The Political Economy of Alternative Realities*

Adam Szeidl
Central European University and CEPR

Ferenc Szucs
Stockholm University

November 28, 2024

Abstract

We build a new theory of populism in which a politician can persuade voters of a false alternative reality that serves to discredit the intellectual elite. In the alternative reality, elite members conspire to criticize the competence of a politician whose ideology they dislike. If believed, the alternative reality reduces accountability; inverts the effect of the elite's message, so that elite criticism helps the politician; and leads to strengthening misbeliefs. Alternative realities are often conspiracy theories which solve a collective action problem; resist evidence; and invite bad policies that trigger elite criticism. These results explain previously unexplained facts.

Keywords: populism, propaganda, misbeliefs, conspiracy theory, distrust

JEL codes: D03, D72, D82, D83

*Emails: szeidla@ceu.edu, ferenc.szucs@su.se. We thank Nageeb Ali, 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, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement number 724501 and Horizon Europe grant agreement number 101163526 for funding.

1 Introduction

The supporters of populist leaders often harbor salient misbeliefs about politics and society. In the Trump era, over 40% of Republicans believe that human activity does not contribute a great deal to climate change, and that the 2020 U.S. presidential election was not conducted fairly.[1] Misbeliefs persist despite being widely contradicted by experts in the news media. At the same time, populist leaders often avoid political accountability for acts that would normally be extremely damaging. Despite being convicted of a felony, Donald Trump won the 2024 presidential election. The presence of misbeliefs and reduced accountability under populism is not well understood.

We propose a new theory of populism which explains both of these facts. In our theory, populist leaders can use propaganda to persuade voters of a coherent system of misbeliefs—an *alternative reality*—that protects them from accountability. As we argue below, politically supplied alternative realities in the U.S., Hungary, and Israel share a common narrative: that a conspiracy of the intellectual elite attacks the competence of a politician for purely ideological reasons. Such an alternative reality can reduce accountability even in the presence of an independent media, because it discredits the criticism coming from that media.

We formalize these ideas using a model, which contributes to prior work on misinformation in politics—including Guriev and Treisman (2020) and Eliaz and Spiegler (2020)—by explicitly incorporating the false alternative reality in which the elite can conspire. Beliefs in this alternative reality generate strategic interaction between the (imagined) conspiring elite and the actors in the objective reality. Through this interaction, the model explains the joint emergence of misbeliefs and reduced accountability. It also predicts that beliefs in the alternative reality can *invert* the effect of the elite's message about the politician, so that elite criticism can increase voter support. And it implies that alternative realities often feature conspiracy theories and resist evidence. These and related predictions explain previously unexplained facts about populism.

In Section 2 we present our model: a principal-agent setting in which the incumbent politician and the intellectual elite attempt to influence voters. The politician has a type dimension, e.g.,

[1] See Moessner and Berg (2023) and Murray (2022). 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.

competence, along which she can be good or bad. Voters prefer a good politician but do not directly observe the type. Both the politician and the elite send messages to influence voter beliefs. First the politician chooses whether to send propaganda about the alternative reality. Then the elite, having observed the politician's choice of propaganda, and having received an informative signal about the politician's type, sends a message that reports on that signal.

Our key assumption is that a share α of voters are receptive to propaganda: the politician's message exogenously and counterfactually increases their prior belief in the alternative reality. We formalize the alternative reality by introducing the notion of *reality types.* We assume that the elite has an alternative reality (AR) type which does in fact conspire, and the politician also has an AR type which believes in the conspiracy. These types have zero objective probability, but receptive voters, if reached by propaganda, assign a non-negligible positive probability to them. Our notion of perfect Bayesian equilibrium requires that the AR types—though they only exist in the receptive voter's mind—act strategically to maximize their own payoffs, creating a coherent alternative reality which engages in strategic interaction with the voters.

The main difference between the reality and alternative reality types lies in the motives of the elite. In reality, the elite consists of many small actors who individually cannot influence voters and hence prefer to report about the politician's type truthfully. But in the alternative reality members of the elite can coordinate—effectively conspire—and thus the elite can send its message strategically to influence voters. It follows that if the AR elite sufficiently dislikes the politician—because they disagree about ideology—she will always report that politician bad to influence voters. Intuitively, in the alternative reality the "liberal 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 the criticism of the elite.

We analyze the model in Section 3. We characterize equilibrium under two main assumptions: that the elite dislikes the politician's ideology to a sufficient extent, and that propaganda can move voter priors to a sufficient extent. The structure of the equilibrium is the following. (i) In the objective reality the politician sends propaganda if and only if she is bad, and the elite always reports truthfully. (ii) In the alternative reality the politician always sends propaganda, and the elite always criticizes the politician. Intuitively, in reality the good politician has no reason to send propaganda as she expects praise from the elite. The bad politician, who expects criticism from the elite, wants to send propaganda if that succeeds in discrediting elite criticism. Discrediting criticism requires an alternative explanation for that criticism, which is the elite conspiracy of the alternative reality. That explanation is coherent because of our assumption that the elite sufficiently dislikes the politician's ideology that—if they could act collectively—they would want to criticise her. The narrative of the alternative reality then is that the conspiring elite always criticizes, leaving the politician no choice but to spread propaganda to counter those lies.

This equilibrium has a number of theoretical implications. First, by discrediting elite criticism, propaganda enables bad politicians to remain in power. This is a new mechanism through which populist propaganda can reduce political accountability.

Second, propaganda *inverts* the effect of the elite on the receptive voter. In the presence of propaganda, elite criticism will *increase* receptive voters' beliefs that the politician is good. To see the logic, consider the receptive voter who has observed propaganda. 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 increases his posterior of the alternative reality, and with it, his posterior that the politician is good. Observing praise has the opposite effect. This inversion result overturns standard intuitions about the impact of information in political economics.

Third, the politician's choice of when to send propaganda *amplifies* receptive voters' misbelief in the alternative reality. Intuitively, the outcomes of propaganda and elite criticism are more likely in the alternative reality than in the observed reality, and thus increase posterior beliefs in the alternative reality.[2] The receptive voter does not "undo" this on-average belief drift because he fails to account for the state-dependence of his prior. He ignores that his prior assigns positive probability to the alternative reality precisely when events tend to be consistent with that alternative reality. The resulting amplification implies that even propaganda that plants a small initial misbelief can have a large effect on society.

[2] This amplification parallels Schwartzstein and Sunderam's (2021) result that models more consistent with the data are more persuasive. Our contribution is showing that politically-supplied conspiracy beliefs are amplified.

These implications help explain previously unexplained facts. First, the implication that propaganda reduces accountability is consistent with evidence that democratic populism is associated with reduced accountability (Funke, Schularick and Trebesch 2023) and that propaganda is an underlying channel (Guriev and Treisman 2022). These facts seem unexplained by prior models. In leading models of populism, populist policies act as a positive signal, thus these models cannot easily explain reduced accountability (Acemoglu, Egorov and Sonin 2013, Bellodi, Morelli, Nicolò and Roberti 2023). In the leading theory of propaganda, Guriev and Treisman (2020), propaganda reduces accountability by painting the politician in a misleading positive light. But that mechanism only works in autocracies where the politician can censor or co-opt the elite: otherwise the elite's message would correct voter beliefs. Our model explains why propaganda reduces accountability in democracies, and also explains that reduced accountability is associated with misbeliefs.

Second, the implication on inversion explains a key unexplained fact in contemporary U.S. politics: that the four criminal indictments against Trump in 2023 were accompanied by an *increase* in his support among Republican voters (Swan, Igielnik, Goldmacher and Haberman 2023). This distrust in the legal system by 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, i.e., elite criticism. This prediction is in line with survey evidence that Republicans claimed to increase support for Trump as a result of the indictments. It is also in line with new evidence we present that scandals by Republican House candidates caused an increase in the donations they received from Trump-supporting Republicans. The model's mechanism for inversion, increased beliefs in the alternative reality, is also consistent with evidence: following the indictments, Republicans sharply increased their beliefs in the conspiracy theory that the 2020 election was stolen. And our model explains the contrast between Trump and Nixon, through the logic that only Trump was anti-elite. Nixon, representing the more educated party (Republicans around 1970), could not credibly argue that the intellectual elite conspired to remove him.

Third, although we do not have direct evidence on amplification, we argue that in the U.S., Hungary, and Israel, the elite conspiracy alternative reality is supplied precisely in situations in which it matches headline facts. This enables Bayesian updating to amplify misbeliefs.

In Section 4 we develop two applications. In each application, we incorporate novel features of the environment without changing the core assumptions concerning alternative realities. 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 alternative reality 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. As a result, the ability to conspire creates higher-powered incentives for the elite to lie, making them more powerful. This makes the conspiracy alternative reality more attractive to the politician, because it can explain away more credible critical evidence through the more powerful elite fabricating that evidence. We believe that this is the first formal explanation for the widespread nature of political conspiracy theories.

Our explanation predicts that alternate realities are often resistant to evidence. This follows because in response to more credible critical evidence, the politician can "upgrade" the alternative reality: 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 elite's messages may be driven by its interests even in domains where individual-level lying is not incentive compatible. This logic helps explain why misbeliefs under populism are wide-ranging, including Republicans' general distrust in science.

In our second application, we investigate the effect of alternative realities on the quality of government policy. We find that propaganda-spreading politicians will not adopt policies supported by the intellectual elite (e.g., vaccination), even if they know that non-adoption reduces welfare. The intuition follows from the inversion result. Since elite criticism increases the support of receptive voters, the politician chooses policies that invite elite criticism. This prediction highlights an important social cost of propaganda, and is consistent with new evidence we present that Republican governors were less likely than Democrats to introduce mask mandates or vaccinate publicly.

Our paper builds on overlapping literatures in political, behavioral, and information economics. Most directly, we build on work studying the supply of misinformation in politics. Early contributions include Glaeser (2005) on the supply of hatred and Besley and Prat (2006) on media capture. More recently, Guriev and Treisman (2020) model the supply of positive propaganda while Ash, Mukand and Rodrik (2021) model the supply of "worldview politics." A conceptual framework underlying much of the work on political misinformation is Bayesian persuasion (Kamenica and Gentzkow 2011).[3] We contribute to this work by modeling misinformation as a strategic alternative reality, an approach potentially portable to other settings; and with new implications, including reduced accountability in democracies, inversion, and the emergence of conspiracy theories.

We also build on theories of populism and identity politics (Acemoglu et al. 2013, Bonomi, Gennaioli and Tabellini 2021, Besley and Persson 2021, Agranov, Eilat and Sonin 2023, Bellodi et al. 2023). Our contribution is a new model of populism which explains the unexplained facts of misbeliefs and reduced accountability through the logic that populist propaganda discredits the elite. In this model, distrust in the elite is the consequence rather than the cause of populism.

Our modelling approach builds on theories of learning and interaction under model misspecification (Berk 1966, Jehiel 2005, Esponda and Pouzo 2016). Model misspecification has been applied in political economics to study groupthink (Bénabou 2013); persuasion (Bénabou, Falk and Tirole 2018, Galperti 2019, Schwartzstein and Sunderam 2021, Aina 2023); political narratives (Eliaz and Spiegler 2020, Eliaz, Galperti and Spiegler 2022); and political dynamics (Levy, Razin and Young 2022). We contribute to this work with a model-misspecification-based approach to model the elite conspiracy alternative reality 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, Uscinski, Sutton, Cichocka, Nefes, Ang and Deravi (2019). Our main contribution to this research is a formal model of the elite conspiracy alternative reality.

[3] Egorov and Sonin (2020) provide a useful review of Bayesian models of political persuasion.

	Trump	Orban	Netanyahu
A. Supplied narrative			
Politician attacked about	Criminality	Authoritarianism	Corruption
By conspirators	Deep state and media	Soros network	Judiciary and media
Ideological issue	Cultural values	Immigration	Palestinian conflict
B. Misbeliefs			
Supporters' beliefs	2020 Election stolen	Soros-plan	Corruption charges false

Table 1: Alternative reality narratives and voter misbeliefs

2 Model

2.1 Motivation

Our model is motivated by two observations. First, in several populist democracies, politicians supply alternative realities that share the following common narrative. The politician is the victim of a coordinated attack about his competence, by a group of conspirators who are members of the intellectual elite, because these conspirators disagree with the politician about an ideological issue. Three examples of this narrative, summarized in Panel A of Table 1, are as follows.

- In the United States, Trump claims to be the victim of a coordinated attack about his criminality, by the deep state and the media, 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, parents attending school board meetings"; and suggests that the goal of the conspiracy is to limit the influence of conservative values in America "they want to silence me because I will never let them silence you" (Corasaniti and Gabriel 2023).
- In Hungary, Orban claims to be the victim of a coordinated attack about his dismantling of checks and balances, by the Soros network—which includes Brussels and the media—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 to be the victim of a coordinated attack about his corruption by the judiciary and the media, because they find him too anti-Palestinian. As Horovitz (2020) explains, Netanyahu's core 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 (summarized in Panel B of the Table).

- In the US, the majority of Republicans believe that Biden did not win the 2020 election legitimately (Murray 2022). 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.

Although we lack direct evidence that these misbeliefs are supply-driven, there is clear overlap between the content of the supply and the content of the misbelief.[4] More broadly, there is much evidence that the political supply of misinformation changes beliefs (Yanagizawa-Drott 2014, Adena, Enikolopov, Petrova, Santarosa and Zhuravskaya 2015, Blouin and Mukand 2019, Barrera, Guriev, Henry and Zhuravskaya 2020, Ajzenman, Cavalcanti and Da Mata 2023) and that populists value

[4] A concern with the evidence on misbeliefs may be "expressive responding:" that survey measures of beliefs reflect political identity rather than true beliefs. But although incentives help correct beliefs about directly verifiable statements (Bullock, Gerber, Hill and Huber 2013), incentives have null effects on beliefs about statements only verifiable via expert opinion (Berinsky 2018), suggesting that the conspiracy beliefs we discuss here are genuine.

the supply sufficiently to capture media (Mcmillan and Zoido 2004, Szeidl and Szucs 2021). These observations motivate our model of the supply of alternative realities that shape voter beliefs.

2.2 Setup

We consider a principal-agent model in which two principals, the intellectual elite and the politician, attempt to influence the average beliefs of voters.[5] There are three classes of actors, the politician, the intellectual elite, and the voters, where we say classes of actors because both the elite and the voters consist of a unit mass of members. We think about the elite as the news media. Each elite member has an audience of voters which is representative of the population of voters but has measure zero. This ensures that individual elite members cannot influence average voter beliefs. Every voter has access to exactly one elite member, i.e., consumes exactly one news media.[6]

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. $\theta_c = 1$ 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. After observing her type, with probability $1 - \beta$ the politician has an opportunity to send propaganda p. We assume that $0 < \beta < 1$. Only the politician knows whether she has the opportunity to send propaganda. $p = 1$ means that the politician sends propaganda; $p = 0$ means that she does not, either because she did not have the opportunity or because she 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 $\hat{\theta}_c$ about the politician's type. This signal is correct with probability $0.5 < \pi < 1$. Every elite member receives the same signal; voters do not receive a signal. We think of π as relatively high. Each member of the elite j then sends a message $s_j \in \{0, 1\}$ about the signal to its zero measure of voters, where

[5] We depart from standard political economic theories which treat the voter as the principal and the politician as the agent, because our focus is to understand how the politician influences the voter.

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

Stage	0	1	2
Politician	θ_c	$\hat{p}$	$\hat{s}$
Media		$\hat{p},\ \hat{\theta}_c$	$\hat{s}$
Receptive voter		$\hat{p}$	$\hat{s}$
Unreceptive voter			$\hat{s}$

Table 2: Timing and allocation of information

$s_j = 1$ means 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: they observe both the elite's message and propaganda. The remaining share $1 - \alpha$ are unreceptive: they only observe the elite's message, but not propaganda. Intuitively, unreceptive voters cannot distinguish propaganda from "politics as usual." By assumption, each elite member j has an audience consisting of a share α receptive and a share $1 - \alpha$ unreceptive voters.

Trembles. 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.

Table 2 summarizes the timing and allocation of information in the model.

Alternative reality. To model alternative realities, we allow receptive voters to entertain two theories of the world, denoted by R (reality) and AR (alternative reality). These theories differ in the beliefs and motives of the principals. R describes the true reality while AR describes a counterfactual reality. Beliefs in the AR will be triggered by propaganda. To formalize these ideas, we introduce (i) types for the principals that represent their beliefs and motives in the R and the AR, and (ii) types for the receptive voters that represent the probability they assign to the AR.

Regarding the principals, we assume that the politician and all members of the elite have the same reality type $\theta_r \in \Theta_r = \{R, AR\}$. The true prior probability of $\theta_r = AR$ is zero. Each R principal believes that the other principals are R, and each AR principal believes that the other

Type	Values (probabilities)	Interpretation
A. Politician		
Competence (θ_c)	1 (q_c), 0 ($1 - q_c$)	1=Good
B. Elite		
Signal ($\hat{\theta}_c$)	θ_c (π), $1 - \theta_c$ ($1 - \pi$)	1=Probably good
C. Politician and Elite		
Reality (θ_r)	R (q_r), AR (q_{ar})	AR=Alternative reality
D. Receptive voter		
Mind (θ_m)	N (if $p = 0$), P (if $p = 1$)	P=Persuaded by propaganda

Table 3: Types and interpretations

principals are AR. Other than these beliefs about θ_r, the AR principals' priors are correct. We introduce the differences in motives between the R and AR principals below.

Turning to receptive voters, we assume that each such voter i has a mind type $\theta_{mi} \in \{N, P\}$ where N represents normal and P represents persuaded. A normal receptive voter thinks that the prior probability of $\theta_r = AR$ is zero. A persuaded receptive voter thinks that the prior probability of $\theta_r = AR$ is $q_{ar} > 0$. We let $q_r = 1 - q_{ar}$. Receptive voter i's initial mind type at the beginning of the game is $\theta^0_{mi} = N$. His eventual mind type remains N if he does not encounter propaganda, and switches to P if he does. A receptive voter forms his posterior, based on the messages his observes, from the prior encoded by his mind type. Since either all receptive voters encounter propaganda or none of them, their mind types are the same, and we will denote it by θ_m. Finally, unreceptive voters never observe propaganda and will never believe in the alternative reality. Thus, the model's type vector is $(\theta_c, \hat{\theta}_c, \theta_r, \theta_m) = \theta$. We summarize the types in Table 3.

Motives. We assume that voters' average posterior belief about θ_c determines the payoffs of both principals. We do this for presentational purposes only, to avoid specifying voters' preferences and actions in the main text. In Appendix A.2 we provide microfoundations based on the idea that voters' beliefs govern the probability that the politician stays in power, which is the true determinant of payoffs. We define voters' average posterior belief about θ_c as

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

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 fist 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_{mi} (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.[7]

Having defined voter beliefs, we can define the preferences of the elite. In both the R and the AR, each elite member j has preferences

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

Here $\theta_c - \kappa$ reflects the elite's valuation of the incumbent politician. The elite likes competence θ_c but dislikes the incumbent by κ. Intuitively, κ measures the degree of 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.

The key difference between the R and the AR elite is that members of the R elite cannot, but members of the AR elite can coordinate. Formally, each R elite member sends her message independently, but the AR elite acts as a single decision maker which chooses an identical message for all of its members to maximize the sum of their utilities.[8] These assumptions imply that (i) members of the R elite, because they influence a zero measure of voters and do not impact the average belief, always send a truthful message; while (ii) members of the AR elite, because they internalize the effect they have on each other and can impact the average belief, send a message to

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

[8] In our microfoundation in Appendix A.2, sending an identical message will be the optimal strategy for a coordinating elite.

influence voters. In both cases, members of the elite send the same message which we denote by 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 when reality is AR, represents the collective interests of the AR elite. The second term, active when reality is R, represents that each elite member acts to tell the truth. Note that because $\hat{s}_j = \hat{s}$ for all elite members j, we can represent the second argument of $\bar{\mu}$ with $\hat{s}$.

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, independently of her type, 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 her type. The second term captures the cost f of sending propaganda. This cost represents both the material cost of designing and disseminating the alternative reality and the opportunity cost of having less time for policy making.

Timing. In summary, the timing of events is organized into the following stages.

0. The politician's type is realized and observed by the politician.

1. With probability β the politician cannot send propaganda; with probability $1 - \beta$ she decides on propaganda $p \in \{0, 1\}$. Propaganda is subject to trembles. The elite observes the realized propaganda $\hat{p}$ and receives a signal 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' mind type changes to $\theta_m = P$. Voters form posterior beliefs.

2.3 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 in both the objective and the alternative reality 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 who may be persuaded by propaganda. We assume that in stage 2, the posterior of each receptive voter with mind type $\theta_m = N, P$ is computed from the prior associated with that mind type. Thus, if the voter is persuaded by propaganda, his posterior is computed from the prior which assigns probability $q_{ar} > 0$ to the AR. 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 Appendix.

Equilibrium selection. Our game has four players and in total 11 player types.[9] Given this complexity we expect multiple equilibria, and introduce the following criteria for equilibrium selection. First, we focus on equilibria which are *politician-pure*: in which all politician types use pure strategies. Second, among these equilibria, we focus on *politician-optimal* equilibria, which maximize the ex ante expected utility of the incumbent R politician. We refer to equilibria satisfying these conditions as *PPO* equilibria.

2.4 Discussion of model assumptions

We already discussed our assumptions about the supply of and the form of the alternative reality in Section 2.1. Here we discuss some more specific modeling choices and interpretations.

Modeling alternative realities. Our approach is to take seriously voters' misbeliefs and explicitly model the false alternative reality they believe in. A key feature of the alternative reality we study

[9] Specifically, $\theta_c \in \{0,1\}$ and $\theta_r \in \{R, AR\}$ for the politician, $\hat{\theta}_c \in \{0,1\}$ and $\theta_r \in \{R, AR\}$ for the elite, $\theta_m \in \{N, P\}$ for the receptive voter, and the unreceptive voter.

is that it contains optimizing agents. These agents impose constraints on real-world outcomes that parallel the out-of equilibrium constraints of perfect Bayesian equilibrium. Perfection requires that agents, even at information sets never reached, must behave optimally; whereas we require that agents, even if they are imaginary, must behave optimally. This approach of explicitly modeling a strategic alternative reality—also used by Bénabou (2013) in a different setting—may be portable to study other systems of misbeliefs in economics.

Elite conspiracy. The results of our baseline model would follow in a framework that does not feature a conspiracy, in which the key difference between the R and the AR elite is that the latter has a lower lying cost. We chose to model the conspiracy both because it is realistic (Douglas et al. 2019) and because it highlights that allowing for coordination fundamentally alters the equilibrium. But we endogenize the conspiracy in Section 4.1 by showing that when the politician can choose between a lying cost and a conspiracy narrative, she will often prefer the latter. Intuitively, by solving their collective action problem, the conspiracy makes the AR elite more powerful. This leads to a formal explanation for the emergence of political conspiracy theories.

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. Intuitively, unreceptive voters do not follow politics closely and thus cannot distinguish propaganda from "politics as usual." For example, in April 2024, 42% of Americans claimed that they were not following election news about the 2024 election closely, and 57% claimed that they mostly got political news because they happened to come across it (Eddy 2024). Given their limited news consumption, these voters may not put together the conspiracy narrative of the politician, and thus remain protected from both its manipulative and information effects. In the conspiracy narratives of Section 2.1, unreceptive voters are the "gullible masses" who do not see through the conspiracy and are thus persuaded by the elite's lies. In the model, these voters—since they are not influenced by propaganda—follow the elite's message, and thus can be mislead by it, providing the incentives for the AR elite to criticize.[10]

[10] Assuming that the voters who are not manipulated by propaganda do not observe propaganda at all is for tractability. The key is that they do not fully infer from propaganda that the politician is bad. As a result, elite criticism continues to reduce their valuation of the politician, incentivizing the conspiring elite to criticize.

Although in the main text we assume that $\alpha < 0.5$, i.e., receptive voters are a minority, in Appendix A.4 we show that our main results hold for $\alpha > 0.5$ as well, albeit may require mixed strategies. Further, 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. At a high level, this assumption is supported by the evidence in Section 2.1 about the impact of misinformation. However, the completely exogenous nature of the supply-induced belief change is somewhat unsatisfactory, as one expects beliefs to be also shaped by demand. To address this issue, in Appendix A.6 we develop a simple model of the demand for misbeliefs based on the idea of motivated beliefs. This model provides microfoundations for the reduced-form framework presented here and yields novel comparative statics we discuss below.

AR type for the politician. Our model has an AR type not only for the elite but also for the politician, and the AR politician believes that $\theta_r = AR$. This is simply a coherence assumption for the alternative reality, which states that if the true state of the world is AR then both principals know this. As we explain below, this assumption is necessary for propaganda to work, because it allows the voter to believe that even good politicians (in the AR) send propaganda.

3 Results

3.1 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 her 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 $\theta_c - \kappa < 0$ always, it implies that the elite—if it can influence voters—wants to minimize voters' average belief that the politician is good. This assumption will be used to ensure the incentive compatibility of the AR elite's equilibrium strategy.

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 equilibrium posterior belief, after observing propaganda and criticism, that the politician is good, in the limit when π converges to one. Assumption 2 then 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.

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

1. *In the reality (R):*
   - *The elite reports the common type truthfully,*
   - *The politician sends propaganda if she can and is bad.*
2. *In the alternative reality (AR):*
   - *The elite always reports that the politician is bad,*
   - *The politician sends propaganda if she can.*

All proofs are in the 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 she 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.

We now turn to more fully flesh out the workings of the equilibrium. We first derive voter beliefs and then explain how these ensure the incentive compatibility of the principals' strategies.

Voter beliefs. We derive voters' posterior beliefs in the proposed equilibrium. The receptive voter's posterior, *absent propaganda* ($\hat{p} = 0$), can be written as a function of the elite's message $\hat{s}$ as

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

On the left-hand side, note the receptive voter's mind type: because $\hat{p} = 0$, he is normal ($\theta_m = N$) 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 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.[11]

The receptive voter's posterior, *in the presence of propaganda*, can be written as a function of

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

the elite's message $\hat{s}$ as

$$\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s}, \theta_m = P) = (1 - \hat{s})\frac{q_{ar}q_c}{q_{ar} + q_r\pi(1 - q_c)}. \tag{5}$$

On the left-hand side, note that because $\hat{p} = 1$, the receptive voter becomes persuaded ($\theta_m = P$) and assigns prior $q_{ar} > 0$ to the alternative reality. To understand the right-hand side, consider first the event when the elite sends criticism ($\hat{s} = 0$) so that the $1 - \hat{s}$ term is one. Beliefs are given by the voter's subjective probability that the politician is good conditional on propaganda and criticism. The numerator measures the joint probability that (i) the politician is good (q_c), (ii) propaganda, which because the politician is good only happens in the AR (q_{ar}), and (iii) criticism, which is automatic in the AR.[12] The denominator measures the probability of propaganda and criticism. In the AR (q_{ar}) the politician always sends propaganda (if she 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. Consider next the event when the elite sends praise ($\hat{s} = 1$). The expression is equal zero: in the AR the elite never sends praise, and in the R only the bad politician sends propaganda.

Finally, the beliefs of the unreceptive voter are similar to those of the receptive voter absent propaganda (4), 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.

Implications of voter beliefs. A key lesson from these belief expressions is that for π high, propaganda deflects elite criticism: receptive voters' beliefs after criticism are higher with than without propaganda. Indeed, taking π to one, (5) implies that beliefs after criticism *with propaganda* converge to

$$\hat{q}_c = \frac{q_{ar}q_c}{q_{ar} + q_r(1 - q_c)} \tag{6}$$

which is bounded away from zero, while, as noted after (4), beliefs after criticism *without propaganda* converge to zero. Intuitively, after propaganda the voter updates from criticism about the politician only partially, because when reality is AR, he expects criticism even when the politician is good.

[12] 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.

Note that the $\hat{q}_c$ defined in (6)—posterior beliefs after propaganda and criticism—is the term in Assumption 2. This term is a natural measure of the extent to which propaganda deflects criticism.

A second lesson from the belief expressions is that propaganda not only alters the effect of elite criticism, it also alters the effect of elite praise, so that among receptive voters it ends up *inverting* the effect of the elite's message: elite criticism becomes good news. This follows directly from (5): with elite criticism ($\hat{s} = 0$) beliefs are positive, while with elite praise ($\hat{s} = 1$) beliefs are zero. Intuitively, because in the AR the elite always criticizes, whereas in the R (since $\pi < 1$) it only sometimes criticizes, observing criticism increases the receptive voter's belief in the AR, and with it, his belief that the politician is good. Conversely, observing praise reduces the receptive voter's belief that reality is AR, and with it, his belief that the politician is good.

Incentive compatibility of the elite. Having characterized beliefs, we turn to explain why the principals are willing to follow the proposed equilibrium. We begin with the elite. The behavior of the R elite is straightforward: because its members are atomistic and cannot influence voter beliefs, they prefer to report truthfully. For the AR elite, which acts as a single actor and wants to lower voter beliefs, sending criticism is incentive compatible if

$$(1-\alpha)\left[\frac{\pi q_c}{\pi q_c + (1-\pi)(1-q_c)} - \frac{(1-\pi)q_c}{(1-\pi)q_c + \pi(1-q_c)}\right] > \alpha \frac{q_{ar} q_c}{q_{ar} + q_r \pi (1-q_c)}. \tag{7}$$

The left-hand side is the AR 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 (4), with the difference that the denominators do not have the β factors. The right-hand side is the AR elite's loss from criticism: the increased beliefs of the α receptive voters who "see through" the conspiracy. This loss is computed by differencing (5) between $\hat{s} = 1$ and $\hat{s} = 0$. This loss is a reflection of the inversion effect: in the presence of propaganda, elite criticism actually helps the politician among receptive voters.

If π is large, then the left hand side of (7) is close to $1-\alpha$ while the right-hand side is close to $\alpha\hat{q}_c$. Therefore, if $\alpha < 0.5$, which is 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 AR elite; and if there are enough of them, then their impact dominates the inversion effect on receptive voters and incentivizes the AR elite to criticize.

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

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

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 (4) to the expected value of beliefs in the presence of propaganda (5). 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 AR politician's incentive compatibility constraint. Both the good and the bad AR politician types believe that the elite is AR and always sends criticism. This means that their incentive compatibility constraint is implied by (8) 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 AR politicians'choice of propaganda is key for the updating of the voter, who, if propaganda is to be effective, should not be able to infer from observing it that the politician is bad. That holds here, because in the AR, even the good politician sends propaganda. This logic underlies equation (5) and prevents the full revelation of the bad R politician's type.

The arguments so far establish that the proposed profile is an equilibrium. We show that it is the unique PPO equilibrium in Appendix A.4, exploiting that PPO requires the politician to use a pure strategy, by checking each possible pure strategy of all politician types.

Relaxing the constraint on receptive voters. Proposition 1 focuses on the case when only a share $\alpha < 0.5$ of voters are receptive. However, as we show Appendix A.4, the model has a unique

PPO equilibrium which features propaganda even when $\alpha > 0.5$. This profile is identical to the equilibrium of Proposition 1 in the behavior of all politician types. But for α high, the behavior of the AR elite is more complex: they now mix between criticizing and praising the politician. The logic follows from the inversion result, combined with the fact that for α high the beliefs of receptive voters dominate the incentives of the elite. If the AR elite were to always criticize, then—by inversion—receptive voters would interpret praise as proof that the politician is bad. Since receptive voters dominate, the AR elite could profitably deviate to praise. Conversely, if the AR elite were to always praise, then receptive voters would interpret criticism as proof that reality is R and conclude that the politician is likely bad. Thus, the AR elite could profitably deviate to criticize. It follows that there is no pure strategy equilibrium. In the Appendix we show that the unique PPO equilibrium is the same as that of Proposition 1 except that it involves mixing by the AR elite. The voter beliefs implied by mixing make the AR elite indifferent.

The core intuition here 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 see through their lies? They should instead praise the politician and thus disprove the conspiracy theory. 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, so that any elite message is consistent with the conspiracy theory. These complex narratives emerge as the conspiracy theory becomes widespread. Importantly, our main qualitative predictions, including that propaganda deflects criticism and generates an inversion effect, continue to hold in this more complex equilibrium.

In a way, the mixed equilibrium resembles the Russian “firehose of falsehood” propaganda technique characterized by the simultaneous broadcasting of contradictory messages (Paul and Matthews 2016). Here the contradictory messages are that the elite is truthful and that the elite is lying. Of course, 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.

3.2 Theoretical implications of equilibrium

We develop several theoretical implications of the model's equilibrium. These implications shed light on the workings of the model and lead to new testable predictions.

Deflection. An immediate implication of the result that propaganda is used in equilibrium is that propaganda succeeds in deflecting elite criticism.[13]

Corollary 1. *Suppose that Assumptions 1 and 2 hold, $\pi > \bar{\pi}$, and $\alpha < 0.5$. In the PPO equilibrium, the bad politician's choice to send propaganda improves voter beliefs:*

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

Note that $E[.]$ represents objectively correct expectations. Thus, propaganda enables bad politicians to remain in power, and by doing so, reduces accountability. Importantly, this result applies in a democratic context, in the presence of uncensored independent media that provides reliable information on the politician's type. It follows that our model can be interpreted as a new theory of democratic populism, in which ideological disagreement with the elite can be used to reduce accountability. This new theory leads to new implications, as we now describe.

Inversion. Perhaps the most surprising new implication is that among receptive voters, propaganda *inverts* the effect of the elite's message. We now state this point formally.

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: $\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 1) < \mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0)$.*

As noted above, inversion is a consequence of equation (5). It follows through two steps. First, because in the AR (but not in the R) the elite always criticizes, elite criticism increases beliefs in the AR. Second, because in the AR (but not in the R) propaganda may also come from a good politician, the increased belief in the AR increases support for the politician. As this logic makes it clear, $\pi < 1$ is necessary for inversion, because it relies on the elite criticizing more often in the AR (with probability 1) than in the R (with probability π). We also note that inversion can be

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

broken down into two predictions: that the receptive voter's beliefs after propaganda (i) increase in response to elite criticism, and (ii) decrease in response to elite praise. Note that (i) cannot hold without (ii): since the receptive voter, given his subjective beliefs, updates as a Bayesian, his prior must be a convex combination of his posteriors.

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 she is anti-elite. Conversely, they may infer from elite praise not that the politician is incompetent, but that she is pro-elite. Although we find this alternative mechanism empirically plausible (and discuss it further in Section 3.3), to keep our model simple we did not incorporate the additional type dimension necessary to model it. Both mechanisms seem to predict that elite criticism increases receptive voters' beliefs in the AR and support for the politician.

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*, i.e., not conditional on the elite's message. The key result 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 experiencing 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)

$$\hat{q}_{ar} = \frac{q_{ar}}{1 - q_r q_c} > q_{ar}. \tag{9}$$

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 eliminates that state from the voter's mind, increasing his beliefs in the AR (for π large) by a factor $1/(1-q_r q_c)$. Intuitively, the voter neglects the correlation between the state of the world and his prior, i.e., that his prior is large precisely when the likely outcomes are propaganda and criticism. The reason correlation neglect amplifies AR beliefs is that these likely outcomes are more consistent with the AR than with the R.

The key lesson here is 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

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

Here q_{ar} is the initial false belief, while $1/(1 - q_r q_c)$, as discussed above, is the amplification.[14]

The mechanisms identified here are related to mechanisms studied in recent work on misbeliefs. Our timing assumption that first propaganda changes the prior and then the voter updates from the new prior 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. The logic of amplification is related to the effect identified in the research on persuasion with models that models more consistent with the data are more persuasive (Schwartzstein and Sunderam 2021,

[14] As the formula shows, amplification increases in q_c. We discuss this point in Corollary 5 below.

Aina 2023). Our formal approach is different from that of these papers because the persuaded voter simultaneously accounts for both realities and reasons about the incentives of the sender. Nevertheless, the main contribution of Corollary 3 is the applied lesson 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 $1 - \pi < \kappa < \pi$, 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 she 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. Motivated by equations (9) and (10) 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 after 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 (9). Intuitively, when the politician is expected to be good (q_c high), propaganda and criticism are unlikely in the R but likely in the AR. Thus, the AR is a relatively better explanation for that message profile and hence will have a higher posterior. In economic terms, 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 (10). There are two forces. First, as just noted, with q_c higher, propaganda induces a larger belief amplification towards the AR. Second, belief in the AR is better for the politician, because in that state she 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 Appendix A.6 we develop a microfounded model of the demand side, which is based on the idea of motivated beliefs in political economics (Brunnermeier and Parker 2005, Levy 2014). 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 (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 4 in the Appendix that our equilibrium is robust to incorporating the demand for misbeliefs, but features an endogenously chosen $q_{ar} > 0$.[15] Moreover, we establish new comparative static results: that q_{ar} is increasing in q_c and in the extent to which the voter likes the incumbent politician (a parameter λ). The intuition is that both q_c and λ increase the

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

probability that the incumbent is reelected, and thus increase the incentive for the voter to believe that the incumbent is competent, which is made possible by believing in the AR. These results also imply that in the equilibrium with the endogenous q_{ar}, the predictions of Corollaries 2, 3, and 4 are unchanged; while the comparative statics of Corollary 5 are strengthened because a higher q_c also increases q_{ar} through the demand side, acting to further amplify the belief in the AR.

3.3 Evidence

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

Propaganda lowers accountability in democracies. The result of Corollary 1, that propaganda reduces accountability, is consistent with evidence that (i) in democracies, populism is associated with reduced accountability (Funke et al. 2023) and (ii) even in democracies and hybrid regimes, propaganda is a channel to reduce accountability (Guriev and Treisman 2022). To our knowledge, existing theories do not explain these facts. In the leading models of populism, including Acemoglu et al. (2013) and Bellodi et al. (2023), populist policies are a positive signal about the politician, thus these models cannot easily explain reduced accountability. The leading theory of propaganda, Guriev and Treisman (2020), explains reduced accountability through the logic that propaganda paints the politician in a misleading positive light. However, that mechanism is only operational in autocracies, because it requires that the politician censors or co-opts the intellectual elite: otherwise elite criticism would correct beliefs and undo the effects of propaganda.

Our model features a mechanism—discrediting elite criticism—which is operational even in democracies and hybrid regimes where part of the media is free. Thus, our model can explain the above facts. In particular, our model can explain why propaganda appears to lower accountability in our example settings of the U.S., Hungary, and Israel (Ott and Dickinson 2020, Sükösd 2022, Rogenhofer and Panievsky 2020). Moreover, our model predicts that this reduced accountability should be accompanied by misbeliefs, consistent with the motivating evidence of Section 2.1.

Propaganda inverts the elite's effect on receptive voters. A key fact in contemporary U.S. politics is that during 2023, the growing body of critical evidence against Donald Trump, including four criminal indictments, was accompanied by an *increase* in popular support for Trump among

	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

Table 4: Impact of indictment on Trump's support by Republicans intending to vote in primary

Republican voters (Swan et al. 2023). Since the indictments were produced by the U.S. legal system, the apparent distrust in them among supporters of the presumptive party of law and order is puzzling. The increase in support for Trump 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 these facts.

Our model can explain these facts through its inversion and comparative statics predictions. We first describe how inversion explains the effects of the Trump indictments and present evidence in support of this explanation; and then describe how the comparative statics explain why we do not observe the same patterns for Nixon and Adams. The explanation for Trump builds on our argument in Section 2.1 that Trump disseminated the elite conspiracy alternative reality. Given this, and assuming Republicans correspond to the model's receptive voters, Corollary 2 explains the evidence by predicting a causal effect: that the indictments—elite criticism—*causally* increased Republican support for Trump.

We now present two pieces of evidence that support this causality. First, in Table 4 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 is 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 hypothetical behavior. For evidence on actual behavior, we turn to the impact of scandals on campaign contributions. We take Wikipedia's list of political scandals of Republican House candidates during 2017-2022, 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 agree, thus they correspond to θ_c. Moreover, since scandals are disseminated by the news media, they correspond to the model's notion of elite criticism. We combine these data with donation data from the Federal Election Commission.[16]

We estimate difference-in-differences regressions of the effect of a scandal on donations that come from Trump supporters and 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 5 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 scandals, plausibly diffused by the news media, seemed to generate an increase in political support among Trump-supporter voters. Since these voters are most likely to be receptive to the alternative reality, the evidence supports the causal link between elite criticism and political views predicted by our model.

A possible alternative explanation for these results is that the scandal increased the competitiveness of the election, and competitiveness made Republicans donate more. Two pieces of evidence speak against this explanation. First, as Table 5 shows, the effect is concentrated among Trump-supporter Republicans, and it is not clear why they should care more about the election outcome. Second, in Appendix A.9 we show that when the election of Republican candidates becomes more competitive because of redistricting, there is no analogous impact on donations. Thus, the effect

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

	Trump donors	Trump donors	Other donors
	Share	Amount (1000 dollars)	
Scandal effect	0.075***	20.33**	-9.80
	(0.009)	(9.88)	(16.59)
Representative and quarter f.e.	yes	yes	yes
Control mean	0.065	16.12	119.0
Observations	3,397	4,387	4,387

Note: Observations are representative-quarter cells. The treatment group is observations of treated representatives in a one-year window around the scandal; the control group is observations of non-treated representatives in the 2017-2022 period. Column 1 is restricted to observations with non-zero total donations. The dependent variable in column 1 is the share of donations from Trump-supporters; in columns 2 and 3 the total volume of donations from Trump-supporters and from other Republican donors, respectively. Standard errors clustered by state in parentheses.

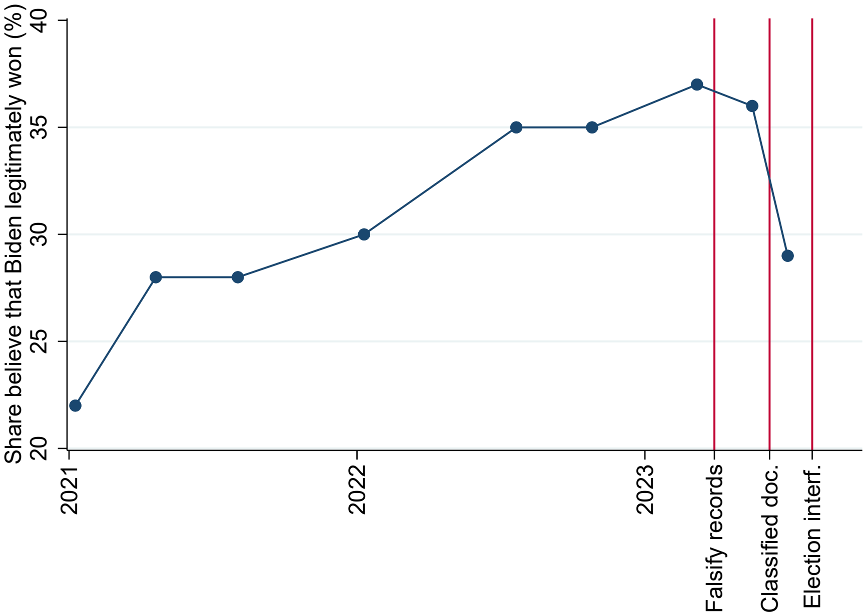
Table 5: Impact of scandals on contributions from Trump-supporter and other donors

we document seems to be driven by elite criticism rather than increased competition.

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 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 mechanism of 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, and journalists interpreted the relationship as causal (Agiesta and Edwards-Levy 2023).

Inversion predicts not only an increase in receptive voters' support in response to elite criticism, but also a decrease in their support in response to elite praise. Since elite praise in the presence

Figure 1: Share of Republican-leaning voters who believe that Biden legitimately won in 2020



Source: Nine CNN opinion polls conducted by SSRS between January 2021 and August 2023 (SSRS 2023a, 2023b, 2023c). 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.

of propaganda is rare, we could not find systematic evidence for this prediction. But one piece of anecdotal evidence comes from Governor Ron DeSantis' presidential campaign. In 2022, DeSantis was considered a real contender for the Republican presidential candidacy (Flegenheimer 2022). But then he received positive reception, at least in comparison to Trump, from academics and the news media (Gift 2023, Zelizer 2022), and even George Soros expressed preference for a scenario in which DeSantis wins the Republican nomination (Soros 2023). Our model suggests that this "relative praise" from the intellectual elite should have led to reduced support from Republican voters. DeSantis indeed experienced reduced support and dropped out of the campaign in early 2024. Importantly, elite endorsements and conspiracy beliefs may have played a role: in this period, conspiracy theories that "Ron DeSoros" was "a tool of the Deep State" became widespread in far-right circles (Thompson 2023).[17]

[17] In this example the plausible mechanism through which elite praise hurt DeSantis was not beliefs about his

Why did receptive voters respond differently to the elite criticism of Trump than they did to the elite criticism of Nixon or Adams? The answer suggested by the model is propaganda. Corollary 2 only predicts inversion when the politician sends propaganda about the elite conspiracy. We are not aware of evidence that either Nixon or Adams spread such propaganda. Our model also explains why they did not do so. Corollary 4 says that propaganda emerges in equilibrium only 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)—making it implausible that the intellectual elite conspired 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. Although corrective information can sometimes be interpreted as such criticism, often that is not the natural interpretation. For example, corrective information about the false claim that "Donald Trump Sent His Own Plane to Transport 200 Stranded Marines" is not elite criticism of Trump, since Trump could well be an excellent president despite not taking the claimed action.[18] In fact, our model suggests a comparative static in the backfiring effect: that corrective information backfires when the salient interpretation of that information is elite criticism.

Politically supplied misbeliefs are amplified by outcomes. Corollary 3 predicts that outcomes amplify beliefs in the alternative reality. We do not have direct evidence for amplification. But we note that its key ingredient—that outcomes should be relatively more likely in the alternative reality—is consistent with the conditions in which the elite conspiracy narrative was supplied in the contexts of Section 2.1. In the US, the deep state conspiracy was supplied around the time

competence but beliefs about his anti-elite status. As we noted after Corollary 2, that is an alternative mechanism for the inverted effect of elite praise.

[18] This is one of the claims corrected in the study by Clayton, Blair, Busam, Forstner, Glance, Green, Kawata, Kovvuri, Martin, Morgan et al. (2020).

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. More broadly, these examples suggest that by constructing an alternative reality that matches the evidence, the politician can take advantage of Bayesian updating to do the heavy lifting in creating misbeliefs. This logic may help explain why misbeliefs in alternative realities are so widespread.

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 (Acemoglu et al. 2013, Engler and Weisstanner 2021, Stoetzer, Giesecke and Klüver 2023). The main novelty relative to prior work is that our model emphasizes disagreement with the intellectual, rather than the political or business elite. As we discussed in Section 2.1, the cleavage with the intellectual elite is indeed a key feature of the populist narratives in the U.S., Hungary, and Israel. Part (ii), as we noted above, helps explain the presence of inversion with Trump and its absence with Nixon and Adams, through the logic that the latter politicians represented pro-elite parties.

4 Applications

We turn to develop two applications of our model: endogenizing the nature of the alternative reality, and studying the impact of propaganda 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.

4.1 Endogenous alternative reality

In our model, the AR features a conspiracy only by assumption. Moreover, for our qualitative results, this assumption is not strictly necessary: we could obtain our main 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. In this application we argue that conspiracy theories are a natural implication of our framework, justifying our modeling approach and helping to explain why real-world alternative realities often feature conspiracies.

Our basic insight 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 reducing 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, the elite has a lower lying cost but does not have the ability to coordinate; in the second, the elite also has the ability to coordinate. We also 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.

We entertain two types of 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.

If the receptive voter receives lying cost (conspiracy) propaganda, his prior puts q_{ar} weight on the lying cost (conspiracy) AR, $1 - q_{ar}$ weight on R, and zero weight on the other possible AR. The politician in the R and in both types of the AR can send either type of propaganda. This is the natural generalization of our basic model to the setting with multiple alternative realities. We will denote the lying-cost AR by AR1 and the conspiracy AR by AR2.

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

$$U_{ej} = (\theta_c - \kappa) \cdot \bar{\mu} - \chi_f \cdot 1_{\{s_j \neq \hat{\theta}_c\}} - \chi_h \cdot 1_{\{\theta_r = R\}} 1_{\{s_j \neq \hat{\theta}_c\}} - \chi_o \cdot 1_{\{\theta_r \neq AR2\}} 1_{organize_j}. \tag{11}$$

The first term involves the average voter belief

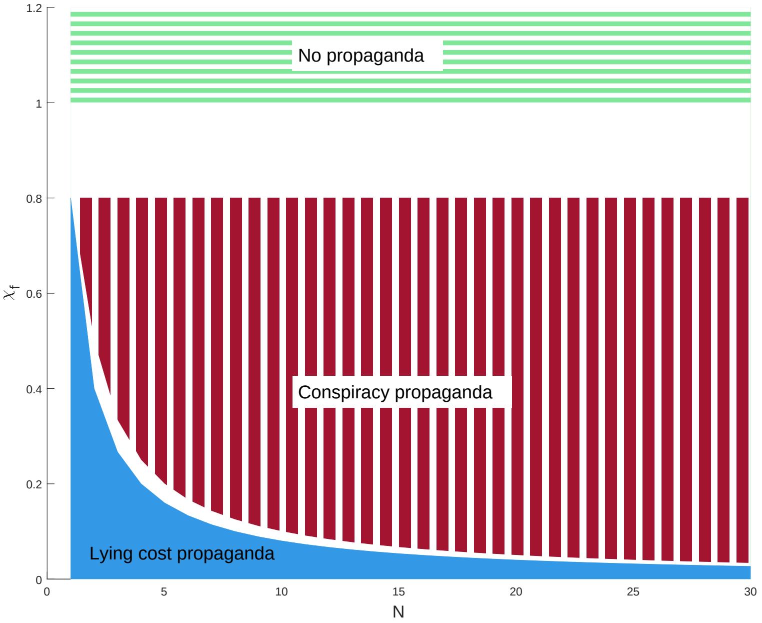
$$\bar{\mu} = \frac{\sum_{j=1}^{N} \bar{\mu}_j(\hat{p}, \hat{s}_j)}{N}$$

where $\bar{\mu}_j(\hat{p}, \hat{s}_j)$ is the average belief among the mass of voters influenced by elite member j. We assume $N > 1$. The cost terms in (11) 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.

Proposition 2. *Under Assumptions 1-2, if $\alpha < 0.5$ then, in the following ranges for χ_f, for π large enough, in the unique PPO equilibrium*

1. *If $\chi_f < (1 - 2\alpha)/N$, then the bad R politician sends lying cost propaganda;*
2. *If $1/N < \chi_f < (1 - 2\alpha)$, then the bad R politician sends conspiracy propaganda;*

Figure 2: Endogenous AR



3. If $1 < \chi_f$*, then no politician sends propaganda.*

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.

As illustrated by the white areas between the three regions in the Figure, the Proposition does not cover the full range of possible χ_f values. In the intermediate ranges mixed equilibria are possible. Since these mixed equilibria did not seem central to our message, we chose to focus on the ranges where the equilibrium is in pure strategies.[19]

Implications and evidence. The Proposition has three main implications. First, it predicts that misbeliefs should often feature conspiracy theories. Conspiracy theories indeed appear common (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 improve beliefs. This is because the politician can respond to an increase in the fabrication cost χ_f by escalating the alternative reality. Intuitively, in response to more credible evidence, the politician can upgrade from a lying cost AR to a conspiracy AR, and explain away the more credible evidence with the narrative of a more powerful elite.[20] 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 has to build 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

[19] We also note that the π large condition in the Proposition is required by χ_f, rather than uniformly.

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

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, Boxell, Conway, Gentzkow, Thaler and Yang (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, Helios, Morelli and Sonno 2023). Our chain of causality suggests that eliminating propaganda should improve trust in science.

4.2 Government policy in the shadow of propaganda

In our final application we explore the question of how spreading alternative realities shapes the quality of governance. The new insight our model offers is that maintaining beliefs in the alternative reality constrains government policy. In particular, the politician has an incentive to follow harmful policies that trigger elite criticism and thus strengthen beliefs in the alternative reality.

To incorporate government policy to the model, we assume that the bad politician can take a policy action that makes her bad type more visible to the elite. This action captures the key aspect of a "bad policy" for our purposes that it invites elite criticism.[21] Formally, in stage 1, the bad politician, simultaneously with her propaganda decision, can take an action $e \in \{0, 1\}$ that has vanishingly small cost and increases the probability that the elite's signal about her type is correct to $\pi' > \pi$. Neither the elite nor the voters observe this action.

We make one other substantive departure from the basic model: we assume that the politician

[21] A microfoundation that makes the role of the policy action explicit is to assume that the politician's common type measures the frequency with which she knows the right policy. With a low probability a bad politician knows it too, but even if she does, she can choose not to follow it.

cares more than the elite about the beliefs of receptive voters. Formally, the politician maximizes

$$U_p = \tilde{\mu}(\theta_c = 1|\hat{p}, \hat{s}) - f \cdot p, \tag{12}$$

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

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

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 AR 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 AR elite, this misperception creates the desired wedge. Another interpretation is that the politician also cares about winning a primary election, and hence overweights the beliefs of her core (receptive) supporters. And a third interpretation is that the AR 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 AR elite to gain from elite criticism.

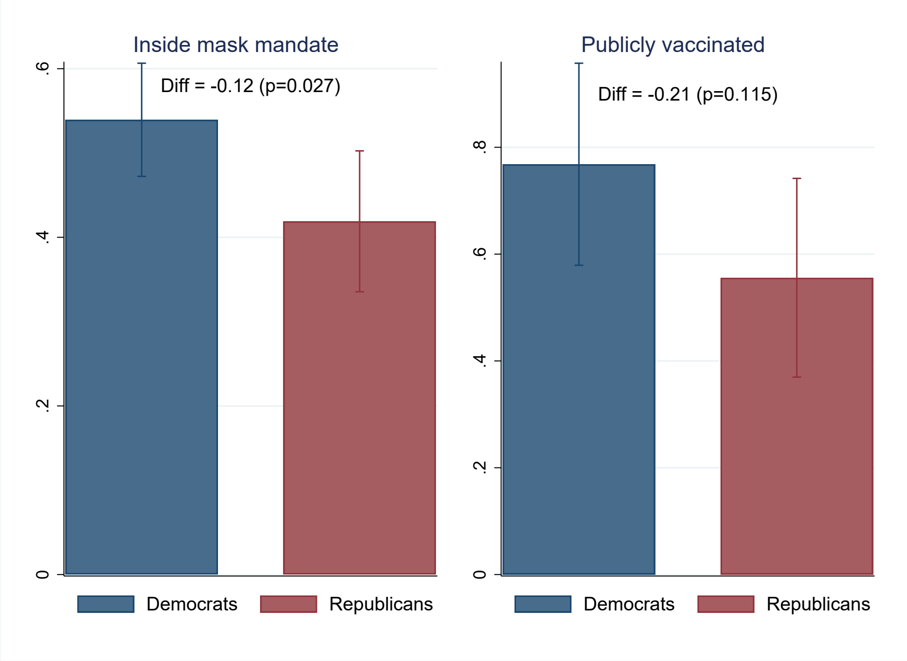
Proposition 3. *Under Assumptions 1 and 2, if $\alpha < 0.5$, and $\alpha' > 1/(1 + \hat{q}_c)$, then for π large enough and $\pi' > \pi$, in the unique PPO equilibrium*

1. *All propaganda and message choices are as in Proposition 1.*

2. *The bad politician chooses to increase the precision of the elite's signal if and only if reality is R and she can send propaganda.*

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 (7) it is a dominant strategy for the AR elite 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

Figure 3: Impact on policy



action and trigger criticism, without upsetting the narrative that the conspiring elite is incentivized to send that criticism.

Evidence. The key empirical prediction is that propaganda-spreading politicians will set policies that invite elite criticism. It is not just that these politicians ignore expert opinion: they actively contradict it. Since major societal issues, e.g., in the health and environmental domain, often require government action, the prediction highlights an important cost of propaganda.

Figure 3 presents evidence on this prediction in the Covid context. The left panel shows that across U.S. states and over time, controlling for the severity of the epidemic, Republican governors introduced indoor mask mandates 12 percentage points less often than Democratic governors. The right panel shows that in the cross-section of states, controlling for the severity of the epidemic, Republican governors vaccinated themselves publicly 21 percentage points less often than Democratic governors.[22] A possible alternative explanation for these facts is pandering to voters'

[22] In the left panel we use monthly data for U.S. states in 2020-21 and control for the number of Covid related cases, hospitalizations and deaths per 100,000 inhabitants in each state-month cell. In the right panel we control for the cumulative—up to October 2021—number of Covid related hospitalizations and deaths per 100,000 inhabitants

demand: Republican governors may avoid indoor masks mandates because these violate personal freedoms important to their voters. Although we cannot conclusively rule out that explanation, we find it implausible. Republican voters often prioritize the sanctity of life over personal freedom, as demonstrated by their opposition to abortion; indoor mask mandates are actually pro-business and thus align with that component of Republican values (Zhao, Yao, Thomadsen and Wang 2023); and public vaccination is hardly a violation of personal freedoms. Thus, in our opinion, the most likely explanation for Figure 3 is that Republican governors preferred not to follow the expert consensus.

5 Conclusion

In this paper we built a new model of populism in which a politician can supply an alternative reality to discredit the criticism of the intellectual elite. The alternative reality is a conspiracy theory in which members of the intellectual elite jointly attack the politician to advance their own interests. This model leads to several new predictions that help explain previously unexplained facts. We summarize the eight main predictions in Table 6. Seven of these predictions are to our knowledge are new, while one parallels an existing prediction but features a new mechanism. As we described in the body of the paper, six of the predictions are supported by consistent evidence, and one by causal evidence. This also means that for most predictions we lack causal evidence. Future work could evaluate the validity of these predictions.

Our approach of explicitly modeling a false alternative reality may be portable to other systems of misbeliefs. One possible example, in the context of international conflict, is misbeliefs about the intentions of an other country. Aiming to deflect criticism or initiate collective action, political leaders may demonize the citizens of the other country, as contemporary Russian propaganda demonizes “the West.” This *threatening enemy* alternative reality may be modeled using the tools developed in this paper. Such a model may lead to predictions about the escalation of conflict, through the logic that persuaded citizens may misinterpret defensive measures of the other country as offensive. Exploring the implications of alternative realities for conflict and other settings is an interesting avenue for future work.

in each state.

Prediction	New result	New mechanism	Evidence
Basic model			
1. AR propaganda lowers democratic accountability	yes	yes	consistent
2. Elite's effect inverted with propaganda	yes	yes	causal
3. AR beliefs amplified by subsequent events	yes	no	consistent
4. Societal division causes propaganda	no	yes	consistent
Endogenous alternative reality			
5. Alternative realities feature conspiracy theories	yes	yes	consistent
6. Credible evidence makes AR conspiratorial	yes	yes	none
7. Propaganda creates distrust and non-adoption	yes	yes	consistent
Government policy			
8. Propaganda causes harmful policies	yes	yes	consistent

Table 6: New predictions

Perhaps the main limitation of our approach is that the demand side of misbeliefs is underdeveloped. Building a fuller theory of why voters adopt alternative realities, and linking the demand side to economic and social conditions, is another interesting topic for future research.

References

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin, " A Political Theory of Populism ," *The Quarterly Journal of Economics*, 02 2013, *128* (2), 771–805.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya, " Radio and the Rise of The Nazis in Prewar Germany," *The Quarterly Journal of Economics*, 07 2015, *130* (4), 1885–1939.

Agiesta, Jennifer and Ariel Edwards-Levy, "CNN Poll: Percentage of Republicans who think Biden's 2020 win was illegitimate ticks back up near 70%," CNN Politics, `https://edition.cnn.com/2023/08/03/politics/cnn-poll-republicans-think-2020-election-illegitimate/index.html`, 2023.

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

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

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

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

Allen, Jonathan, "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`, 2023.

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

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

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

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

____, Armin Falk, and Jean Tirole, "Narratives, imperatives, and moral reasoning," Technical Report, National Bureau of Economic Research 2018.

Berinsky, Adam J, "Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey responding," *The Journal of Politics*, 2018, *80* (1), 211–224.

Berk, Robert H., "Limiting Behavior of Posterior Distributions when the Model is Incorrect," *The Annals of Mathematical Statistics*, 1966, *37* (1), 51–58.

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

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

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

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini, "Identity, Beliefs, and Political Conflict," *The Quarterly Journal of Economics*, 09 2021, *136* (4), 2371–2411.

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

Bullock, John G, Alan S Gerber, Seth J Hill, and Gregory A Huber, "Partisan bias in factual beliefs about politics," Technical Report, National Bureau of Economic Research 2013.

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

Clayton, Katherine, Spencer Blair, Jonathan A Busam, Samuel Forstner, John Glance, Guy Green, Anna Kawata, Akhila Kovvuri, Jonathan Martin, Evan Morgan et al., "Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media," *Political behavior*, 2020, *42*, 1073–1095.

Corasaniti, Nick and Trip Gabriel, "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`, 2023.

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

Eddy, Kirsten, "More than half of Americans are following election news closely, and many are already worn out," Pew Research Center, `https://www.pewresearch.org/short-reads/2024/05/28/more-than-half-of-americans-are-following-election-news-closely-and-many-are-already-worn-out/`, 2024.

Egorov, Georgy and Konstantin Sonin, "The political economics of non-democracy," Technical Report, National Bureau of Economic Research 2020.

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

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

Engler, Sarah and David Weisstanner, "The threat of social decline: income inequality and radical right support," *Journal of European Public Policy*, 2021, *28* (2), 153–173.

Esponda, Ignacio and Demian Pouzo, "Berk-Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models," *Econometrica*, 2016, *84* (3), 1093–1130.

Flegenheimer, Matt, "Is Ron DeSantis the Future of the Republican Party?," The New York Times Magazine, `https://www.nytimes.com/2022/09/13/magazine/ron-desantis.html`, 2022.

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

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

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

Gift, Thomas, "Trump v DeSantis: how the two Republican presidential heavy-hitters compare," The Conversiation, `https://theconversation.com/trump-v-desantis-how-the-two-republican-presidential-heavy-hitters-compare-202010`, 2023.

Glaeser, Edward L., "The Political Economy of Hatred," *The Quarterly Journal of Economics*, 02 2005, *120* (1), 45–86.

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

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

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

Horovitz, David, "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/`, 2020.

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

hvg.hu, "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`, 2017.

Jehiel, Philippe, "Analogy-based expectation equilibrium," *Journal of Economic Theory*, 2005, *123* (2), 81–104.

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

Koçak, Korhan, "Sequential updating: A behavioral model of belief change," in "Tech. Rep., Technical report, Working Paper" 2018.

Kocsis, Eva, "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`, 2017.

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

Levy, Gilat, Ronny Razin, and Alwyn Young, "Misspecified Politics and the Recurrence of Populism," *American Economic Review*, March 2022, *112* (3), 928–62.

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

McFadden, Alyce and Jeffery C. Mays, "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`, 2024.

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

Moessner, Christopher and Jennifer Berg, "Many Americans believe that climate change is mostly caused by human activity, but few report making changes to help limit it," Ipsos, `https://www.ipsos.com/en-us/many-americans-believe-climate-change-mostly-caused-human-activity-few-report/-making-changes-help`, 2023.

Murray, Mark, "Poll: 61% of Republicans still believe Biden didn't win fair and square in 2020," NBC News, `https://www.nbcnews.com/meet-the-press/meetthepressblog/poll-61-republicans-still-believe-biden-didnt-win-fair-square-2020-rcna49630`, 2022.

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

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

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

Ott, Brian L and Greg Dickinson, "The Twitter presidency: How Donald Trump's tweets undermine democracy and threaten us all," *Political Science Quarterly*, 2020, *135* (4), 607–636.

Paul, Christopher and Miriam Matthews, "The Russian "firehose of falsehood" propaganda model," *Rand Corporation*, 2016, *2* (7), 1–10.

Rogenhofer, Julius Maximilian and Ayala Panievsky, "Antidemocratic populism in power: Comparing Erdoğan's Turkey with Modi's India and Netanyahu's Israel," *Democratization*, 2020, *27* (8), 1394–1412.

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

Soros, George, "Remarks Delivered at the 2023 Munich Security Conference," https://www.georgesoros.com/2023/02/16/remarks-delivered-at-the-2023-munich-security-conference/, 2023.

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

——, "CNN Poll: March 8-12, 2023," https://s3.documentcloud.org/documents/23706881/cnn-poll-most-republicans-care-more-about-picking-a-2024-gop-nominee-who-agrees-with-them-on-issues-than-one-who-can-beat-biden.pdf, 2023.

——, "CNN Poll: May 17-20, 2023," https://s3.documentcloud.org/documents/23823977/cnn-poll-trump-leads-2024-gop-primary-field-but-voters-are-open-to-supporting-other-candidates.pdf, 2023.

Stoetzer, Lukas F, Johannes Giesecke, and Heike Klüver, "How does income inequality affect the support for populist parties?," *Journal of European Public Policy*, 2023, *30* (1), 1–20.

Sükösd, Miklós, "Victorious victimization: Orbán the orator—deep securitization and state populism in Hungary's propaganda state," in "Populist rhetorics: case studies and a minimalist definition," Springer, 2022, pp. 165–185.

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

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

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

Thompson, Stuart A., "'Ron DeSoros'? Conspiracy Theorists Target Trump's Rival.," The New York Times, https://www.nytimes.com/2023/05/05/technology/ron-desantis-conspiracy-theorists.html, 2023.

Uscinski, Joseph E., Karen Douglas, and Stephan Lewandowsky, "Climate Change Conspiracy Theories," 09 2017.

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

Yanagizawa-Drott, David, " Propaganda and Conflict: Evidence from the Rwandan Genocide," *The Quarterly Journal of Economics*, 11 2014, *129* (4), 1947–1994.

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

Zelizer, Julian, "Opinion: What gives DeSantis an edge over Trump," cnn.com, https://edition.cnn.com/2022/06/23/opinions/desantis-trump-2024-zelizer/index.html, 2022.

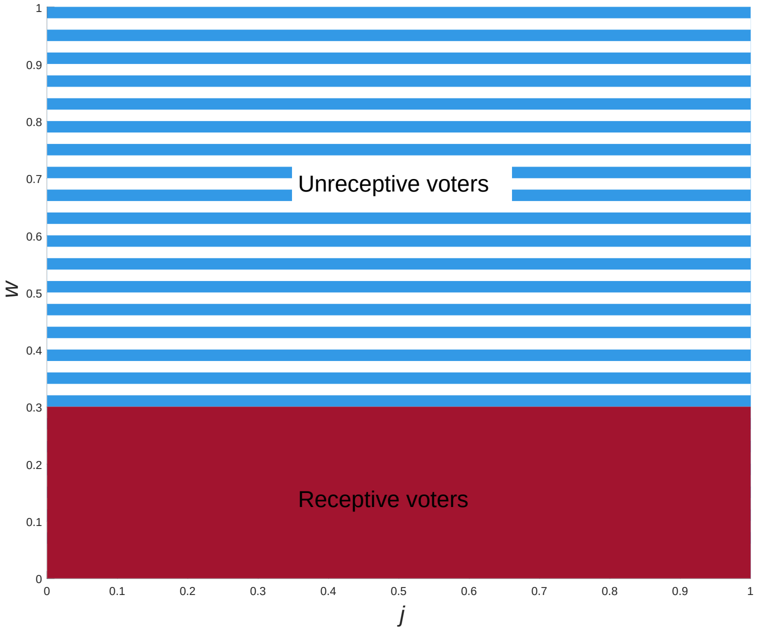
Zhao, Nan, Song Yao, Raphael Thomadsen, and Chong Bo Wang, "The Impact of Government Interventions on COVID-19 Spread and Consumer Spending," *Management Science*, 2023, *0* (0), null.

Appendix for Online Publication

A Definitions and proofs

A.1 Formal model of audiences

Figure 4: Media audiences



Our baseline model has a unit mass of voters distributed uniformly on the unit square. We index voters by $i = (j, w)$, thus

$$\int_0^1 \int_0^1 1 dj dw = 1$$

As it is shown in Figure 4, the first α share of voter along dimension w—$\alpha = 0.3$ in the figure—are receptive to propaganda, while the rest are unreceptive. We index media—both elite and new—by j and its audience is given by

$$\text{Audience of media } j = \{(z, w) \in [0, 1]^2 : z = j\}$$

A.2 Microfoundation of principals' objectives

In the main model we assume that the AR elite and the incumbent politician care about the average voter belief about the politician's type. Here we provide microfoundations for this assumption using a probabilitic voting model, in which after stage 2 of the game an election takes place between the incumbent and a challenger. The challenger is good with probability q_c^c. Voter i chooses between the incumbent and a challenger to maximize utility

$$U_{v,i} = c\tilde{\theta}_c + \lambda \cdot 1_{\{\text{Incumbent}\}} + \epsilon + \eta_i, \tag{A1}$$

where $v \in \{rec, un\}$; $\tilde{\theta}_c$ is the competence of the elected politician; λ is an additional preference component of the voter about the incumbent, which reflects ideological alignment; and ϵ and η_i are mean-zero, independent, uniformly distributed common and individual preference shocks, which have supports $[-\bar{g}, \bar{g}]$ and $[-\bar{h}, \bar{h}]$, constant densities $g = 1/(2\bar{g})$ and $h = 1/(2\bar{h})$. We assume that $\bar{h} > c + \lambda + \bar{g}$ and $\bar{g} > c + \lambda$ to avoid corner outcomes.

Elite members' preferences are given by

$$\tilde{U}_{e,j} = c\tilde{\theta}_c - \lambda \cdot 1_{\{\text{Incumbent}\}} \tag{A2}$$

reflecting that their ideology is the opposite of the voters'. Thus, $\lambda > 0$ corresponds to the incumbent being ideologically pro-voter, while $\lambda < 0$ corresponds to the incumbent being ideologically pro-elite.

The incumbent politician's preferences are given by

$$\tilde{U}_p = E \cdot 1_{\{\text{In office}\}} - \tilde{f} \cdot p, \tag{A3}$$

where E is an ego rent and $\tilde{f}$ is the cost of propaganda.

The following Lemma shows that the preferences in this microfounded model are equivalent to those in the model in the main text, implying that the two models have the same equilibria.

Lemma 1. *In this model, the expected utilities of the elite and the politician, conditional on the politician's type θ_c and the message profile $(\hat{s}, \hat{p})$, are positive affine transformations of the utility* *functions introduced in the main text*

$$U_{e,j}(\theta_c, \hat{p}, \hat{s}) = (\theta_c - \kappa)\bar{\mu}(\theta_c = 1|\hat{p}, \hat{s})$$
$$U_p(\theta_c, \hat{p}, \hat{s}) = \bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) - f \cdot p,$$

where $\kappa \equiv q_c^c + \frac{\lambda}{c}$ *is the cost of reelecting the incumbent for the elite, and* $f \equiv \frac{\tilde{f}}{E \cdot g \cdot c}$ *is the normalized cost of propaganda.*

Proof of Lemma 1. The probability, conditional on a fixed common shock ϵ, that voter i votes for the incumbent is

$$\Pr\left[c(q_c^c - \mu_{v,i}(\theta_c|\hat{p}, \hat{s}_{j(i)})) - \lambda - \epsilon < \eta_i|\epsilon\right] = 0.5 - h\left[c(q_c^c - \mu_{v,i}(\theta_c|\hat{p}, \hat{s}_{j(i)})) - \lambda - \epsilon\right]$$

because η_i has a uniform distribution with a density h. The incumbent wins the election if she gets the majority of votes:

$$\int \left\{0.5 - h\left[c(q_c^c - \mu_i(\theta_c = 1|\hat{p}, \hat{s}_{j(i)})) - \lambda - \epsilon\right]\right\} di > 0.5$$
$$c(q_c^c - \bar{\mu}(\theta_c = 1|\hat{p}, \hat{s})) - \lambda < \epsilon,$$

where voters' average posterior belief of the politician's type is given by

$$\begin{aligned}\bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) \equiv \int \mu_i(\theta_c = 1|\hat{p}, \hat{s}_{j(i)})di &= \int\int \mu_{(j,w)}(\theta_c = 1|\hat{p}, \hat{s}_j)dwdj \\ &= \int \left[\alpha\mu_{rec,j}(\theta_c = 1|\hat{p}, \hat{s}_j)dj + (1-\alpha)\mu_{un,j}(\theta_c = 1|\hat{p}, \hat{s}_j)\right] dj.\end{aligned}$$

In the second line we use the notation that $\mu_{rec,j}$ and $\mu_{un,j}$ is the average belief of all receptive voters and all unreceptive voters, respectively, in the audience of elite member j. Because voters within the audience of elite member j and voter type (receptive/unreceptive) access the same signals, their beliefs are the same, so both of these averages are averaging a constant. Moreover, since functions $\mu_{rec,j}(\theta_c = 1|\cdot)$ and $\mu_{un,j}(\theta_c = 1|\cdot)$ are the same for each elite member j, the integral is maximized by the same value of $\hat{s}_j$ for all j. Thus, the optimal behavior of the AR elite is to choose the same message s for all members, and therefore below simply denote $\hat{s}_j$ by $\hat{s}$.

The incumbent's probability of winning is thus

$$\begin{aligned}P &\equiv \Pr\left[c(q_c^c - \bar{\mu}(\theta_c = 1|\hat{p}, \hat{s})) - \lambda < \epsilon\right] \\ &= g \cdot c \cdot \bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) + g(\lambda - c \cdot q_c^c) + 0.5.\end{aligned} \tag{A4}$$

Now consider the AR elite. Her conditional expected utility is

$$\begin{aligned}E[\tilde{U}_e|\theta_c, \hat{p}, \hat{s}] &= P(c\theta_c - \lambda) + (1-P)cq_c^c \\ &= g \cdot c^2(\theta_c - \kappa)\bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) \\ &+ [g(\lambda - cq_c^c) + 0.5][c(\theta_c - q_c^c) - \lambda] + cq_c^c \\ &= L_e[(\theta_c - \kappa)\bar{\mu}(\theta_c = 1|\hat{p}, \hat{s})]\end{aligned}$$

where $\kappa \equiv q_c^c + \frac{\lambda}{c}$ and L_e is a positive affine transformation, as claimed. Note that L_e depends on the state θ_c, but this is not a problem because the state is exogenous from the perspective of all actors.

Next consider the politician. Her expected utility is

$$\begin{aligned}E(\tilde{U}_p|\hat{p}, \hat{s}) &= E \cdot P - \tilde{f} \cdot p \\ &= E\left[g \cdot c \cdot \bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) + g(\lambda - c \cdot q_c^c) + 0.5\right] - \tilde{f} \cdot p \\ &= E \cdot g \cdot c\left[\bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) - f \cdot p\right] + E[g(\lambda - c \cdot q_c^c) + 0.5] \\ &= L_p[\bar{\mu}(\theta_c = 1|\hat{p}, \hat{s}) - f \cdot p],\end{aligned}$$

where L_p is a positive affine transformation, as claimed.

A.3 Definition of equilibrium

We start with introducing notation. We define the politician's type to be $\theta_p = (\theta_c, \theta_r)$. We define the elite's type to be $\theta_e = (\hat{\theta}_c, \theta_r)$, which differs from the politician's type only because the elite does not observe θ_c directly, only a signal $\hat{\theta}_c$ on it. We define the receptive voter's type to be $\theta_{rec} = \theta_m$ because his priors depend on θ_m. Note that the types of different actors are correlated. We denote the action of actor k in stage $t \in \{1, 2\}$ by a_k^t. We let $\hat{a}_k^t$ stand for the realized action after Nature's tremble, and $\hat{a}^t$ for the realized action profile. The history at stage t is denoted by $\hat{h}^t = (\hat{a}^1, ..., \hat{a}^t)$.

We define strategies as probability distributions over actions at the stages where an actor gets to move. Because the politician and the elite only move in stage 1, their strategies only depend on their type, and are denoted by $\sigma_p(a_p^1|\theta_p)$ respectively $\sigma_e(a_e^1|\theta_e)$. As the receptive voter moves in stage 2 after observing $\hat{a}^1 = (\hat{s}, \hat{p})$, his strategy depends on $\hat{a}^1$ and is denoted by $\sigma_{rec}(a^2_{rec}|\theta_{rec}, \hat{a}^1)$. The unreceptive also moves in stage 2 but only observes $\hat{s}$, not $\hat{p}$. Thus, his strategy only depends on $\hat{s}$, but for ease of notation we will denote it by $\sigma_{un}(a^2_{un}|\theta_{un}, \hat{a}^1)$. We let $\hat{\sigma}$ denote perturbed strategies that incorporate Nature's trembles. We denote the prior belief of actor k of type θ_k by $\mu^0_k(\theta|\theta_k)$, and the posterior belief after history $\hat{h}^t$ by $\mu^t_k(\theta|\theta_k, \hat{h}^t)$. We allow beliefs to depend on types, both because the types of different actors are correlated so that the type of k has information about the types of $-k$, and because different types can have different priors.

Our equilibrium concept is a version of perfect Bayesian equilibrium that recognizes our framework's departure from common priors and full rationality. As usual, equilibrium requires that actors best respond and form consistent beliefs. We begin with beliefs. We first note that because of the trembles beliefs will be always well defined. Belief consistency does not impose any condition on principals, because they move only at stage 1 where they know only their priors. Belief consistency for the receptive voter requires that he follows Bayesian updating at the end of stage 1:

$$\mu^1_{rec}(\theta_{-rec}|\theta_{rec}, \hat{a}^1) = \frac{\mu^0_{rec}(\theta_{-rec}|\theta_{rec}) \cdot \hat{\sigma}^1_{-rec}(\hat{a}^1|\theta_{-rec})}{\sum_{\theta'_{-rec}} \mu^0_{rec}(\theta'_{-rec}|\theta_{rec}) \cdot \hat{\sigma}^1_{-rec}(\hat{a}^1|\theta'_{-rec})} \tag{A5}$$

where $\mu^0_{rec}(\theta_{-rec}|\theta_{rec})$ is the prior of the receptive voter of type θ_{rec} about the types of the other actors $\theta_{-rec} = (\theta_c, \hat{\theta}_c, \theta_r)$. This definition accounts for the model's deviation from rationality that the receptive voter's mind type and beliefs may change in stage 1, by computing the posterior for each mind type $\theta_m = N, P$ using the prior associated with that mind type. In particular, if the receptive voter is reached by propaganda and becomes persuaded, (A5) computes his posterior from the prior of the persuaded voter $\mu^0_{rec}(.|\theta_m = P)$. Intuitively, because the persuaded voter uses Bayes rule, he infers from the presence of propaganda about the politician's type; but because propaganda also influences his type, this inference is based on the prior modified by propaganda. Implicit in this is that when the receptive voter receives messages $\hat{a}^1 = (\hat{s}, \hat{p})$, first propaganda $\hat{p}$ changes his mind type and prior, and then he updates from his new prior based on the information content of $\hat{a}^1$. Finally, the unreceptive voter performs standard Bayesian updating based on observing $\hat{s}$.

We next formulate the best-response condition. To do so, we introduce subjective expected utility. In the model presented in the main text only the principals derive utility, while in the microfoundation presented above the voters also derive utility. In both cases, each actor who maximizes utility, at each stage where it moves, has a subjective probability distribution over final outcomes, where the final outcome is mean voter beliefs in the model presented in the main text. This distribution can differ from the objectively correct distribution because the persuaded voter has an incorrect prior about θ. Actor k at stage t uses its subjective probability distribution over outcomes to compute its subjective expected utility, denoted $U_k(\sigma|\hat{h}^t, \theta_k, \mu_k(\theta|\theta_k, \hat{h}^t))$. For the unreceptive voter who does not observe the full history, we use the same notation to represent his expected utility conditional on only the part of history $\hat{h}^t$ that he does observe. Then the best-response property of equilibrium is that at each stage t at which k has a move, for all actions σ'_k available to k,

$$U_k(\sigma|\hat{h}^t, \theta_k, \mu_k(.|\theta_k, \hat{h}^t)) \geq U_k((\sigma'_k, \sigma_{-k})|\hat{h}^t, \theta_k, \mu_k(.|\theta_k, \hat{h}^t)).$$

Finally, we need to define what we mean by a mixed equilibrium in this model with an infinitesimal lying cost. We say a mixed equilibrium respects the lying cost if (a) it is a mixed equilibrium; and (b) for any $\varepsilon > 0$ there exists $\delta > 0$ such that for a lying cost χ below δ there exists an equilibrium in which all mixing probabilities are within ε of the original equilibrium. We only consider equilibria that respect the lying cost.

A.4 Proof of Proposition 1

Going beyond the result stated in the main text, we characterize the unique PPO equilibrium for all values of α. We start with the definitions of two equilibrium profiles.

Definition 1. A strategy profile has the *simple propaganda form* if

1. In the reality (R):
   - The elite reports the common type truthfully,
   - The politician sends propaganda if she can and she is bad.
2. In the alternative reality (AR):
   - The elite always reports that the politician is bad,
   - The politician sends propaganda if she can.

Definition 2. A strategy profile has the *complex propaganda form* if the AR elite, when the signal is good, randomizes between the good and the bad message, while all other principal types behave as in the simple propaganda profile.

We now prove the following generalization of Proposition 1.

Proposition 1'. *Under Assumptions 1 and 2 there exists $\bar{\pi} < 1$ such that for $\pi > \bar{\pi}$ there exists $\alpha(\pi) > 0.5$ such that*

1. *For $\alpha < \alpha(\pi)$ the unique PPO equilibrium has the simple propaganda form.*

2. *For $\alpha > \alpha(\pi)$ the unique PPO equilibrium has the complex propaganda form.*

Proof. Because the proof is long, we have broken it into several numbered steps.

1. Voter beliefs in the simple propaganda profile

We first derive voters' posterior beliefs assuming that play follows the simple propaganda profile. These formulas will be key for the analysis. The $1-\alpha$ share of unreceptive voters have the following posterior beliefs, irrespective of propaganda, as a function of the elite's message:

$$\mu_{un}(\theta_c = 1|\hat{s}) = \hat{s}\frac{\pi q_c}{\pi q_c + (1-\pi)(1-q_c)} + (1-\hat{s})\frac{(1-\pi)q_c}{(1-\pi)q_c + \pi(1-q_c)}. \tag{A6}$$

This expression follows by straightforward Bayesian updating from the elite's message, under the assumption (made by these voters) that the elite's message equals her signal and hence is correct with probability π.

Consider next the share α of receptive voters. In the absence of propaganda, their beliefs are given by (4), which we repeat here for convenience

$$\mu_{rec}(\theta_c = 1|\hat{p} = 0, \hat{s}, \theta_m = N) = \hat{s}\frac{\pi q_c}{\pi q_c + (1-\pi)(1-q_c)\beta} + (1-\hat{s})\frac{(1-\pi)q_c}{(1-\pi)q_c + \pi(1-q_c)\beta}. \tag{A7}$$

This equation is also the result of straightforward Bayesian updating. The difference relative to (A6) is that these voters, since they are capable of observing it, learn from the fact that there is no propaganda. This mechanism explains the terms involving β in the denominators, since β is the probability with which the bad politician is unable to send propaganda. Thus, a good message, absent propaganda, can reflect a bad politician, an incorrect elite signal, and the inability to send propaganda, captured in the denominator in the first term; and a bad message, absent propaganda, can reflect a bad politician, a correct elite signal, and the inability to send propaganda, captured in the denominator of the second term.

An implication is that because $\beta > 0$, the voter does not fully infer from the absence of propaganda that the politician is good, so that the elite's message is still informative for his updating. In fact, (A7) implies that for π large (holding fixed β) beliefs are primarily determined by the elite's message $\hat{s}$, so that they are near one when $\hat{s} = 1$ and near zero when $\hat{s} = 0$. Intuitively, even though the absence of propaganda is informative, the elite's message is a more informative signal.

Finally, the beliefs of the receptive voter in the presence of propaganda are given by (5), which we repeat here for convenience

$$\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s}, \theta_m = P) = (1 - \hat{s})\frac{q_{ar}q_c}{q_{ar} + q_r\pi(1 - q_c)}. \tag{A8}$$

This formula too follows from Bayesian updating. Consider first a bad elite message. Since propaganda changed the voter's prior, he thinks that the politician may be good in the AR, explaining the numerator. However, propaganda and a bad signal can also emerge in the AR if the politician is bad, and in R if the politician is bad and the elite's message is correct, explaining the denominator. Consider next a good elite message. The profile of praise and propaganda is only possible in R and proves that the politician is bad.

2. Cutoff α value

We turn to characterize the condition on α under which the simple propaganda equilibrium exists. This will turn on whether, in the simple propaganda profile, the AR elite finds it optimal to criticize after propaganda. Since the goal of the AR elite is to minimize voter beliefs, it follows from the above expressions that she chooses to criticize if and only if

$$(1 - \alpha)\left[\frac{\pi q_c}{\pi q_c + (1 - \pi)(1 - q_c)} - \frac{(1 - \pi)q_c}{(1 - \pi)q_c + \pi(1 - q_c)}\right] > \alpha\frac{q_{ar}q_c}{q_{ar} + q_r\pi(1 - q_c)}. \tag{A9}$$

The left-hand side is the gain from worsening the beliefs of non-receptive voters, and is obtained by differencing (A6) between $\hat{s} = 1$ and $\hat{s} = 0$. The right-hand side is the loss from improving the beliefs of receptive voters, and is obtained by differencing (A8) between $\hat{s} = 1$ and $\hat{s} = 0$. It is straightforward to check that this inequality yields a threshold $\bar{\alpha}(\pi)$, such that for $\alpha < \bar{\alpha}(\pi)$ the AR elite strictly prefers to criticize the politician. Moreover, for π approaching 1, the term multiplying $1 - \alpha$ on the left hand side approaches 1, while the term multiplying α on the right hand side approaches $\hat{q}_c < 1$, implying that for π large enough, $\bar{\alpha}(\pi) > 0.5$. As a result, when π is large, for $\alpha < 0.5$ we are always in the range corresponding to the simple propaganda equilibrium.

We now turn to show that the proposed equilibrium exists, separately in the ranges below and above $\bar{\alpha}(\pi)$.

3. Equilibrium existence for $\alpha < \bar{\alpha}(\pi)$

We establish that the simple propaganda profile is an equilibrium using backward induction. The R elite always reports truthfully after any history to minimize the lying cost. The AR elite, absent propaganda, will (for π large) always send a bad message, because that minimizes the posterior of both the non-receptive voter by (A6) and the receptive voter by (A7). The AR elite, following propaganda, will send a bad message because $\alpha < \alpha(\pi)$ means that (A9) holds. Thus, all elite types find it optimal to follow the strategies in the proposed profile.

We next consider the politician types. Start with the good R politician. For π high, she expects to be praised by the R elite, and is thus getting a payoff close to the highest possible in the game. Since the cost of propaganda f is bounded away from zero, for π high she does not send propaganda.

Consider the bad R politician. She will prefer to send propaganda if and only if

$$\alpha \left[\pi \left(\frac{q_{ar} q_c}{q_{ar} + q_r \pi (1 - q_c)} - \frac{(1-\pi) q_c}{(1-\pi) q_c + \pi (1 - q_c) \beta} \right) + (1 - \pi) \cdot \frac{-\pi q_c}{\pi q_c + (1-\pi)(1 - q_c) \beta} \right] > f. \tag{A10}$$

The left hand side measures the expected gain from propaganda. Propaganda only has an effect on the share of voters α who observe propaganda. For these voters, if the elite sends a bad message (with probability π), then propaganda changes beliefs to the value given in (A8) for $s = 0$, from the value given in (A7) for $s = 0$. This explains the first term. If the elite sends a good message (with probability $1 - \pi$), then propaganda changes beliefs to the value given in (A8) for $s = 1$, which is zero, from the value given in (A7) for $s = 1$. This explains the second term.

Observe that the limit of the left-hand side, as π goes to one, is

$$\alpha \frac{q_{ar} q_c}{q_{ar} + q_r(1 - q_c)} = \alpha \hat{q}_c.$$

Thus, Assumption 2 implies that for π sufficiently large the bad R politician will prefer to send propaganda.

Consider next the good and the bad AR politicians. They prefer to send propaganda if

$$\alpha \left[\frac{q_{ar} q_c}{q_{ar} + q_r \pi (1 - q_c)} - \frac{(1 - \pi) q_c}{(1 - \pi) q_c + \pi (1 - q_c) \beta} \right] > f. \tag{A11}$$

This is slightly different from condition (A10), because while in the R the elite sends a bad message only with probability π, in the AR it sends a bad message with probability 1. However, it remains true that in the limit as π goes to one, the left-hand side converges to $\alpha \hat{q}_c$, so that Assumption 2 implies that the AR politicians too will prefer to send propaganda.

4. Equilibrium existence for $\alpha > \bar{\alpha}(\pi)$

We prove that there exists an equilibrium that has the complex propaganda profile: following propaganda, the AR elite sends a bad message after a bad signal and plays a mixed action after a good signal, while all other players follow the simple propaganda profile. We proceed by backward induction. As before, the R elite reports truthfully to avoid the lying cost.

Now consider the condition for the AR elite's indifference after propaganda. Suppose that the mixing probability of sending a good report after a good signal is r. Voters' average belief after a good report is given by

$$\begin{aligned}
&\bar{\mu}(\theta_c = 1 | \hat{p} = 1, \hat{s} = 1) \\
&= \alpha \mu_{rec}(\theta_c = 1 | \hat{p} = 1, \hat{s} = 1, \theta_m = P) + (1 - \alpha) \mu_{un}(\theta_c = 1 | \hat{s} = 1) \\
&= \alpha \frac{q_{ar} q_c \pi r}{q_{ar} q_c \pi r + q_{ar}(1 - q_c)(1 - \pi) r + q_r (1 - q_c)(1 - \pi)} + (1 - \alpha) \frac{\pi q_c}{\pi q_c + (1 - \pi)(1 - q_c)}.
\end{aligned}$$

The first term follows from Bayesian updating by a receptive voter influenced by propaganda, who accounts for the fact that the AR elite randomizes after a good signal. This means that a good politician can be consistent with a good report and propaganda, if reality is AR, the elite's signal was good, and the elite randomized to follow that signal, explaining the numerator. However, the profile of propaganda and a good signal can also emerge in the AR if the politician is bad, the elite's signal was incorrect (good), and the elite randomized to follow it; and in the R if the politician is bad and the elite's signal was incorrect. This explains the denominator. The second term is the belief of the unreceptive voter and comes from (A6).

Voters' average belief after a bad report is given by

$$\begin{aligned}
&\bar{\mu}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0)\\
&= \alpha\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0, \theta_m = P) + (1-\alpha)\mu_{un}(\theta_c = 1|\hat{s} = 0)\\
&= \alpha\frac{q_{ar}\pi q_c(1-r) + q_{ar}(1-\pi)q_c}{q_{ar}\pi q_c(1-r) + q_{ar}(1-\pi)q_c + q_{ar}\pi(1-q_c) + q_{ar}(1-\pi)(1-q_c)(1-r) + q_r\pi(1-q_c)}\\
&+ (1-\alpha)\frac{(1-\pi)q_c}{(1-\pi)q_c + \pi(1-q_c)}.
\end{aligned}$$

The first term is the update of the receptive voter after propaganda and a bad message. This profile can be consistent with a good politician if reality is AR, and either the elite's signal was good and she randomized not to follow it, or was bad in which case she always follows it, explaining the numerator. However, this profile can also arise: in the AR if the politician is bad and the elite's signal was correct (bad); in the AR if the politician is bad, the elite's signal was incorrect (good) but she randomized to send a bad message; and in R if the politician is bad and the elite's signal was correct. This explains the denominator. The second term is the update of the unreceptive voter and comes from (A6).

It is tedious but straightforward to compute the partial derivatives of these beliefs with respect to r, and to sign them for $r \in [0,1]$:

$$\frac{\partial\bar{\mu}(\theta_c = 1|\hat{p} = 1, \hat{s} = 1)}{\partial r} = \frac{\pi(1-\pi)q_c(1-q_c)q_{ar}q_r}{[q_{ar}q_c\pi r + q_{ar}(1-q_c)(1-\pi)r + q_r(1-q_c)(1-\pi)]^2} > 0$$

and

$$\begin{aligned}
&\frac{\partial\bar{\mu}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0)}{\partial r} =\\
&-\frac{q_{ar}q_c(1-q_c)[\pi^2 q_r + q_{ar}(2\pi-1)]}{[q_{ar}\pi q_c(1-r) + q_{ar}(1-\pi)q_c + q_{ar}\pi(1-q_c) + q_{ar}(1-\pi)(1-q_c)(1-r) + q_r\pi(1-q_c)]^2} < 0.
\end{aligned}$$

Thus, for $r \in [0,1]$ the mean belief after praise is strictly increasing, while the mean belief after criticism is strictly decreasing in r.[23] Direct substitution implies that for $r = 0$ the former mean belief is smaller than or equal than the latter mean belief if and only if

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

which is the opposite of (7), implying that it holds since we assume $\alpha > \bar{\alpha}(\pi)$. For $r = 1$ the former mean belief is larger than the latter mean belief if and only if

$$\begin{aligned}\alpha \frac{q_{ar}q_c\pi}{q_{ar}q_c\pi + q_{ar}(1-q_c)(1-\pi) + (1-q_{ar})(1-q_c)(1-\pi)} + (1-\alpha)\frac{q_c\pi}{q_c\pi + (1-q_c)(1-\pi)} \\ > \alpha \frac{q_{ar}q_c(1-\pi)}{q_{ar}q_c(1-\pi) + q_{ar}(1-q_c)\pi + (1-q_{ar})(1-q_c)\pi} \\ + (1-\alpha)\frac{q_c(1-\pi)}{q_c(1-\pi) + (1-q_c)\pi}.\end{aligned}$$

To evaluate this inequality, note that (i) the left-hand side is increasing in π and (ii) we obtain the right-hand side from the left-hand side by replacing π with $1-\pi$. Thus, the inequality follows from $\pi > 1-\pi$ which holds since $\pi > 0.5$. It follows that there is a unique mixing probability r that makes the AR elite indifferent after propaganda between praise and criticism. This establishes the optimality of the AR elite's behavior.

To establish optimality for the politician, we first need to characterize r for π approaching one. To do this, consider the indifference condition

$$\begin{aligned}\alpha\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 1, \theta_m = P) + (1-\alpha)\mu_{un}(\theta_c = 1|\hat{s} = 1) \\ = \alpha\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0, \theta_m = P) + (1-\alpha)\mu_{un}(\theta_c = 0|\hat{s} = 0).\end{aligned}$$

Combining this condition with the fact that $\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0, \theta_m = P) \leq 1$ allows us to derive the inequality

$$\begin{aligned}\frac{q_{ar}q_c\pi r}{q_{ar}q_c\pi r + q_{ar}(1-q_c)(1-\pi)r + (1-q_{ar})(1-q_c)(1-\pi)} \\ \leq 1 - \frac{1-\alpha}{\alpha}\left(\frac{q_c\pi}{q_c\pi + (1-q_c)(1-\pi)} - \frac{q_c(1-\pi)}{q_c(1-\pi) + (1-q_c)\pi}\right)\end{aligned}$$

[23] In fact, these expressions also show that as π approaches 1, the first partial derivative approaches zero, while the second remains bounded away from zero, which is a property we will use later.

which can be further rewritten as

$$\frac{q_{ar}(1-q_c)(1-\pi)r + q_r(1-q_c)(1-\pi)}{q_{ar}q_c\pi r + q_{ar}(1-q_c)(1-\pi)r + (1-q_{ar})(1-q_c)(1-\pi)} \geq \frac{1-\alpha}{\alpha}\left(\frac{q_c\pi}{q_c\pi + (1-q_c)(1-\pi)} - \frac{q_c(1-\pi)}{q_c(1-\pi) + (1-q_c)\pi}\right)$$

and, increasing the left-hand-side, as

$$\frac{(1-q_c)(1-\pi)}{q_{ar}q_c\pi r} \geq \frac{1-\alpha}{\alpha}\left(\frac{q_c\pi}{q_c\pi + (1-q_c)(1-\pi)} - \frac{q_c(1-\pi)}{q_c(1-\pi) + (1-q_c)\pi}\right).$$

This implies

$$\frac{1}{r} \geq \frac{q_{ar}q_c\pi}{(1-q_c)(1-\pi)}\frac{1-\alpha}{\alpha}\left(\frac{q_c\pi}{q_c\pi + (1-q_c)(1-\pi)} - \frac{q_c(1-\pi)}{q_c(1-\pi) + (1-q_c)\pi}\right).$$

This expression implies that, uniformly in $\alpha \geq \bar{\alpha}(\pi)$, as π goes to one r goes to zero.

With this result in hand, we can establish the optimality of the politician's proposed behavior for π large. Consider the limit, as π approaches one, of the mean voter belief after a bad report and propaganda. Given that r goes to zero uniformly in α, the limit, uniformly in $\alpha \geq \bar{\alpha}(\pi)$, is

$$\alpha\frac{q_{ar}q_c}{q_{ar} + q_r(1-q_c)} = \alpha\hat{q}_c.$$

It follows that under Assumption 2, for π sufficiently large (independently of $\alpha \geq \bar{\alpha}(\pi)$) the R politician and both AR politicians will find it optimal to send propaganda. And the good R politician still prefers not to, because absent propaganda her payoff approaches the possible maximum (as π goes to one), while with propaganda she pays a non-negligible cost f. We conclude that the mixed equilibrium exists for π sufficiently high and $\alpha \geq \bar{\alpha}(\pi)$.

We conclude this existence proof by showing that this mixed equilibrium respects the lying cost. For any small lying cost χ, the indifference condition is distorted by a small additive constant. The argument for the mean beliefs after praise and propaganda are strictly increasing and decreasing, respectively, continues to be valid. Thus, it remains true that for any $\alpha > \bar{\alpha}(\pi)$, for a lying cost small enough there exists a mixing probability that ensures indifference. Moreover, as the lying cost approaches zero, the implied mixing probability approaches that corresponding to a zero lying cost. This follows from the observation made in footnote 23 that the slope in r of the belief after criticism remains bounded away from zero (while that after praise approaches zero), which implies that a small wedge between the two beliefs can be compensated for by a small change in r. It follows that for any given π, when the lying cost is sufficiently small, the payoffs of all parties are going to be close to those in the original equilibrium. Now in the original equilibrium the AR elite after a good signal is indifferent and mixes, the AR elite after a bad signal is indifferent but sends a bad message, and all other parties strictly prefer their equilibrium action. In the new profile of the game with lying cost the AR elite after a good signal is indifferent by construction; therefore the AR elite after a bad signal—given the lying cost—strictly prefers to send a bad message and does so, generating the same action as in the original game. By continuity all other parties have a strict preference to take their prescribed action. Thus this new profile is indeed close to the original profile and is an equilibrium of the game with a small lying cost.

5. Equilibrium selection for $\alpha < \bar{\alpha}(\pi)$

We use the politician pure refinement, that is, we only consider equilibria in which all politician types play pure strategies. Our goal is to identify the politician pure equilibrium which is optimal for the R politician. Our proof strategy is to check for all possible pure strategy profiles of all politician types. We go through the politician types one-by-one.

[Good R politician.] In any equilibrium, for π sufficiently high, the good R politician never sends propaganda. This is because with a high π probability her good type is revealed, in which case her utility is maximized, and propaganda has cost f bounded away from zero.

[Bad R politician.] Any equilibrium in which the bad R politician does not send propaganda, or is indifferent to not sending propaganda, is dominated by our preferred equilibrium. This is because payoffs absent propaganda would be the same for the bad R politician in all equilibria; and in our preferred equilibrium the bad R politician strictly prefers to send propaganda, implying that she earns a higher payoff from doing so. Thus, it suffices to consider equilibria in which the R politician strictly prefers to send propaganda if she is bad.

[Good AR politician.] Suppose that the good AR politician does not send propaganda. This means that propaganda reveals that the politician is bad. Hence, propaganda cannot be worthwhile for the bad R politician, a contradiction. Thus, the good AR politician must send propaganda.

[Bad AR politician.] This is the last step in the proof, but it is a complicated step. It will be useful to start this step by considering the behavior of the AR elite. It is immediate that after no propaganda, the AR elite always sends a bad message. After propaganda, we need to consider what the AR elite does as a function of the signal she receives.

- Propaganda and a good signal. Then, the AR elite must send a bad message with positive probability. Otherwise, a bad message after propaganda will prove that the elite received a bad signal (both in the R and the AR), implying that (for π high) the bad R politician will not want to send propaganda.
- Propaganda and a bad signal. Then, the AR elite must send a bad message with probability one. This follows because of the infinitesimal lying cost: since she weakly prefers a bad message after a good signal, when it is a lie, she must strictly prefer it after a bad signal, when it is not a lie.

It follows that the AR elite always sends a bad message after a bad signal, but has two qualitatively different strategies after a good signal: she either randomizes or sends a good message.

We now return to the strategy of the bad AR politician. We have four subcases: whether the bad AR politician does not or does send propaganda, and whether the AR elite after a good signal randomizes or always sends a bad message.

Subcase (1i): The bad AR politician does not send propaganda, and conditional on propaganda, after both signal realizations (good or bad) the AR elite criticizes. Since in this profile the bad AR politician is always criticized, she has even stronger incentives than the bad R politician to send propaganda. Indeed, the latter can sometimes get a good message, which reduces the payoff of propaganda and increases the payoff of no propaganda. Since the bad R politician prefers propaganda, so should the bad AR politician, a contradiction.

Subcase (1ii): The bad AR politician does not send propaganda, and conditional on propaganda, the AR elite mixes after a good signal and sends a bad message after a bad signal. In this subcase, ignoring the infinitesimal lying cost, the AR elite must be indifferent between the two messages, implying that the voters' mean beliefs after propaganda and a good message must be the same as after propaganda and a bad message. But then any politician type has the same payoff from propaganda: they may face a different distribution of elite messages, but mean beliefs after propaganda and any elite message at the same. Moreover, not sending propaganda is worse for the bad AR politician than for the bad R politician, since the latter sometimes gets a good message. Thus, propaganda should generate a strictly higher payoff gain for the bad AR politician than for the bad R politician, and since the latter prefers it, so should the former. This is a contradiction.

Subcase (2i): Both the good and the bad AR politician sends propaganda, and conditional on propaganda, after both signal realizations (good or bad) the AR elite criticizes. This is the structure of our preferred equilibrium, and the existence proof shows that given $\alpha < \alpha(\pi)$ this is an equilibrium.

Subcase (2ii): Both the good and the bad AR politician sends propaganda, and conditional on propaganda, the AR elite mixes after a good signal and sends a bad message after a bad signal. In this candidate equilibrium, relative to our preferred equilibrium, propaganda and a bad message are worse while propaganda and a good message are better for the politician. Indeed, in this candidate equilibrium propaganda and a bad message are stronger evidence that the politician is bad (because they arise with a lower probability when the AR politician is good) while propaganda and a good message are weaker evidence that the politician is bad (because they arise with a higher probability when the AR politician is good). Since $\alpha < \alpha(\pi)$ ensures that the AR elite prefers to criticize in our preferred equilibrium, it follows that she will strictly prefer to criticize in this candidate equilibrium, a contradiction.

6. Equilibrium selection for $\alpha > \bar{\alpha}(\pi)$

Since in the proof for $\alpha < \alpha(\pi)$ we used that $\alpha < \alpha(\pi)$ only in subcases (2i) and (2ii), the previous steps continue to hold. It follows that in any equilibrium meeting our selection criteria, both the good and the bad AR politicians send propaganda, the elite after a bad signal sends a bad message, and the elite after a good signal either mixes or sends a bad message. Since $\alpha > \alpha(\pi)$, the elite after a good signal cannot be sending a bad message. Thus, she must be mixing. The existence proof characterizes the unique mixing probability that makes this profile an equilibrium.

A.5 Proofs of Corollaries

Proof of Corollary 1. Under Assumptions 1 and 2, Proposition 1' implies that for $\pi > \bar{\pi}$ the bad R politician strictly prefers to send propaganda, which implies that

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

and hence that the left-hand-side is positive.

Proof of Corollary 2. Under Assumptions 1 and 2, Proposition 1' implies that depending on the value of α the unique PPO equilibrium takes either the simple or the complex propaganda form. In the simple propaganda equilibrium, after a history of propaganda, the beliefs of the receptive voter are given by equation (5), which immediately implies that a bad message improves the perception of the politician among receptive voters. In the complex propaganda equilibrium, the AR elite, after observing a good signal, is indifferent between reporting good and reporting bad. Since sending a bad message (relative to a good message) harms the politician's perceived competence among non-receptive voters, to make the elite indifferent, that bad message must improve the politician's perceived competence among receptive voters.

Proof of Corollary 3. Under Assumptions 1 and 2, by Proposition 1', the unique PPO equilibrium takes either the simple or the complex propaganda form. In either equilibrium, the receptive voter's posterior about the AR, after updating from propaganda, but before updating from the elite's message, is

$$\mu_{rec}(AR|\hat{p} = 1, \theta_m = P) = \frac{q_{ar}}{1 - q_r q_c}.$$

This follows because the receptive voter has a new prior q_{ar}, and from this prior and the observation of propaganda, he infers that the state (θ_r, θ_c) is either (R,bad), or (AR, bad), or (AR, good). The unconditional joint probability of the AR states is q_{ar}, but the total probability of these three states is just $1 - q_r q_c$. Observe that this expression is larger than q_{ar}.

Now consider the voter's posterior after observing the elite's message as well. In a normal Bayesian setting, the expected value of that posterior would equal the belief we just computed.

That is not the case here, because the voter misunderstands the distribution of the elite's signals. Nevertheless, for π approaching one the expected posterior will equal the above expression. This is because for π approaching one, the voter expects that message to be almost always negative (since even in the complex propaganda equilibrium r approaches zero) and hence his posterior after a bad message will be close to his posterior after propaganda. Moreover, it is also the case in the objective reality that the elite's message is almost always negative, implying that the objective expected posterior of the receptive voter will also be close to his post-propaganda posterior.

Proof of Corollary 4. The proof is organized in numbered steps.

1. Voter beliefs in the no propaganda profile

We say that a strategy profile has the *no propaganda form* if no politician type sends propaganda and all elite types report truthfully. We start by computing voter beliefs in this profile. First consider the history of no propaganda. Since no politician type sends propaganda in the equilibrium profile, then neither the attentive nor the inattentive voters update form the absence of propaganda. Therefore, both voters have the same posterior as the inattentive voter in Proposition 1 given by equation (A6). Next, consider the off-equilibrium history of propaganda. The attentive voter attributes propaganda to a tremble, and since he knows that all elite types are trustworthy he forms the same beliefs as the inattentive voter in equation (A6).

2. Equilibrium existence

We establish that the no propaganda profile is an equilibrium using backward induction. The R elite reports truthfully after any history to minimize lying cost. The AR elite will report truthfully after any history too. This is because in the proposed equilibrium voters form beliefs according to equation (A6), increases in $\hat{s}$ for large π, and the AR elite wants to maximize the average voter's posterior after a good signal and minimize it after a bad signal. In stage 1, no politician types chooses to send propaganda, since propaganda is costly but does not change any voter's posterior beliefs about the politician's type.

3. Equilibrium selection

Here we prove that the no propaganda equilibrium is the unique PPO equilibrium. The good

R politician does not send propaganda in any equilibrium, since propaganda is costly and, as π approaches one, her good type is almost completely revealed by the elite message.

The good AR politician does not send propaganda either. To see why, suppose she does, and consider a history without propaganda in the AR. Since there was no propaganda, the voter remains normal and follows the elite's signal.[24] Given this, the AR elite, who prefers to keep the good AR politician, will send a good message after a good signal. Thus, the AR politician will get a payoff near her first best for π close to one. As a result, she does not engage in costly propaganda.

Finally, since no good politician type uses propaganda, no bad type uses it either to avoid revealing their type.

Proof of Corollary 5. We focus on the $\alpha < 0.5$ case in which we have the simple propaganda equilibrium. Begin with claim 1. In this equilibrium, the persuaded voter's posterior after observing propaganda and criticism is

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

This follows because in the AR the voter expects propaganda and criticism explaining the numerator; and in addition in the R he expects it if the politician is bad and the elite's signal is correct, explaining the denominator. This expression is clearly increasing in q_c.

Now consider claim 2. The difference between the unreceptive voter's beliefs with versus without propaganda is zero. To express the difference between the receptive voter's beliefs with versus without propaganda, note that his expected belief, *absent propaganda*, when the politician is bad, is

$$E[\mu_v(\theta_c=1|\hat{p}=0,\theta_m=N)|\theta_c=0]=\pi\frac{(1-\pi)q_c}{(1-\pi)q_c+\pi(1-q_c)}+(1-\pi)\frac{\pi q_c}{\pi q_c+(1-\pi)(1-q_c)}.$$

The receptive voter's expected belief, *with propaganda*, when the politician is bad, is

$$E[\mu_v(\theta_c=1|\hat{p}=1,\theta_m=P)|\theta_c=0]=\pi\frac{q_{ar}q_c}{q_{ar}+q_r\pi(1-q_c)}.$$

[24] We can compute the receptive voter's belief explicitly. Denoting by y whether in the candidate profile the bad R politician sends propaganda, it has the following form, which is increasing in $\hat{s}$ for π large

$$\mu_{rec}(\theta_c=1|\hat{p}=0,\hat{s},\theta_m=N)=\hat{s}\frac{\pi q_c}{\pi q_c+(1-\pi)(1-q_c)(1-y+y\beta)}+(1-\hat{s})\frac{(1-\pi)q_c}{(1-\pi)q_c+\pi(1-q_c)(1-y+y\beta)}.$$

It is straightforward to verify that for π large enough the difference is increasing in q_c. For a more direct argument, note that for π converging to 1 the difference converges to $\hat{q}_c = q_c \cdot \hat{q}_{ar}$, which is strictly increasing in q_c because both terms are increasing in q_c. Since all of the functions here are complex analytic, convergence of the function implies convergence of the derivative, implying that for π large the difference is increasing in q_c.

A.6 Microfundation of demand for alternative reality

Model setup. We present a simple model in which the receptive voter's propaganda-induced prior misbelief is endogenized. This model extends the probabilistic voting model presented in Appendix A.2. We assume that the receptive voter can only entertain the alternative reality proposed to him by the politician through propaganda, but—in the spirit of the idea of motivated beliefs—he can decide whether and to what extent to believe in it. More specifically, we assume that the receptive voter after observing propaganda, in stage 1, chooses how much prior belief $q_{ar} \in [0, 1]$ to put on the AR presented by the politician. Receptive voter i's objective function at this stage is

$$V_{rec,i} = \tilde{E}_{q_{ar}}[U_{rec,i}|\hat{p} = 1] - E[C(\mu_{rec,i}(AR|\hat{p}, \hat{s}, q_{ar}))|\hat{p} = 1]. \tag{A12}$$

We use the notation that $\tilde{E}_{q_{ar}}[.]$ computes the receptive voter's subjective expectation given his choice of prior belief q_{ar}. while $E[.]$ computes the objectively correct expectation. The first term is the receptive voter's subjective expectation of his utility defined by equation (A1). The second term represents the cost of holding incorrect posterior beliefs about the nature of reality in terms of subsequent outcomes. As common in the literature on motivated beliefs, this term is computed using the objectively correct expectations (Brunnermeier and Parker 2005). We condition on $\hat{p} = 1$ in both terms because we assume that the receptive voter chooses beliefs only when he receives propaganda, so that the expectations are taken over the realization of $\hat{s}$. As in Levy (2014), the cost is modeled in a reduced-form fashion; here it is a function of the voter's subjective posterior belief in the AR, $\mu_{rec,i}(AR|\hat{p}, \hat{s}, q_{ar})$, which depends on the voter's choice of prior q_{ar}. Assuming that the cost is a function of beliefs in the AR reflects that the cost is the result of taking bad personal decisions, such as not taking up vaccinations. The cost does not depend on beliefs about the politician's quality: voter i understands that being infinitesimal he does not have an impact on the election outcome. This formulation is similar to Brunnermeier and Parker (2005) in that voters choose their optimal beliefs balancing between the benefit of optimism and the cost of worse decision making, but differs in that—for simplicity—we do not model the latter explicitly. We assume that $C'(\cdot)$ is convex, $C'(0) = 0$, and $\lim_{x\to 1} C'(x) = \infty$.

Analysis. We assume that the conditions stated in Proposition 1 hold, and we will study the equilibrium identified in that Proposition when q_{ar} is endogenously chosen. We leave the question of whether other equilibria emerge for future work. Thus, in the derivations that follow we assume that strategies are as specified in our preferred equilibrium, and we will later confirm that those strategies continue to constitute an equilibrium.

Substituting in from (A1), the receptive voter's utility can be written as

$$U_{rec,i} = P(\hat{s}, \hat{p}) \cdot c \cdot \mu_{rec,i}(\theta_c = 1|\hat{p}, \hat{s}, q_{ar}) + (1 - P(\hat{s}, \hat{p}))c \cdot q_c^c + P(\hat{s}, \hat{p}) \cdot \lambda. \tag{A13}$$

Here $P(\hat{s}, \hat{p})$ is the probability that the incumbent wins the election, defined by equation (A4), except that here we made explicit its dependence on $\hat{p}$ and $\hat{s}$. Note that $P(\hat{s}, \hat{p})$ is exogenous from the individual voter's perspective, because it is determined by other voters' beliefs about the AR. Since q_c^c denotes the probability that the challenger is good, the first two terms measure the subjective expected value of the politician being good. The last term measures the subjective expected value of the politician being ideologically pro-voter.

We now turn to compute the subjective expected utility of voter i, that is, the subjective expected value of (A13). This requires some preliminaries. First, we note that although the maximization problem of voter i is with respect to q_{ar}, we will find it convenient to treat it as a maximization problem with respect to $\hat{q}_{ar} = q_{ar}/(q_{ar} + q_r\pi(1 - q_c))$ which is the receptive voter's posterior belief in the AR after propaganda and criticism. This is an equivalent reformulation because q_{ar} is a strictly monotone transformation of $\hat{q}_{ar}$. A consequence of this approach is that we will express terms of interest as functions of $\hat{q}_{ar}$.

Second, because the last two terms in (A13) will not contribute to the economics of the results, we introduce the notation

$$(1 - P(\hat{s}, 1))c \cdot q_c^c + P(\hat{s}, 1) \cdot \lambda = k(\hat{s}).$$

Third, to compute the subjective expected value of (A13), note that $\pi + q_{ar}(1-\pi)$ is voter i's subjective probability of observing elite criticism conditional on propaganda. We introduce the notation

$$\rho(q_{ar}) = \frac{\pi + q_{ar}(1-\pi)}{\pi}$$

so that the subjective probability of observing criticism is $\pi\rho(q_{ar})$. We note for future reference that (i) with a slight abuse of notation we will often treat ρ as a function of $\hat{q}_{ar}$, which is valid since q_{ar} is a strictly monotonic transformation of $\hat{q}_{ar}$; (ii) as π converges to one, $\rho(\hat{q}_{ar})$ converges to one uniformly in $\hat{q}_{ar}$; (iii) $\rho(\hat{q}_{ar})$ is a ratio of polynomials in $\hat{q}_{ar}$ and hence is (more precisely, can be extended into) a complex analytic function of $\hat{q}_{ar}$, which implies that as π goes to one, all derivatives of ρ with respect to $\hat{q}_{ar}$ converge to zero uniformly.

With these preliminaries, we can write the subjective expected value of (A13) as

$$\tilde{E}_{q_{ar}}[U_{rec,i}|\hat{p}=1] = \pi\rho(\hat{q}_{ar}) \cdot P(0,1) \cdot c \cdot \hat{q}_{ar} \cdot q_c + \pi\rho(\hat{q}_{ar}) \cdot [k(0) - k(1)] + k(1). \tag{A14}$$

Substituting back into the receptive voter's objective (A12) and noting that the cost is a function of the posterior AR belief $\hat{q}_{ar}$ yields

$$V_{rec,i} = \pi\rho(\hat{q}_{ar}) \cdot P(0,1) \cdot c \cdot \hat{q}_{ar} \cdot q_c + \pi\rho(\hat{q}_{ar}) \cdot [k(0) - k(1)] + k(1) - \pi C(\hat{q}_{ar}).$$

We maximize this with respect to $\hat{q}_{ar}$ by taking the first order condition, which yields

$$P(0,1)cq_c \cdot \rho(\hat{q}_{ar}) + \rho'(\hat{q}_{ar}) \cdot [P(0,1)cq_c\hat{q}_{ar} + k(0) - k(1)] = C'(\hat{q}_{ar}). \tag{A15}$$

This condition characterizes the equilibrium $\hat{q}_{ar}$. There are two important points to note. First, as mentioned above, when π approaches one ρ converges to one and ρ' converges to zero, so that the first order condition converges to the much simpler form $P(0,1)cq_c = C'(\hat{q}_{ar})$. With that "approximate first-order condition" all the remaining analysis would follow easily. Much of the work below is showing that the results also obtain just before the limit. The second point to note is that the $P(0,1)$ on the left-hand-side also depends on $\hat{q}_{ar}$ in equilibrium (even if it was exogenous to voter i), because we know from equation (A4) that $P(0,1)$ is an increasing linear function of receptive voters' average posterior about the politician's type, that is $\mu_{rec}(\theta_c = 1|\hat{s} = 0, \hat{p} = 1) = \hat{q}_{ar}q_c$.

We now analyze this first-order condition, first under the (false) assumption that $\rho \equiv 1$ (which implies $\rho' \equiv 0$), and then properly. If $\rho \equiv 1$ were true, then we could directly trace the two sides of the approximate first-order-condition $P(0,1)cq_c = C'(\hat{q}_{ar})$ as a function of $\hat{q}_{ar}$. For $\hat{q}_{ar} = 0$, the left-hand-side is positive given the definition of $P(0,1)$, while the right-hand-size is zero by assumption. As $\hat{q}_{ar}$ increases, the left-hand-side traces out an increasing linear function, while the right-hand-side an increasing convex function which asymptotes to infinity. Thus, there is a unique point of equilibrium.

Relaxing the false assumption, but taking π large so that the deviations from the approximate first-order condition are small, it is still the case that the left-hand-side starts from a positive value while the right-hand-side starts from zero. Moreover, given the properties of ρ highlighted above, the left-hand side remains arbitrarily close to a increasing linear function, and its derivative remains arbitrarily close to the positive constant slope of that function. The right-hand-side is still a smooth convex function, thus there is at least one point of intersection. Since the intersection requires that the right-hand-side "catches up" to the left-hand-side, in its neighborhood the slope of the right-hand-side must be strictly higher than the constant slope of $P(0,1)cq_c$. Thus, for π sufficiently large, the slope of the right-hand-side will be strictly higher than the slope of the left-hand-side (which is arbitrarily close to the aforementioned constant). It follows that there cannot be a second intersection. We conclude that for π large there is a unique q_{ar}. Moreover, the arguments also imply that as π converges to 1, that q_{ar} converges to the solution of the approximate first-order condition $P(0,1)cq_c = C'(q_{ar})$.

Assumption 3. Assumption 2 holds with the unique solution q^*_{ar} of $P(0,1)cq_c = C'(\hat{q}_{ar})$.

Proposition 4. *Suppose that Assumptions 1 and 3 hold and $\alpha < 0.5$. For π sufficiently large, the equilibrium of Proposition 1 remains an equilibrium with a unique endogenously chosen q_{ar}. Moreover, q_{ar} is increasing in the voter's preference for an incumbent government λ and in the voter's prior probability of a good politician q_c.*

Proof of Proposition 4. Consider the proposed equilibrium profile. In that profile, for π large, the unique optimal q_{ar} will satisfy Assumption 2. As a result, Proposition 1 shows that the profile is an equilibrium.

To establish the comparative statics, we need two preliminary steps. First, (A4) implies that the probability the incumbent politician remains in power, conditional on propaganda and criticism, is

$$P(0,1) = q \cdot c \cdot \left[\alpha \hat{q}_{ar} q_c + (1-\alpha) \frac{(1-\pi) q_c}{(1-\pi) q_c + \pi(1-q_c)} \right] + g(\lambda - c \cdot q_c^c) + 0.5,$$

where $\hat{q}_{ar} q_c$ is the receptive and $(1-\pi)q_c/[(1-\pi)q_c + \pi(1-q_c)]$ is the non-receptive voter's posterior belief. Second, if we rearrange equation (A15) and define

$$F \equiv \rho(\hat{q}_{ar}) \cdot P(0,1) c q_c + \rho'(\hat{q}_{ar}) \cdot [P(0,1) c q_c \hat{q}_{ar} + k(0) - k(1)] - C'(\hat{q}_{ar})$$

then

$$\frac{\partial F}{\partial \hat{q}_{ar}} = \rho'(\hat{q}_{ar}) \cdot P(0,1) c (1 + q_c) + \rho''(\hat{q}_{ar}) \cdot [P(0,1) c q_c \hat{q}_{ar} + k(0) - k(1)] - C''(\hat{q}_a r),$$

and because when π approaches one both $\rho'(\hat{q}_{ar})$ and $\rho''(\hat{q}_{ar})$ converge uniformly to zero, while $C''(\hat{q}_{ar})$ is by definition positive, we have that for π large $\partial F / \partial \hat{q}_{ar} < 0$.

Given these preliminaries, we can apply the Implicit Function Theorem to obtain

$$\begin{aligned} \frac{\partial \hat{q}_{ar}}{\partial \lambda} = -\frac{\partial F / \partial \lambda}{\partial F / \partial \hat{q}_{ar}} &= \frac{\rho(\hat{q}_{ar}) \frac{\partial P(0,1)}{\partial \lambda} \cdot c q_c + \rho'(\hat{q}_{ar}) \cdot \frac{\partial}{\partial \lambda} [P(0,1) \cdot c q_c \hat{q}_{ar} + k(0) - k(1)]}{-\partial F / \partial \hat{q}_{ar}} \\ &\xrightarrow[\pi \to 1]{\text{unif.}} \frac{\frac{\partial P(0,1)}{\partial \lambda} \cdot c q_c}{C''(\hat{q}_{ar})} = \frac{g \cdot c q_c}{C''(\hat{q}_{ar})} > 0 \end{aligned}$$

which implies that for π large enough $\hat{q}_{ar}$ is increasing in λ. And then q_{ar} is also increasing in λ because q_{ar} is an increasing transformation of $\hat{q}_{ar}$.

The intuition for the result is that λ increases the probability $P(0,1)$ that the incumbent remains in power. Intuitively, the voter, who enjoys being optimistic, want to protect his positive belief about the politician who is likely to win the election.

The second comparative static also follows from the implicit function theorem:

$$\begin{aligned} \frac{\partial \hat{q}_{ar}}{\partial q_c} = -\frac{\partial F / \partial q_c}{\partial F / \partial \hat{q}_{ar}} &= \frac{\rho(\hat{q}_{ar}) \frac{\partial}{\partial q_c} [P(0,1) c q_c] + \rho'(\hat{q}_{ar}) \frac{\partial}{\partial q_c} [P(0,1) c q_c \hat{q}_{ar} + k(0) - k(1)]}{-\partial F / \partial \hat{q}_{ar}} \\ &\xrightarrow[\pi \to 1]{\text{unif.}} \frac{\frac{\partial}{\partial q_c} [P(0,1) c q_c]}{C''(\hat{q}_{ar})} = \frac{c[g \cdot c \cdot \alpha \hat{q}_{ar} q_c + P(0,1)]}{C''(\hat{q}_{ar})} > 0 \end{aligned}$$

which proves the result. The intuition here operates through two channels. First, q_c directly increases the benefit of believing in the alternative reality, since the AR allows the voter to maintain the pleasurable prior belief q_c that the politician is good. Second, q_c increases the incumbent's probability of reelection; and the voter prefer to maintain a favorable opinion about the likely winner of the election.

A.7 Proof of Proposition 2

Denote the lying cost AR by AR1 and the conspiracy AR by AR2.

Case 1: $\chi_f < (1-2\alpha)/N$.

Behavior in any equilibrium. We begin by characterizing the behavior of some actors in any large-π equilibrium. Since the R elite's reputation costs are prohibitively large, the R elite is truthful in any profile. Given this, for π large enough, the good R politician does not send propaganda.

Fix an equilibrium and consider a member j of the AR1 elite after some history of propaganda $\hat{p}$. The impact on $\hat{\mu}$ of reporting good rather than bad after a good signal is

$$\frac{(1-\alpha)}{N} \cdot [\mu_{un,i(j)}(\hat{s}_j = 1) - \mu_{un,i(j)}(\hat{s}_j = 0)] + \frac{\alpha}{N} \cdot [\mu_{rec,i(j)}(\hat{p}, \hat{s}_j = 1) - \mu_{un,i(j)}(\hat{p}, \hat{s}_j = 0)].$$

In the limit as π approaches one the elite signal becomes perfectly informative and the first term approaches $(1-\alpha)/N$. The second term, since beliefs are always between zero and one, is always bounded from below by $-\alpha/N$. Thus, as long as

$$\frac{1-\alpha}{N} - \frac{\alpha}{N} > \chi_f$$

holds, for π large enough elite member j—who cares about reducing $\bar{\mu}$ but has a cost χ_f from lying—will report bad after a good signal. Since we are in Case 1, this condition holds. Thus, the AR1 elite always criticizes after a good signal. Since after a bad signal the gain from criticism is the same and the cost of criticism (relative to praise) becomes $-\chi_f$, the AR1 elite always criticizes after a bad signal as well.

Consider the AR2 elite. Since $N > 1$, we have $1 - 2\alpha > \chi_f$, and an analogous argument shows that the AR2 elite (as π approaches 1), when reporting bad rather than good after a good signal, gains $1-\alpha$ from unreceptive voters but loses at most α from receptive voters. Thus, the AR2 elite always criticizes after a good signal; and then it always criticizes after a bad signal as well.

Existence of candidate equilibrium. We now show that the following strategy profile is an equilibrium: the R elite is truthful; the good R politician does not send any propaganda; the bad R politician sends AR1; both AR politicians send AR1; and the elite in both ARs always criticizes. We have already established that the R elite is truthful, that the good R politician does not send propaganda, and that the elite in both ARs criticizes. It remains to characterize the behavior of the bad R politician and the AR politicians.

To do this, note that in the proposed equilibrium the belief of the voter who observed no propaganda continues to be given by (A7), while the belief of the voter who observed AR1 is

$$\mu_v(\theta_c|\hat{s}, \hat{p} = AR1) = (1 - \hat{s})\frac{q_{ar}q_c}{q_{ar} + q_r(1 - q_c)\pi} \tag{A16}$$

This expression is derived analogously to our basic model. Propaganda and praise ($\hat{s} = 1$) conclusively prove that the politician is bad. For propaganda and criticism, the numerator reflects that in the AR a good politician always sends propaganda and gets criticism, while the denominator reflects that propaganda and criticism can also arise in the R if the politician is bad.

The belief of the voter who observed AR2 is

$$\mu_v(\theta_c|\hat{s}, \hat{p} = AR2) = \hat{s}\frac{q_c\pi}{q_c\pi + (1 - q_c)(1 - \pi)} + (1 - \hat{s})\frac{q_{ar}q_c + q_r(1 - \pi)q_c}{q_{ar} + q_r[(1 - \pi)q_c + (1 - q_c)\pi]}. \tag{A17}$$

The first term represents beliefs after observing AR2 propaganda and praise by the elite. This term is no longer zero because the outcome is attributed to a tremble. More precisely, propaganda shifts the prior to put a positive weight on AR2, but, because AR2-propaganda is not observed on the equilibrium path, it is attributed to a tremble and does not generate updating. Since praise never occurs in AR2, given praise the voter updates that reality is R, thinks that the AR2 propaganda was a tremble, and forms beliefs based on the signal only. The second term represents beliefs after observing AR2 propaganda and criticism from the elite. As in the first term, the voter puts a positive weight on the AR2, but since AR2 propaganda never happens on the equilibrium path, it is attributed to a tremble and does not generate updating. Therefore, the numerator reflects that in the AR2 a q_c share of politicians are good and in the R a good politician is only criticized if the elite receives a bad signal (which happens with probability $1 - \pi$). The denominator reflects that the elite always sends a bad message in AR2, while in reality she criticizes the incumbent if the politician is good but she received an incorrect signal or if the politician is bad and she received a correct signal.

Similarly to the basic model, (A16) implies that on the proposed equilibrium path, as π converges to one, the R politician's return to successful AR1 propaganda is governed by $\hat{q}_c$. Hence, by Assumption 2, for the bad R politician AR1 propaganda is better than no propaganda. Moreover, AR1 propaganda is better than AR2 propaganda because in the limit as π goes to one, (A16) and (A17) imply that the return to AR1 is the same as that to AR2, but AR1 has a lower cost. The same logic implies that the AR politicians—in both AR1 and AR2—choose to send AR1 propaganda. This confirms that the proposed profile is an equilibrium.

Equilibrium selection. We show that for π large the proposed equilibrium is the unique PPO equilibrium. Recall that PPO implies that the politician uses pure strategies. We already characterized the behavior in any equilibrium of the R and AR elites and the good R politician. Our preferred equilibrium is better than any equilibrium in which the bad R politician refrains from propaganda, because here she strictly prefers to send AR1 propaganda and thus doing so improves her payoff. Thus, in any PPO equilibrium, the bad R politician must send either AR1 or AR2 propaganda. We consider these cases in turn.

[Bad R politician sends AR1 propaganda.] Then the AR1 politician must also send AR1 propaganda, because otherwise observing AR1 would lead the persuaded voter, who now has a positive prior on R and AR1 (but not on AR2) to conclude that reality is R, which cannot be profitable for the bad R politician. This already shows that the equilibrium path is the same as in our preferred equilibrium. We now show that for π large enough the equilibrium is also the same. It is not optimal for the AR2 politician to send no propaganda, since the R politician, who gets criticized less often, sends propaganda. Suppose that the AR2 politician sends AR2 propaganda. Then the persuaded voter's beliefs after AR2 propaganda are that reality is AR2 and the politician is good with probability q_c. Deviating to AR1 propaganda would instead generate beliefs that are identical to those that emerge after the AR1 politician sends AR1 propaganda, as given by (A16). Thus, except for the knife-edge case of indifference, which can only happen for one value of $\pi < 1$ given the strict monotonicity of (A16) in π, if the AR2 politician prefers to send AR2 propaganda, then so does the AR1 politician, a contradiction. It follows that for π large enough in any PPO equilibrium the AR2 politician sends AR1 propaganda. This is our preferred equilibrium.

[Bad R politician sends AR2 propaganda.] Then the AR2 politician must also send AR2 propaganda. Consider the AR1 politician. No propaganda cannot be optimal for her, since the R politician, who gets criticized less often, sends AR2 propaganda. If she sends AR1 propaganda, the voter will conclude that reality is AR1 and she is good with probability q_c. This is better than AR2 propaganda, which is more expensive and leads to worse beliefs, so she sends AR1. Given this, the AR2 politician also prefers to send AR1, a contradiction.

Case 2: $1/N < \chi_f < (1 - 2\alpha)$.

Behavior in any equilibrium. We begin by characterizing the behavior of some actors in any equilibrium. As in Case 1, the assumption that χ_h is prohibitively large implies that the R elite is truthful. Therefore, for π large the good R politician does not send propaganda. For π large the AR1 elite is also truthful. This is because, in the limit as π goes to one, the maximal gain from changing the perception of her audience is $1/N$, which, since we are in Case 2, is smaller than her lying cost of χ_f. However, for π large the AR2 elite always sends a bad message after a good signal, because doing so generates a gain of $1 - \alpha$ in the limit from unreceptive voters, and a loss of at most α from persuaded voters, and in Case 2 we have that $1 - 2\alpha > \chi_f$.

Existence of candidate equilibrium. We now show that the following strategy profile is an equilibrium. The R and the AR1 elite are truthful; the AR2 elite always criticizes; the good R politician does not send any propaganda; the bad R politician sends AR2; both AR politicians send AR2. Given the results above, we only need to verify the optimality of the behavior of the bad R and the AR politicians.

Observe that no politician sends AR1 propaganda. This follows from the fact that the AR1 elite is truthful, which implies that AR1 propaganda has no effect on the voter's interpretation of the elite's message, while having a positive cost. However, sending AR2 propaganda is optimal for the bad R politician, for the same reason that propaganda is optimal in the basic model. Indeed, since AR1 is off the table, the setup is identical to that of the basic model, and by Assumption 2, for π sufficiently high the benefit of propaganda exceeds the cost. The same logic implies that sending AR2 propaganda is optimal for the AR1 and the AR2 politician.

Equilibrium selection. In any equilibrium weakly better for the bad R politician that the one proposed here, she has to send AR2 propaganda. This is because the proposed equilibrium yields a higher payoff than that of not sending propaganda, and sending AR1 propaganda—as established in the previous paragraph—is not useful given that the AR1 elite is truthful. Since the bad R politician is sending AR2, the AR2 politician must also be sending AR2, otherwise the voter learns from observing AR2 (and having a positive prior on R and AR2) that reality must be R. Finally, the AR1 politician must also prefer to send AR2 propaganda, since doing so is more attractive than sending no propaganda, and sending AR1, as established above, is even worse than sending no propaganda.

Case 3: $1 < \chi_f$.

We prove that in the unique equilibrium the elites in all realities are always truthful and the politicians never send propaganda. As before, the R elite is truthful. The assumption that $1 < \chi_f$ implies that the gain to any AR elite from fully influencing the entire electorate is smaller than the fabrication cost. It follows that telling the truth is optimal for them as well. Since neither propaganda changes the interpretation of the elite's message, no politician chooses propaganda.

A.8 Proof of Proposition 3

Key to the proof is that for π large, both when $e = 0$ and when $e = 1$, the elite's signal is almost perfectly informative. As a result, the large-π arguments used in the proof of the main result also apply here.

Behavior in any equilibrium. We begin by characterizing the behavior of some actors in any large-π equilibrium. Begin with the elite. As in the basic model, since its members have no impact on the outcome, the R elite is always truthful. Consider the AR elite. In the absence of propaganda they always send a bad message. In the presence of propaganda, the gain from sending a bad rather than a good message, as π approaches one, approaches $1-\alpha$, because the $1-\alpha$ share of unreceptive types believe (for π large) that the elite's message is almost perfectly informative. The loss from sending a bad rather than a good message is at most α because in the worst case the share α of receptive voters react in the exact opposite way to her message. Since $\alpha < 0.5$, for π large enough the AR elite always sends a bad message.

Now consider the good R politician. For π large enough, she earns close to the maximal payoff absent propaganda, and hence refrains from costly propaganda.

Existence. We turn to establish that the proposed profile constitutes an equilibrium. Given the above results, to prove existence, we only need to focus on the bad R politician and the good and bad AR politicians. First consider their decisions about propaganda. For π large, the bad R politician, and the good and bad AR politician all prefer to send propaganda by Assumption 2. This is for the same logic as in the main result. Absent propaganda the elite (i) almost certainly sends a bad message (both when $e = 0$ and when $e = 1$), and (ii) is perceived by all voters to be almost fully informative. Hence expected average beliefs about competence become approximately zero. In the presence of propaganda, because the elite almost always sends a bad message, the expected weighted average belief μ' approximates $\alpha' \hat{q}_c$, since receptive voters' belief approximates $\hat{q}_c$ (for the same reason as in our main setting) while unreceptive voters' beliefs approximate zero. Since $\alpha' > \alpha$, the result follows from Assumption 2.

Now consider the bad politician' decision about e. The bad AR politician, since she expects to be criticized no matter what she does, is indifferent between more or less precise elite signals and chooses $e = 0$. The bad R politician who can not send propaganda (which happens with a probability β) expects, for π large, that voter beliefs will be close to zero after a bad elite message and close to one after a good elite message. Thus, she would like to minimize the probability of a bad elite message and chooses $e = 0$.

Finally, consider the bad R politician who can send propaganda. At this step we need to explicitly calculate voters' beliefs after propaganda. In the proposed equilibrium the politician chooses $e = 1$, making the elite's signal correct with probability π'. Therefore the belief of the receptive voter after propaganda, as a function of the elite's message, is

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

As in the basic model, when the elite praises the politician ($\hat{s} = 1$), posterior beliefs are that the politician is bad. When the elite criticizes, posterior beliefs are a function of the probability of criticism when the politician is good, which can only happen in the AR ($q_{ar}q_c$), relative to the probability of criticism, which always happens in the AR (q_{ar}) and happens in R for the bad politician if the elite's message is bad ($q_r(1-q_c)\pi'$). Note that the last term accounts for the fact that the bad R politician chooses a more precise signal.

To compute the belief of the unreceptive voter about the politician, we introduce $\hat{\pi} = \beta\pi + (1-\beta)\pi'$, which is the unreceptive voter's belief about the precision of the elite's signal. This holds because in the proposed path the bad R politician sets $e = 1$, implying precision π', precisely in the β probability event in which she can send propaganda. The beliefs of the unreceptive voter after propaganda are given by

$$\mu_{un}(\theta_c = 1|\hat{p} = 1, \hat{s}) = \hat{s}\frac{\pi q_c}{\pi q_c + (1-\hat{\pi})(1-q_c)} + (1-\hat{s})\frac{(1-\pi)q_c}{(1-\pi)q_c + \hat{\pi}(1-q_c)}.$$

The first term says that when observing a good signal, posterior beliefs are governed by the probability of that good signal under a good politician, πq_c, relative to the probability of a good signal under a good or a bad politician $\pi q_c + (1-\hat{\pi})(1-q_c)$, where the $\hat{\pi}$ reflects the probability of a correct signal under a bad politician. The intuition for the second term, which expresses posterior beliefs after a bad signal, is similar.

The condition that the bad R politician prefers $e = 1$ is

$$\begin{aligned} &\alpha' \cdot [\mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0) - \mu_{rec}(\theta_c = 1|\hat{p} = 1, \hat{s} = 1)] \\ &+(1-\alpha') \cdot [\mu_{un}(\theta_c = 1|\hat{p} = 1, \hat{s} = 0) - \mu_{un}(\theta_c = 1|\hat{p} = 1, \hat{s} = 1)] > 0. \end{aligned} \tag{A18}$$

Indeed, the left-hand-side is a weighted average of the belief changes of the unreceptive and receptive voter in response to improving the precision of the signal, which here means that a positive signal is turned into a negative signal. The weights are those that the politician assigns to the two classes of voters. Substituting in the above expressions for the beliefs, we obtain

$$\alpha'\frac{q_{ar}q_c}{q_{ar} + q_r\pi'(1-q_c)} + (1-\alpha')\left[\frac{(1-\pi)q_c}{(1-\pi)q_c + \hat{\pi}(1-q_c)} - \frac{\pi q_c}{\pi q_c + (1-\hat{\pi})(1-q_c)}\right] > 0.$$

It is straightforward to check that as π and π' approach one, the condition collapses to $\alpha' > 1/(1+\hat{q}_c)$. Thus, for any such α', we can find π large enough that the result holds.

Equilibrium selection. We show that for π large the proposed equilibrium is the unique PPO equilibrium. We already characterized in any equilibrium the behavior of the R and AR elites and the good R politician. Our preferred equilibrium is better than any equilibrium in which the bad R politician refrains from propaganda, because here she strictly prefers to send propaganda and thus doing so improves her payoff. Thus, in any PPO equilibrium, the bad R politician must send propaganda. But then the good AR politician must also send propaganda, since otherwise propaganda would reveal that the politician is bad, in which case it would not be worth it for the bad R politician. At this step we used the fact that we are looking for a PPO equilibrium, so that the good AR politician is not mixing. And then the bad AR politician must also send propaganda, since she faces a worse portfolio of elite messages (always criticism) than the bad R politician (often criticism). Thus, the propaganda decisions are uniquely pinned down.

We now turn to the policy decision. The AR politician, since she is always criticized anyway, chooses $e = 0$. The bad R politician who cannot send propaganda, since she would like to minimize the probability that the elite sends a bad message, chooses $e = 0$. Finally, consider the bad R politician who can send propaganda. Consider a candidate equilibrium in which this politician sets $e = 0$. Then the equilibrium path, including actions and beliefs, is exactly identical to the simple propaganda equilibrium of the basic model. Thus, we can evaluate the condition that the bad R politician prefers $e = 1$ by substituting in the beliefs from (A6) and (5) into (A18). It is straightforward to check that as π approaches one, the condition approaches $\alpha' > 1/(1 + \hat{q}_c)$, which holds by assumption. Thus, setting $e = 1$ is optimal, a contradiction. The only remaining case is our preferred equilibrium.

A.9 Evidence

A possible alternative explanation for the scandal effects documented by Table 5 is that scandals increase donations because they intensify electoral competition. We provide evidence gainst this explanation by exploiting the redistricting of congressional districts before the 2022 midterm elections. We combine data on predicted Democratic vote margins for both the old and the new districts of Republican representatives from FiveThirtyEight with donations data from the Federal

	Trump donors	Trump donors	Other donors
	Share	Amount (1000 dollars)	
Δ predicted Dem margin	0.001	-1.07	1.43
	(0.001)	(1.60)	(3.57)
Old predicted Dem margin	0.001	0.402	5.36***
	(0.0006)	(0.454)	(1.05)
Constant	0.109***	49.7***	346.4***
	(0.017)	(14.1)	(38.2)
Observations	266	296	296

Table A1: Impact of redistricting on contributions from Trump-supporter and other donors

Elections Commission. We estimate

$$y_i = \text{const} + \beta \Delta DVM_i + \gamma DVM_i^{old} + \varepsilon_i, \tag{A19}$$

where y_i measures donations received by candidate i in the quarter of the 2022 midterm elections; DVM_i^{old} is the predicted Democratic vote margin of candidate i in their electoral district in the period 2011-2020; and $\Delta DVM_i = DVM_i^{new} - DVM_i^{old}$ is the change in predicted Democratic vote margin between the new and the old district.

Table A1 reports the results. Column 1 shows that a reduction in the chance of winning—induced by an unfavorable change in the electoral map—has a small and insignificant effect on the Trump-supporter share, while columns 2 and 3 document small impacts on the volume of donations. Thus, a decline in the electoral prospects of Republican house candidates changes neither the volume nor the composition of donations.