beliefs, sending criticism is incentive compatible if

$$(1 - \alpha)\left[\frac{\pi q_c}{\pi q_c + (1 - \pi)(1 - q_c)} - \frac{(1 - \pi)q_c}{(1 - \pi)q_c + \pi(1 - q_c)}\right] > \alpha\frac{q_{ar}q_c}{q_{ar} + q_r\pi(1 - q_c)}. \tag{8}$$

The left-hand side is the elite's gain from criticism: the reduced beliefs of the $1 - \alpha$ unreceptive voters. Inside the brackets, we have the difference between unreceptive voters' belief after praise versus criticism. As noted above, these expressions are similar to those of the receptive voter absent propaganda (6), with the difference that the denominators do not have the β factors. The right-hand side is the elite's loss from criticism: the increased beliefs of the α receptive voters who "see through" the conspiracy. This loss is computed by differencing (7) between $\hat{s} = 1$ and $\hat{s} = 0$.

If π is large, then the left hand side of (8) is close to $1 - \alpha$ while the right-hand side is close to $\alpha\hat{q}_c$. Therefore, if $\alpha < 0.5$, as assumed in Proposition 1, then for π large the inequality holds. Intuitively, unreceptive voters—who do not entertain the alternative reality—are manipulable by the elite; and if there are enough of them, then their impact incentivizes the AR elite to criticize.

Incentive compatibility of the politician. Finally, we turn to the politician. In

[9] When taking the limit in the second term, we used that $\beta < 1$.

R, as noted above, the good politician who expects praise from the elite has no reason to send propaganda. For the bad politician, sending propaganda is incentive compatible if

$$\begin{aligned}\alpha\Bigg[\pi\left(\frac{q_{ar}q_c}{q_{ar}+q_r\pi(1-q_c)}-\frac{(1-\pi)q_c}{(1-\pi)q_c+\pi(1-q_c)\beta}\right)\\ +(1-\pi)\cdot\frac{-\pi q_c}{\pi q_c+(1-\pi)(1-q_c)\beta}\Bigg]>f.\end{aligned}\tag{9}$$

The left hand side measures the expected gain from propaganda. Propaganda only has an effect on receptive voters (α). For them, propaganda changes expected beliefs about competence θ_c, from the expected value of beliefs absent propaganda (6) to the expected value of beliefs in the presence of propaganda (7). These expectations are computed over the distribution of the elite's message that $\hat{s}=0$ with probability π and $\hat{s}=1$ with probability $1-\pi$. The formula then follows by direct substitution. For the politician to prefer propaganda, this expected gain has to exceed the cost of propaganda f. As π approaches one, the second and third fractions on the left-hand side vanish and the remaining terms converge to $\alpha\hat{q}_c$. By Assumption 2, $\alpha\hat{q}_c > f$, thus for π sufficiently large the bad R politician's incentive compatibility constraint holds.

Next consider the politician's incentive compatibility constraint in the AR. Both the good and the bad types observe the state and know that the elite always sends criticism. This means that their incentive compatibility follows from (9) because they expect more criticism. Formally, on the left-hand side the weight on the first two terms increases from π to 1 while the weight on the third term decreases to zero. It follows that for large π, the AR politicians also send propaganda.

Note that the good politician's choice of propaganda in the AR is key for the updating of the voter, who, if propaganda is to be effective, should not be able to infer from it that the politician is bad. This logic underlies equation (4) and prevents the revelation of the bad R politician's type.

Relaxing the constraint on receptive voters. Proposition 1 focuses on the case when only a share $\alpha < 0.5$ of voters are receptive. However, we show in online Appendix A.4 that the model has a unique PPO equilibrium which features propaganda even when $\alpha > 0.5$. This equilibrium is identical to that of Proposition 1 in the behavior of the politician. But for α high, the behavior of the elite in the AR is more complex: they now mix between criticizing and praising the politician. The core intuition is that for α high, the conspiracy theory has to address an internal consistency problem: why should elites lie once they know that most people (receptive voters) see through their lies? In our model, the alternative reality evolves to address this problem by making the elites more cunning. Elites now sometimes tell the truth, to confuse voters and ensure that voters no longer see through their lies. Importantly, our main qualitative predictions continue to hold in this more complex equilibrium.[10]

C. Theoretical implications of equilibrium

The equilibrium has a number of new theoretical implications which lead to testable predictions.

Deflection. A core implication is that propaganda can deflect elite criticism.[11]

COROLLARY 1: *Suppose that Assumptions 1 and 2 hold, $\pi > \bar{\pi}$, and $\alpha < 0.5$. In the PPO equilibrium, propaganda by the bad politician increases voters' beliefs that the politician is good:*

$$E[\bar{\mu}(\theta_c = 1|\hat{p} = 1, \hat{s})|\theta_c = 0] > E[\bar{\mu}(\theta_c = 1|\hat{p} = 0, \hat{s})|\theta_c = 0].$$

Note that $E[.]$ represents objectively correct expectations. The result follows essentially from equation (4), which showed (for π large) that under propaganda, elite criticism does not persuade receptive voters. Thus, propaganda enables bad politicians to remain in power, and by doing so, reduces political accountability. Importantly, this result applies in the presence of independent media that provides reliable information on the politician's type. It follows that in our model, populist propaganda reduces accountability.

Inversion. A second implication of the model is that among receptive voters, propaganda *inverts* the effect of the elite's message: elite criticism increases, while elite praise decreases receptive voters' support for the politician.

COROLLARY 2: *Suppose that Assumptions 1 and 2 hold, $1 > \pi > \bar{\pi}$, and $\alpha < 0.5$. In the PPO equilibrium, in the presence of propaganda, elite criticism strictly increases the receptive voter's support for the politician: $\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 1) < \mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0)$.*

Formally, this result follows from (7), which shows that the persuaded voter's belief after elite criticism remains bounded away from zero, but after elite praise becomes equal to zero. The former statement holds because propaganda discredits elite criticism; the latter holds because elite praise is not possible in the alternative reality, so that observing it punctures the alternative reality.

The underlying intuition is that for the persuaded voter, the elite's message is primarily informative about the nature of reality, not the politician's type. This force emerges because in our model the voter questions the motives of the elite (i.e., the nature of reality). In particular, because in the AR the elite always criticizes, while in the R (for $\pi < 1$) the elite only sometimes criticizes, elite criticism is more consistent with the AR and increases beliefs in the AR. Then,

[10] We note that conspiracy theories could resolve the internal consistency problem in other ways too, such as by falsely claiming that α is low, i.e., that only a minority are aware of the conspiracy.

[11] For consistency with the Proposition we focus on the $\alpha < 0.5$ case in stating our corollaries, but we show in the online Appendix that Corollaries 1, 2 and 3 hold for all values of α.

because in the AR propaganda may come from a good politician, while in the R it always comes from a bad politician, the increased belief in the AR increases beliefs that the politician is good. As this logic makes it clear, $\pi < 1$ is necessary for inversion, because it ensures that elite criticism is relatively more consistent with the AR. We conclude that populist propaganda inverts the standard effect of elite information on receptive voters' beliefs.[12]

Although in our model inversion is driven by beliefs about the politician's competence, in practice there can be an alternative mechanism driven by beliefs about the politician's anti-elite nature. That is, persuaded voters may infer from elite criticism not that the politician is (relatively) competent, but that he is anti-elite. We find this alternative mechanism plausible, but for simplicity we did not incorporate the additional type dimension necessary to model it.

Amplification. The previous result explored how beliefs in the presence of propaganda vary with the elite's message. We next characterize beliefs in the presence of propaganda *on average.* The key insight is that propaganda-induced AR beliefs are amplified by Bayesian updating.

COROLLARY 3: *Suppose that Assumptions 1 and 2 hold, $\pi > \bar{\pi}$, and $\alpha < 0.5$. In the PPO equilibrium, even though signals are generated by R, the receptive voter's expected posterior, relative to his (propaganda-induced) prior, moves towards the AR: $E[\mu_{rec}(AR|\hat{p}, \hat{s})|\hat{p} = 1] > q_{ar}$.*

This result seems counterintuitive from a Bayesian perspective. A standard Bayesian with the wrong prior, as long as his prior assigns positive probability to the truth, should form posteriors that on average drift towards the truth. Corollary 3 says that here, even though R is always included in the receptive voter's prior, he forms beliefs that on average drift away from the truth.

To understand the result, recall that propaganda increases the receptive voter's prior about the AR to q_{ar}. He then uses this new prior to interpret the sequence of messages he receives. For π large, this sequence is very likely to be propaganda and criticism, which leads him to increase his posterior belief about the AR to (approximately)

$$\hat{q}_{ar} = \frac{q_{ar}}{1 - q_r q_c} > q_{ar}. \tag{10}$$

This increase is non-standard: the sequence of propaganda and criticism contains (for π large) essentially no new information beyond the fact that the voter's prior changed. It emerges because the voter with the new prior still entertains the state of the world that reality is R and the politician is good (which has probability $q_r q_c$) even though the fact that his prior changed already ruled it out. Updating

[12] We note that inversion can be broken down into two predictions: that the persuaded voter's beliefs (i) increase after elite criticism, and (ii) decrease after elite praise. Note that (i) cannot hold without (ii): since the persuaded voter updates as a Bayesian, his prior must be a convex combination of his posteriors.

eliminates that state from the voter's mind, increasing his beliefs in the AR (for π large) by a factor $1/(1-q_rq_c)$. Intuitively, the voter neglects the fact that his prior is large precisely when the likely outcomes are propaganda and criticism. His correlation neglect amplifies AR beliefs because these likely outcomes are more consistent with the AR than with the R.

These arguments imply that the success of propaganda is determined by two factors: (i) planting initial misbeliefs; (ii) a plausible narrative that amplifies misbeliefs by explaining the observed reality better than the true narrative. We note that these two factors are already reflected in Assumption 2. The assumption requires that the posterior belief about the politician's quality $\hat{q}_c$ is large enough, and we can express that posterior, normalized by the prior q_c, as

$$\frac{\hat{q}_c}{q_c} = \frac{q_{ar}}{1-q_rq_c}. \tag{11}$$

Here q_{ar} is the initial false belief, while $1/(1-q_rq_c)$, as discussed above, is the amplification.

The mechanisms identified here are related to two lines of research on misbeliefs. First, amplification depends on our timing assumption that first propaganda changes the prior and then the voter updates from the new prior, which is related to the sequential updating assumption of Cheng and Hsiaw (2022) and Koçak (2018) that a person first updates about the credibility of a source and then about the information provided by that source. A key difference is that in our setting the change in the prior is driven by the supply side, leading to predictions about when misbeliefs arise. Second, the logic of amplification is related to the research on persuasion with models, in which models more consistent with the data are found to be more persuasive (Schwartzstein and Sunderam, 2021; Aina, 2023). Although we also differ in our formal approach, our key contribution to this work is the applied finding that beliefs in politically-supplied conspiracy theories are amplified by observed outcomes.

Comparative statics of presence of propaganda. Proposition 1 shows that bad politicians always choose propaganda, but focuses on the case when the elite strongly dislikes the incumbent politician: $\kappa > 1$ by Assumption 1. We now relax that assumption.

COROLLARY 4: *Suppose that Assumption 2 holds, $\pi > \bar{\pi}$, and $\alpha < 0.5$. If $1-\pi < \kappa < \pi$, then there is a unique PPO equilibrium, and in that equilibrium no politician sends propaganda.*

The point here is that when κ is (somewhat) lower than 1, propaganda is no longer used. This is because when κ is in the specified range, the elite wants to keep the politician if and only if he is good. But then the narrative in which the elite conspires to remove a good politician is not plausible and will not be believed by voters. The politician then has no reason to send propaganda. It follows that successful propaganda requires the ideological disagreement κ between the politician and the elite to be sufficiently large.

Comparative statics of effectiveness of propaganda. We turn to explore the conditions under which propaganda is more or less effective. We focus on the effect of q_c, the prior probability that the politician is good, which may be reflected in the politician's baseline popularity.

COROLLARY 5: *Suppose that Assumptions 1 and 2 hold, $\pi > \bar{\pi}$, and $\alpha < 0.5$. In the PPO equilibrium,*

1) *The receptive voter's belief in the AR conditional on observing propaganda and criticism, $\mu_{rec}(AR|\hat{p} = 1, \hat{s} = 1)$, is increasing in q_c.*

2) *The bad R politician's expected gain from propaganda*

$$E[\bar{\mu}(\theta_c = 1|\hat{p} = 1) - \bar{\mu}(\theta_c = 1|\hat{p} = 0)|\theta_c = 0]$$

is increasing in q_c.

The first result is a comparative static of the amplification in Corollary 3 and follows essentially from equation (10). When the politician is more likely to be good (q_c higher), propaganda and criticism are less likely in the R but more likely in the AR, and thus increase more the posterior of the AR. Intuitively, when the politician is more likely to be good, a conspiracy is a more plausible explanation for elite criticism.

The second result, that propaganda is more valuable for the politician when q_c is higher, follows essentially from equation (11). There are two forces. First, as just noted, with q_c higher, propaganda induces a larger belief change towards the AR. Second, belief in the AR is better for the politician, because in that state he is perceived to be good with a higher q_c probability.

An interesting implication of the second result is that as a politician looses popularity, propaganda becomes less effective in shoring up support. When perceived competence falls, the conspiracy becomes a relatively less plausible explanation for elite criticism; and even if it does discredit elite criticism, it does not undo voters' direct perception of the politician's incompetence.

Demand for misbeliefs. A weakness of our theoretical model is that we take the demand for misbeliefs as given. To address this, in online Appendix A.7 we develop a microfounded model of the demand side, which is based on the idea of motivated beliefs (Brunnermeier and Parker, 2005; Bénabou and Tirole, 2006). In this model, a voter who experiences propaganda becomes aware of the elite conspiracy alternative reality, and chooses his prior belief q_{ar} in that alternative reality. This choice is made before the voter updates from propaganda and the elite's message. The voter chooses q_{ar} by trading off his subjective expected utility from the election outcome against a cost of changing his prior. The utility from the election outcome comes from our microfoundation of voter behavior (online

Appendix A.2). The cost of changing the prior is a function of the expected posterior belief in the AR, capturing the idea that beliefs in the AR may impair decisions in other domains.

We show in Proposition 1 in the online Appendix that our equilibrium is robust to incorporating the demand for misbeliefs, but features an endogenously chosen $q_{ar} > 0$.[13] We also show that in this equilibrium the predictions of Corollaries 1-5 continue to hold.

D. *Evidence*

We turn to discuss evidence on the model's implications, focusing on Corollaries 1-4.

Propaganda lowers accountability in democracies. A growing body of evidence documents that populism is associated with reduced accountability. Funke, Schularick and Trebesch (2023) show that populism reduces GDP per capita and consumption by 10% relative to a plausible non-populist benchmark; yet populists stay in power for twice as long as non-populists. More anecdotally, Donald Trump won the 2024 US election despite being a convicted felon, while Hungary's Orban and Turkey's Erdogan stayed in power for extended periods despite evidence that they eroded democratic institutions (Guriev and Treisman, 2022).

To our knowledge, existing theories do not explain populism's association with reduced accountability. Models of populism, including Acemoglu, Egorov and Sonin (2013) and Bellodi et al. (2023), work through the logic that populism is a positive signal about the politician, hence cannot easily explain reduced accountability. Models that do predict reduced accountability do so in non-democracies, through the mechanisms of repressing voters or silencing the media (Egorov and Sonin, 2024; Guriev and Treisman, 2020). For example, in Guriev and Treisman (2020), propaganda paints the politician in a false positive light, and independent media cannot correct this false view because it is silenced. These mechanisms cannot easily explain reduced accountability in democracies that have independent media.

Our model explains reduced accountability with Corollary 1, which predicts that populist propaganda reduces accountability. The reason propaganda works despite an independent media is that it discredits that media. The mechanism of discrediting—a conspiracy theory—is also consistent with evidence: as we discussed in Section I.A, the populist narrative is often centered around an elite conspiracy. We note that although in this paper we focus on democracies, our mechanism may also play a role in autocracies and hybrid regimes, where it can complement other anti-media strategies such as censorship and media capture.

Propaganda inverts the elite's effect on receptive voters. A key fact in contemporary US politics is that during 2023, the growing body of critical evidence against Donald Trump, including four criminal indictments, was accompanied by

[13] We no longer show that the equilibrium is unique.

Table 2—Impact of indictment on Trump's support by Republicans intending to vote in primary

	All	Moderate	Conservative
More likely to vote for him	41%	24%	44%
Less likely to vote for him	4%	13%	3%
Not affect whether you vote for him	55%	63%	53%
Observations	488	80	408

Source: CBS New poll between June 7 and 10, 2023 (YouGov, 2023).

an *increase* in his support among Republican voters (Swan et al., 2023). Since the indictments were produced by the US legal system, this reaction by the presumptive party of law and order is puzzling. It is even more puzzling when compared to two other salient legal cases against leading politicians. President Richard Nixon, following the Watergate scandal, and New York mayor Eric Adams, following his 2024 criminal indictment, both experienced large reductions in popular support even among supporters of their own party (Franklin, 2018; McFadden and Mays, 2024). We not aware of other formal models that explain why, following their legal challenges, support increased for Trump but declined for Nixon and Adams.

Our model can explain these facts through its inversion and comparative statics predictions. We first describe how inversion explains the increase in support for Trump and present supporting evidence; and then describe how the comparative statics explain the opposite pattern for Nixon and Adams. The explanation for Trump follows directly from Corollary 2, under the assumptions that Trump spreads propaganda and that Republicans correspond to the model's receptive voters. Then, the Corollary predicts that the indictments—elite criticism—*causally* increase Republicans' support for Trump.

We present two pieces of evidence that support this causality. First, in Table 2 we show results from a 2023 poll investigating the impact of the indictments on Trump's political support (YouGov, 2023). Among registered Republicans intending to vote in the primaries, 41% claimed that they would be more likely, and only 4% claimed that they would be less likely, to vote for Trump if he was indicted in the matter of handling classified documents. The effects were large even among moderate Republicans. Thus, Republicans anticipated that their own support would increase in response to critical evidence.

But this evidence is about voters' hypothetical behavior. For evidence on voters' actual behavior, we turn to the impact of Republican politicians' scandals on campaign contributions. Scandals, diffused by the news media, are an example of elite criticism, thus the logic of our model predicts that—given the presence of propaganda—they should increase campaign contributions from receptive voters. To explore this effect, we take Wikipedia's list of political scandals of Republican House candidates during 2017-2022 (Wikipedia, 2024), and select the 11 scandals that are related to sexual misconduct, financial misconduct, election fraud, or violence. These are issues on which probably most voters and elite members

Table 3—Impact of scandals on contributions from Trump-supporter and other donors

	Trump donors	Trump donors	Other donors
	Share	Amount (1,000 dollars)	
Scandal	0.075***	20.35**	-9.86
	(0.009)	(9.88)	(16.59)
Representative and quarter f.e.	yes	yes	yes
Control mean	0.065	16.06	119.0
Observations	3,393	4,382	4,382

Note: Observations are house representative by quarter cells. Representatives with a scandal during 2017-22 are in the sample in a one-year window around their scandal. Representatives without a scandal are in the sample in all quarters in which they were in office during 2017-22. Scandal is one for representatives that have a scandal in the quarters after the scandal. Column 1 is restricted to observations with non-zero total donations. In column 1, the dependent variable is the share of donations from Trump supporters; in columns 2 and 3 it is the volume of donations from Trump supporters and from other Republicans, respectively. Standard errors are clustered by state.

agree, thus they correspond to θ_c. We combine these data with donation data from the Federal Election Commission (FEC, 2024).[14]

We estimate difference-in-differences regressions of the effect of a scandal on donations that come from Trump supporters and from other donors. We define Trump supporters as individuals who donated to the Make America Great Again PAC in the 2020 election campaign. Our control group includes donations to other Republican House candidates in the same period. Table 3 reports the results. Column 1 shows that relative to a control mean of 6.5 percent, the share of donations coming from Trump-supporter donors increased after the scandal by a significant 7.5 percentage points. Columns 2 and 3 show that this increase was largely driven by a significant increase in Trump-supporters' donations of about $20,000 per quarter, with no significant change in other donors' donations. We conclude that the evidence supports the causal link between elite criticism and political views predicted by our model.[15]

Beyond predicting that elite criticism should increase Republicans' support for Trump, the model also predicts the underlying mechanism: that elite criticism should increase Republicans' beliefs in the alternative reality. This prediction is consistent with survey evidence we present in Figure 1, which plots, over time, the

[14] We use quarterly data on contributions made by private individuals to the election committees of congressional candidates.

[15] A possible alternative explanation is that the scandal increased election competitiveness, and competitiveness affected donations. Two pieces of evidence speak against this. First, as Table 3 shows, the effect is concentrated among Trump-supporters, and it is not clear why they should care more about the election. Second, in online Appendix B we show that when competitiveness increases because of redistricting, there is no analogous impact on donations.

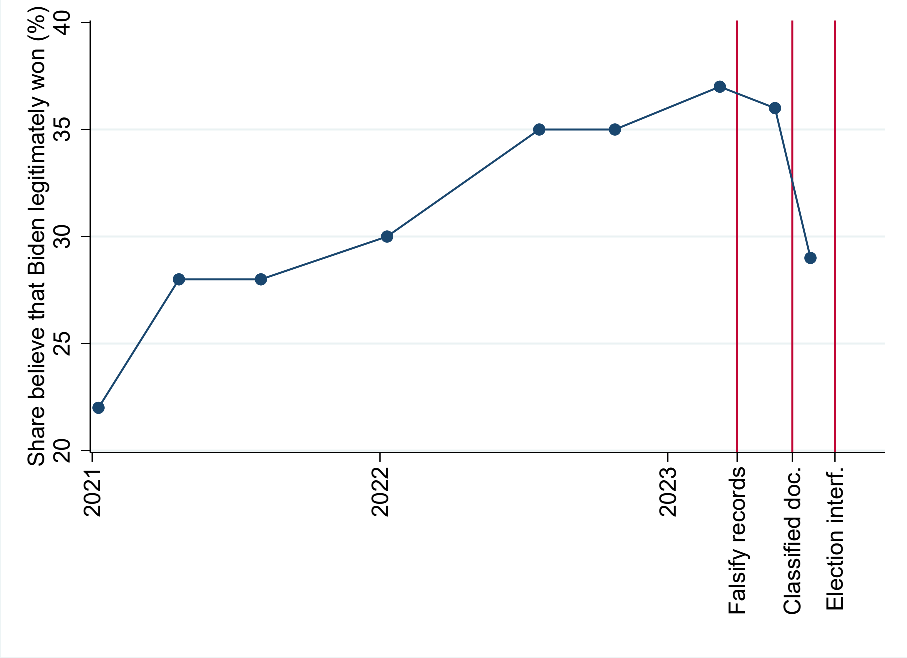

Figure 1. Share of Republicans who believe that Biden legitimately won in 2020

Source: Nine CNN opinion polls conducted by SSRS between January 2021 and August 2023 (SSRS, 2023). Vertical lines indicate dates of indictments against Donald Trump in 2023: (1) in March 30 for falsifying business records; (2) in June 8 for mishandling of classified documents; (3) in August 1 for attempting to overturn the 2020 US presidential election.

share of Republican-leaning voters who believe that Biden legitimately won the 2020 presidential election. The share steadily increases during 2021 and 2022, but sharply drops in the summer of 2023, exactly around the time of the first three Trump indictments. That is, as predicted by the model, beliefs in the conspiracy theory that the 2020 election was stolen (AR beliefs) increased precisely at the time of the indictments (elite criticism). Although this evidence does not conclusively prove that the indictments caused the increase in misbeliefs, the sharp change in the absence of other salient events is suggestive of causality.

Why did receptive voters respond differently to the elite criticism of Trump than they did to the elite criticism of Nixon or Adams? Although both Nixon and Adams attempted to use conspiracy theories to deflect criticism (Shabecoff, 1974; Mays, 2024), they were unsuccessful. Corollary 4 says that propaganda is only successful when the cleavage between the politician and the intellectual elite is sufficiently large. But both Nixon and Adams represented the more educated party—Republicans in the early 1970s, Democrats today (Kuziemko, Marx and Naidu, 2022)—suggesting that the cleave between them and the elite was small, making it implausible that the elite would conspire to remove them from power.

We conclude that the model is consistent with the presence of inversion for Trump and its absence for Nixon and Adams.

Finally, we relate our results on inversion to the research on the "backfire effect" that corrective information can sometimes lead individuals to more strongly endorse a misbelief. Early work showed evidence for this backfire effect, but more recent research suggests that corrective information is usually somewhat effective in correcting beliefs (Nyhan, 2021). Our inversion prediction is different from the backfire effect: it is not about the impact of corrective information but about the impact of elite criticism of the politician. Corrective information is often not elite criticism of the politician. In fact, our model suggests the comparative static that the backfire effect should be stronger when the salient interpretation of corrective information is elite criticism of the politician.

Politically supplied misbeliefs are amplified by outcomes. Beliefs in the alternative reality are surprisingly widespread: e.g., as shown in Figure 1, the overwhelming majority of Republicans believe in the conspiracy theory that Biden did not legitimately win the 2020 elections. Such widespread misbeliefs seem difficult to explain purely with propaganda's ability to move priors. But they may be easier to explain with Corollary 3, which predicts that realized outcomes amplify those prior beliefs. We do not have direct evidence on amplification, but we note that in line with its logic, alternative realities tend to be supplied precisely in conditions in which they are more consistent with realized outcomes. Returning to the contexts of Section I.A, in the US, the deep state conspiracy was supplied around the time of the Trump indictments; in Hungary, the Soros-Brussels conspiracy to import immigrants was supplied around the time of the European migrant crisis; in Israel, the judiciary-media conspiracy was supplied around the time of the legal cases against Netanyahu. In each of these cases, observed outcomes—indictments, immigration, court cases—were almost inevitable in the alternative reality and hence should have strengthened beliefs in that alternative reality.

Propaganda is only used in divided societies by anti-elite politicians. Corollary 4, in combination with Proposition 1, show that propaganda is only used if disagreement between the politician and the elite is large, that is, if (i) there is large division in society, and (ii) the politician is on the other side of the elite in that division. Part (i) highlights a new mechanism that links societal cleavages to populism. A novelty relative to prior work is that the cleavage need not be about differences in income, with the elite being the rich; but may be about differences in cultural values, with the elite being intellectuals. Thus, in the US context, our model formalizes the narrative that Democrats' move to the left enabled right-wing populism (Norris and Inglehart, 2019), through the logic that the resulting increase in cultural disagreement κ made the elite conspiracy narrative plausible. Part (ii), as we noted above, helps explain the presence of inversion under Trump and its absence under Nixon and Adams, through the logic that the latter politicians represented pro-elite parties.

III. Applications

We turn to develop two applications of our model: endogenizing the conspiratorial nature of the alternative reality, and studying the impact of populism on government policy. In each application, we introduce additional assumptions to capture new features of the environment, but do not change our fundamental assumptions concerning the alternative realities.

A. Endogenizing the conspiracy theory

In our basic model, the AR features a conspiracy only by assumption. Moreover, for our qualitative results, this assumption is not strictly necessary: we could obtain our qualitative results in a model without a conspiracy, in which in the AR the elite has a lower cost of lying. Thus, incorporating a conspiracy into the model may seem superfluous. Here, we argue that conspiracy theories are in fact a natural implication of our framework, justifying our approach and helping to explain prevalence of political conspiracy theories.

The basic insight we develop here is that the elite conspiracy solves a collective action problem. This problem arises because a lie about the politician's competence by any given elite member benefits every other elite member, since they all benefit from lower support for the politician. The ability to coordinate allows these externalities to be internalized, strengthening the incentives to lie. As a result, the conspiracy-based alternative reality can explain away a wider range of criticism: even credible evidence like an indictment that individual elite members would not, but collectively the "deep state" might have the incentive to manufacture.

To explore these issues formally, we extend the model to allow for two different types of alternative realities. In the first, elite members have a lower lying cost but do not have the ability to coordinate; in the second, they also have the ability to coordinate. In addition, we introduce a variable that measures the credibility of the evidence the elite provides in support of its message: a publicly known fabrication cost that each elite member has to pay in order to send a false message.

Model. Modeling the lying-cost alternative reality requires that each elite member has some individual-level incentives to manipulate. We thus assume that the elite consists of a finite number of members N, and each of them accesses a mass $1/N$ of voters. We further assume that there is a non-infinitesimal lying cost χ which can be written as the sum of a fabrication cost χ_f and an integrity or honor cost χ_h. The fabrication cost χ_f is the cost of manufacturing the evidence presented in the elite's message—such as videos of intensive care units during Covid—which we assume is known by the voter and cannot be changed by the alternative reality. The honor cost χ_h is the private cost to an elite member for telling a lie. In addition, there is an organizing cost χ_o which each elite member has to pay if they conspire. In the objective reality both χ_o and χ_h are prohibitively high, so that elite members do not conspire and tell the truth.

The politician chooses to send propaganda about one of the following two alternative realities.

1) Lying cost AR. In this AR, χ_o continues to be prohibitively high but $\chi_h = 0$. The cost of sending propaganda to make the voter believe in this AR is $f' < f$.
2) Conspiracy AR. In this AR both $\chi_o = 0$ and $\chi_h = 0$. The cost of sending propaganda to make the voter believe in this AR is f. Since $\chi_o = 0$, we assume that in this AR the elite always coordinates if it is in their joint interest, i.e., there are no coordination problems.

We will denote the lying-cost AR by AR1 and the conspiracy AR by AR2. There are three reality states: $\theta_r \in \{R, AR1, AR2\}$ such that the objective probability of $\theta_r = R$ is 1. If the receptive voter receives propaganda on AR1, his prior of AR1 increases to q_{ar}; if he receives propaganda on AR2, his prior of AR2 increases to q_{ar}. His prior of the other possible AR remains zero. The politician in any reality state can send either AR1 or AR2 propaganda. This is the natural generalization of our model to the setting with multiple alternative realities.

Since the elite has a finite number of members, average voter beliefs become

$$\bar{\mu} = \frac{\sum_{j=1}^{N} \bar{\mu}_j(\hat{p}, \hat{s}_j)}{N}$$

where $\bar{\mu}_j(\hat{p}, \hat{s}_j)$ is the average belief among voters influenced by elite member j. We assume $N > 1$.

Given the non-infinitesimal lying costs, elite member j's utility becomes

$$(12)\quad U_{ej} = (\theta_c - \kappa) \cdot \bar{\mu} - \chi_f \cdot 1_{\{s_j \neq \hat{\theta}_c\}} - \chi_h \cdot 1_{\{\theta_r = R\}} 1_{\{s_j \neq \hat{\theta}_c\}} - \chi_o \cdot 1_{\{\theta_r \neq AR2\}} 1_{organize_j}.$$

The first term captures the impact of the average voter belief $\bar{\mu}$. The remaining terms reflect that the elite must pay a fabrication cost χ_f for fabricating a lie in all realities, an honor cost χ_h for not telling the truth in R, and an organizing cost χ_o for attempting to organize in R and AR1. Since χ_o and χ_h are prohibitively high, in equilibrium the last two costs are never paid. The politician's utility is still given by (3) with $\bar{\mu}$ governing voter beliefs.

PROPOSITION 2: *Under Assumptions 1-2, if $\alpha < 0.5$ then, in the following ranges for χ_f, for π large enough, the unique PPO equilibrium is such that*

1) *If $\chi_f < (1 - 2\alpha)/N$, then in reality the bad politician sends lying cost propaganda;*
2) *If $1/N < \chi_f < (1 - 2\alpha)$, then in reality the bad politician sends conspiracy propaganda;*
3) *If $1 < \chi_f$, then no politician sends propaganda.*

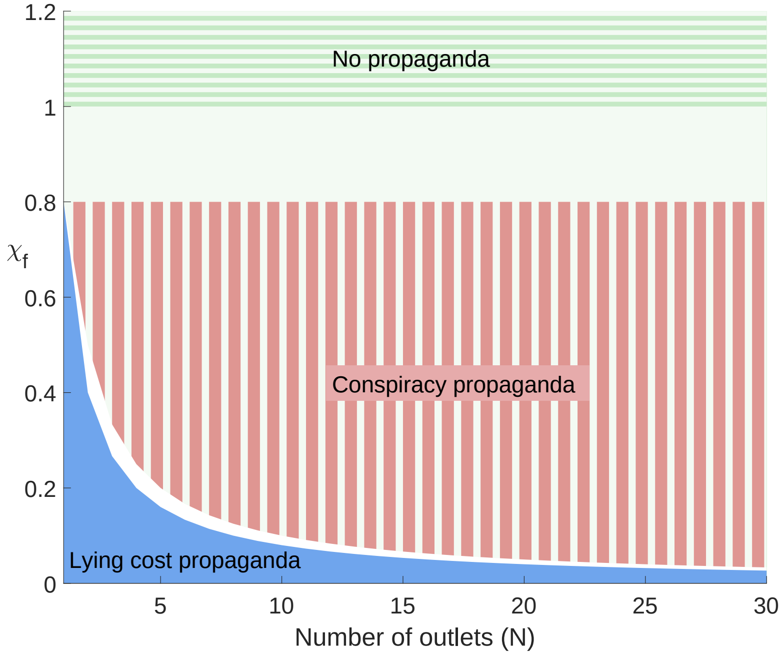

FIGURE 2. ENDOGENOUS AR

Figure 2 is helpful for understanding the result. The horizontal axis is N, the number of elite members, and the vertical axis is χ_f, the publicly known fabrication cost. The first part of the Proposition says that for χ_f low, i.e., when the evidence supporting the elite's message is easy to fabricate, lying cost propaganda is sufficient. In this case, the narrative that elite members have "no honor" is sufficient to explain away the weak evidence. More precisely, because each elite member influences a share $1/N$ of voters, each has a non-negligible gain from manipulating these voters. Hence, absent an honor cost and given the low fabrication cost, each is willing to fabricate a fake message. This range corresponds to the blue (solid) region in the Figure.

The second part of the Proposition says that for χ_f in the middle range, the equilibrium uses conspiracy propaganda. In this range lying cost propaganda no longer works: the individual-level gain to each elite member no longer covers the fabrication cost. But conspiracy propaganda works, because if elite members act collectively, then the individual-level gains increase by a factor of N. Intuitively, each elite member now internalizes that her action benefits all other elite members, and thus has higher-powered incentives to fabricate her message. This is the equilibrium in the red (vertical stripes) region in the Figure. Observe that the higher N, i.e., the more fragmented the elite, the wider the range of the conspiracy equilibrium. Since in practice N is likely to be large, the Proposition suggests that the conspiracy AR is a likely outcome.

The third part of the Proposition says that for χ_f high, corresponding to the green (horizontal stripes) region, propaganda is not used. At such a high cost, even a collectively acting elite does not have sufficient incentives to fabricate lies.[16]

Implications and evidence. The Proposition has three main implications. First, it predicts that misbeliefs should often feature conspiracy theories. Conspiracy theories are indeed prevalent (Douglas et al., 2019) and we are not aware of other formal theories that explain their emergence.

Second, the Proposition predicts that increasing the credibility of evidence need not correct beliefs. This is because the politician can respond to an increase in the fabrication cost χ_f by escalating the alternative reality from a lying cost AR to a conspiracy AR, and explain away the more credible evidence with a more powerful elite.[17] Through this logic, alternative realities can resist evidence. Thus, we formalize the argument of Sunstein and Vermeule (2009) that maintaining a conspiracy theory in the face of contradictory evidence requires an ever-widening conspiracy.

Third, the proposition implies that propaganda—because it often uses a conspiracy theory—can lead to distrust in science and the non-adoption of best practices. This is because the conspiracy narrative makes the elite more powerful in other domains too, which affects the behavior of the voter in those domains. Once the voter believes that the elite can conspire, he will suspect that even seemingly credible elite messages in the health or climate domains may be driven by the elite's private interest. For example, reports about climate change by scientists, which seem prohibitively expensive to fabricate individually, may be driven by their collective desire to control the population (Uscinski, Douglas and Lewandowsky, 2017).

This last point helps explain the evidence that misbeliefs under populism go beyond politics, including Republicans' attitudes in the health and climate domains. For example, Allcott et al. (2020) show that under Covid Republicans were less likely to engage in social distancing; Wallace, Goldsmith-Pinkham and Schwartz (2022) show that they had higher excess death rates attributable to Covid; and Hotez (2023) shows the persistence of Covid-denialism in the face of credible evidence. Our model explains these facts through the logic that populism causes distrust in the elites. This is in contrast to prior work that emphasized the causality from distrust to populism (Bellodi et al., 2023; Guiso et al., 2023). Our chain of causality suggests that eliminating propaganda should improve trust in

[16] As illustrated by the white areas between the three regions in the Figure, the Proposition does not cover the full range of χ_f values. In the intermediate ranges mixed equilibria are possible, which did not seem central to our message. We also note that the π large condition in the Proposition is required by χ_f, rather than uniformly.

[17] More generally, and outside our current model, the politician could escalate the scale of the conspiracy theory by claiming that it involves more actors.

science.

B. Government policy under conspiratorial populism

In our second application we explore how conspiratorial populism shapes government policy. This is a fundamental question since much evidence shows that populism is associated with large economic and non-economic costs (Guriev and Papaioannou, 2022) which presumably derive from harmful policies. Our model identifies two mechanisms through which populism leads to harmful policies. First, there is a direct effect of reduced accountability: populism enables incompetent, corrupt, or authoritarian politicians to maintain political power, and implement incompetent, corrupt, or authoritarian policies. Second, on top of this basic effect, our model predicts that populist politicians will often choose harmful policies *purely to trigger elite criticism* and thereby strengthen beliefs in the alternative reality. We turn to formally develop this second insight.

To incorporate government policy into the model in a simple way, we assume that the bad politician can set policies that make his bad type more visible to the elite. Formally, in stage 1, the bad politician, simultaneously with his propaganda decision, can take an action $e \in \{0, 1\}$, where $e = 1$ represents his intent to set a bad policy. We assume that $e = 1$ has vanishingly small cost to the politician. Although e is not directly observable, it reduces the quality of the policy mix, and thereby increases the probability that the elite's signal $\hat{\theta}_c$ is correct to $\pi' > \pi$.[18] The key here is that setting $e = 1$ invites elite criticism.

We make one other substantive departure from the basic model: we assume that the politician cares more than the elite about the beliefs of receptive voters. Formally, the politician maximizes

$$U_p = \tilde{\mu}(\theta_c = 1|\hat{p}, \hat{s}) - f \cdot p, \tag{13}$$

where $\tilde{\mu}$ is the weighted average of receptive and unreceptive voters' beliefs with a new weight α'

$$\tilde{\mu}(\theta_c = 1|\hat{p}, \hat{s}) = \alpha' \cdot \mu_{rec}(\theta_c = 1|\hat{p}, \hat{s}) + (1 - \alpha') \cdot \mu_{un}(\theta_c|\hat{s})$$

and we will assume that $\alpha' > \alpha$ by a sufficiently large margin. This assumption creates a wedge between the incentives of the politician (who weight receptive voters with α') and those of the elite (who weight receptive voters with α). One interpretation of this wedge, already suggested before, is that receptive voters incorrectly perceive that $\alpha < 0.5$, even though the true α is larger. Since the perceived α governs the incentives of the elite in the AR, this misperception creates the desired wedge. Another interpretation is that the politician also cares

[18] Formally, define the realized policy as $\theta_c - \phi e + \xi$ where ϕ is the effectiveness of the politician's action and ξ is a random policy shock. The elite's signal is $\tilde{\theta}_c = 1\{\theta_c - \phi e + \xi > \tau\}$, where the τ threshold is such that $\Pr[\xi > \tau - 1] = \pi$, $\Pr[\xi > \tau] = 1 - \pi$, and $\Pr[\xi > \tau + \phi] = 1 - \pi'$.

about winning a primary election, and hence overweights the beliefs of his core (receptive) supporters.[19] And a third interpretation is that the elite cares disproportionately about some part of the audience, such as foreigners or donors, who are plausibly less receptive to propaganda. Under all three interpretations, the wedge allows both the politician and the conspiring elite to gain from elite criticism.

PROPOSITION 3: *Under Assumptions 1 and 2, if $\alpha < 0.5$ and $\alpha' > 1/(1+\hat{q}_c)$, then for π large enough and $\pi' > \pi$, in the unique PPO equilibrium*

1) *All propaganda and message choices are as in Proposition 1.*
2) *The bad politician chooses to set a bad policy ($e = 1$) if and only if reality is R and he sends propaganda.*

That is, populist propaganda drives bad policies. To see the intuition, first note that because $\alpha < 0.5$ we are in the parameter range of Proposition 1. In this range, by equation (8), it is a dominant strategy for the elite in the AR to always criticize the pro-voter politician, because among the majority $1 - \alpha$ of unreceptive voters elite criticism reduces voter beliefs. In contrast, as Corollary 2 demonstrated, among the minority α of receptive voters, elite criticism increases voter beliefs. Thus, for a politician who puts a sufficiently large weight on receptive voters, elite criticism is beneficial. This force induces the politician to take the bad policy action and trigger criticism.

Implications and evidence. The key empirical prediction is that populist politicians, both because their type is bad and in order to invite elite criticism, choose harmful policies. Harmful policies are also a prediction in the Acemoglu, Egorov and Sonin (2013) model of populism, where populists signal their independence from the elite using policies that disproportionately harm the elite. The key difference is that our model does not make any assumption about the nature of the bad policy, and is thus consistent with harmful policies that *do not* disproportionately harm the elite.

Because of this difference, our model can help explain the puzzling fact that populists, despite their pro-people rhetoric, do not appear to be siding with "the people:" their policies seem to hurt the non-elite at least as much as they hurt the elite. Macro evidence on this comes from Funke, Schularick and Trebesch (2023), who show not only that populists reduce GDP per capita, but also that they fail to reduce inequality. Thus, populists seem to cause equal harm to the non-elite and the elite.

Populists also favor specific policies that seem to disproportionately harm the non-elite. Perhaps the most direct example is corruption. Populism's erosion of democratic institutions (Funke, Schularick and Trebesch, 2023) can enable

[19] The assumption that the politician's core supporters are the receptive voters seems realistic in that Republicans are the voters who believe in the elite conspiracy.

large-scale corruption, and indeed, populism is associated with a large increase in executive corruption (Zhang, 2024). In turn, stealing government funds plausibly harms those the most who rely on government services the most, i.e., the non-elite. A second example is tariffs, which are commonly used by populists (Funke, Schularick and Trebesch, 2023) including Donald Trump. Although tariffs do not necessarily hurt the non-elite, there are good reasons why they might: they raise import prices, and induce retaliatory tariffs targeted at the populist's non-elite supporters. Consistent with this logic, Fajgelbaum et al. (2019) find in a model-based evaluation that the tariffs of the first Trump administration, due to tariff retaliations, affected tradeable sector workers in heavily Republican countries the most negatively.[20] A third example is climate policy. The Biden Administration's climate bill, the Inflation Reduction Act, was opposed by Republican representatives and suspended by President Trump, despite its widely acknowledged benefits for blue-collar workers in Republican-leaning states (Friedman, Plumer and Stevens, 2025).

To our knowledge, this evidence is not explained by prior models. Although our model does not predict which bad policies get implemented, it is consistent with bad policies that harm the non-elite as much as they harm the elite, and in this sense can explain the evidence.[21] We conclude that the current wave of populism may generate substantial harmto both the elite and the non-elite.

IV. Conclusion

In this paper we built a new model of populism as a conspiracy theory. In our model, a politician can supply a false alternative reality claiming that members of the elite conspire to attack him because they disagree with his ideology. We show that, among voters receptive to it, this alternative reality can discredit the elite's truthful message about the politician. In turn, discrediting is beneficial for a "bad" politician because it enables him to remain in power. Through this logic, our model explains two previously unexplained facts about populism: the political use of conspiratorial narratives, and their association with reduced accountability.

A key prediction of the model is that conspiratorial propaganda inverts the effect of elite criticism among receptive voters, so that elite criticism increases these voters' support for the populist. The underlying intuition is that for receptive voters, the elite's message is primarily informative about the nature of reality, and elite criticism is more consistent with the alternative reality. This result explains Republicans' increased support for Trump after the indictments. The model also explains the absence of such an increase for Nixon after Watergate, through the logic that Nixon did not have sufficient ideological cleavage with

[20] Grossman and Helpman (2021) explain populism-induced tariffs in an identity-based model of trade, but in their model tariffs do not materially harm the non-elite.

[21] Neither does our model predict which of its two mechanisms—lack of accountability or the desire to trigger the elite—is responsible for specific policies. But we note that one channel for the second mechanism may be the populist's personnel policy. Selecting experts for key policy roles can invite elite praise and should therefore be avoided; and the selection of non-experts leads to harmful policies.

the elite to make an ideology-motivated attack plausible. Another prediction of the model is that beliefs in the alternative reality tend to strengthen in response to realized outcomes, because the alternative reality was chosen in anticipation of those outcomes. This result helps explain why beliefs in political conspiracy theories are widespread.

We then developed two applications of the model. In the first, we showed that alternative realities often endogenously feature conspiracies in order to better resist evidence. Thus, our model provides a formal explanation for the emergence of political conspiracy theories. An implication is that increasing the credibility of evidence can make the alternative reality conspiratorial, which in turn can create distrust in the elite beyond politics. This result helps explain Republicans' general distrust in science.

In our second application, we studied government policy under conspiratorial populism. We found that populists may set harmful policies that disproportionately harm the non-elite, both because—given reduced accountability—they can, and because doing so triggers elite criticism. This result helps explain why populism is associated with economic underperformance without meaningful reductions in inequality, as well as policy choices such as corruption, tariffs, and anti-environmentalism which may disproportionately harm the non-elite. We conclude that our theoretical results shed light on a number of key facts about populism.

Our model studied the supply side of populism. But much empirical research shows that populism is more likely to emerge following economic crises (Guriev and Papaioannou, 2022), suggesting that the demand side also plays a role. Building a psychologically realistic model of the demand side that informs this evidence is an important topic for future work.

In this paper, we used the framework of a false alternative reality to study populist ideology. The same framework may be used to study other ideologies as well. One possible example is nationalism. Aiming to deflect criticism or initiate collective action, political leaders may demonize the citizens of the other country, an alternative reality that captures some elements of nationalism. Modeling this alternative reality may lead to predictions about the emergence and persistence of conflict, based on the idea that nationalistic ideology leads to a misinterpretation of other countries' actions. More generally, formalizing other ideologies as strategic alternative realities is a potentially important avenue for future research.

REFERENCES

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2013. " A Political Theory of Populism ." *The Quarterly Journal of Economics*, 128(2): 771–805.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2015. " Radio and the Rise of The Nazis in Prewar Germany." *The Quarterly Journal of Economics*, 130(4): 1885–1939.

Agranov, Marina, Ran Eilat, and Konstantin Sonin. 2023. "Information Aggregation in Stratified Societies." National Bureau of Economic Research Working Paper 31510.

Aina, Chiara. 2023. "Tailored Stories." Working Paper, Harvard University.

Ajzenman, Nicolás, Tiago Cavalcanti, and Daniel Da Mata. 2023. "More than words: Leaders' speech and risky behavior during a pandemic." *American Economic Journal: Economic Policy*, 15(3): 351–371.

Allcott, Hunt, Levi Boxell, Jacob Conway, Matthew Gentzkow, Michael Thaler, and David Yang. 2020. "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic." *Journal of public economics*, 191: 104254.

Allen, Jonathan. 2023. "Awaiting possible indictment, Trump rallies in Waco and vows to 'destroy the deep state'." NBC News, `https://www.nbcnews.com/politics/awaiting-possible-indictment-trump-rallies-waco-rcna75684`.

Angelucci, Charles, and Andrea Prat. 2024. "Is journalistic truth dead? Measuring how informed voters are about political news." *American Economic Review*, 114(4): 887–925.

Ash, Elliott, Sharun Mukand, and Dani Rodrik. 2021. "Economic Interests, Worldviews, and Identities: Theory and Evidence on Ideational Politics." National Bureau of Economic Research Working Paper 29474.

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya. 2020. "Facts, alternative facts, and fact checking in times of post-truth politics." *Journal of Public Economics*, 182: 104123.

Bellodi, Luca, Massimo Morelli, Antonio Nicolò, and Paolo Roberti. 2023. "The shift to commitment politics and populism: Theory and evidence." *BAFFI CAREFIN Centre Research Paper*, , (204).

Bénabou, Roland. 2013. "Groupthink: Collective delusions in organizations and markets." *Review of economic studies*, 80(2): 429–462.

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018. "Narratives, imperatives, and moral reasoning." National Bureau of Economic Research.

Besley, Tim, and Torsten Persson. 2021. "The rise of identity politics." Working paper, London School of Economics and Stockholm School of Economics.

Besley, Timothy, and Andrea Prat. 2006. "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability." *American Economic Review*, 96(3): 720–736.

Blouin, Arthur, and Sharun W. Mukand. 2019. "Erasing Ethnicity? Propaganda, Nation Building, and Identity in Rwanda." *Journal of Political Economy*, 127(3): 1008–1062.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini. 2021. "Identity, Beliefs, and Political Conflict." *The Quarterly Journal of Economics*, 136(4): 2371–2411.

Brunnermeier, Markus K, and Jonathan A Parker. 2005. "Optimal expectations." *American Economic Review*, 95(4): 1092–1118.

Bénabou, Roland, and Jean Tirole. 2006. "Belief in a Just World and Redistributive Politics*." *The Quarterly Journal of Economics*, 121(2): 699–746.

Cheng, Haw, and Alice Hsiaw. 2022. "Distrust in experts and the origins of disagreement." *Journal of economic theory*, 200: 105401.

Corasaniti, Nick, and Trip Gabriel. 2023. "Trump Tells Supporters His Criminal Indictments Are About 'You'." The New York Times, `https://www.nytimes.com/2023/08/08/us/politics/trump-indictments-2024-campaign.html`.

Douglas, Karen M, Joseph E Uscinski, Robbie M Sutton, Aleksandra Cichocka, Turkay Nefes, Chee Siang Ang, and Farzin Deravi. 2019. "Understanding conspiracy theories." *Political Psychology*, 40: 3–35.

Egorov, Georgy, and Konstantin Sonin. 2024. "The Political Economics of Non-democracy." *Journal of Economic Literature*, 62(2): 594–636.

Eliaz, Kfir, and Ran Spiegler. 2020. "A Model of Competing Narratives." *American Economic Review*, 110(12): 3786–3816.

Eliaz, Kfir, Simone Galperti, and Ran Spiegler. 2022. "False Narratives and Political Mobilization."

Fajgelbaum, Pablo D, Pinelopi K Goldberg, Patrick J Kennedy, and Amit K Khandelwal. 2019. "The Return to Protectionism*." *The Quarterly Journal of Economics*, 135(1): 1–55.

FEC. 2024. "Campaign Finance Data." `https://www.fec.gov/data/browse-data/?tab=committees`.

Franklin, Charles. 2018. "Nixon, Watergate and Partisan Opinion." `https://medium.com/@PollsAndVotes/nixon-watergate-and-partisan-opinion-524c4314d530`.

Friedman, Lisa, Brad Plumer, and Harry Stevens. 2025. "Trump Is Freezing Money for Clean Energy. Red States Have the Most to Lose." The New York Times, `https://www.nytimes.com/2025/02/10/climate/trump-clean-energy-republican-states.html`.

Funke, Manuel, Moritz Schularick, and Christoph Trebesch. 2023. "Populist leaders and the economy." *American Economic Review*, 113(12): 3249–3288.

Galperti, Simone. 2019. "Persuasion: The Art of Changing Worldviews." *American Economic Review*, 109(3): 996–1031.

Glaeser, Edward L. 2005. "The Political Economy of Hatred." *The Quarterly Journal of Economics*, 120(1): 45–86.

Grossman, Gene M, and Elhanan Helpman. 2021. "Identity politics and trade policy." *The Review of Economic Studies*, 88(3): 1101–1126.

Guiso, Luigi, Herrera Helios, Massimo Morelli, and Tommaso Sonno. 2023. "Economic insecurity and the demand of populism in Europe." *Economica.*

Guriev, Sergei, and Daniel Treisman. 2020. "A theory of informational autocracy." *Journal of Public Economics*, 186: 104158.

Guriev, Sergei, and Daniel Treisman. 2022. *Spin Dictators: The Changing Face of Tyranny in the 21st Century.* Princeton University Press.

Guriev, Sergei, and Elias Papaioannou. 2022. "The Political Economy of Populism." *Journal of Economic Literature.*

Horovitz, David. 2020. "Victim of a left-wing coup? Why Netanyahu's conspiracy theory is foul and absurd." The Times of Israel, `https://www.timesofisrael.com/victim-of-a-left-wing-coup-why-netanyahus-conspiracy-theory-is-foul-and-absurd/`.

Hotez, P.J. 2023. *The Deadly Rise of Anti-science: A Scientist's Warning.* Johns Hopkins University Press.

hvg.hu. 2017. "A magyarok csaknem fele nem is hisz a Soros-tervben." hvg.hu, `https://hvg.hu/itthon/20171020_A_magyarok_kozel_fele_nem_is_hisz_a_Sorostervben`.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *American Economic Review*, 101(6): 2590–2615.

Koçak, Korhan. 2018. "Sequential updating: A behavioral model of belief change."

Kocsis, Eva. 2017. "Orban Viktor a Kossuth Radio '180 perc' cimu musoraban." [Radio broadcast transcript] Website of the Hungarian Government, `https://2015-2019.kormany.hu/hu/a-miniszterelnok/beszedek-publikaciok-interjuk/orban-viktor-a-kossuth-radio-180-perc-cimu-musoraban-20171006`.

Kuziemko, Ilyana, Nicolas Longuet Marx, and Suresh Naidu. 2022. "'Compensate the Losers?'Economy-Policy Preferences and Partisan Realignment in the US."

Levy, Raphaël. 2014. "Soothing politics." *Journal of Public Economics*, 120: 126–133.

Le Yaouanq, Yves. 2023. "A model of voting with motivated beliefs." *Journal of Economic Behavior & Organization*, 213: 394–408.

Mays, Jeffrey C. 2024. "Mayor's Public Defense Leans on Conspiracy Theories and Race." The New York Times, `https://www.nytimes.com/2024/09/26/nyregion/eric-adams-defense-conspiracy-theories-race.html`.

McFadden, Alyce, and Jeffery C. Mays. 2024. "69 Percent of New Yorkers Think Eric Adams Should Resign, Poll Shows." The New York Times, `https://www.nytimes.com/2024/10/04/nyregion/eric-adams-resign-poll.html`.

Mcmillan, John, and Pablo Zoido. 2004. "How to Subvert Democracy: Montesinos in Peru." *Journal of Economic Perspectives*, 18(4): 69–92.

Mudde, Cas. 2004. "The populist zeitgeist." *Government and opposition*, 39(4): 541–563.

Navot, Doron. 2022. "Corruption in Israel." In *The Palgrave International Handbook of Israel.* 1–14. Springer.

Norris, Pippa, and Ronald Inglehart. 2019. *Cultural backlash: Trump, Brexit, and authoritarian populism.* Cambridge University Press.

Nyhan, Brendan. 2020. "Facts and myths about misperceptions." *Journal of Economic Perspectives*, 34(3): 220–236.

Nyhan, Brendan. 2021. "Why the backfire effect does not explain the durability of political misperceptions." *Proceedings of the National Academy of Sciences*, 118(15): e1912440117.

Schwartzstein, Joshua, and Adi Sunderam. 2021. "Using Models to Persuade." *American Economic Review*, 111(1): 276–323.

Shabecoff, Philipe. 1974. "A Secondary Defense of Nixon." The New York Times, `https://www.nytimes.com/1974/07/16/archives/a-secondary-defense-of-nixon-ziegler-presents-thesis.html`.

SSRS. 2023. "CNN Poll: July 1-31, 2023." `https://www.documentcloud.org/documents/23895856-cnn-poll-on-biden-economy-and-elections`.

Sunstein, Cass R, and Adrian Vermeule. 2009. "Conspiracy theories: Causes and cures." *Journal of political philosophy*, 17(2): 202–227.

Swan, Jonathan, Ruth Igielnik, Shane Goldmacher, and Maggie Haberman. 2023. “How Trump Benefits From an Indictment Effect.” The New York Times, https://www.nytimes.com/2023/08/13/us/politics/trump-indictment-effect.html,.

Szeidl, Adam, and Ferenc Szucs. 2021. “Media Capture Through Favor Exchange.” *Econometrica*, 89(1): 281–310.

Uscinski, Joseph E., Karen Douglas, and Stephan Lewandowsky. 2017. “Climate Change Conspiracy Theories.”

Wallace, Jacob, Paul Goldsmith-Pinkham, and Jason L Schwartz. 2022. “Excess death rates for Republicans and Democrats during the COVID-19 pandemic.” National Bureau of Economic Research.

Wikipedia. 2024. “List of federal political scandals in the United States.” https://en.wikipedia.org/wiki/List_of_federal_political_scandals_in_the_United_States.

Yanagizawa-Drott, David. 2014. “ Propaganda and Conflict: Evidence from the Rwandan Genocide.” *The Quarterly Journal of Economics*, 129(4): 1947–1994.

YouGov. 2023. “CBS News Pol.” https://docs.cdn.yougov.com/3aamn30mjr/cbsnews_20230611_1.pdf.

Zhang, Dong. 2024. “Draining the Swamp? Populist leadership and corruption.” *Governance*, 37(4): 1141–1161.