The Political Economy of Alternative Realities[*]

Adam Szeidl
Central European University and CEPR

Ferenc Szucs
Stockholm University

January 2, 2024

Abstract

We build a model 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 inverts the effect of the elite's message, so that criticism helps the politician. This force leads to reduced accountability, distrust in experts, and bad policies that invite elite criticism. Alternative realities feature conspiracies which solve a collective action problem, evolve in response to evidence, and create demand for new media that reinforce them.

Keywords: misbeliefs, manipulation, propaganda, populism, adoption of best practices, media

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 for funding.

1 Introduction

Misbeliefs about politics and society are widespread. Contrary to the experts' consensus, 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 and accurately.[1] Such misbeliefs are often organized into a coherent "alternative reality" centered around the sinister motives of powerful actors. For example, 15% of Americans believe, and 79% of Republicans do not reject, that the government and the media are controlled by a cabal of Satan-worshipping pedophiles.[2] As we demonstrate below, salient alternative realities, in contexts including the United States, Hungary, and Israel, share a common narrative: that a conspiracy of the intellectual elite attacks the competence of a politician, because they disagree with the ideology of that politician. Beliefs in such alternative realities are potentially important, but their causes, mechanisms, and implications are not well understood.

In this paper we build a model in which the politician can persuade voters of a coherent but false alternative reality. Our approach builds on prior work about misinformation in politics, including Glaeser (2005), Guriev and Treisman (2020), and Eliaz and Spiegler (2020), and contributes by explicitly modeling the above-outlined alternative reality, in which members of the intellectual elite can conspire to attack a politician to advance their own interests. Beliefs in this alternative reality generate strategic interaction between the (imagined) conspiring elite and the actors in the objective reality. A key implication is that persuading the voter of the alternative reality *inverts* the effect of the elite's message about the politician, so that elite criticism increases, while elite praise reduces voters' support. Through this and related mechanisms, the model generates several new political-economic implications which explain previously unexplained evidence.

In Section 2 we present our model. Our basic framework is a principal-agent model in which the incumbent politician and the intellectual elite are the principals and the median voter is the agent. Both principals can send messages to influence the voter, who then decides whether to keep or replace the politician. The politician has two payoff-relevant types. (i) A "common"

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

[2] See Public Religion Research Institute (2021). Alesina, Miano and Stantcheva (2020) argue that partisan differences in beliefs reflect different perceptions of the objective reality.

type, over which the voter and the elite have the same preference, and along which the politician can be good or bad. Examples include competence or honesty. The voter does not observe this type dimension. (ii) A "divisive" type, capturing ideology, over which the voter and the elite have different preferences, and along which the politician can be pro-voter or pro-elite. Examples include cultural values or economic redistribution. All actors observe this type dimension.

Since the voter does not directly observe the politician's common type (good vs bad), both the elite and the politician send messages to influence his beliefs. The elite sends a message which simply reports whether the politician is likely to be good or bad. At the same time, the politician can also send a message—which we call propaganda—which exogenously, and counterfactually, increases the voter's prior probability of 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 the voter reached by propaganda assigns a non-negligible positive probability to them. Our notion of perfect Bayesian equilibrium requires that the AR types—though they only exist in the voter's mind—act strategically to maximize their own payoffs, creating a coherent alternative reality which engages in strategic interaction with the voter, and through him with all other actors.

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 have no impact on the voter's belief, and hence prefer to report about the politician's common 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 the voter. It follows that if the AR elite sufficiently dislikes the pro-voter politician (because they disagree on the divisive issue), she will always report that politician bad in the common dimension, hoping to influence the voter's opinion and hence the election outcome. 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, the voter who is persuaded by propaganda partially believes this alternative reality and distrusts the criticism of the elite.

We analyze the model in Section 3. We begin by characterizing equilibrium. We show that (1) propaganda is only used if disagreement about the divisive issue is sufficiently large, and (2) in that case, the politician sends propaganda if and only if she is pro-voter on the divisive dimension and bad on the common dimension. Both results hinge on the plausibility of the alternative reality. The intuition for (1) is that the alternative reality in which the elite criticizes because of disagreement about the divisive issue can only be plausible if the disagreement is sufficiently large. The intuition for (2) is that even then, the alternative reality is only plausible if the politician and the elite do disagree, i.e., the politician is pro-voter; and that only the bad politician needs to discredit the elite.

The key mechanism underlying this equilibrium, and the reason propaganda succeeds in discrediting elite criticism, is that propaganda *inverts* the effect of the elite on the voter. In the presence of propaganda, elite praise will reduce, while elite criticism will increase voter beliefs that the politician is good. To see the logic, consider the 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, elite criticism is relatively more consistent with the alternative reality. It follows that observing criticism increases the voter's 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 and drives several of our main implications.

We emphasize three empirical implications of the model. The most fundamental is the just-described inverse effect of elite criticism in the presence of propaganda. This implication resolves what we view to be a key puzzle in contemporary U.S. politics: that during 2023, the growing body of critical evidence against Donald Trump, including four criminal indictments, was accompanied by an *increase* in his popular support (Swan, Igielnik, Goldmacher and Haberman 2023). Our model not only explains this fact, but predicts that the relationship is causal, i.e., that the indictments caused the increase in support. This prediction is in line with survey evidence we review showing that Republicans claimed to increase their support for Trump in response to the indictments. It is also in line with new evidence we present that scandals by Republican House candidates—

plausibly diffused by the news media—caused an increase in the donations they received from Trump-supporting Republicans. These patterns are directly tied to the core mechanism of our model, and to our knowledge are not explained by other models.

A second implication is that propaganda enables bad politicians to remain in power. This prediction is consistent with Guriev and Treisman's (2022) narrative evidence that modern "informational autocrats" use propaganda to stay in power. Existing models, including Guriev and Treisman (2020), obtain this prediction through the mechanism that propaganda sends a positive message about the politician. Our model has a different mechanism, based on propaganda inverting the effect of the elite, which is consistent with the evidence described above. Our model also predicts that positive propaganda, absent discrediting the elite, would not work, as the elite's message would correct beliefs.

A third implication is that propaganda generates spillovers to non-political domains: distrust in science and the non-adoption of expert-recommended best practices. Intuitively, once the voter contemplates an elite conspiracy, he fears that the elite's messages may be driven by its interests even in non-political domains. This prediction, which has not been formalized in prior work, is consistent with Republicans' distrust in science and unwillingness to follow best practices such as social distancing (Allcott, Boxell, Conway, Gentzkow, Thaler and Yang 2020).

In Section 4 we apply the model to three new domains. In each domain, we incorporate novel features of the environment, but do not change the core assumptions concerning alternative realities. First, 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., mask mandates), even if they know that non-adoption is universally harmful. This follows directly from propaganda's inversion effect: because elite praise reduces and elite criticism increases voter support, the politician will prefer policies that invite elite criticism. This prediction highlights a potentially major 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.

Second, we investigate the reason that alternative realities often feature a conspiracy. 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 alternative reality often dominates, because it solves a collective action problem of the elite. Intuitively, each elite member's lie about the politician's quality 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 manufacturing even that evidence. Thus, our model explains the emergence of political conspiracy theories. In addition, our logic predicts that improving the credibility of critical evidence need not correct beliefs, because in response, the politician can "upgrade" the alternative reality. This upgrade from a lying elite to a conspiring elite is socially harmful, because it increases distrust in the elite in other domains.

Third, we study the implications of alternative realities for the media market. It is a salient fact that many non-traditional media outlets, most prominently Fox News, spread alternative realities. This fact is not easily explained by existing theories, which predict that media slant the presentation of facts (Mullainathan and Shleifer 2005, Gentzkow and Shapiro 2006), but not that they present non-truths and alternative realities. Our model provides an explanation based on the idea that the more discredited the elite media, the higher the demand for non-traditional media. Formally, we add a new (pro-voter) media outlet to the model, and show that it can create demand for itself by falsely reporting that the propaganda-spreading politician is good. Doing so strengthens beliefs in the alternative reality, and shifts audiences away from the elite media. Beyond explaining why new media may lie, this framework predicts that new media can further erode trust in experts, consistent with the evidence on the harmful effects of Fox News on social distancing and Covid deaths (Bursztyn, Rao, Roth and Yanagizawa-Drott 2020, Simonov, Sacher, Dubé and Biswas 2020).

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. Foundational contributions include Glaeser (2005) on the supply of hatred and Besley and Prat (2006) on media capture. More recently, Ash, Mukand and Rodrik (2021) model the supply of “worldview politics”, and Guriev and Treisman (2020) model the supply of propaganda that is intended to convince the public of politicians’ competence. 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 formalizing misinformation with an alternative reality which serves to discredit the elite, and with new implications, especially about the inverse effect of elite criticism, the political demand for bad policy, the popularity of conspiracies, and the spillover to new media.

Our modelling approach builds on a theories of learning and interaction under model misspecification, including Berk (1966), Jehiel (2005), and Esponda and Pouzo (2016). Model misspecification has been applied in political economics, to study persuasion by Bénabou, Falk and Tirole (2018), Galperti (2019), Schwartzstein and Sunderam (2021), and Aina (2023); political narratives by Eliaz and Spiegler (2020) and Eliaz, Galperti and Spiegler (2022); political dynamics by Levy, Razin and Young (2022); and trust in media by Gentzkow, Wong and Zhang (2021). Our alternative reality is also a misspecified model, but is different from these approaches in that the misspecification is not about a feature of the environment but about the payoff-relevant types of other agents. This results in strategic interaction between actors in the alternative and the objective reality. We believe that these interactions are new to political economics, and they drive our new predictions.

We also build on theories of populism and identity politics, including Acemoglu, Egorov and Sonin (2013), Bonomi, Gennaioli and Tabellini (2021), Besley and Persson (2021), Agranov, Eilat and Sonin (2023), and Bellodi, Morelli, Nicolò and Roberti (2023). Our contribution to this work is to model populism based on an alternative reality that serves to discredit the elite, and the new implications, especially about the inverse effect of elite criticism, the political demand for bad policy, the popularity of conspiracies, and the spillover to new media.[4]

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

[4] Another strand of this literature, reviewed by Guriev and Papaioannou (2022), studies empirically the demand-side determinants of populism.

	Trump	Orban	Netanyahu
Politician attacked about	Election interference	Checks and balances	Corruption
By conspirators	Deep state and media	Soros network	Judiciary and media
Ideological issue	Cultural values	Immigration	Palestinian conflict

Table 1: Alternative reality narratives

2 Model

2.1 Motivation

Our model is motivated by two observations. First, in several democracies, salient (false) alternative realities share the same common narrative, which is the following. 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 Table 1, are as follows.

- In the United States, Trump claims to be the victim of a coordinated attack about his interference in the election, 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” (Kilgore 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".

A natural question is whether these alternative realities emerge because of voters' demand or politicians' supply. Our second observation is that—although demand surely also plays a role—supply is important. This is for two reasons. First, even when there is demand, there is a cost to building and disseminating alternative realities. Since the politician benefits from them, he has an incentive to build and disseminate them. Consistent with this logic, populist politicians are known to value the supply of misinformation sufficiently that they are willing to capture a large section of the independent media (Mcmillan and Zoido 2004, Szeidl and Szucs 2021). Second, much evidence shows that the political supply of misinformation changes beliefs in domains that include hatred, extremism, inter-ethnic attitudes, immigration, and Covid practices (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).

These observations about the supply of alternative realities motivate our model. To formalize the alternative reality, we introduce a hypothetical conspiratorial type of the intellectual elite. And to formalize supply, we allow the politician to increase the voter's prior probability of this type.

2.2 Setup

We build a principal-agent model in which two principals, the intellectual elite and the politician, attempt to influence an agent, the voter.[5] We start by presenting the neoclassical (non-behavioral)

[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.

part of the model. There are three classes of actors, the politician p, the intellectual elite e, and the voters v. We say classes of actors because both the elite and the voters consist of a unit mass of identical members. We think about the elite as the news media, and assume a one-to-one correspondence between elite members and voters, so that each elite member has exactly one voter as its audience. This assumption ensures that individual elite members cannot affect the election outcome. As we will see below, because of symmetry, we can represent all elite members, and all voters, as a single actor each. We let i stand for any class of actors.

At the beginning of the game the neoclassical types are realized. Only the politician has such types, along two dimensions. The first represents a *common* issue, $\theta_c \in \{0, 1\}$ and $\theta_c = 1$ with probability q_c, where common means that the preferences of the voter and the elite on the issue agree. $\theta_c = 1$ implies that the politician is "good" and increases the voters' and elite members' per capita consumption by c. We assume that θ_c is observed by the politician, and that members of the elite receive a signal $\hat{\theta}_c$ on it which is correct with probability $\pi \leq 1$. Every elite member receives the same signal, and voters do not receive a signal. We think of π as relatively high. The politician's second type dimension represents a *divisive* ideological issue, $\theta_d \in \{0, 1\}$ and $\theta_d = 1$ with probability q_d, where divisive means that the preferences of the voter and the elite on the issue differ. $\theta_d = 1$ means that the politician is pro-voter, i.e., her preferences about the divisive issue align with that of the voter, while $\theta_d = 0$ means that the politician is pro-elite, i.e., her preferences align with that of the elite. We assume that θ_d is observed by all actors, and that θ_d and θ_c are drawn independently.

After observing the signal $\hat{\theta}_c$, each elite member j sends a message $s_j \in \{0, 1\}$ to its voter, where $s_j = 1$ means that the signal they received is good. We sometimes refer to $s_j = 1$ as praise and $s_j = 0$ as criticism. Simultaneously, the politician decides whether to send propaganda $p \in \{0, 1\}$. Propaganda is successful in reaching the voters with probability $\alpha < 1$. Each voter observes the message of its elite member and, if it is successful, propaganda. Then, each voter votes on whether to reelect the politician. If the politician is not reelected, a new one is drawn from the prior distribution of objective types. Note that in this neoclassical setup, propaganda plays no role.

Alternative reality types and beliefs. To model alternative realities, we allow the voter to enter-

Type	Values (probabilities)	Interpretation
A. Politician		
Common (θ_c)	1 (q_c), 0 ($1 - q_c$)	1=Good
Divisive (θ_d)	1 (q_d), 0 ($1 - q_d$)	1=Pro-voter
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. Voter		
Mind (θ_m)	N (if $p = 0$), P (if $p = 1$)	P=Persuaded by propaganda

Table 2: Types and interpretations

tain two theories of the world, denoted by R (reality) or AR (alternative reality). These theories differ in the beliefs and motives of the principals. Formally, we introduce (i) types for the principals that represent their beliefs and motives in the R and the AR, and (ii) types for the voter that represent the probability they assign to the AR. Regarding the principals, the politician and all members of the elite have the same reality type $\theta_r \in \Theta_r = \{R, AR\}$, where 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 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.

Regarding the voter, we assume that he has a "mind" type $\theta_m \in \Theta_m = \{N, P\}$ where N represents normal and P represents persuaded. The normal voter thinks that the prior probability of $\theta_r = AR$ is zero; the persuaded voter thinks that the prior probability of $\theta_r = AR$ is $q_{ar} > 0$. We let $q_r = 1 - q_{ar}$. The voter's initial mind type at the beginning of the game is $\theta_m^0 = N$. His eventual mind type, θ_m, is N if he is not reached by propaganda and P otherwise. The voter forms his posterior, based on the messages his observes, from the prior encoded by his mind type. Thus, the model's type vector is $(\theta_d, \theta_c, \hat{\theta}_c, \theta_r, \theta_m) = \theta$. We summarize the types in Table 2.

Motives. Begin with the preferences of the elite. In both the R and the AR, each elite member j has preferences over the type of the politician after the election:

$$U_{ej} = c\tilde{\theta}_c - \lambda\tilde{\theta}_d \tag{1}$$

where $\tilde{\theta}_c$ and $\tilde{\theta}_d$ are the common and divisive types of the politician who wins the election.[6] Here $c > 0$ measures the importance of the common issue, and $\lambda > 0$ measures the strength of disagreement on the divisive issue.[7] Thus, the elite derives utility c from a politician who is good on the common issue, but disutility λ from a politician who is pro-voter on the divisive issue. We further assume that each elite member has a small preference for sending a truthful message, 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 elite members to maximize the sum of their utilities.[8] These assumptions imply that (i) members of the R elite, because they influence a single voter and have no impact on the election outcome, always send a truthful message; while (ii) members of the AR elite, internalizing the effect they have on each other, act as a single strategic player that can influence voters. In both cases, members of the elite send the same message which we denote by s. It follows that for the purposes of characterizing behavior, we can represent the elite as a single player which maximizes

$$U_e = 1_{\{\theta_r = AR\}} \cdot (c\tilde{\theta}_c - \lambda\tilde{\theta}_d) + 1_{\{\theta_r = R\}} \cdot 1_{\{s = \theta_c\}}. \tag{2}$$

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.

The preferences of the politician, independently of her type, are given by

$$U_p = E \cdot 1[\text{reelected}] - f \cdot p \tag{3}$$

[6] We omit preferences about the incumbent politician in the current period, as her type cannot be changed.

[7] Thus, λ represents the product of the importance of the divisive issue and the misalignment in the preferences—difference between the ideal points—of the elite and the voter.

[8] In our baseline model with identical voters, it would suffice to require that a small majority of elite members act uniformly. For simplicity we assume that the entire elite does so, which would be their optimal strategy in the presence of idiosyncratic voter-level taste shocks as in Section 4.2.

where E measures the utility from being in power, and f is the cost of propaganda $p \in \{0, 1\}$.

Every voter has the utility function

$$U_v = c\tilde{\theta}_c + \lambda\tilde{\theta}_d + \epsilon, \tag{4}$$

where, as before, $c > 0$ measures the benefit from the good politician and $\lambda > 0$ the benefit from a pro-voter politician (i.e., the misalignment between elite and voter preferences). In addition, ϵ is a common mean-zero uniform popularity shock with support $[-\bar{g}, \bar{g}]$ and density $g = 1/(2\bar{g})$. We assume $\bar{g} > c+\lambda$ so that with positive probability ϵ dominates the utility for any realization of $\tilde{\theta}_c$ or $\tilde{\theta}_d$. Note that ϵ only affects the preferences of the voter. Because voters' preferences are identical, we focus on equilibria in which they behave identically and represent them as a single actor.

Trembles. We assume that both the elite's message s 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 actual message s_j sent; and with independent probability ε_p, the realized propaganda message $\hat{p}$ is the opposite of the actual propaganda p sent. We let ε_e and ε_p go to zero and characterize the equilibrium in the limit.

Timing. The timing of events is organized into the following stages.

0. The politician's type is realized. The voter observes her divisive type θ_d, the elite also observes a signal on her common type $\hat{\theta}_c$.

1. The elite sends message $s \in \{0, 1\}$ and the politician decides on propaganda $p \in \{0, 1\}$. Both messages are subject to trembles and all actors observe the realized messages $(\hat{s}, \hat{p})$. If $\hat{p} = 1$ then the voter's mind type changes to $\theta_m = P$.

2. The voter decides whether to reelect the politician. If the politician is not reelected, a new politician with randomly drawn divisive and common types is elected.

3. Payoffs realize.

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 voter who may be persuaded by propaganda. We assume that after stage 1, the posterior of each voter 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 three players and in total 20 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 *coordination-proof* in the sense that the AR elite, after observing the signal realization $\hat{\theta}_c$, cannot propose a joint deviation to the AR politician that would improve the expected payoff of both. Intuitively, since members of the AR elite can coordinate among themselves, it seems plausible that—when they share interests—they can also coordinate with the AR politician. Second, we focus on equilibria which are *politician-pure*: in which all politician types use pure strategies. This assumption is only used in the application of Section 4.2 in which the politician can spin different alternative realities, and we introduce it here only to be consistent throughout the paper. In that application, the assumption rules out strategies in which the politician, while spinning one narrative, has to explain that with positive probability she is lying and the reality is another narrative. Third, among these equilibria, we focus on *politician-optimal* equilibria, which maximize the ex ante expected utility of the incumbent R

[9] Specifically, $\theta_c \in \{0,1\}$, $\theta_d \in \{0,1\}$, $\theta_r \in \{R, AR\}$ for each principal, $\theta_d \in \{0,1\}$ and $\theta_m \in \{N,P\}$ for the voter.

politician. We refer to equilibria satisfying these conditions as *CPO* equilibria.

2.4 Discussion of model assumptions

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

Examples of common and divisive types. Examples of the common type, i.e., issues on which voters and the elite agree, include competence or honesty. Examples of the divisive type, i.e., issues on which voters and the elite disagree, include the ideological issues listed in Table 1, as well as other ideological issues such as economic redistribution.

Elite conspiracy. The results of our baseline model would plausibly 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 partly because it is realistic (Douglas, Uscinski, Sutton, Cichocka, Nefes, Ang and Deravi 2019). But we provide a microfoundation in Section 4.2 by showing that when the politician can choose between a lying cost and a conspiracy narrative, she will often favor the latter. Intuitively, by solving their collective action problem, the conspiracy makes the AR elite more powerful, which allows the alternative reality to explain away even credible critical evidence.

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.

R elite is truthful. This assumption is a natural starting point since the puzzle we want to explain is that voters do not believe the message of the intellectual elite.

Belief changes. We assume that propaganda changes beliefs about the elite's AR type, not about the politician's common type. We do this because in our framework the beliefs about the elite are more important. Changing the belief that the politician is good, absent changing the belief about the elite, would be ineffective, since (for π high) the elite's message would essentially correct beliefs. We also assume that only the politician, but not the elite, can move priors. This is natural since the politician is a single decision maker while the elite is atomistic. In addition, it is probably easier to create than to eliminate beliefs in a conspiracy theory.

Equilibrium concept. As is standard in economics, our equilibrium concept assumes that actors know each others' strategies. In our setting with manipulable priors, equilibrium does not seem easily justifiable with learning. However, although a formal foundation is beyond the scope of this work, a natural informal justification is based on persuasion and introspection. The propaganda-spreading politician may explain the narrative of the equilibrium to make propaganda persuasive (Shiller 2017, Eliaz and Spiegler 2020). And the voter may fill in any gaps in the politician's narrative by thinking through the motives of the other actors.[10]

3 Results

3.1 Equilibrium behavior

Key mechanism. The logic of the equilibrium will be that propaganda is effective because it partially deflects the elite's criticism. We start by explaining how deflection works. Suppose that π is near one and that the politician is pro-voter. Consider the profile in which, in R, only the bad politician sends propaganda, while in AR, both the good and the bad politician send propaganda. This is the profile emerging in the main equilibrium.

Consider the beliefs of the voter after observing propaganda and criticism. If the voter were normal, i.e., assigned zero probability to AR, he would learn that the politician is bad, for two reasons: the R elite's critical message is truthful, and in R propaganda is only sent by the bad politician. In contrast, the persuaded voter, who assigns probability $q_{ar} > 0$ to the AR, believes (for π approaching one) that the politician is good with probability

$$\hat{q}_c = \mu_v(\theta_c = 1 | \hat{p} = 1, \hat{s} = 0, \theta_d = 1, \theta_m = P) = \frac{q_{ar} q_c}{q_{ar} q_c + (1 - q_c)} > 0. \tag{5}$$

[10] The process of thinking through the motives of others requires beliefs about the beliefs of others, which are not straightforward without common priors. We take the view that agents agree to disagree: they are aware of differences in prior beliefs, and reason about others taking into account these differences. Importantly, while higher-order beliefs matter for our informal equilibrium justification, they are not needed for the equilibrium definition or the analysis.

Start with the numerator. In the AR both propaganda and criticism are expected irrespective of whether the politician is good, so the probability of observing both and having a good politician is $q_{ar}q_c$. Now consider the denominator. Propaganda and criticism will also arise if the politician is bad, in both R and AR, explaining the additional term $1 - q_c$. The key is that $\hat{q}_c > 0$: the voter updates from propaganda and criticism about the common type only partially, because when reality is AR, he expects these messages even for the good politician. It follows that propaganda deflects criticism, and $\hat{q}_c$ measures the extent to which it does so.[11]

The flip side of propaganda altering the effect of elite criticism is that propaganda also alters the effect of elite praise. More specifically, propaganda ends up *inverting* the effect of the elite's message. To see why, continue the analysis of the voter's beliefs after observing propaganda. As we have just seen, when the elite criticizes, the voter believes that the politician is good with positive probability. In contrast, when the elite praises, the voter believes that the politician is good with zero probability. This is because in the AR the elite always criticizes, whereas in the R it only sometimes criticizes (as long as π is not yet equal to one). Thus, elite praise proves that reality is R, and in R only the bad politician sends propaganda. It follows that elite praise reduces and elite criticism increases voter beliefs, an inversion effect that will be key for our results.

Equilibrium. We turn to introduce the main strategy profiles that emerge as equilibria.

Definition 1. A strategy profile has the *no propaganda form* if no politician type sends propaganda and all elite types report truthfully.

This is the profile that would emerge if propaganda had no effect on beliefs.

Definition 2. 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 and only if she is pro-voter and bad.

[11] The logic that the persuaded voter updates differently than the normal voter from the elite's signal parallels the intuition in Alesina et al. (2020) that identical information translates into different political preferences depending on existing perceptions.

2. In the alternative reality (AR):
    - The elite reports that the politician is bad if and only if the politician is pro-voter,
    - The politician sends propaganda if and only if she is pro-voter.

This is they main equilibrium profile emerging from the model. In the objective reality R, the politician uses propaganda to deflect criticism, but only when she is bad and pro-voter. In the alternative reality AR, as we assumed above, the politician always sends propaganda, irrespective of whether she is good or bad. Finally, the behavior of the AR elite is "simple" in that elite members always criticize the pro-voter politician, irrespective of whether that politician is good or bad.

Definition 3. A strategy profile has the *complex propaganda form* if the AR elite, when the politician is pro-voter and the signal is good, randomizes between the good and the bad message, while all other principal types behave as in the simple propaganda profile.

This profile is similar to the simple propaganda equilibrium in that only the bad pro-voter politician sends propaganda. But it differs in that the behavior of the AR elite is more complex: instead of automatically criticizing the pro-voter politician, elite members sometimes praise her. As we explain in more detail below, this more complex alternative reality can "explain away" the joint outcome of elite praise and propaganda that destroys the simple alternative reality.

Assumption 1. When the elite is fully informative ($\pi = 1$), the benefit to the bad pro-voter politician from partially hiding her common type is higher than the cost of propaganda:

$$E \cdot \alpha \cdot \hat{q}_c \cdot c \cdot g > f.$$

Recall that E is the utility from being in power, α is the probability that propaganda is successful, $\hat{q}_c$ is the expected improvement from propaganda in the voter's belief that the politician is good, c is the benefit of having the good politician and g is the density of the politician's popularity shock. The assumption thus ensures that—for π close enough to one, in which case the formula for $\hat{q}_c$ is approximately correct—spreading the alternative reality is profitable.

Let $\overline{\lambda} = c \cdot \max\{(1-q_c)/(1-q_d), q_c/q_d\}$ and $\underline{\lambda} = c \cdot \min\{(1-q_c)/(1-q_d), q_c/q_d\}$. It is easy to verify that $\lambda > \overline{\lambda}$ means that the divisive issue is sufficiently important that the elite wants to keep even the bad pro-elite politician, but wants to remove even the good pro-voter politician. In contrast, $\lambda < \underline{\lambda}$ means that the divisive issue is sufficiently unimportant that the elite wants to remove even the pro-elite bad politician, and wants to keep even the pro-voter good politician.

Proposition 1. *Under Assumption 1 there exists $\bar{\pi} < 1$, independent of λ, such that for $\pi > \bar{\pi}$*

1. *If the divisive issue is unimportant, $\lambda < \underline{\lambda}$, the unique CPO equilibrium has no propaganda.*
2. *If the divisive issue is important, $\lambda > \overline{\lambda}$, there exists $\alpha(\pi) > 0.5$ such that*
   (a) *For $\alpha < \alpha(\pi)$ the unique CPO equilibrium has the simple propaganda form.*
   (b) *For $\alpha > \alpha(\pi)$ the unique CPO equilibrium has the complex propaganda form.*

All proofs are in the Appendix. We unpack the result and its intuition in steps. Part (1) states that when λ is sufficiently small, propaganda is never used in equilibrium. This is because $\lambda < \underline{\lambda}$ ensures that even a conspiring elite would not want to remove a pro-voter politician who is good. Thus, the AR elite reports about the politician's type truthfully, and hence the politician has no reason to increase beliefs in the AR. Intuitively, elite manipulations are only believable if the elite has a conceivable reason to want to remove the politician.

In contrast, part (2) states that when λ is sufficiently large, propaganda *is* used in equilibrium. Here $\lambda > \overline{\lambda}$ ensures that the AR elite would want to remove even a good pro-voter politician, so that an active elite conspiracy is potentially believable. When this holds, the level of α determines whether propaganda is simple or complex, but in both cases, the politician uses propaganda if and only if she is pro-voter and bad. Intuitively, because the politician is pro-voter (not pro-elite), it is believable that the elite, were it able to conspire, would act to remove her. And because the politician is bad, by Assumption 1 she would gain from discrediting the message of the elite. In contrast, the pro-elite or the good R politician would never send propaganda: the former cannot exploit disagreement with the elite since they are on the same side, and the latter has no incentive to discredit the elite's truthful message.

To understand the inner logic of the equilibrium, and the form of propaganda, it is helpful to think through the behavior of the other actors. Begin with the case when $\alpha < \alpha(\pi)$. The behavior of the R elite is straightforward: because its members are atomistic and cannot influence the voter, they prefer to report truthfully. Now consider the AR actors. In the simple propaganda equilibrium, the AR elite—which wants to remove the pro-voter politician—always criticizes. This follows because α is not too high, so that with a large enough $1 - \alpha$ probability propaganda fails to reach the voter, in which event he takes the criticism by the AR elite at face value. Anticipating this, the AR elite criticizes to reduce voter beliefs.

Importantly, in the event in which propaganda does reach the voter, the AR elite would prefer a good message. This is because of the inversion effect. On the proposed path the AR elite always criticizes, so that in the presence of propaganda, a good message can only occur in R (in the $1 - \pi$ probability event in which the elite's signal is incorrect). Thus, a good message would prove to the voter that the politician is bad. This force incentivizes the AR elite to send a good message, but when α is not too high, the incentive is not sufficient to overturn the simple propaganda equilibrium.

Still in the case of $\alpha < \alpha(\pi)$, consider the behavior of the AR politician. Both the good and the bad type believe that the elite is AR and criticizes their competence. Therefore, both send propaganda to deflect this criticism. This behavior 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.

Consider next the case when $\alpha > \alpha(\pi)$. Recall that in the simple propaganda equilibrium, elite praise, when the voter is reached by propaganda, conclusively proves that reality is R. When α is high, so that the voter is likely to be reached by propaganda, this effect creates a strong incentive for the AR elite to send praise: by doing so, elite members "outsmart" the conspiracy-believing voter and make him think that reality is R. This logic overturns the simple propaganda equilibrium, and leads to a mixed strategy profile in which the AR elite sometimes tells the truth.

Intuitively, as the conspiracy theory becomes more mainstream (α high), it 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. The conspiracy theory evolves to address this problem by making the elites more sophisticated. Elites now sometimes tell the truth to confuse voters, so that any elite message is consistent with the conspiracy theory. The Proposition predicts that these complex narratives emerge as the conspiracy theory becomes widespread.

Inversion effect. As we noted earlier, the key mechanism through which propaganda deflects elite criticism is the that it *inverts* the effect of the elite's message. We now state this point formally.

Corollary 1. *Suppose that Assumption 1 holds, $\lambda > \overline{\lambda}$, and the politician is pro-voter. In the CPO equilibrium, if $\hat{p} = 1$, then elite criticism strictly increases support for the politician.*

In the simple propaganda equilibrium, we get inversion because in the alternative reality the (conspiring) elite always criticizes, while in the objective reality the (honest) elite only sometimes criticizes. Thus, observing elite criticism will increase the voter's posterior belief of the alternative reality, and with it, his posterior belief that the politician is good. In the complex propaganda equilibrium we get inversion because the AR elite chooses to praise with positive probability only if doing so sometimes hurts the politician. Since praise helps the politician when propaganda does not succeed, it must be that praise hurts the politician when propaganda succeeds.[12]

Updating. Given the central role of belief updating for our results, it is helpful to characterize at a more abstract level how updating in our model differs from that in standard Bayesian frameworks.

Corollary 2. *Suppose that Assumption 1 holds, $\lambda > \overline{\lambda}$, and the politician is pro-voter. In the CPO equilibrium:*

1. *Even though signals are generated by R, the voter's average posterior, relative to his average (post-propaganda) prior, moves towards the AR.*

2. *Even though the voter's prior about θ_c is correct in both R and AR, his average posterior is too positive.*

Both of these results seem 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. Statement (1) says that here, even though R

[12] The logic that elite praise decreases voters' support for the politician is related to Ali, Mihm and Siga (2018) who show that the support of many others can decrease voters' support for desirable redistributive policies.

is always included in the voter's prior, he forms beliefs that on average drift away from the truth. Moreover, in a standard Bayesian framework with a correct prior, the average posterior should agree with the prior. Statement (2) says that here the average posterior departs from the prior.

Both of these results are driven by the fact that the prior—because it is strategically chosen by the politician—is correlated with the state of the world and the voter does not account for this correlation. Start with (1). The key is that the voter gets a prior that puts a positive weight on the AR precisely in the state of the world that has a signal profile (propaganda and bad report) relatively more likely in the AR. Although the voter correctly accounts for the correlation between the state of the world and the signal profile, he does not account for the correlation between the state of the world and his prior. Statement (2) follows through a similar logic: the prior allows for the AR precisely in the signal profile which is less bad for the politician in the AR than in the R. These intuitions are reminiscent of intuitions emerging in the context of persuasion with models (Schwartzstein and Sunderam 2021, Aina 2023), but are driven by a different mechanism: the principal's state-dependent choice of prior.

3.2 Implications and evidence

We now discuss and present evidence on several empirical implications of the model. Some of these implications are new to the literature, while others parallel those of existing models but highlight a new mechanism.

Elite has inverted effect in the presence of propaganda. Perhaps the most important implication, summarized in Corollary 1, is that propaganda inverts the effect of the elite's message: elite criticism can increase popular support for the politician. This prediction resolves what we view to be a key puzzle in contemporary U.S. politics: that during 2023, the growing body of critical evidence against Donald Trump, including four criminal indictments, was accompanied by an *increase* in his popular support (Swan et al. 2023). Although it is conceivable that in the future this pattern reverses, the remarkable resilience of Trump's support in the face of evidence is still puzzling, and we are not aware of other formal models that explain it.

Corollary 1 explains this fact by predicting a causal link, i.e., that the indictments caused the

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

increase in support. We now present two pieces of evidence that are consistent with this causality. First, in Table 3 we show results from a recent poll investigating the impact of the indictments on Trump's political support (YouGov 2023). Among registered Republicans intending to vote in the primaries, 41% claim that they would be more likely, and only 4% claim that they would be less likely, to vote for Trump if he is indicted in the matter of handling classified documents. Even among moderate Republicans in this group, 24% would be more likely, and only 13% would be less likely to vote for Trump. Thus, Republicans anticipate 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 the model's common type. 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.[13]

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. All specifications

[13] 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 4: Impact of scandals on contributions from Trump-supporter and other donors

include representative and quarter fixed effects. Table 4 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 mainly 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, seem to generate an increase in political support among Trump-supporter voters. Since these voters are most likely to believe in 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 increases the competitiveness of the election, and this increase in competitiveness makes Republicans donate more. Two pieces of evidence speak against this explanation. First, as Table 4 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.2 we show that when the election of Republican congressional candidates becomes more competitive because of redistricting, there is no analogous impact on donations. Thus, the effect we document seems to be driven by elite criticism rather than increased competition, consistent with our model.

Propaganda lowers accountability. A second implication of our model is that propaganda increases the re-election probability of the bad (pro-voter) politician and hence lowers accountability. This implication is to be expected in any model of propaganda, and is a key result in Guriev and Treismann (2020), where it obtains through "positive propaganda", i.e., propaganda that sends a positive message about the politician. Our model highlights a different mechanism, which is based on propaganda inverting the effect of the elite's message. This mechanism is consistent with the evidence we just discussed and offers an alternative explanation for the political success of informational autocrats who stay in power by means of propaganda (Guriev and Treisman 2022). Our model also predicts that positive propaganda, absent discrediting the elite, would not work: the elite's truthful message would immediately correct beliefs.

Propaganda creates distrust in science and the non-adoption of best practices. A third implication is that propaganda creates distrust in the scientific consensus. Intuitively, once the voter believes that the elite can conspire to advance their own goals, he will suspect that the elite's message in any other domain may be driven by their private interest.[14] For example, reports about Covid by a conspiring elite may be driven by their desire for financial gain from drug companies, rather than their interest in telling the truth. We are not aware of prior work formalizing the negative effect of propaganda on trust in the elites.

This implication is consistent with the beliefs and behavior of Republicans in the health and climate domains. For example, Allcott et al. (2020) show that Republicans are less likely to engage in social distancing, and Wallace, Goldsmith-Pinkham and Schwartz (2022) shows that they have higher excess death rates attributable to Covid. Prior work on populism has emphasized an opposing chain of causality: that distrust in the elites creates demand for populism (Bellodi et al. 2023, Guiso, Helios, Morelli and Sonno 2023). These two chains of causality have different policy implications: ours suggests that stopping propaganda should improve trust in science.

[14] It is straightforward to extend our framework to formalize this implication.

Propaganda is only used in divided societies by the pro-voter politician. A fourth implication is predictable variation in propaganda. Proposition 1 shows that propaganda is only used (i) if disagreement is large, and (ii) by the pro-voter politician. Prediction (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 novelties relative to the prior work are that in our model the relevant cleavage is ideological disagreement rather than economic inequality; and that our model emphasizes disagreement with the intellectual, rather than the political or business elite. Prediction (ii) suggests that the nature of the divisive issue may determine the nature of the alternative reality. When the divisive issue is cultural values, in which case the voter is plausibly to the right of the elite, the pro-voter politician is right-wing and we should observe right-wing propaganda. When the divisive issue is economic redistribution, in which case the voter is plausibly to the left of the elite, we should observe left-wing propaganda. Although the negative correlation between the political leaning of the elite and that of the populist can be explained without our model, we highlight the role of alternative realities as a mechanism.

4 Applications

We turn to develop three applications of our model: the impact of propaganda on government policy, endogenizing the nature of the alternative reality, and the impact of propaganda on media markets. 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 Government policy in the shadow of propaganda

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, to avoid praise from the discredited elite that would puncture 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 common type more visible to the elite. This action captures the key aspect of a "bad policy" for our purposes that it invites elite criticism.[15] With this modeling choice, the bad policy is in the same domain as the common type of the politician, but we believe that similar results could be obtained if the policy was in a different domain.

Formally, the bad politician, after observing whether propaganda is successful, can take an action 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 voter observe this action. Our assumption that the politician chooses the policy action after observing the success of propaganda reflects our view that she has an informational advantage relative to the elite about the popularity of her message. Stage 1 of the game then becomes the following:

1. (a) The politician decides on propaganda $p \in \{0, 1\}$. Propaganda is successful with probability α and subject to trembles.

   (b) The politician observes the realized propaganda message $\hat{p}$, and, if her type is bad, decides whether to take a policy action $e \in \{0, 1\}$ which increases the probability that the elite's signal about her type is correct.

   (c) The elite observes the signal $\hat{\theta}_c$ and sends message $s \in \{0, 1\}$, which is subject to trembles. All actors observe the realized messages $(\hat{s}, \hat{p})$. If $\hat{p} = 1$ then the voter's mind type changes to $\theta_m = P$.

The remaining stages of the model are unchanged. We have the following result.

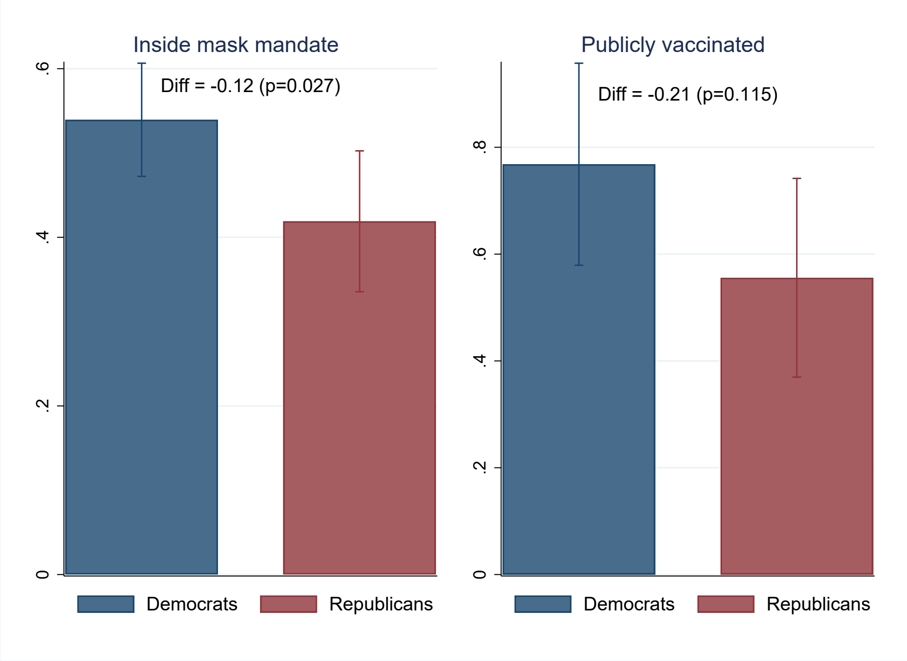
Proposition 2. *Suppose that Assumption 1 holds, $\lambda > \bar{\lambda}$ and $\alpha < 0.5$. There exists $\bar{\pi} < 1$, such that for $\pi' > \pi > \bar{\pi}$, in the unique CPO equilibrium,*

1. *Elite message and propaganda take the simple propaganda form.*

2. *The politician adopts the "bad policy" if and only if she is pro-voter and bad, reality is R, and propaganda was successful.*

[15] 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.

Figure 1: Impact on policy



To see the intuition, first note that because $\alpha < 0.5$ we are in the parameter range of Proposition 1 in which the equilibrium has the simple propaganda form. In this range, it is a dominant strategy for the AR elite to always criticize the pro-voter politician, because in the likely event in which propaganda is unsuccessful, elite criticism sharply reduces voter beliefs. In contrast, as Corollary 1 demonstrated, in the event in which propaganda is successful, elite criticism increases voter beliefs. Thus, for a politician who expects propaganda to be successful, elite criticism is beneficial. This force induces her to take the bad policy action and reveal her bad type. Intuitively, once the politician knows that discrediting is bound to succeed, she wants to invite more criticism from the "lying New York Times" because voters interpret that as further evidence for an elite conspiracy.

Implications and evidence. The key empirical prediction is that propaganda-spreading politicians will set policies that invite criticism from the intellectual elite. It is not just that these politicians ignore expert opinion: they actively contradict it. To our knowledge, this prediction is new to the literature. Since major societal issues, e.g., in the health and environmental domain, often require government action, the prediction highlights a potentially major cost of propaganda.

Figure 1 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 documents in the cross-section of states that, controlling for the severity of the epidemic, Republican governors vaccinated themselves publicly 21 percentage points less often than Democratic governors.[16]

These facts are explained by Proposition 2. There is a possible alternative explanation, based on political ideology: for example, Republican governors may avoid indoor masks mandates because these violate personal freedoms. Although we cannot conclusively rule out that explanation, we find it implausible. Republicans often prioritize national security over personal freedom, as demonstrated by the Patriot Act (Rackow 2002); indoor mask mandates are actually pro-business and thus align with that component of Republican ideology (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 1 is that Republican governors preferred not to follow the expert consensus.

4.2 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. The collective action problem arises because a lie about the politician's common type by any given elite member benefits every other elite member, since they all benefit from reducing support for the politician.

[16] 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 in each state.

The ability to coordinate allows these externalities to be internalized, strengthening the incentives to lie and thus making the elite more powerful. It follows that an alternative reality which features the ability to coordinate 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 a component of the cost of lying to the voter: a publicly known fabrication cost that each elite member has to pay in order to send a false message. The fabrication cost will allow us to speak about the credibility of evidence—such as videos from intensive care units during Covid—that the elite provides in support of its message. To accommodate these new features, we also make some adjustments to the modeling framework, but none of the adjustments affect our fundamental assumptions.

Assume that the elite consists of a finite number of members N, and each of them accesses a mass $1/N$ of voters. Thus, unlike in our atomistic baseline model, each elite member has some individual-level incentives to manipulate. Assume further that the lying cost is non-infinitesimal, and can be written as the sum of a fabrication cost χ_f and a reputation cost χ_r. The fabrication cost χ_f is the cost of manufacturing the evidence presented in the elite's message, which we assume is verifiable by the voter and cannot be changed by the alternative reality. This is the key variable with respect to which we will study comparative statics. The reputation cost χ_r 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_r 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 $\chi_r = 0$. 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 $\chi_r = 0$. 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 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.

Because now each elite member influences a positive measure of voters, it is convenient and standard (Persson and Tabellini 2002) to introduce an idiosyncratic preference shock for voters. This ensures that each elite member's effect on the voting outcome is independent of the tightness of the election. Let the utility of voter i be

$$U_{vi} = c\theta_c + \lambda\theta_d + \epsilon + \eta_i,$$

where the new term η_i is a mean-zero uniformly distributed individual preference shock, independent across voters, with support $[-\bar{h}, \bar{h}]$ and constant density $h = 1/(2\bar{h})$. Similarly to the common popularity shock, we assume $\bar{h} > c + \lambda$, so that the voting outcome is always interior.

Let $\Delta U_e = g \cdot c \cdot [\lambda(1 - q_d) - c(1 - q_c)]$ denote the gain to the elite, when $\pi = 1$, from a maximal (unit) change in voters' perception of the common type of a politician who is in fact good. Intuitively, $g \cdot c$ is the impact of a unit change in voter beliefs on the probability of winning, and the term in parentheses is the effect on elite utility of removing a good pro-voter politician.

Proposition 3. *Suppose that Assumption 1 holds, $\lambda > \overline{\lambda}$, $\alpha < 0.5$ and the politician is pro-voter. Then, in the following ranges for χ_f, for π large enough there is a unique CPO equilibrium that has the simple or the no propaganda form, and*

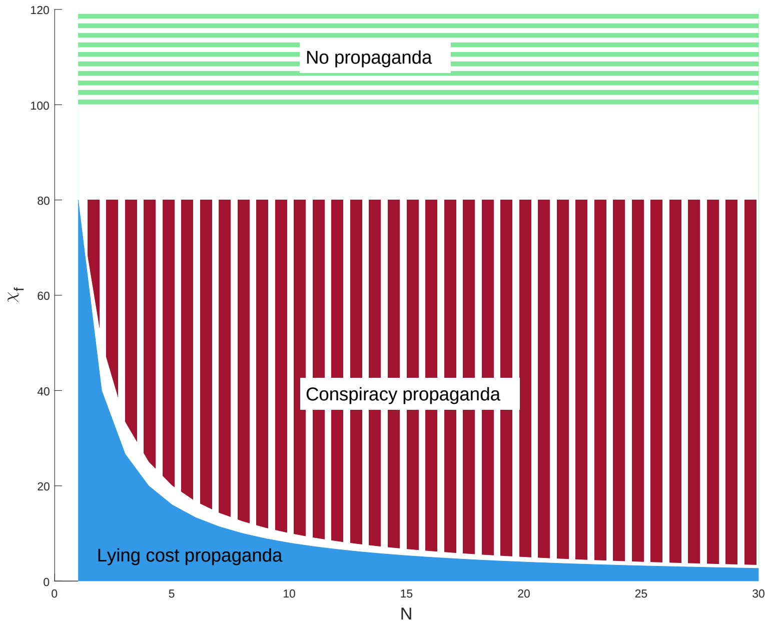
1. *If $\chi_f < (1 - 2\alpha)\Delta U_e/N$, then the bad R politician sends lying cost propaganda;*

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

3. *If $\Delta U_e < \chi_f$, 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

Figure 2: Endogenous AR



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 claim that elite members have no reputation cost of lying 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 proportional to $\Delta U_e/N$ from manipulating these voters, and hence—absent a reputation cost of lying—each is willing to fabricate a fake message as long as the cost of doing so does not exceed this gain. 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, on the order of $\Delta U_e/N$, 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 similar to the complex propaganda equilibrium 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.[17] Moreover, in this (and only this) result of the paper, we make use of the politician-pure criterion of our CPO equilibrium definition, which rules out equilibria that involve an AR politician randomizing between multiple ARs. In such equilibria, the politician–who we imagine is the person explaining the equilibrium to the voter—would have to say that although she now claims one AR, the voter should know that she may be lying and the truth may be a different AR. We do not find such narratives plausible, and the politician-pure criterion eliminates them.

Implications and evidence. The Proposition has two main implications, both of which, to our knowledge, are new. First, it explains the popularity of alternative realities that feature conspiracies. Such alternative realities indeed appear common (Douglas et al. 2019) and we are not aware of formal theories that explain their emergence.

Second, the Proposition predicts that increasing the credibility of evidence need not improve voter beliefs. This is because the politician can respond to an increase in the fabrication cost χ_f by escalating the alternative reality. Intuitively, to explain away the more credible evidence, the politician can upgrade from a lying cost AR to a conspiracy AR, and while the former cannot, the latter can coherently explain even the more credible evidence because it features a more powerful elite.[18] This finding strengthens our earlier point about the inverted effect of elite criticism, by

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

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

showing that even credible elite criticism may be deflected by the endogenously chosen alternative reality. Our logic formalizes the informal account of the evolution of conspiracy theories in Sunstein and Vermeule (2009): that maintaining a conspiracy theory in the face of disconfirming evidence requires an ever-widening conspiracy.

One consequence of these ideas is that providing more credible evidence, by escalating the alternative reality, can actually lead to worse societal outcomes. This is because the conspiracy alternative reality makes the AR elite more powerful in other domains too, and thus enables it to pursue its own interests more effectively in those domains. This leads to more distrust in the elite and hence worse voter decisions in other domains. This logic may help explain the persistence of Covid-denialism in the face of credible evidence (Hotez 2023).

4.3 New media and beliefs in the alternative reality

A salient fact about U.S. media is that several non-traditional outlets, most prominently Fox News, spread and reinforce the false alternative realities propagated by Republican politicians.[19] This fact appears to be unexplained by existing theories. It cannot be easily explained by theories of captured media (Besley and Prat 2006) since there is no evidence that non-traditional outlets are controlled by politicians. Nor can it be easily explained by theories of independent media (Mullainathan and Shleifer 2005, Gentzkow and Shapiro 2006), which predict that media slant the presentation of facts, but not that they present non-truths and alternative realities. Here we propose an explanation based on the idea that demand for non-traditional media arises because of audiences' distrust in the elite media, implying that it is in the interest of the non-traditional media to sustain that distrust by reinforcing the alternative reality.

To formalize this idea, we add to our basic model a mass of new media outlets which can also send a signal about the common type of the politician. Paralleling our model of the elite media, we assume that each new media outlet is too small to influence elections. Formally, there is a continuum of new outlets linked by a one-to-one mapping to voters, so that each voter consumes

[19] Examples of Fox News spreading alternative realities include false claims about the 2020 election (Gabbatt 2022) or about immigration (Confessore 2022). Non-traditional outlets exist in essentially all media markets: in cable television they include the One America News Network, in radio the programs of Rush Limbaugh and Alex Jones, in online media Breitbart and NewsWars, and among local newspapers, The Tennessee Star and The New Boston Post.

exactly one elite and one new media outlet. Because the new outlets have identical incentives, we only consider equilibria in which they have identical strategies, and treat them as a single decision maker. For simplicity, we assume that both the elite and the new media observe the politician's common type perfectly ($\pi = 1$) and then both send a message about it to the voter. But the new media is still slightly less informative than the elite: formally, its message is subject to a vanishing tremble that is arbitrarily larger than that of the elite media.[20]

We introduce media competition by assuming that the elite media wants to maximize the voter's belief that reality is R, while the new media wants to maximize the voter's belief that reality is AR. This assumption captures the essence of competition for audiences. Indeed, if reality is R, then, because the elite media is more informative (has a smaller tremble), the voter should prefer it in the future; but if reality is AR, then, because the elite media conspires, the voter should prefer the new media in the future. We assume that the elite and the new media trade off this audience-seeking motive against a positive cost of lying. Formally, the objectives of the elite and new media are

$$\begin{aligned} U_e &= 1_{\{\theta_r = R\}}[\mu_v(R|\hat{s}^e, \hat{s}^n, \hat{p}) - \chi \cdot 1_{\{s^e \neq \theta_c\}}] + 1_{\{\theta_r = AR\}}(c\tilde{\theta}_c - \lambda\tilde{\theta}_d) \\ U_n &= \mu_v(AR|\hat{s}^e, \hat{s}^n, \hat{p}) - \chi \cdot 1_{\{s^n \neq \theta_c\}} \end{aligned}$$

where $\hat{s}^e$ and $\hat{s}^n$ are the realized messages of the elite and the new media. Start with the elite. The first term, active when reality is R, has two parts. The first part captures audience-seeking, measured with the voter's posterior belief that reality is R, denoted $\mu_v(R|\hat{s}^e, \hat{s}^n, \hat{p})$. The second part captures the cost of lying, including both reputation and fabrication costs, denoted χ. The second term, active when reality is AR, is the same as in the basic model, and reflects the elite media's policy preferences in the AR in which it can act collectively.[21] The objective of the new media simply reflects its audience-seeking preferences and the lying cost. For the new media we do not include policy preferences even in the AR, because we assume that the new media is pro-voter, implying that it does not conspire and its policy preferences do not affect behavior.

We further assume that the new media is successful in entering the market with probability $\gamma < 1$, and that it does not reach audiences and earns zero utility when it fails to enter. This

[20] At the cost of additional notation we could have captured the difference in informativeness in the signals as well.

[21] Implicit here is that for the AR elite audience-seeking and truth-telling are not important: its electoral preferences are dominant so these other terms can be ignored.

assumption captures that the new media is less established than the elite media.

In summary, relative to the baseline model, the game changes in stages 0 and 1 as follows.

0. The politician's type is realized. All players observe her divisive type θ_d. The elite and the new media also observe her common type.

1. The elite and new media send messages $s = (s^e, s^n) \in \{0,1\}^2$ and the politician decides on propaganda $p \in \{0,1\}$. Propaganda and the new media's message reach the voter with probability α and γ respectively. All messages are subject to trembles, and the elite's trembles are smaller than the new media's. If $\hat{p} = 1$ then the voter's mind type changes to $\theta_m = P$.

We need some parametric assumptions for our result.

Assumption 2. Propaganda is beneficial even if it only works in the absence of the new media:

$$E \cdot \alpha \cdot (1 - \gamma) \cdot \hat{q}_c \cdot c \cdot g > f.$$

This is a strengthening of Assumption 1. The left hand side now includes $1 - \gamma$, ensuring that propaganda is sufficiently attractive even if it only affects beliefs when the new media fails.

Assumption 3. Being truthful is profitable for the R elite but not for the new media

$$\alpha\gamma < \chi < \alpha.$$

The first inequality ensures that for the R elite, the potential audience-seeking gains from lying—earned in the $\alpha\gamma$ probability event in which propaganda is successful and the new media enters—are lower than the cost of lying χ. The second inequality ensures that for the new media, lying is attractive if, in the α-probability event in which propaganda succeeds, it fully moves beliefs. The reason the same lying cost χ incentivizes the elite media to be truthful and the new media to be lying is that the elite, which already established a market, cares less about competition.

Proposition 4. *Suppose that Assumptions 2 and 3 hold, $\lambda > \bar{\lambda}$, $\alpha < 0.5$, and the politician is pro-voter. Then, in any CPO equilibrium*

1. *The politician and the elite behave as in the simple propaganda equilibrium.*

2. (a) The new media in R, after a bad signal, sends a good message with positive probability;

(b) That good message, relative to a bad message, increases the voter's posterior of AR.

The result establishes that the new media lies—reports the bad politician good—with positive probability, and that doing so increases the voter's belief in the AR. For an intuition, consider first a setting in which the voter assigns zero probability to the AR. Then the new media cannot change beliefs, and thus reports truthfully to avoid the lying cost. But when the voter assigns positive probability to the AR, lying can be profitable. Indeed, assume the equilibrium profile is that the new media is always truthful. Then, in the presence of propaganda, the new media's report that the politician is good would conclusively prove to the voter that reality is AR, because only in that state of the world does the good politician send propaganda. Since the new media benefits from the voter believing that reality is AR, this creates an incentive for lying. Moreover, by the second inequality of Assumption 3, this incentive is sufficiently strong to break the honest equilibrium. At the same time, the first inequality of Assumption 3 ensures that the elite media does not have strong enough incentives to lie, because doing so would only matter when the new media succeeds, which happens with a sufficiently low γ probability. Thus, in equilibrium the new media lies with positive probability, establishing result 2(a). Since doing so is beneficial, it must be that lying moves voter beliefs, establishing result 2(b).

Implications and evidence. The result that new media lie with positive probability, and that doing so increases beliefs in the alternative reality, helps explain our motivating fact that new media like Fox News spread false alternative realities. The result also yields a new implication: that new media affect not only political preferences (DellaVigna and Kaplan 2007) but also beliefs in the alternative reality. This can increase distrust in science and further limit the adoption of health and climate best practices. This prediction is in line with evidence showing that the consumption of Fox News reduced social distancing and increased mortality during the Covid pandemic (Bursztyn et al. 2020, Simonov et al. 2020).

Prediction	New result	New mechanism	Evidence
Basic model			
1. Elite's effect inverted with propaganda	yes	yes	causal
2. Propaganda lowers accountability	no	yes	consistent
3. Propaganda creates distrust and non-adoption	yes	yes	consistent
4. Societal division causes propaganda	no	yes	consistent
Government policy			
5. Propaganda causes harmful policies	yes	yes	consistent
Endogenous alternative reality			
6. Alternative realities feature conspiracy theories	yes	yes	consistent
7. Credible evidence makes AR conspiratorial	yes	yes	none
New media			
8. Propaganda causes new media to spread lies	yes	yes	consistent
9. New media generate non-adoption of practices	yes	yes	causal

Table 5: New predictions

5 Conclusion

In this paper we built a model in which a politician can supply an alternative reality to discredit the criticism of the intellectual elite. Key to our approach is to formalize the alternative reality with a misspecification about the payoff-relevant type of the elite. This results in the voter reasoning about and responding strategically to the imagined behavior of the elite, which generates the key mechanism that the alternative reality inverts the effect of the elite's message on voter beliefs. This mechanism, and more broadly requiring that the alternative reality remains consistent with evidence, drives our results about the politician's resilience to criticism, the choice of bad policies, the emergence of conspiracy theories, and the spillovers to private media.

We summarize the nine main predictions of the model in Table 5. Seven of these predictions are to our knowledge are new, while two parallel those made in prior work but feature a new mechanism. As we described in the body of the paper, six of the predictions are supported by consistent evidence, and two by causal evidence. However, for some key predictions, especially #5 about the effect of propaganda on government policy, and #7 about the effect of the quality of evidence on the nature of the alternative reality, we lack causal evidence. Future work could evaluate the validity of these predictions.

It is useful to contrast our modeling approach with that of other work studying narratives and misspecified models in political economics, including Aina (2023), Bénabou et al. (2018), Eliaz and Spiegler (2020), Eliaz et al. (2022), Galperti (2019), Gentzkow et al. (2021), Levy et al. (2022), and Schwartzstein and Sunderam (2021). A key difference relative to all of this work is that in our setting the misspecification is about the payoff-relevant type of another actor. This results in strategic interaction between actors in the objective and the alternative reality, and this interaction drives the predictions in Table 5. For example, the inversion effect emerges because the voter reasons about the incentives of the elite in both the objective and the alternative realities.

Perhaps the main limitation of our approach is that it is silent about the demand side. Developing a behavioral-economic theory of why voters are willing to believe in alternative realities is an important avenue 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.

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.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva, "The Polarization of Reality," *AEA Papers and Proceedings*, May 2020, *110*, 324–28.

Ali, S Nageeb, Maximilian Mihm, and Lucas Siga, "Adverse selection in distributive politics," *Available at SSRN 3579095*, 2018.

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

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.

Bursztyn, Leonardo, Aakaash Rao, Christopher P Roth, and David H Yanagizawa-Drott, "Misinformation during a pandemic," Technical Report, National Bureau of Economic Research 2020.

Confessore, Nicholas, "How Tucker Carlson Stoked White Fear to Conquer Cable," The New York Times, `https://www.nytimes.com/2022/04/30/us/tucker-carlson-gop-republican-party.html`, 2022.

DellaVigna, Stefano and Ethan Kaplan, "The Fox News effect: Media bias and voting," *The Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234.

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.

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.

Gabbatt, Adam, "Fox and friends confront billion-dollar US lawsuits over election fraud claims," The Guardian, `https://www.theguardian.com/media/2022/jul/04/fox-oan-newsmax-lawsuits-election-fraud-claims`, 2022.

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

Gentzkow, Matthew and Jesse M. Shapiro, "Media Bias and Reputation," *Journal of Political Economy*, 2006, *114* (2), 280–316.

___, Michael B. Wong, and Allen T. Zhang, "Ideological Bias and Trust in Information Sources," Working paper, Stanford, MIT, Harvard 2021.

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.

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.

Kilgore, Ed, "Trump Tells Latino Voters He's the Victim of Communists," Intelligencer, `https://nymag.com/intelligencer/2023/11/trump-telling-latinos-persecuted-by-communists.html`, 2023.

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.

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

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.

Mullainathan, Sendhil and Andrei Shleifer, "The market for news," *American economic review*, 2005, *95* (4), 1031–1053.

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.

Persson, Torsten and Guido Tabellini, *Political Economics: Explaining Economic Policy*, 1 ed., Vol. 1, The MIT Press, 2002.

Public Religion Research Institute, "Understanding QAnon's Connection to American Politics, Religion, and Media Consumption," `https://www.prri.org/research/qanon-conspiracy-american-politics-report/`, 2021.

Rackow, Sharon H., "How the USA Patriot Act Will Permit Governmental Infringement upon the Privacy of Americans in the Name of "Intelligence" Investigations," *University of Pennsylvania Law Review*, 2002, *150* (5), 1651–1696.

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

Shiller, Robert J., "Narrative Economics," *American Economic Review*, April 2017, *107* (4), 967–1004.

Simonov, Andrey, Szymon K Sacher, Jean-Pierre H Dubé, and Shirsho Biswas, "The persuasive effect of fox news: non-compliance with social distancing during the covid-19 pandemic," Technical Report, National Bureau of Economic Research 2020.

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.

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.

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.

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.

A Appendix for Online Publication

A.1 Definitions and proofs

Definition of equilibrium. We start with introducing notation. We define the politician's type to be $\theta_p = (\theta_d, \theta_c, \theta_r)$. We define the elite's type to be $\theta_e = (\theta_d, \hat{\theta}_c, \theta_r)$, 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 voter's type to be $\theta_v = (\theta_d, \theta_m)$ because he observes θ_d and his priors depend on θ_m. Note that the types of different actors are correlated. We denote the action of actor i in stage $t \in \{1, 2\}$ by a_i^t. We let $\hat{a}_i^t$ 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 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 $\sigma_v(a_v^2|\theta_v, \hat{a}^1)$. We let $\hat{\sigma}$ denote perturbed strategies that incorporate Nature's trembles. We denote the prior belief of actor i of type θ_i by $\mu_i^0(\theta|\theta_i)$, and the posterior belief after history $\hat{h}^t$ by $\mu_i^t(\theta|\theta_i, \hat{h}^t)$. We allow beliefs to depend on types, both because the types of different actors are correlated so that the type of i has information about the types of $-i$, 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. To formulate the best-response condition, we first introduce subjective expected utility. For each actor, at each stage where it moves, its beliefs and the strategy profile generate a probability distribution over final outcomes. This distribution can differ from the objectively correct distribution because the persuaded voter has an incorrect prior about θ. Actor i at stage t uses its subjective probability distribution over outcomes to compute its subjective expected utility, denoted $U_i(\sigma|\hat{h}^t, \theta_i, \mu_i(\theta|\theta_i, \hat{h}^t))$. Then the best-response property of equilibrium is that at each stage t at which i has a move, for all actions σ_i' available to i,

$$U_i(\sigma|\hat{h}^t, \theta_i, \mu_i(.|\theta_i, \hat{h}^t)) \geq U_i((\sigma_i', \sigma_{-i})|\hat{h}^t, \theta_i, \mu_i(.|\theta_i, \hat{h}^t)).$$

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 voter requires that he follows Bayesian updating at the end of stage 1:

$$\mu_v^1(\theta_{-v}|\theta_v, \hat{a}^1) = \frac{\mu_v^0(\theta_{-v}|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{a}^1|\theta_{-v})}{\sum_{\theta'_{-v}} \mu_v^0(\theta'_{-v}|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{a}^1|\theta'_{-v})} \tag{6}$$

where $\mu_v^0(\theta_{-v}|\theta_v)$ is the prior of the voter of type θ_v about the types of the other actors $\theta_{-v} = (\theta_c, \hat{\theta}_c, \theta_r)$. This definition accounts for the model's deviation from rationality that the 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 voter is reached by propaganda and becomes persuaded, (6) computes his posterior from the prior of the persuaded voter $\mu_v^0(.|\theta_d, P)$. 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 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 $\hat{a}^1$.

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.

Proof of Proposition 1. Our proof identifies the CPO equilibrium and shows that it has the properties described in the proposition.

Case 1: $\lambda < \underline{\lambda}$.

Existence. We verify that the proposed profile is an equilibrium. The R elite, because it cannot coordinate, ignores its effect on the voter and given its preference for truth-telling always sends an honest report s. The normal voter updates as follows

$$\mu_v(\theta_c|\hat{p} = 0, \hat{s}, \theta_m = N) = \hat{s}\frac{q_c\pi}{q_c\pi + (1-q_c)(1-\pi)} + (1-\hat{s})\frac{q_c(1-\pi)}{q_c(1-\pi) + (1-q_c)\pi}. \tag{7}$$

In particular he is more likely to keep the politician after a good message. The AR elite knows that no politician sends propaganda and can thus influence the normal voter directly. She prefers to keep the good and remove the bad politician, and will thus report truthfully. The persuaded voter understands this and will thus update the same way as the normal voter. Finally, all politician types know that the elite reports truthfully in both R and AR, and thus they have no reason to send propaganda.

The equilibrium is CPO.

Lemma 1. *Under Assumption 1, if parameters are such that the AR elite after a good signal wants to keep the politician, then for π high enough any coordination-proof equilibrium has no propaganda.*

Proof of Lemma. Consider an equilibrium in which the good AR politician sends propaganda with positive probability. Because propaganda has a cost, this must be because the AR elite after a good signal sends a bad message with probability bounded away from zero as π goes to one. In turn, the reason to send a bad message after a good signal must be that the AR politician sends propaganda with probability bounded away from zero as π goes to one.

But then the AR elite after a good signal will initiate a deviation to the profile in which the politician does not send propaganda and the AR elite sends a good message. For π large this profile gets near the first best for both the politician and the elite. Hence it is strictly better to the politician than any profile in which she is at least indifferent to sending propaganda, because propaganda has a cost bounded away from zero. It is also strictly better to the AR elite than the supposed equilibrium profile in which, with probability bounded away from zero, the outcome is propaganda and a bad message. The only way such a profile would be near the first best is if voter beliefs after propaganda and the bad message were close to certain that the politician is good. But then the bad R politician, by Assumption 1, would also use propaganda, which means that after propaganda and a bad message voter beliefs cannot be near certain that the politician is good, a contradiction.

Finally, since the good AR politician does not send propaganda, and for π high the good R politician does not send propaganda either, propaganda can only come from the bad politician, implying that it is never used in equilibrium.

The Lemma implies that any CPO equilibrium has no propaganda. In any such equilibrium, the AR elite will always report her signal, i.e., the equilibrium has the no propaganda form.

Case 2: $\lambda > \overline{\lambda}$.

Subcase 1: Incumbent politician is pro-voter.

Cutoff α *value.* We first characterize the conditions under which the simple propaganda equilibrium may exist. Suppose actors play that profile, then the persuaded voter's beliefs are as follows

$$\mu_v(\theta_c|\hat{p}=1,\hat{s},\theta_m=P)=(1-\hat{s})\frac{q_c q_{ar}}{q_{ar}+q_r\pi(1-q_c)}. \tag{8}$$

Intuitively, when the message is good, it reveals that reality is R and the politician is bad. When the message is bad, the voter updates in a Bayesian fashion, taking into account that a bad message can come in R if the politician is bad and the message is correct, as well as in AR for all politician types.

We now turn to the behavior of the principals. Since $\lambda > \overline{\lambda}$, the AR elite wants to remove both the good and the bad pro-voter politician. The AR elite prefers to criticize if the voter is responsive to her message in this equilibrium, which is the case if

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

It is a bit tedious but straightforward to check that this inequality yields threshold $\bar{\alpha}(\pi)$, such that for $\alpha < \bar{\alpha}(\pi)$ the AR elite finds it optimal to criticize the politician, so that the simple propaganda equilibrium has a chance to be an equilibrium. It is also straightforward to check that $\lim_{\pi\to 1}\bar{\alpha}(\pi)=\frac{1}{1+\hat{q}_c}$.

Subsubcase 1: $\alpha < \bar{\alpha}(\pi)$.

Existence. We show that the simple propaganda equilibrium exists for this case. We have already characterized the beliefs of the voter and the behavior of the AR elite. Consider now the bad R politician. The gain from successful propaganda, in terms of its impact on the voter's belief, is

$$
\begin{aligned}
&E[\mu_v(\theta_c|\hat{p},\hat{s})|\theta_c=0,\hat{p}=1] - E[\mu_v(\theta_c|\hat{p},\hat{s})|\theta_c=0,\hat{p}=0] \\
&= \pi\cdot[\mu_v(\theta_c|\hat{p}=1,\hat{s}=0)-\mu_v(\theta_c|\hat{p}=0,\hat{s}=0)] + (1-\pi)\cdot[\mu_v(\theta_c|\hat{p}=1,\hat{s}=1)-\mu_v(\theta_c|\hat{p}=0,\hat{s}=1)] \\
&= \pi\cdot\left[\frac{q_{ar}q_c}{q_{ar}+q_r\pi(1-q_c)} - \frac{q_c(1-\pi)}{q_c(1-\pi)+(1-q_c)\pi(1-\alpha)}\right] - (1-\pi)\cdot\left[\frac{q_c\pi}{q_c\pi+(1-q_c)(1-\pi)(1-\alpha)}\right].
\end{aligned}
$$

As π converges to 1 this converges to $\hat{q}_c$, so Assumption 1 ensures that for π sufficiently high, propaganda is optimal for the bad R politician. It is also optimal for both the good and the bad AR politician, because both expect criticism with full probability (not just with probability π like the bad R politician) and hence profit even more from propaganda.

The good R politician, by avoiding propaganda, ensures that the voter's expected belief about her is at least the prior q_c. In contrast, if she sends successful propaganda, she forces the voter's expected belief below q_c: if the elite reports her good the voter will believe her bad with certainty, and if the elite reports her bad the voter will believe that she is bad with a higher-than $1-q_c$ probability because more bad than good types send propaganda.

We have confirmed that in the proposed equilibrium the elite and the politician best respond, and we have characterized the beliefs and hence behavior of the voter. To conclude, we clarify that our above arguments also cover off-equilibrium information sets. Such information sets only happen at stage 2, i.e., after the message profile is realized: because propaganda determines the voter type, the normal voter after propaganda and the persuaded voter absent propaganda can never occur in this game. Still, Bayesian updating in our equilibrium definition, (6), specifies beliefs based on how a voter of the given type would update from the information content of the messages he observes. In particular, the normal voter (who believes reality is R) will conclude from observing propaganda that the politician is bad. And it is straightforward to verify that the persuaded voter absent propaganda will form the following beliefs

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

These beliefs also pin down voting behavior.

The equilibrium is CPO. We first show that when the politician is pro-voter the simple propaganda equilibrium is coordination-proof. This follows because the AR elite and the AR politician have opposing incentives: the former wants the politician out, while the latter wants to stay. Thus there cannot be a joint deviation that improves the payoff of both.

We next show that this equilibrium is politician optimal. This proof will proceed through several steps. In any equilibrium that is weakly better for the bad R politician than our preferred, she must be sending propaganda and must prefer to do so, as otherwise her payoff is lower than in our preferred equilibrium.

Lemma 2. *In any equilibrium in which the bad R politician strictly prefers propaganda, both AR politicians send propaganda with positive probability, and after both signals the AR elite criticizes with positive probability.*

Proof of Lemma. If the bad AR politician does not send propaganda, then the AR elite after a bad signal will report her bad. Given that the bad R politician prefers propaganda, she must then also prefer propaganda. So it has to be that the bad AR politician sends propaganda with positive probability. This in turn implies that the AR elite after a bad signal must send a bad message with positive probability. The same reasoning applies for the good AR politician, who must also send propaganda with positive probability, which then implies that the AR elite after a good signal must send a bad message with positive probability.

Since this equilibrium is weakly better than our preferred, it must be that in the limit as π goes to one, the expected belief after propaganda and a bad signal is weakly better than in our equilibrium. This means that either (i) the probability of a bad message and propaganda is strictly higher when the AR politician is good than when the AR politician is bad, or (ii) these probabilities are equal to each other and equal to one.

Start with case (i). Assume that in the limit as π goes to one, r_1 (r_2) is the probability of a bad message after a good (bad) signal, and t_1 (t_2) is the probability of propaganda by a good (bad) AR politician. Then we have $r_1 t_1 > r_2 t_2$. Assume that the good AR politician uses propaganda more often than the bad AR politician, $t_1 > t_2$. It implies that the AR elite after a bad signal strictly prefers sending a bad message, since after a good signal she is at least indifferent and after a bad signal she faces propaganda less often. Therefore $r_2 = 1$, which then implies that the bad AR politician finds optimal to always use propaganda, $t_2 = 1$. This contradicts our initial assumption.

Thus we conclude that $t_1 \leq t_2$.

Now assume that the AR elite criticizes more often after a good than after a bad signal, $r_1 > r_2$. It follows that the good AR politician strictly prefers and always sends propaganda, $t_1 = 1$, since the bad AR politician facing a lower rate of criticism is at least indifferent, so she has to strictly prefer propaganda. This implies that either $t_1 > t_2$ or $t_1 = t_2 = 1$. We have already seen that there is no equilibrium where $t_1 > t_2$, so we must have both AR politician types use propaganda with full probability. Now recall that $r_1 > r_2$ says that the AR elite after a bad signal sends a good message with a strictly higher probability. But this means that in a nearby equilibrium with a lying cost (and π near 1) she must have an extra incentive to send a good message to compensate for her lying cost. That extra incentive could only come from the bad AR politician doing more propaganda than the good AR politician; which is not possible since, as we have seen, the good AR politician strictly prefers propaganda. This contradiction rules out case (i).

In case (ii), for π near one, the strategy profile is near the simple propaganda equilibrium. But this implies that both AR politicians will strictly prefer sending propaganda, since they do so in the simple propaganda equilibrium by Assumption 1. This then implies, using an argument based on the lying cost analogous to that in the previous paragraph, that the AR elite after a bad signal must send a bad message with full probability. Thus the only remaining case has the structure of the complex propaganda equilibrium. Consider this structure and let r be the AR elite's mixing probability of a good message after a good signal. We are interested in whether there is an r such that sending a good message is as bad for the politician as sending a bad message. Voter beliefs, as a function of r, are lowest after a good message when $r = 0$, because in that case a good message and propaganda can only come in R and prove that the politician is bad. Similarly, voter beliefs are highest after a bad message when $r = 0$ because then propaganda and a bad message include the event of the good AR politician with full probability. Thus the mixed equilibrium cannot exist when, even with $r = 0$, the bad message leads to worse beliefs than the good message. The condition for this is

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

We conclude that when this holds, the simple propaganda equilibrium is optimal.

Subsubcase 2: $\alpha > \bar{\alpha}(\pi)$.

Existence. We prove that there exists an equilibrium that has the complex propaganda form. Begin with the condition for AR elite's indifference. Suppose mixing probability of sending good report is r. Expected belief after good report

$$\begin{aligned}&\alpha\mu_v(\theta_c|\hat{p}=1,\hat{s}=1)+(1-\alpha)\mu_v(\theta_c|\hat{p}=0,\hat{s}=1)\\&\quad=\alpha\frac{q_{ar}q_c\pi r}{q_{ar}q_c\pi r+q_{ar}(1-q_c)(1-\pi)r+q_r(1-q_c)(1-\pi)}+(1-\alpha)\frac{q_c\pi}{q_c\pi+(1-q_c)(1-\pi)(1-\alpha)}\end{aligned}$$

and expected belief after bad report

$$\begin{aligned}&\alpha\mu_v(\theta_c|\hat{p}=1,\hat{s}=0)+(1-\alpha)\mu_v(\theta_c|\hat{p}=0,\hat{s}=0)\\&\quad=\alpha\frac{q_{ar}q_c\pi(1-r)+q_{ar}q_c(1-\pi)}{q_{ar}q_c\pi(1-r)+q_{ar}q_c(1-\pi)+q_{ar}(1-q_c)\pi+q_{ar}(1-q_c)(1-\pi)(1-r)+q_r(1-q_c)\pi}\\&\qquad\qquad+(1-\alpha)\frac{q_c(1-\pi)}{q_c(1-\pi)+(1-q_c)\pi(1-\alpha)}.\end{aligned}$$

Observe that for $r\in[0,1]$ the former is strictly increasing while the latter is strictly decreasing in r. For $r=0$ the former is smaller or equal than the latter if and only if

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

which holds when $\alpha\geq\alpha(\pi)$—α is at least as big as the threshold defined in the small α case.

For $r=1$ the former is larger than the latter if and only if

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

which will hold if

$$\frac{q_{ar}q_c(1-\pi)+q_{ar}(1-q_c)\pi+q_r(1-q_c)\pi}{q_{ar}q_c(1-\pi)}>\frac{q_{ar}q_c\pi+q_{ar}(1-q_c)(1-\pi)+q_r(1-q_c)(1-\pi)}{q_{ar}q_c\pi}$$

or equivalently

$$\frac{\pi}{1-\pi}>\frac{1-\pi}{\pi}$$

which always holds since $\pi > 0.5$. It follows that there is a unique mixing probability that makes the AR elite indifferent.

The indifference condition implies

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

or

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

and, increasing the left-hand-side

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

which implies

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

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

Now consider the limit, as π goes to one, of the 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

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

It follows that under Assumption 1, 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 views about her are above q_c while with propaganda are near $\hat{q}_c$ and hence below q_c. We conclude that the mixed equilibrium exists for π sufficiently high and $\alpha \geq \bar{\alpha}(\pi)$.

We next show 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 beliefs being increasing respectively decreasing is the same, and thus, under the same condition for α, for a lying cost small enough, there exists a mixing probability that ensures indifference. Moreover, when the lying cost is small, the implied mixing probability is close to that with zero lying cost, because the slope of the second belief in r is bounded away from zero uniformly in π. It follows that for any given π, when the lying cost is sufficiently small, the payoffs of all parties are going to be very 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; and 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.

The equilibrium is CPO. It is immediate that the equilibrium is CP when the AR politician is pro-voter, because they have opposing incentives with the AR elite. We next show that this is the optimal equilibrium for the bad R politician. First, this is better than any equilibrium in which the bad R politician does not send propaganda with certainty, because here she strictly prefers to send propaganda and thus doing so improves her payoff. Thus by Lemma 2 both AR politicians send some propaganda, and the AR elite reports bad after both types of signals with positive probability.

There cannot be a politician-optimal equilibrium in which the AR elite always reports bad after a good signal. In such an equilibrium, the good AR politician would strictly prefer and always send propaganda since the bad R politician, who gets a bad report with a weakly lower probability, does so. Consider the bad AR politician. If she randomizes, then the AR elite after a bad signal must also randomize, since we know that the bad R politician who gets criticized after a bad signal strictly prefers propaganda. But the reason the AR elite randomizes must be that in the presence of propaganda a good message is sufficiently worse than a bad message. Given this, the best response of the AR elite after a good signal must be to strictly send a good message, since the good AR politician always sends propaganda. This is a contradiction. It remains to check the case in which the bad AR politician always sends propaganda. Suppose that the AR elite after a bad signal sends a good message with positive probability. Since the AR elite after a good signal always reports a bad message, this will contradict the lying cost. More specifically, with a positive lying cost the AR elite after a bad signal needs some extra incentive to send a good message, which has to be that she faces more propaganda, which can only happen if the good AR politician sends less propaganda in the lying cost game, which contradicts that the good AR politician strictly prefers propaganda. Thus the only possible case is when the structure of the equilibrium has the simple propaganda form. However, that profile is not an equilibrium because the AR elite has then an incentive to send a good message when

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

holds.

Now consider the bad AR politician. There are three cases. First, if she sends propaganda with higher probability than the good AR politician. In that case the AR elite, which is indifferent after a good signal, will expect a higher probability of propaganda after a bad signal, and hence will want to report the politician good. This is because she was indifferent after a good signal, and since reporting good is clearly bad for her absent propaganda, it must be good for her with propaganda. In turn, it cannot be an equilibrium for the AR elite to always report good after a bad signal, because then with π high enough the bad AR politician will stop sending propaganda. Second, if the bad AR politician sends propaganda with lower probability than the good AR politician. Then, by a similar reasoning, the AR elite will want to send a bad message after a bad signal. Given that, for π high the bad AR politician will always want to send propaganda, a contradiction. Third, if the bad AR politician sends propaganda with the same probability as the good AR politician. First suppose that this probability is interior. Then the simultaneous indifference of the good and the bad AR politician requires that the AR elite sends a good message with the same probability after a good and a bad signal.

In this indifference case there are two subcases. The first subcase is when, even with π high, the probability with which the AR politicians send propaganda is bounded away from one. In that subcase, following a message profile of propaganda and a bad report, the bad R politician is pooled with the good and bad AR politicians in which the weight on the AR types is bounded away from one. As a result, the bad R politician is perceived to be worse than in our preferred mixed equilibrium. This cannot be a politician-optimal equilibrium. The second subcase is when, as π approaches one, the mixing probability of propaganda approaches one. In this case the mixing probability of the AR elite for reporting bad must also approach one: otherwise a message profile of (propaganda,good) will signal that reality is AR and that the politician is the average type, whereas a message profile of (propaganda, bad) will not exclude the bad R politician and hence signal that the politician is worse than average by a non-vanishing margin. This contradicts the indifference of the AR elite. But if for π high the AR elite almost always sends a bad message, then the payoff of the bad R politician is very similar to the payoff of the bad AR politician. Since the bad AR politician is indifferent to sending propaganda, the bad R politician must also be almost indifferent. This cannot lead to an equilibrium better than our preferred one.

The final case is when the good and the bad AR politicians send propaganda with full probability. Since the AR elite is indifferent after a good signal, it must be also indifferent after a bad signal. Suppose that the AR elite randomizes after a bad signal. This is the main remaining case to rule out. Now consider a nearby equilibrium with a small lying cost. In that nearby equilibrium the AR elite must still be randomizing after both a good and a bad signal. But then it has to be that the bad AR politician is sending propaganda with a higher probability than the good AR politician: otherwise, having the AR elite indifferent to sending a bad message after a good signal would imply that the AR elite strictly prefers sending a bad message after a bad signal to save on the lying cost. In turn, this implies that in the nearby equilibrium the good AR politician is sending propaganda with less than full probability, so that in the original equilibrium the good AR politician is indifferent.

This in turn implies that the AR elite after a good signal has to send a good message with probability bounded away from zero: if it almost always sends a bad message then—since the bad R politician strictly prefers propaganda—the good AR politician will also strictly prefer propaganda. But this means that propaganda and a good message arise in the AR with probability bounded away from zero when the politician is in fact good, and they arise almost never in the R. This is a force towards propaganda and a good message generating positive updating. This cannot happen in equilibrium as then the AR elite would never send a good signal. To counter it, the AR elite after a bad signal must send a good message with a strictly higher probability than after a good signal. The reason this probability must be strictly higher is that if they send a good message with an equal or lower probability, then (with π high) propaganda and a good signal just mean the type is average or better, while propaganda and a bad signal mean the type is strictly below average already in the AR, and even worse when the bad R politician is mixed in. But if the AR elite sends a good message with higher probability after a bad than after a good signal, then the bad AR politician must strictly prefer to avoid propaganda since the good AR politician was indifferent. This contradicts our starting assumption that the bad AR politician sends propaganda.

It follows that the good and bad AR politicians send propaganda with full probability, and that the AR elite always sends a bad message after a bad signal. This is our preferred mixed equilibrium.

Subcase 2: Incumbent politician is pro-elite.

Existence. We show that the profile in which (i) the AR elite always reports the politician good, and (ii) no politician sends propaganda, is an equilibrium.

Let's start with characterizing voters' beliefs. The normal voter, absent propaganda, holds beliefs given by equation 7. The persuaded voter, after propaganda, attributes propaganda to a tremble but changes his prior, and hence holds beliefs

$$\mu_v(\theta_c|\hat{s}, \hat{p}=1, \theta_m = P) = \hat{s}\frac{q_{ar}q_c + q_r q_c \pi}{q_{ar} + q_r q_c \pi + q_r(1-q_c)(1-\pi)} + (1-\hat{s})\frac{q_c(1-\pi)}{q_c(1-\pi)+(1-q_c)\pi}. \tag{9}$$

Start with the AR elite. Because $\lambda > \overline{\lambda}$ she wants to keep both types of politicians. She expects no propaganda and a normal voter, and hence the return on praise is

$$\frac{q_c\pi}{q_c\pi + (1-q_c)(1-\pi)} - \frac{q_c(1-\pi)}{q_c(1-\pi)+(1-q_c)\pi},$$

which is positive for large π. Therefore it is optimal for the AR elite to praise all politician types.

Consider the AR politician. Her gain from successful propaganda is

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

which is negative for large π. So neither common types of the AR politician engages in propaganda. Consider the R politician next. When she is praised then successful propaganda has the same return as above. When she is criticized then voter learns the reality is R and propaganda has no effect. As a result, no R politician engages in propaganda either.

Next consider off-path information sets. Like in Subcase 1, these only occur in stage 2: we need to deal with the normal voter after propaganda, and—since propaganda is off the equilibrium path—the persuaded voter after any history. The normal voter after propaganda, because it is off the equilibrium path, will attribute propaganda to a tremble and form beliefs and behavior just like the normal voter absent propaganda. Since propaganda is off the path, the persuaded voter after propaganda attributes it to a tremble and form beliefs and behavior just like the persuaded voter absent propaganda. These beliefs are given by equation (9). Finally, his belief pins down his voting behavior: as this is the last stage of the game, he faces a binary decision problem which has a solution, and chooses that solution. Indifference has zero probability because the preference shock has a smooth distribution.

This equilibrium is CPO. Since the AR politician wants to keep all politician types, then Lemma 1 implies that no coordination-proof equilibrium has propaganda. Since no politician uses propaganda, then it is optimal for the AR elite to always send a good signal.

Proof of Corollary 2. Under Assumption 1, if the division is large ($\lambda > \bar{\lambda}$) and the politician is pro-voter, then, by Proposition 1, the unique CPO equilibrium takes either the simple or the complicated propaganda form.

Start with the first half of the statement. The normal voter has correct prior beliefs on the type of reality, while the persuaded voter puts a q_{ar} weight on the AR reality. Therefore the expected prior is $(1-q_c)\alpha q_{ar}$.

In the absence of propaganda, normal voter's posterior belief remains the same as his prior. In both propaganda equilibria, as π tends to one, the persuaded voter's belief about reality observing propaganda and criticism converges to $q_{ar}/[q_{ar}+q_r(1-q_c)]$. As a result the expected posterior on the type of reality is

$$\mu_v(\theta_r = AR) \to (1-q_c)\alpha \frac{q_{ar}}{q_{ar}+q_r(1-q_c)},$$

which is larger than the expected prior.

Now, consider the second part of the statement. In these equilibria, both the normal and the persuaded voters' prior beliefs are that the politician's common type is good with a probability q_c. This implies that the voter's expected prior is also q_c.

In any equilibrium the normal voter's posterior beliefs are given by equation 7. In both propaganda equilibria, as $\pi \to 1$, the persuaded voter's belief on the politician's common type observing a bad signal converges to $\hat{q}_c$.

Therefore the voter's expected posterior, as $\pi \to 1$, converges to

$$\mu_v(\theta_c) \to q_c + (1-q_c)\alpha\hat{q}_c,$$

which is larger than the expected prior q_c.

Proof of Proposition 2. Our proof identifies the unique CPO equilibrium and shows that it has the properties described in the proposition.

Case 1: Incumbent politician is pro-voter.

Existence. We first show that the strategy profile in which (i) the AR elite always reports the politician bad; (ii) the good and bad AR politician and the bad R politician all use propaganda; (iii) bad R politician makes her common type visible if and only if propaganda was unsuccessful, while the bad AR never does it; constitutes an equilibrium.

The normal voter's beliefs are characterized by equation (7). The persuaded voter's beliefs are similar to equation (8) with π replaced by π'.

We now turn to the behavior of the principals. Since $\lambda > \overline{\lambda}$, the AR elite wants to remove both the good and the bad pro-voter politician. In any equilibrium, the impact of a good report rather than a bad on the voter's belief of the politician's quality is at least

$$(1-\alpha)\left[\frac{q_c\pi}{q_c\pi+(1-q_c)(1-\pi)(1-\alpha)} - \frac{q_c(1-\pi)}{q_c(1-\pi)+(1-q_c)\pi(1-\alpha)}\right] - \alpha,$$

because with at least $1-\alpha$ probability there is no propaganda and the voter takes the elite massage at face value, whereas with at most α probability there is propaganda, and then reporting the politician good is at most changes the voter's belief of the politician by 1. Since $\alpha < 0.5$, for π high enough, this is positive thus the AR elite improves the politician's reelection chances by praise. Therefore, sending a bad message is the AR elite's dominant strategy after any signal.

The bad R and the good and the bad AR politicians use propaganda for the same reason as in the baseline model, since for π' $(> \pi)$ propaganda yields a positive return to the politician. The good R politician, by avoiding propaganda, ensures that the voter's expected belief about her is at least the prior q_c. In contrast, if she sends successful propaganda, she forces the voter's expected belief below q_c: if the elite reports her good the voter will believe her bad with certainty, and if the elite reports her bad the voter will believe that she is bad with a higher-than $1 - q_c$ probability because more bad than good types send propaganda.

If the propaganda was successful, the bad R politician wants to make her common type more visible since

$$\mu_v(\theta_c|\hat{p} = 1, e = 1) - \mu_v(\theta_c|\hat{p} = 1, e = 0) = (\pi' - \pi)\frac{q_{ar}q_c}{q_{ar} + q_r(1 - q_c)\pi'} > 0$$

for $\pi' > \pi$.

However if the propaganda is unsuccessful then keeping her type less visible yields

$$\begin{aligned}&\mu_v(\theta_c|\hat{p} = 0, e = 0) - \mu_v(\theta_c|\hat{p} = 0, e = 1)\\ &= (\pi' - \pi)\left[\frac{q_c\pi}{q_c\pi + (1 - q_c)(1 - \pi)} - \frac{q_c(1 - \pi)}{q_c(1 - \pi) + (1 - q_c)\pi}\right],\end{aligned}$$

which is positive for large π if $\pi' > \pi$.

We have confirmed that in the proposed equilibrium the elite and the politician best respond, and we have characterized the beliefs and hence behavior of the voter. To conclude, we clarify that our above arguments also cover off-equilibrium information sets. Such information sets only happen at stage 2, i.e., after the message profile is realized: because propaganda determines the voter type, the normal voter after propaganda and the persuaded voter absent propaganda can never occur in this game. Still, Bayesian updating in our equilibrium definition, (6), specifies beliefs based on how a voter of the given type would update from the information content of the messages he observes. In particular, normal voter after propaganda believes that the politician is bad, while the persuaded voter believes that

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

These beliefs also pin down voting behavior.

The equilibrium is CPO. It is immediate that the equilibrium is coordination-proof, since the AR elite and the AR politician has opposing incentives. The equilibrium is also politician optimal. To see this, first notice that there is no equilibrium where the AR elite does not strictly prefer to send a bad message. Since in all equilibria, the AR elite criticize, no matter what signal she received, then manipulating the elite's signal is pointless, thus no AR politician makes her type visible. After the history of no (successful) propaganda the dominant strategy of the bad R politician is not to make her type visible.

In any equilibrium that is weakly better for the bad R politician than our preferred, she must be sending propaganda and must prefer to do so, as otherwise her payoff is lower than in our preferred equilibrium. However, if the bad R politician strictly prefers to send propaganda, then the AR politicians, facing an even higher chance of a bad message, should also prefer to use it. Finally, for π (and π') sufficiently high the good R politician always prefer to restrain from propaganda. So we conclude that in any equilibrium as good for the bad R politician as our preferred, bad R and the good and bad AR use propaganda with a probability one, while the good R politician does not use it. In any such strategy profile it is optimal for the bad R politician to make her common type more visible. This concludes that our preferred equilibrium is politician optimal.

Case 2: Incumbent politician is pro-elite.

Existence. We first show that the strategy profile in which (i) the AR elite always reports the politician good; (ii) no politician uses propaganda; (iii) no politician makes her common type visible; constitutes an equilibrium.

Beliefs of the normal, respectively, the persuaded voter is given by equations (7) and (9). Similarly to *Case 1*, the impact of a good report on the reelection chances of the politician is positive in all equilibria. Since the AR elite always reports good, then for π and π' high enough, no AR politician uses propaganda. This implies that the bad R politician does not use propaganda either since it would reveal her type.

Similarly to *Case 1*, since the AR elite's message is not contingent on her signal, there is no incentive for the bad AR to make her type more visible. Finally, since the bad R politician does not use propaganda, then she best responds by not making her type visible either.

This equilibrium is CPO. For π and π' high enough, no matter what the elite signal precision is, Lemma 1 implies that in any coordination-proof equilibrium no politician sends propaganda, and given that it is optimal for the AR elite to always send a good signal. Also, the bad R politician finds it optimal to keep her type less visible.

Proof of Proposition 3. *Voting behavior.* Consider voter i the reader of newpaper j. He votes for the incumbent if

$$c\mu_v(\theta_c|\hat{s}_j) + \lambda + \epsilon + \eta_i > cq_c + \lambda q_d$$

The incumbent wins the election if she gets the majority of votes

$$\frac{1}{N}\sum_j \{0.5 - h[c(q_c - \mu_v(\theta_c|\hat{s}_j)) - \lambda(1 - q_d) - \epsilon]\} > 0.5$$

The probability of winning is

$$\Pr\left[\frac{1}{N}\sum_j c(q_c - \mu_v(\theta_c|\hat{s}_j)) - \lambda(1 - q_d) < \epsilon\right].$$

This formula implies that a unit change in the beliefs of all voters result in a change in the probability of winning of cg, which is what we used in the definition of ΔU_e in the main text.

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

Case 1: $\chi_f \leq (1 - 2\alpha)\Delta U_e/N$. We begin by characterizing the behavior of some actors in any large-π equilibrium. We start with the observation that in the limit as π goes to one, in any profile, the utility gain to the elite from reporting bad after a good signal is at most ΔU_e. This is because in any profile the maximum they can move beliefs is by one, and the gain from doing so approaches $cg[\lambda(1 - q_d) - c(1 - q_c)]$ for π going to one. Assuming that the R elite's reputation costs are large enough, $\Delta U_e/N - \chi_f < \chi_r$, for large π the R elite is truthful in any profile. Given this, for π large enough, the good R politician does not send propaganda.

Because we are in case 1, we have

$$(1 - \alpha)\frac{1}{N}\Delta U_e - \alpha\frac{1}{N}\Delta U_e > \chi_f.$$

As π converges to one, the left hand side converges to a lower bound for the payoff to the AR elite from reporting bad after a good signal. This is because when no propaganda is observed, $(1/N)\Delta U_e$ is the limiting payoff to the AR elite from reporting a good politician bad; when propaganda is observed, reporting bad may change the voter's beliefs by at most one, and the effect of that on the elite approaches $cg[\lambda(1-q_d)-c(1-q_c)]$ for π going to one; and the probability that propaganda is observed is α. It follows that for π large, in any equilibrium, the AR elite in both ARs always criticizes the good politician.

Existence. 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 AR1 and AR2, respectively, is

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

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

The first expression is derived analogously to our basic model: propaganda and praise ($\hat{s}=1$) only happens in R when the politician is bad, while propaganda and criticism can happen in the AR, in R if the politician is bad, or in R if the politician is good but the signal is bad. In the second expression, the first term represents beliefs after observing AR2 propaganda and praise by the elite. Propaganda shifts the prior to put a positive weight on AR2. Because AR2 propaganda is not observed on the equilibrium path, it is attributed to a tremble and does not generate updating. But praise never occurs in AR2, thus the voter updates to reality being R, and then forms beliefs accordingly. The second term in the second expression is that same as the first expression.

These beliefs imply that on the equilibrium path, as π converges to one, the R politician's return to successful AR1 propaganda is the same as in the baseline model—proportional to $\hat{q}_c$—

which means that for large enough π propaganda is better than no propaganda. Moreover, AR1 propaganda is better than AR2 propaganda because the return to choosing AR2 is also proportional to $\hat{q}_c$ in the limit, but AR2 is more expensive than AR1. The same logic implies that the AR politicians also choose to send AR1 propaganda. This confirms that the proposed profile is an equilibrium.

The equilibrium is CPO. We show that any politician optimal pure strategy equilibrium has the same equilibrium path, and in fact, generically, this is the unique CPO equilibrium. Because the good R politician is getting the highest possible payoff, it suffices to focus on the expected payoff of the bad R politician.

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, with positive probability, refrains from propaganda, because here she strictly prefers to send AR1 propaganda and thus doing so improves her payoff.

Suppose that the bad R politician sends AR1 propaganda. Then the AR1 politician must also send AR1 propaganda, because otherwise observing AR1 would lead the voter 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 generically 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 voter 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 the AR1 politician sending AR1 propaganda. Thus, except for the knife-edge case of indifference, which is non-generic, if the AR2 politician prefers to send AR2 propaganda, then so does the AR1 politician, a contradiction. It follows that generically in any CPO equilibrium the AR2 politician sends AR1 propaganda, which is our preferred equilibrium.

Suppose that the 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. The fabrication cost is medium, $\Delta U_e/N < \chi_f < (1-2\alpha)\Delta U_e$. We begin by characterizing the behavior of some actors in any equilibrium. As in Case 1, we assume the reputation costs are large enough that the R elite is truthful, therefore the good R politician does not send propaganda. For π large the AR1 elite is also truthful, since given we are in Case 2, in the limit as π goes to one the maximal gain from changing the perception of her audience is smaller than her lying cost

$$\Delta U_e/N < \chi_f.$$

However, for π large the AR2 elite always sends a bad message after a good signal because

$$(1-\alpha)\Delta U_e - \alpha\Delta U_e > \chi_f$$

since we are in Case 2.

Existence. 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 focus on the behavior of the bad R and the AR politicians.

Observe that no politician sends AR1 propaganda since, as we just established, the AR1 elite is truthful, hence AR1 propaganda has no effect on the voter's interpretation of the elite's message, but it has a positive cost. However, sending AR2 propaganda is optimal for the bad R politician, because of the same arguments we applied in the proof of Proposition 1: by Assumption 1, for π sufficiently high the benefit of propaganda exceeds the cost. Given this, doing so is also optimal for the AR2 and for the AR1 politician.

The equilibrium is CPO. In any equilibrium better for the bad R politician that the one proposed here, she has to send AR2 propaganda: sending AR1 propaganda is not worth it, and the equilibrium proposed is better than that without propaganda. Given this, the AR2 politician must also be sending AR2 propaganda, otherwise the bad R politician's type is revealed by propaganda. Then the equilibrium path is the same as in the proposed equilibrium, and then the AR1 politician also prefers to send propaganda.

Case 3. The fabrication cost is large $\chi_f > \Delta U_e$. 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. Telling the truth is also the dominant strategy of the elite in both ARs since, by assumption, the gain from fully influencing the whole electorate is smaller than the fabrication cost. Since neither of the propaganda yields any gain, no politician chooses propaganda.

Proof of Proposition 4. First we show that there is an equilibrium in which the claimed properties hold. Consider the profile in which all principals of the baseline model behave as in the simple propaganda equilibrium. It is straightforward to verify that these are the dominant strategies of these principals. For the AR elite, criticizing is optimal because $\alpha < 0.5$, so that their effect on voter perception when propaganda does not come through always dominates any negative effect they may have on voter perception when it does. For the R elite, truth-telling is optimal because the new elite interferes with sufficiently low probability γ by the first inequality of Assumption 3. For the good R politician, no propaganda is optimal because they are earning the highest possible payoff. For the bad R politician, and for the AR politicians, propaganda is optimal because by the second inequality of Assumption 3 it is worth it even if it only improves beliefs in the event in which the new media is not active.

We continue constructing our equilibrium by characterizing the behavior of the new media. In doing this, it will be important to take trembles seriously. To start, suppose that the new media always praises the politician. This is not yet our equilibrium, but it is a step to construct it. In this profile, in the event of propaganda and a bad elite message, beliefs about the AR after a good new media message are given by

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

But beliefs after a bad new media message are also given by $\hat{q}_{ar}$ because a bad new media message is a tremble. Let $\Delta\mu$ be the belief gap, in a strategy profile, following propaganda and a bad elite message, between beliefs about AR after a good versus a bad new media message. The belief gap is zero in the profile we just considered. This also implies that the profile is not an equilibrium: the new media, after a bad signal, prefers to report honestly. Now start increasing the probability r_1 that the new media, in R, after a bad signal, reports bad. This will make it relatively more likely that a bad signal comes in the R, but given non-zero trembles does not prove that it does. As r_1 increases, the belief gap will increase, and when $r_1 = 1$, the belief gap becomes close to 1 for small trembles. Since, by Assumption 3, $\alpha/\chi > 1$, for small enough trembles there is an intermediate r_1 for which the belief gap is exactly χ/α. At that belief gap the new media is indifferent, and willing to randomize with interior probability r_1 in R after a bad message. This is an equilibrium that has the claimed properties: the new media sends a good message with positive probability in R after a bad signal, and a good message increases voter's belief about the AR.

Second, we show that in any politician-optimal equilibrium the properties claimed in the proposition hold. As we discussed above, in any equilibrium the principals of the baseline model behave as in the simple propaganda equilibrium. Now consider the new media. Suppose first that the new media, in the AR, after a good signal, sends a bad message with positive probability. Then the new media after a bad signal, in both the R and the AR, must strictly prefer to send a bad message, as doing so is less costly, and they expect the same message profile from the elite media and the politician. Now there are two cases. If the new media in AR after a good signal sends a good message with positive probability, then such a message proves that reality is AR, and hence the new media would strictly prefer to send that message after a good signal, a contradiction. If the new media in the AR after a good signal always sends a bad message, then there is never a good message in equilibrium, so that such a message would be interpreted as a tremble and lead to the same beliefs as a bad message. The good message is then strictly preferred by the new media after a good signal, again a contradiction. We conclude that the new media in the AR after a good signal always sends a good message.

It cannot be a politician-optimal equilibrium that after a bad signal, in R, the new media always sends bad message. Then the new media in R would always reveal the politician's common type, whereas in the equilibrium we just constructed it does not always do so, so our constructed equilibrium is better for the politician. Thus, in any politician-optimal equilibrium, the new media sends a good message with positive probability in R after a bad signal, confirming the first property claimed in the proposition.

Finally, since the new media in R is at least indifferent to sending a good message after a bad signal, it must be that voter beliefs about the AR are higher after a good than after a bad message to compensate the new media for the lying cost of the former. This proves the second property claimed in the proposition.

A.2 Evidence

A possible alternative explanation for the scandal effects documented by Table 4 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 Elections Commission. We estimate

$$y_i = \text{const} + \beta \Delta DVM_i + \gamma DVM_i^{old} + \varepsilon_i, \tag{10}$$

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.

	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