The Political Economy of Alternative Realities[*]

Adam Szeidl
Central European University and CEPR

Ferenc Szucs
Stockholm University

July 5, 2022

Abstract

We build a model in which a politician can persuade voters of an alternative reality that serves to discredit the intellectual elite. The alternative reality states that members of the elite conspire and criticize the politician because she disagrees with them on a divisive issue such as cultural values. The alternative reality is false, but if the voter believes it, he will distrust the elite's criticism of the politician. This model makes several predictions. (1) The alternative reality is spread by low-quality politicians and reduces accountability. (2) The nature of the divisive issue—cultural versus economic—determines whether right-wing or left-wing politicians spread alternative realities. (3) Once the elite has been discredited, the voter will not trust its advice even in unrelated domains such as climate change. (4) The politician will follow policies (e.g., anti-vaccination) that contradict the elite consensus even if she knows those policies to be universally harmful, to avoid the appearance of being in the elite conspiracy. (5) Discrediting the elite creates a niche in the media market filled by non-elite outlets (e.g., Fox news) which, because they are only demanded due to beliefs in the alternative reality, spread misinformation to reinforce those beliefs. We discuss evidence consistent with these predictions.

[*] Emails: szeidla@ceu.edu, ferenc.szucs@ne.su.se. We thank Ruben Enikolopov, Maria Petrova, Giacomo Ponzetto, David Stromberg and conference and seminar audiences for comments, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement number 724501.

1 Introduction

A majority of Republicans with high science knowledge believe, contrary to Democrats with high science knowledge and to the experts' consensus, that human activity does not contribute a great deal to climate change. Similarly, a majority of Republicans believe, contrary to Democrats and to the experts' consensus, that the 2020 U.S. presidential election was not conducted fairly and accurately. These sorts of misbeliefs are often held as part of a larger system of incorrect beliefs, which feature conspiracy theories and form a semi-coherent alternative reality. For example, 15% of Americans believe, and a full 79% of Republicans do not reject outright, that the government, media and financial worlds in the U.S. are controlled by a cabal of Satan-worshipping pedophiles.[1] Beliefs in such alternative realities are likely to be highly consequential for economics and politics. But their causes, mechanisms, and precise implications are not well understood.

In this paper we build a model in which politicians can persuade voters of an alternative reality that ends up influencing their beliefs about a range of issues. Our approach builds on prior work about misinformation in politics, especially Glaeser (2005), Guriev and Treisman (2020), and Ash, Mukand and Rodrik (2021), and contributes by explicitly modeling an internally consistent and strategically interacting alternative reality, which is introduced to discredit the criticism of the intellectual elite. In this alternative reality members of the intellectual elite—including the news media—conspire, and criticize the politician about a commonly important issue such as competence if she disagrees with them about a divisive issue such as cultural values. The alternative reality is false, but if the voter believes it, he will distrust the elite's truthful revelation of the politician's competence. Key to our approach is that the voter thinks coherently about, and responds strategically to, the alternative reality he believes in. This assumption generates a range predictions about politics, the media, and the non-adoption of societal best practices (e.g., in climate and health), which are consistent with evidence we discuss.

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

[1] For beliefs about climate see Funk and Kennedy (2020), for beliefs in conspiracies see Public Religion Research Institute (2021).

agent.[2] Both principals can send messages to influence the voter's electoral behavior, and the voter then decides whether to keep or replace the politician. The politician has two payoff-relevant types. (i) A "common" type, such as quality or honesty, over which the voter and the elite have the same preference, and along which the politician can be good or bad. The voter does not observe this type dimension. (ii) A "divisive" type, over which the elite and the voter have different preferences, and along which the politician can be pro-voter or pro-elite. The two leading examples of the divisive type are cultural values (where a pro-voter politician is right-wing) and economic redistribution (where a pro-voter politician is left-wing). 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 perception of it. The elite sends a message which simply reports whether the politician is 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 convinced by the politician assigns 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 and 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 truthfully the politician's common type. 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

[2] 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.

the election outcome. Intuitively, in the alternative reality the New York Times criticizes Trump not because he is incompetent, but because they disagree on cultural values. In turn, the voter, since propaganda persuaders him to partially believe in the alternative reality, understands this mechanism and distrusts the report of the elite.

A key assumption underlying this framework is that propaganda can "irrationally" manipulate voter's prior beliefs about the elite. This assumption is consistent with well-identified evidence we discuss at the end of the Introduction about the impact of propaganda on a range of outcomes; as well as new suggestive evidence we present in Section 2 that during the past decades of anti-intellectual Republican rhetoric, Republicans' trust in science decreased substantially; and that across 26 European countries, the populist vote share predicts distrust in the media. In addition, the logic of the alternative reality—that elite members act collectively to advance their own goals —is consistent with the narrative of numerous conspiracy theories (Douglas, Uscinski, Sutton, Cichocka, Nefes, Ang and Deravi 2019).

The main result of our model is that the politician will send propaganda if and only if she is bad on the common dimension and pro-voter on the divisive dimension. Intuitively, because she is bad, she can gain from discrediting the elite's (objectively truthful) report that she is bad; and because she is pro-voter (not pro-elite), the alternative reality in which the elite lies to remove her is believable to the voter. In contrast, the good politician does not gain from discrediting the elite which truthfully reports her good; and the pro-elite politician cannot gain from discrediting the elite which is on her side.

The main result has several implications. Most directly, it implies that bad politicians are more likely to use propaganda and that doing so enables them to stay in power. This implication is consistent with the description in Guriev and Treisman (2022) of informational autocracy in countries such as Putin's Russia, Orban's Hungary, Erdogan's Turkey, or Fujimori's Peru, in which autocratic—interpreted as bad in our model—leaders use propaganda to stay in power. Differently from Guriev and Treisman's account, in which propaganda works by improving beliefs about politician, in our model propaganda works by creating distrust in the elite, a mechanism consistent with the suggestive evidence we mentioned above that populism in both the U.S. and Europe is associated with such distrust.

A second implication of the main result is predictable variation in whether propaganda is left-wing or right-wing. Our result says that propaganda is only used by the *pro-voter* politician. Thus, when the main divisive issue between voters and the elite is cultural values—on which the voter is plausibly to the right of the elite—the pro-voter politician is right-wing and we predict right-wing populism. In contrast, when the main divisive is economic redistribution—on which the voter is to the left of the elite—we predict left-wing populism. Intuitively, the alternative reality in which the elite criticizes because of the divisive issue is only credible if the elite and the politician are on opposing sides of that issue. We present new cross-country evidence consistent with this prediction: we document that cultural disagreement between the high- versus low-educated predicts right-wing (but not left-wing) populism, while economic disagreement between the high- versus low-educated predicts left-wing (but not right-wing) populism.

A third implication is that once the elite has been discredited, the voter does not want to follow its advice even in non-political domains, fearing that the elite's messages in those domains too are coordinated and driven by its interests. Thus, the model predicts that propaganda creates distrust in the scientific consensus, and leads to the non-adoption of scientifically approved best practices. Consistent with this prediction, we document that Republicans are less likely to vaccinate against Covid or believe in climate change, and Allcott, Boxell, Conway, Gentzkow, Thaler and Yang (2020) document that they are also less likely to engage in social distancing.

In Section 3 we develop two applications of the model. The first application investigates the effect of the alternative reality on the quality of governance. Our framework makes the sobering prediction that politicians spreading alternative realities will not adopt policies supported by the scientific consensus (e.g., mask mandates), even if they know that non-adoption is universally harmful. Intuitively, such politicians prefer to avoid praise from the discredited elite. To formalize this intuition, we add a new stage to the model which requires the politician's competence along a new dimension, such as Covid containment policies. We show that if the politician has undermined trust in the elite, then getting praise from the elite about the new competence dimension will lead the voter to increase his belief that the politician is also part of the conspiracy (formally, that the politician's divisive type has switched). The politician will then choose to act incompetently to trigger the criticism of the elite and thereby maintain the support of the voter.

Consistent with this logic, we document that—controlling for the severity of Covid—Republican governors were less likely than Democrats to introduce mask mandates or vaccinate publicly. More generally, because addressing health or environmental problems often requires governmental adoption of best practices, the mechanism highlighted here can generate large social costs.

Our second application is motivated by the salient fact that several non-traditional media outlets, most prominently Fox News, seem to supply and reinforce alternative realities. This fact seems unexplained 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 reinforce alternative realities. Our model provides an explanation based on the idea that the more discredited the elite media, the more voters look for other sources of information, and the higher the demand for non-traditional media. To formalize this idea, we add a new media outlet to the model which is pro-voter along the divisive issue and hence cannot be in the conspiracy. We show that this new outlet can create demand for itself by strengthening beliefs in the alternative reality, e.g., by falsely reporting that the populist politician is good on the common dimension. Beyond explaining why non-traditional media reinforce propaganda, this framework predicts that such media reduce trust in the elite and limit the adoption of best practices. Consistent with these predictions, evidence shows that the consumption of Fox News reduced social distancing, and the consumption of some Fox programs increased the number of Covid deaths (Bursztyn, Rao, Roth and Yanagizawa-Drott 2020, Simonov, Sacher, Dubé and Biswas 2020).

Our paper builds on a literature studying misinformation in politics. Foundational contributions include Glaeser (2005) on the supply of hatred and Besley and Prat (2006) on media capture. Closer to our paper, Ash et al. (2021) study the supply of "worldview politics" which alter voters' understanding of how the world works; and Guriev and Treisman (2020) study a model of "informational autocracy" in which politicians use propaganda to convince the public of their competence. We contribute to this work by formalizing misinformation with a strategic model of an alternative reality which serves to discredit the elite, and with the implications about politics, media, and the non-adoption of practices. A complementary approach to persuasion in politics is based on the Bayesian persuasion model of Kamenica and Gentzkow (2011). We depart from that approach by allowing propaganda to manipulate priors in a non-Bayesian way, but preserve the requirement that the agent reasons given those priors in a Bayesian fashion. Empirical work on misinformation includes studies documenting the impact of propaganda on genocide, extremism, inter-ethnic attitudes and immigration (Yanagizawa-Drott 2014, Adena, Enikolopov, Petrova, Santarosa and Zhuravskaya 2015, Blouin and Mukand 2019, Barrera, Guriev, Henry and Zhuravskaya 2020). This evidence supports our key assumption that propaganda can influence voter beliefs.

Our focus on differences in beliefs builds on a literature studying disagreement and polarization in learning, including studies of confirmation bias (Rabin and Schrag 1999), echo chambers (Golub and Jackson 2012), limited memory (Fryer, Harms and Jackson 2018) and inattention (Che and Mierendorff 2019). Closer to our work, Gentzkow, Wong and Zhang (2021) study a dynamic model in which learning shapes trust in news sources and small biases accumulate. Our contribution to this work is a new theory of disagreement based on an explicit model of the alternative reality, and the implications for politics, media, and the adoption of best practices.

Our exploration of the implications of a divisive issue builds on work studying how economic and cultural cleavages generate populism and identity politics, including Acemoglu, Egorov and Sonin (2013) Bonomi, Gennaioli and Tabellini (2021), Besley and Persson (2021), and especially Levy, Razin and Young (2022), who study a model in which one group of voters have a too simple theory of the world. Our contribution is to show how cleavages can be exploited with an alternative reality that serves to discredit the elite, and the implications for politics, media, and the adoption of practices.[3]

Finally, our modelling approach builds on literature studying learning and strategic interaction in the presence of model misspecification, including Berk (1966), Eyster and Rabin (2005), Esponda and Pouzo (2016) and Heidhues, Kőszegi and Strack (2018). Our contribution to this work is to incorporate players' reasoning about their misspecified opponents' motives, and to study the incentives of players to create misspecification.

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

2 A model of the political supply of alternative realities

2.1 Setup

We build a principal-agent model in which two principals, the intellectual elite and the politician, attempt to influence an agent, the voter. The basic framework is the following. The intellectual elite, e.g., the news media, observes whether the politician is good or bad along a dimension commonly important to both the voter and the elite (e.g., quality or corruption). The elite sends a message about this observation to the voter. Simultaneously, the politician can choose to send propaganda to manipulate the voter's interpretation of this message. Based on the report from the elite and the propaganda from the politician, the voter decides whether to reelect the politician or choose an alternative politician randomly drawn from the prior distribution.

We introduce alternative realities to this framework by allowing the voter to entertain two theories of the world, which are formalized through the "reality type" of the principals, R (reality) or AR (alternative reality). The R and AR principals differ in their preferences and beliefs. Importantly, the objective reality is R, and the AR principals do not actually exist: their objective probability is zero. Our key assumption is that propaganda makes the voter believe with positive probability in the AR principals. This belief generates strategic interaction between the voter and the (imagined) AR principals, influencing the voter's objective behavior, and through that others' behavior and outcomes.

Neoclassical setup. To describe the precise model, it is helpful to start by presenting the neoclassical (non-behavioral) part of the setup. 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 election outcome.[4] As we will see below, because of symmetry, in the analysis of the game we can represent all elite members, and all voters, as a single actor each. We let i stand for any

[4] More generally we could allow that each elite member has a zero measure of voters as its audience.

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 type dimension 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" or of the "high type", who increases the voters' and elite members' per capita consumption by c. We assume that θ_c is only observed by members of the elite, but not by the voters. The politician's second type dimension represents a *divisive* 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 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 politician's common type, each elite member j sends a message $s_{cj} \in \{0,1\}$ to its voter, where $s_{cj} = 1$ means that the politician's common type is good. We sometimes refer to the message $s_{cj} = 0$ as criticism. Simultaneously, the politician decides whether to send propaganda $p \in \{0,1\}$ to the voter. Each voter observes the message of its elite member and of the politician, and then decides whether to vote to reelect the politician. If the politician is not reelected, a new politician is drawn from the prior distribution of objective types. Note that in this neoclassical version of the model propaganda plays no role.

We make the assumption that both elite messages s_{cj} and propaganda p are subject to vanishing noise, and that the noise affecting the elite's message is vanishingly smaller. These assumptions serve two roles: they ensure that beliefs are well-defined off the equilibrium path, and that the elite's message contains information over and above the propaganda message.[5] Formally, with probability ε_e, perfectly correlated across elite members, every elite member's realized message $\hat{s}_{cj}$ is the opposite of the actual message s_{cj} intended to be sent; and with independent probability ε_p the realized propaganda message $\hat{p}$ is the opposite of the actual propaganda p sent. We assume that ε_e, ε_p and $\varepsilon_e/\varepsilon_p$ all go to zero, and characterize the equilibrium in the limit.

[5] In particular, even when the elite's and the politician's actions are perfectly correlated, as they will be in equilibrium, the elite's message is strictly more informative.

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. Politician and Elite		
Reality (θ_r)	R (q_r), AR (q_{ar})	AR=Alternative reality
C. Voter		
Mind (θ_m)	N (if $p = 0$), P (if $p = 1$)	P=persuaded by propaganda

Table 1: Types and interpretations

Alternative reality. We formalize the alternative reality through (i) types for the principals that represent their motives in the reality (R) and in the alternative reality (AR), and (ii) types for the agent that represent the probability they assign to the alternative reality. The logic is that the alternative reality (AR) types have zero objective probability, but that the agent, if reached by propaganda, will assign these types positive probability. Here we introduce the types and their beliefs, and below we define their preferences. Concerning the principals, we assume that the politician and all members of the elite have the same reality type $\theta_r \in \Theta_r = \{R, AR\}$ 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 reality types, the AR principals' priors are correct. Concerning 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 the AR principals is zero; the persuaded voter thinks that the prior probability of the AR principals is $q_{ar} > 0$, while that of the R principals is $q_r > 0$ (with $q_r + q_{ar} = 1$). The voter's initial mind type at the beginning of the game, θ_m^0, is normal, while his eventual mind type, θ_m, is normal if the voter is not reached by propaganda and persuaded otherwise. We define the model's type vector to be $(\theta_d, \theta_c, \theta_r, \theta_m) = \theta$. The type dimensions and their interpretations are summarized in Table 1.

Our model also makes a second departure from rationality, which concerns the internal logic of alternative reality, and which allows the persuaded voter to "see through" the conspiracy of the AR elite. We assume that members of the AR elite think propaganda is ineffective: formally, the AR elite believes that the voter's type after propaganda remains normal. This assumption, which we refer to as naivete, ensures that the AR elite is not aware that the voter understands her incentive to manipulate, and allows the voter to interpret the AR elite's message as a direct reflection of the AR elite's preference for the election outcome.

Preferences. We begin with the preferences of the intellectual elite. In both R and 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 $c > 0$ is the benefit from a good politician, and $\lambda > 0$ can be thought of as the product of the importance of the divisive issue times the extent of misalignment in preferences—the difference between the ideal points—of the elite and the voter. 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 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 one AR elite member's message determines all others' messages. It follows that members of the R elite, because they influence a single voter and have no impact on the election outcome, always send a truthful message. In contrast, the AR elite, because its members coordinate and can influence voters, acts as a single strategic player that maximizes the utility function (1). In both cases, members of the elite send the same message which we denote s_c. Moreover, for the purposes of characterizing behavior, we can represent the elite as a single player which maximizes

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

The preferences of the politician, independently of her type, are characterized by the utility

[6] We omit preferences about the incumbent politician as her type cannot be changed by actions in the model.

function

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

where E measures ego utility from being in power after the election, and f is the cost, in the present, of engaging in 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), and ϵ is common a mean-zero uniformly distributed popularity shock with support $[-\bar{g}, \bar{g}]$ and constant density $g = 1/(2\bar{g})$. We assume $\bar{g} > c + \lambda$ so that with positive probability the popularity shock dominates the utility from any realization of the common or divisive type. Note that ϵ only affects the preferences of the voter, not those of the elite or the politician. Because their preferences are identical, we focus on equilibria in which voters behave in the same way and represent them as a single actor.

Timing. The timing of events is the following.

0. The politician's type is realized. The voter observes her divisive type θ_d, the elite also observes her common type θ_c.

1. The elite sends message $s_c \in \{0, 1\}$ and the politician decides on propaganda $p \in \{0, 1\}$. Both messages are subject to trembles and all actors observe the realized messages $(\hat{s}_c, \hat{p})$.

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.

We refer to these periods as the stages of the game.

2.2 Strategies, beliefs and equilibrium

We define the politician's type to be $\theta_p = (\theta_d, \theta_c, \theta_r)$. Because the elite has access to the same information as the politician, it will be convenient to define the elite's type to be $\theta_e = \theta_p$. 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 $\hat{h}^t = (\hat{a}^1, ..., \hat{a}^t)$.

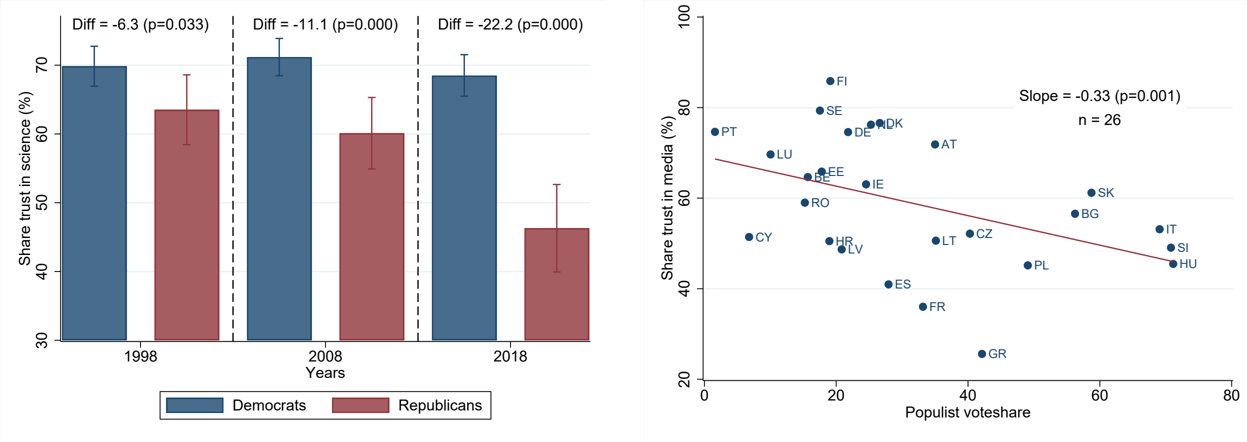
We define strategies as probability distributions over actions at the stages where an actor gets to move. Because the politician and the elite only move in stage 1, their strategies only depend on their type, and are denoted by $\sigma_p(a_p^1|\theta_p)$ respectively $\sigma_e(a_e^1|\theta_e)$. As the voter moves in stage 2 after observing $\hat{a}^1 = (\hat{s}_c, \hat{p})$, 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 departures 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, its understanding of the game, and the strategy profile generate a probability distribution over final outcomes. This distribution can differ from the objectively correct distribution because of our two departures from rationality: (i) the persuaded voter has an incorrect prior about θ; and (ii) the AR elite falsely believes that after propaganda the voter's type remains normal. 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 best-response property of equilibrium is

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

Belief consistency does not impose any condition on principals, because they move only at

Figure 1: Distrust in the intellectual elite



stage 1 where they know only their priors.[7] Belief consistency for the voter requires that he follows Bayesian updating at the end of stage 1:

$$\mu_v^1(\theta_p|\theta_v, \hat{a}^1) = \frac{\mu_v^0(\theta_p|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{a}^1|\theta_p)}{\sum_{\theta_p'} \mu_v^0(\theta_p'|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{a}^1|\theta_p')} \tag{5}$$

where $\mu_v^0(.|\theta_v)$ is the prior of the voter of type θ_v. This definition accounts for the model's first 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, (10) 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.

2.3 Discussion of model assumptions

Departures from rationality. Our first departure from rationality is that propaganda makes the voter assign positive probability to a non-existent alternative reality. The assumption that propaganda affects beliefs is consistent with evidence from different contests that propaganda affects behavior

[7] It is straightforward to characterize the beliefs of principals at all stages, because they either know or (in the case of the AR elite) think they know all types. If the true type profile after history $\hat{h}^{t-1}$ is $\theta^* = (\theta_d^*, \theta_c^*, \theta_r^*, \theta_m^*)$ then the politician and the R elite believe $\mu_i(\theta^*|\theta_i, \hat{h}^{t-1}) = 1$ while the AR elite believes $\mu_e((\theta_d^*, \theta_c^*, \theta_r^*, N)|\theta_e = AR, \hat{h}^{t-1}) = 1$.

and attitudes (Yanagizawa-Drott 2014, Adena et al. 2015, Blouin and Mukand 2019, Barrera et al. 2020). The assumption that it affects beliefs about the elite is consistent with evidence from both sides of the Atlantic. Figure 1A plots the share of Democrats and Republicans—controlling for demographics—who trust in science for 1998, 2008 and 2018, and shows that while Democrats' trust in science remained largely unchanged, Republicans' trust declined during this period of increasing anti-intellectual rhetoric by the Republican party. And Figure 1B shows that across 26 European countries the vote share of populist parties is a strong predictor of distrust in the media.[8] Moreover, the logic of the alternative reality, that elite members act collectively to advance their own actions, is consistent with the narrative of numerous conspiracy theories (Douglas et al. 2019).

An alternative rational modeling approach, paralleling Glaeser (2005), would be to let the elite conspire with positive prior probability, and let the politician's message contain true information about the conspiracy. However, we believe that such a Bayesian model would be calibrationally off: if the prior of AR is small, then the politician's message would have only a small effect on beliefs, while if the prior of AR is large then there should be more conspiracies in reality.

Our second departure from rationality is that the AR elite fails to realize that the voters are persuaded. This assumption reflects our sense that actors in a conspiracy theory are often to some extent caricatures, and that conspiracy believers often think they are in possession of rare information (Douglas et al. 2019); and allows voters to "see through" the conspiracy and interpret the AR elite's message as a direct reflection of her preference. This assumption is substantively used only in our applications, and is made for convenience in the basic model.

Specifics of the model. Beyond the above departures from rationality, our model makes several specific assumptions. First, we create an AR type not only for the conspiratorial elite but also for the politician, and assume that the AR politician believes in the conspiracy. We do this for two reasons: to make the alternative coherent by requiring that if the elite conspires the politician is aware of it; and because it seems plausible that the politician spreading propaganda would want to communicate that she believes in it. Second, we assume that propaganda changes the voter's belief

[8] In Figure 1A we use data from the General Social Survey and control for age, gender, race, and years of schooling. In Figure 1B we compute the share of people who trust in media from the 2016 wave of Eurobarometer, and the vote shares of populist parties from popu-list.org, using the highest populist vote share between 2009 and 2020 in any national or EU parliamentary election.

about the elite, but not about the common type of the politician. We do this because changing the belief about the common type, absent changing it about the elite, would not be effective: the elite's message about the common type would correct beliefs. Third, we assume that the (R) elite is truthful. This is a natural point of departure since the puzzle we want to explain is that voters trust the opinion of the elite too little. However, the key to our story is not that elite always tells the truth, but that it does not act in a coordinated way to advance its goals. Fourth, we assume that only the politician, but not the elite, can move priors. This assumption is natural since the politician is a single decision maker while the elite is atomistic, and consistent with our sense that it is easier to create than to eliminate beliefs in a conspiracy theory, especially by the supposed conspirators.

Equilibrium concept. As is standard in economics, our equilibrium concept assumes that actors know each others' strategies. In our setting, one natural justification is based on introspection and persuasion: the persuaded voter may reason about the behavior of AR principals based on his understanding of the game, and his reasoning may be helped by the propaganda-spreading politician who may spell out the AR elite's optimal strategy in order to persuade the voter. This foundation seems especially plausible when the equilibrium is unique, which will be the case in our setting.

An issue with our equilibrium concept is that it requires the voter to think sophisticatedly about the R politician who spreads alternative realities, while at the same time the voter thinks in simplified terms about the AR elite. This tension can be resolved in two ways. First, the voter's sophisticated thinking may be helped by the politician, who may explain that the elite's conspiracy is the reason for the elite's criticism, effectively revealing why sending propaganda about that conspiracy is beneficial to the politician. Second, the alternative modeling approach in which the voter has the additional bias that he does not make Bayesian inference from the presence of propaganda—effectively ignores that propaganda is sent for strategic reasons—is immune to this tension and we believe would generate the same predictions as the current model. That alternative approach feels more realistic to us, but we chose our current model to keep more closely with the economics tradition.

Real-world analogues of model components. We discuss real-world examples for the common and divisive issue and for the trembles. For the common issue, natural examples include general competence in governing and (the absence of) corruption. For the divisive issue we have two leading examples. In the first, which fits the U.S. and some European countries, the divisive issue represents a collection of cultural concerns related to the treatment of disadvantaged groups, including racism, women's rights, LGBT rights, immigration. In this example the (median) voter is culturally conservative and the elite is culturally liberal. In the second example, which fits some Latin-American countries, the divisive issue is in the economic domain and represents redistribution. In this example the (median) voter is economically liberal while the elite is economically conservative (less in favor of redistribution). Finally, elite trembles can represent elite members observing the same slightly noisy signal about the politician's common type, and propaganda trembles can represent that the propaganda campaign is unsuccessful or that an information campaign unexpectedly acts as propaganda.

2.4 Results

The main result of the model is that under some conditions, (in reality) only the bad pro-voter politician uses propaganda. In preparation for stating this result, consider the strategy profile in which (a) the R politician sends propaganda if and only if her common type is bad, while the AR politician sends propaganda always, and (b) the R elite reports honestly while the AR elite criticizes always. As we show below, this will be the key equilibrium path in the model. If the voter were always normal, i.e., did not entertain the alternative reality, in this profile he would always learn from the elite's message and the presence of propaganda that the politician is bad. But the persuaded voter, who assigns positive probability $q_{ar} > 0$ to the alternative reality, believes conditional on observing propaganda and criticism that the politician is good with probability

$$\hat{q}_c = \mu_v(\theta_c = 1 | \hat{p} = 1, \hat{s}_c = 0) = \frac{q_c q_{ar}}{q_c q_{ar} + (1 - q_c)}. \tag{6}$$

Note that $0 < \hat{q}_c < q_c$: the voter updates from propaganda and criticism only partially, because he expects both messages even for the good politician in the alternative reality. After this preparation, we are ready to state the key conditions for our main result.

Assumption 1. For the bad pro-voter politician, the benefit of partially hiding her common type is higher then the cost of propaganda:

$$E\hat{q}_c cg > f.$$

Recall that E is the utility from being in power, $\hat{q}_c$ from (6) 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. This condition makes the conspiracy theory profitable to the bad politician: It states that if propaganda protects her from the full impact of elite criticism, the politician is willing to use propaganda.

Assumption 2. The elite wants to remove all types of the pro-voter politician but keep all types of the pro-elite politician:

$$(1 - q_c)c < (1 - q_d)\lambda \text{ and } q_c c < q_d \lambda.$$

This condition makes the conspiracy theory believable—in case the politician is pro-voter—to the voter: It states that the disagreement λ on the divisive issue is large enough that, were it able to conspire, the elite would act to remove the pro-voter politician (and keep the pro-elite politician) irrespective of her common type.

Proposition 1. *Under Assumptions 1-2, in the unique equilibrium*

1. *In the reality (R):*
   - *The elite reports on the common type truthfully,*
   - *The politician sends propaganda if and only if she is pro-voter and bad.*
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.*
3. *Propaganda increases the reelection probability of the bad pro-voter politician.*

All proofs are in the Appendix. The core result is that, in the objective reality, only pro-voter bad politician uses propaganda. Intuitively, because the politician is pro-voter, by Assumption 2 the voter finds it believable that the elite, were it able to conspire, would act to remove her; and because she is bad, by Assumption 1 she would benefit from discrediting the message of the elite. It follows that for the pro-voter bad politician propaganda is both believable and profitable: it convinces the voter that the elite may conspire, results in the voter not fully believing the elite's message that the politician is bad, and reduces the probability that the politician is voted out. In contrast, the good R politician and the pro-elite politician never send propaganda: the former benefits from the elite's (truthful) message so has no incentive to discredit, while the latter cannot benefit from discrediting the elite which is on her side.

To more fully understand the equilibrium, consider the behavior of the other actors. The R elite is atomistic, cannot influence the voter, and hence prefers to report truthfully; while the AR elite, since by Assumption 2, the conflict over the divisive issue is sufficiently large, always wants to remove the pro-voter politician and hence always reports her bad. The pro-voter AR politicians (both good and bad), as they believe that the elite is AR and hence always reports them bad, use propaganda to partially deflect this criticism. The pro-elite AR politicians (both good and bad) believe the elite is AR and hence always reports them good, and see no reason to use propaganda that calls attention to this fact. Finally, consider the behavior of the voter. Key to the result that propaganda works is that the voter does not infer from observing it that the politician is bad. The reason for this is that after propaganda the voter assigns positive probability to the alternative reality, and then—since both the good and bad (pro-voter) AR politicians choose propaganda—updates from propaganda about the politician being good only partially (to $\hat{q}_c$), not fully (to zero).

Having characterized the equilibrium path, the prediction that propaganda increases the re-election probability of the bad pro-voter politician follows directly. On the equilibrium path, if propaganda fails because of a tremble, the voter remains normal and will correctly interpret the elite's message that the politician is bad; but if propaganda succeeds, the voter becomes persuaded and will put positive probability on the voter being good and reality being AR.

2.5 Implications

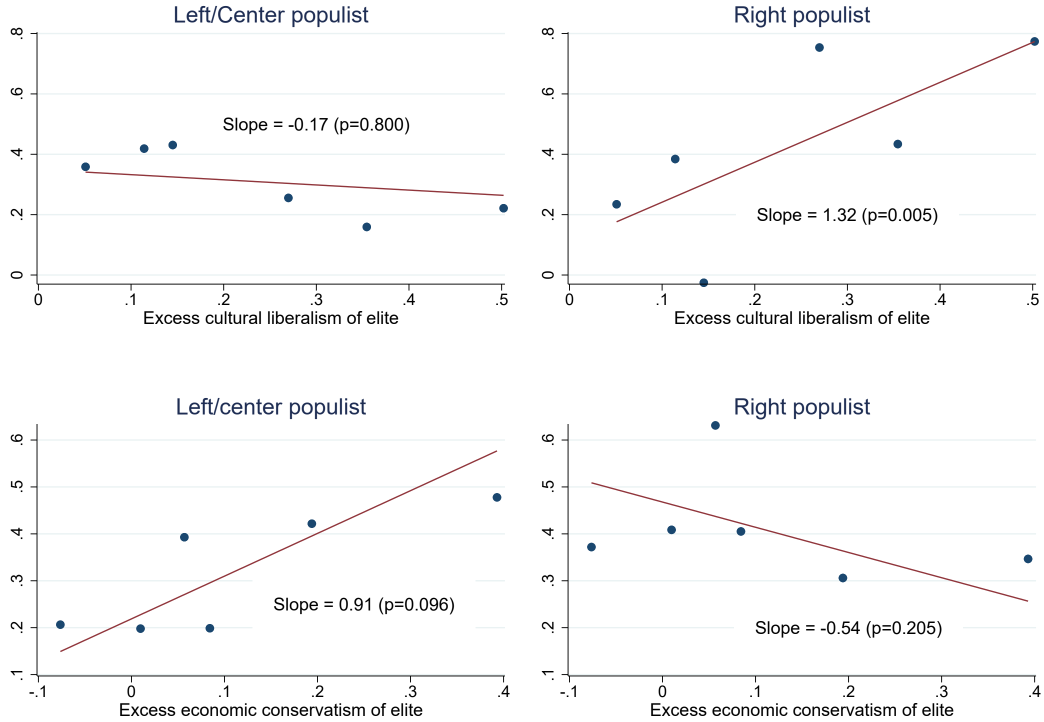
Proposition 1 has several implications which we now discuss.

Propaganda lowers accountability. An immediate implication is that propaganda increases the re-election probability of the bad (pro-voter) politician and hence lowers accountability. This result is broadly consistent with Guriev and Treismann's (2022) account of informational autocrats: autocratic leaders—which we interpret as bad in the common type dimension—who stay in power by means of government propaganda. Guriev and Treisman highlight Putin in Russia, Orban in Hungary, Erdogan in Turkey and Fujimori in Peru as some of the prime examples of such leaders. In the Guriev and Treisman (2019) model, propaganda works by improving the voters' beliefs about the politician's type. In contrast, in our model propaganda works by discrediting the elite, so that the elite's truthful message has limited influence on the voter. In our model, propaganda about the politician's type, absent discrediting the elite, would not work: the elite's truthful message would immediately correct beliefs. Thus our model suggests that discrediting the elite is a necessary first step before propaganda about the politician's type can be effectively employed. Consistent with this logic, populism on both sides of the Atlantic is associated with lower trust in the elite as shown in Figure 1 above.

Propaganda only used by the pro-voter politician. A second key implication is that propaganda is only used by the pro-voter politician. Since what constitutes pro-voter (versus pro-elite) differs across contexts, this result predicts variation in the nature of populism. In particular, when the major conflict between voters and elite is in the cultural domain, related to equal rights to minorities, so that the elite is more liberal than the voter, we expect propaganda from right-wing politicians. But when the major conflict between voters and elite is in the economic domain, related to redistribution, so that the elite is more conservative than the voter, we expect propaganda from left-wing politicians.

Figure 2 presents evidence on this prediction. We use the 7th wave of the World Values Survey to compute measures for 29 democratic countries of the excess cultural liberalism and economic conservatism of the elite versus the non-elite. Our measure of excess cultural liberalism of the elite is the gap in attitudes of people with versus without a masters degree on a set of issues such as

Figure 2: Cultural disagreement predicts right-wing, economic disagreement left-wing populism



immigration and gender inequality. Our measure of excess economic conservatism is the analogous gap in attitudes about income inequality. We correlate these measures with the presence of populist parties as classified by the Global Party Survey (GPS). The top row of Figure 2 documents that the excess cultural liberalism of the elite (controlling for excess economic conservatism) is uncorrelated with the presence of left or center populist parties, but strongly positively correlated with the presence of right-wing populist parties. The bottom row documents that excess economic conservatism of the elite (controlling for excess cultural liberalism) is strongly positively correlated with the presence of left or center populist parties, but if anything, negatively correlated with the presence of right-wing populist parties.[9] These are exactly the patterns our model predicts.

[9] We classify a party right-wing populist if the GPS classify it as an extreme populist party with conservative cultural values and a large focus on the cultural dimension of politics. We classify all extreme populist parties with a liberal cultural ideology as left/center populist. Examples of left/center populist parties include the Five Star

Through the same logic, our model suggests that the emerging salience of beliefs in alternative realities in American politics today may be driven by the growing disagreement on cultural issues between the voter and the elite. In essence, the growing cultural disagreement made it more believable to voters that the elite would want to misinform them about a culturally conservative politician.

Distrust and non-adoption of best practices in other domains. A broader implication of the results is that once the elite has been discredited, the voter will no longer trust it in other non-political domains either. A simple way to formalize this insight is to introduce a new action after stage 2—for example, vaccinating against Covid, or acting in a climate-conscious manner—that the voter can choose to take. The action may be good or bad for the voter with equal probabilities, and independently, good or bad for the elite with equal probabilities. The elite sends a message about whether the action is good for the voter. Payoffs are such that in response to a signal which is perfectly informative with probability $\hat{q}_r$ and uninformative otherwise, the optimal choice for the voter is to not take the action.[10] In this modified game the equilibrium of the base game is unchanged; and in realizations that do not involve propaganda the voter will make the optimal choice about the new action, while in realizations that do involve propaganda the voter will not take the action even when doing so would be beneficial to him.

This insight about propaganda limiting the adoption of best practices even in non-political domains is consistent with the beliefs and behavior of Republicans in the health and climate domains. Figure 3 illustrates this using evidence from U.S. counties. The left panel documents a strong negative relationship, controlling for demographics including education, between vaccination rates and Republican vote share, the right panel reports a similar negative relationship between belief in human-made climate change and the Republican vote share.[11]

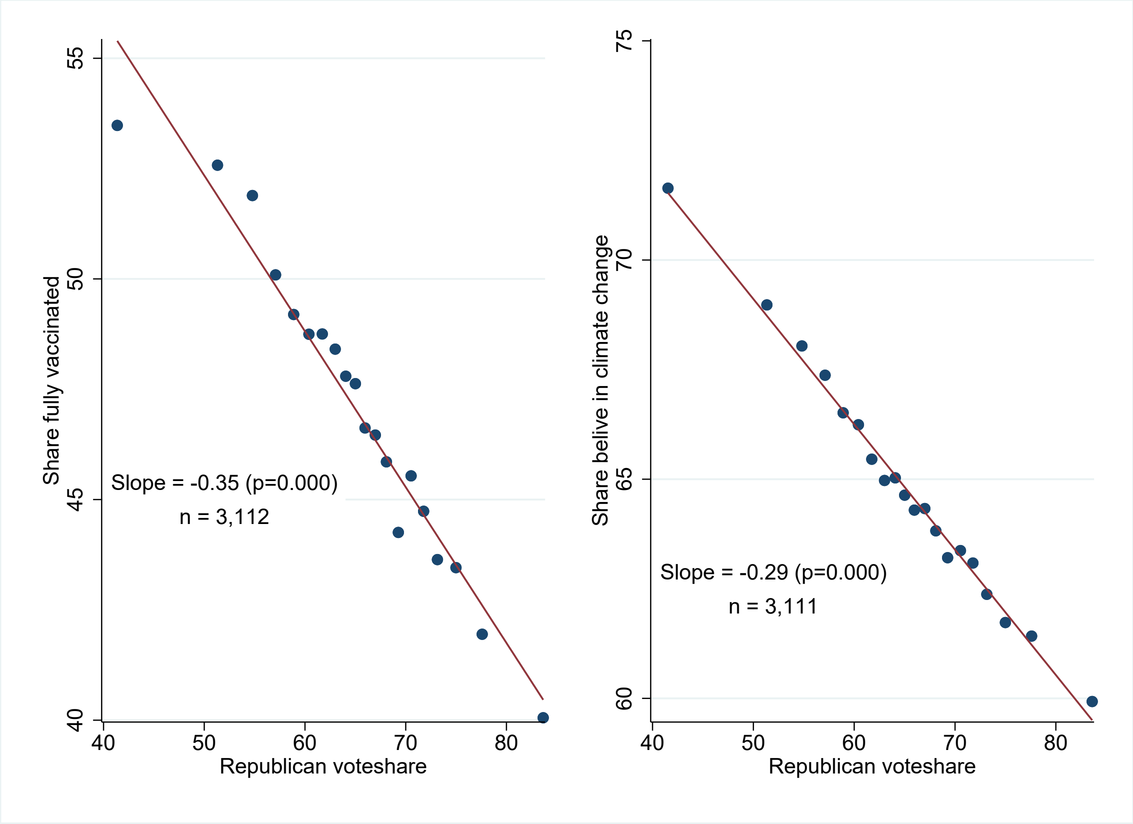
Discrediting versus censorship. From the perspective of the politician, an alternative strategy to discrediting is to silence criticism from the media using censorship (Guriev and Treisman 2020).

Movement in Italy or Syriza in Greece.

[10] Here $\hat{q}_r$ is the posterior probability of reality after observing propaganda on the equilibrium path.

[11] The Republican vote share is measured by the vote share of Donald Trump in the 2020 presidential election. We control for the share of residents with college education, median household income, unemployment rate and state fixed effects.

Figure 3: Distrust in experts in the health and climate domains



Our model offers some insights about this tradeoff. On the one hand, censorship is expensive, vulnerable to deviations by independent media, and may be politically costly or infeasible (Besley and Prat 2006). On the other hand, with censorship there is less need to discredit the elite and hence best practices are more likely to be adopted by the population. This logic makes the new testable prediction that citizens in China trust the scientific consensus more than citizens in Russia. The tradeoff between discrediting and censorship can change if the media gains access to irrefutable evidence which proves without doubt that the politician is bad: then the politician will have a stronger incentive to switch from discrediting to censorship. This logic can help explain why, following the invasion of Ukraine—which plausibly increased the probability of the media gaining access to difficult-to-refute evidence—Russia appears to be shifting towards full state control of the media.

3 Applications to government and media behavior

We develop two applications of our basic framework. First, we explore how propaganda-induced beliefs in the alternative reality constrain government policy in other domains. Second, we investigate how propaganda creates demand for alternative media outlets which reinforce the beliefs in the alternative reality.

3.1 Government policy

In this application we study how the political economy of alternative realities shapes the quality of governance. The core intuition is that her desire to maintain beliefs in the alternative reality constrains the politician in all domains about which the elite can express an opinion. In such domains, the politician has an incentive to follow policies that contradict the elite consensus, in order to avoid praise from the elite and the appearance of being part of the elite conspiracy.

To model this intuition, we add new stages to the model, which require the politician's competence about a new issue such as Covid containment. The politician is competent about this issue ($\theta_k = 1$) with a probability of q_k independently of all other type realizations, and his competence is realized only after the issue emerges. If he is competent, he increases per capita consumption by k relative to an incompetent politician. The politician chooses to act competently or incompetently: a competent politician can act incompetently but not the other way around.[12] As in the basic model, we introduce trembles to pin down beliefs: the politician's competence action (denoted by $\bar{\theta}_k$) is subject to a vanishing tremble, and the realized action after the tremble is denoted by $\hat{\theta}_k$.

In keeping with the notion that the voter learns about the politician's competence through the media, we assume that only the politician and the elite observe $\hat{\theta}_k$. The elite then sends a message s_k about $\hat{\theta}_k$ to the voter, and elite members prefer to be truthful about this message. This message too is subject to a tremble with vanishing probability and the realized message is denoted $\hat{s}_k$.

To capture that the voter may form doubts about the politician's independence from the elite conspiracy, we assume that with small probability ξ the politician's divisive type switches. Observe that in the alternative reality, when a pro-voter politician switches to being pro-elite, the elite media

[12] An indifferent politician prefers to break ties by acting competently.

suddenly wants to praise her: for this reason we view the switch as a metaphor for the politician getting roped in the conspiracy. The switch occurs simultaneously to the competence realisation, in both the R and the AR realities, and is only observed by the politician and the elite. We record the type switch in our notation by denoting the politician's initial divisive type by θ_d^0, and her eventual divisive type by θ_d.

The utility of the elite and the voter modifies to:

$$\begin{aligned} U_e &= 1_{\{\theta_r = AR\}} \cdot (k\tilde{\theta}_k + c\tilde{\theta}_c - \lambda\tilde{\theta}_d) + 1_{\{\theta_r = R\}} \cdot (1_{\{s_c = \theta_c\}} + 1_{\{s_k = \theta_k\}}) \\ U_v &= k\tilde{\theta}_k + c\tilde{\theta}_c + \lambda\tilde{\theta}_d + \epsilon \end{aligned}$$

where $\tilde{\theta}_k$, $\tilde{\theta}_c$ and $\tilde{\theta}_d$ are the competence, common and divisive types of the politician who wins the election. Consider the objective of the elite media. The first term, active when reality is AR, reflects the elite's collective policy preferences and it differs from the basic model by including the competence of the elected politician. The second term, active when the reality is R, captures the elite media's lying costs associated with both messages it sends. Voters' preferences also include the competence of the elected politician. The implicit assumption that voters care about the competence type (θ_k) and not the competence action ($\hat{\theta}_k$) of the future politician can be microfounded with a continuation game without further reelection concerns. We assume that the support of ϵ is large enough that all possible payoff realizations from the types are interior: $\bar{g} > k + c + \lambda$.

The timing of the game is as follows.

0. The politician's initial type is realized. The voter observes the divisive type θ_d^0, the elite also observes the common type θ_c.
1. Simultaneously, the elite sends message $s_c \in \{0, 1\}$ and the politician decides whether to send propaganda $p \in \{0, 1\}$. Both messages are subject to trembles. All actors observe $(\hat{s}_c, \hat{p})$.
2. The politician observes her final divisive type θ_d and her competence type θ_k.
3. The politician chooses her competence action, which is realized with a tremble: $\hat{\theta}_k$. The elite observes θ_d and $\hat{\theta}_k$.

4. The elite sends a message s_k on competence. All actors observe the message after a tremble: $\hat{s}_k$.

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

6. Payoffs realize.

We formally extend the equilibrium notion to this richer model in the Appendix.

Before stating the main result of this application, we formulate some assumptions.

Assumption 3. The elite wants to remove all types of the pro-voter politician but keep all types of the pro-elite politician:

$$(1 - q_c)c + (1 - q_k)k < (1 - q_d)\lambda$$

and

$$q_c c + q_k k < q_d \lambda.$$

This assumption is a revised version of Assumption 2 to the new setting, and plays an analogous role: it ensures the conspiracy theory is believable to the voter. It states that the disagreement λ on the divisive issue is large enough that, were it able to conspire, the elite would act to remove the pro-voter politician (and keep the pro-elite politician) irrespective of both her common type and her competence about the new issue.

Assumption 4. For the voter, a politician who is bad and competent is worse than a politician who is good with probability $\hat{q}_c$ and competent with probability q_k:

$$\hat{q}_c c > (1 - q_k)k.$$

This assumption essentially says that the politician's common type is more important than her competence about the new issue. More precisely, it is more important for the politician to (partially) hide that she has a bad common type than to signal competence about the new issue. This assumption is plausible because we think the common type captures the quality of the politician in a broad sense, while her competence about the new issue refers to one specific and idiosyncratic domain.

Assumption 5. The cost to the politician of being proven pro-elite is higher than the benefit of avoiding suspicion about being bad:

$$\lambda > (q_c - \hat{q}_c)c.$$

This assumption ensures that disagreement λ between the voter and the elite is sufficiently strong that the incentive to avoid appearing to be in the elite conspiracy constrains the politician. By acting competently, a pro-voter politician risks being perceived as captured by the elite. However, she also gains due to a subtle mechanism: she is only perceived as captured in the alternative reality, but in that reality the elite's first-stage message reporting her bad was completely uninformative, so the voter perceives her as more likely to be good. The assumption states that the cost of the former dominates the benefit of the latter.

Proposition 2. *Under Assumptions 1 and 3-5, for generic parameters and ξ sufficiently low, there is a unique pure strategy equilibrium path in which:*

1. *The first stage of the game unfolds as before.*
2. *In the reality (R), after propaganda:*
   - *The elite reports on competence truthfully,*
   - *The politician always acts incompetently.*
3. *In the alternative reality (AR), after propaganda:*
   - *The elite reports incompetence if and only if the politician is pro-voter,*
   - *The politician always acts competently if she can.*
4. *After propaganda, elite criticism about competence increases the reelection probability of the bad pro-voter politician.*

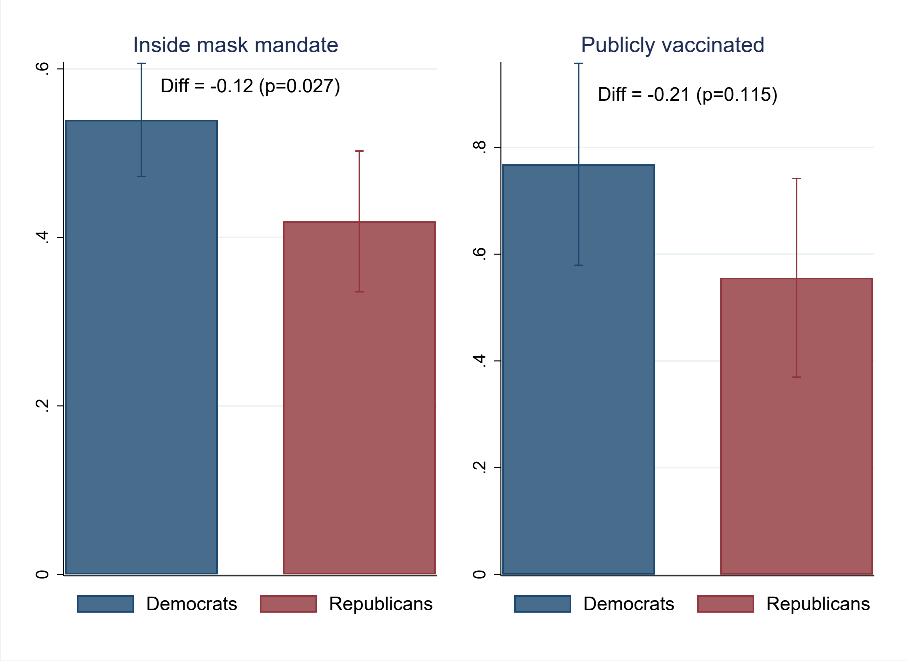
The key part of the result is that after spreading propaganda, the politician will always act incompetently about the new issue. To see why this is an equilibrium, suppose that the politician deviates and acts competently. The elite media reports this. But a message of competence can only come if reality is AR and the politician's type switched: it cannot come if reality is R because on path the politician acts incompetently, and it can only come if reality is AR after a type switch because then the elite reports strategically to support the politician if and only if she is pro-elite (Assumption 3). Thus the voter concludes that the politician is pro-elite. There is an additional subtle effect, that because the voter learns reality is AR, he perceives the politician to be less likely to be bad, since the elite's first-period message is now perceived uninformative. But Assumption 5 says that the cost of being pro-elite is higher than the benefit of being less likely bad, and therefore the politician chooses to act incompetently.

To see the intuition for uniqueness, note that in any equilibrium, the elite's message that the politician acted competently can come in two scenarios: either in reality AR if the politician's type switched, or in reality R if the politician indeed acted competently. Both of these scenarios decrease the value of the politician for the voter. For the former scenario we have already seen this in the previous paragraph. For the latter scenario, the logic is that in reality R the politician must be bad, and Assumption 4 says that being perceived bad is too costly relative to a gain in perceived competence. Thus acting competently is costly in both scenarios, implying that in any equilibrium the propaganda-spreading politician will act incompetently.

Evidence about health policies. The prediction that politicians who use propaganda then choose policies that contradict the scientific consensus is supported by Figure 4. The left panel shows that across U.S. states and over time, controlling for the severity of the epidemic, Republican governors introduce indoor mask mandates 12 percentage points less often than their Democratic counterparts. 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.[13] More generally, because addressing health or environmental problems often requires governmental adoption of best practices, the mechanism highlighted here can generate large social costs.

[13] In the left panel we use a monthly data for all 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.

Figure 4: Impact on policy



3.2 New media and beliefs in the alternative reality

A salient fact about media in the U.S. is that non-traditional media outlets, including Fox News, appear to reinforce and amplify the anti-elite alternative realities propagated by Republican politicians.[14] This fact appears to be unexplained by existing theories. It cannot be explained by theories of media capture (Besley and Prat 2006) since these media do not seem to be controlled by politicians. Nor can it be easily explained by leading theories of independent media (Mullainathan and Shleifer 2005, Gentzkow and Shapiro 2006) which argue that independent media slant the presentation of facts to cater to readers but do not predict that they present non-truths and reinforce alternative realities. In this application we propose an explanation based on the idea that demand for the new media arises because of audiences' distrust in the elite media, implying that it is in the best interest of the new media to sustain that distrust by reinforcing the alternative reality.

[14] Non-traditional outlets exist in essentially all media markets: in cable television, such as the One America News Network, in radio, such as the programs of Rush Limbaugh and Alex Jones, in online media, such as Breitbart and NewsWars, and among local newspapers, such as The Tennessee Star and The New Boston Post.

We introduce to the model a class of new media outlets. Paralleling our modelling of the elite media, we assume that each new media outlet is too small to influence electoral outcomes, and formalize this by assuming that there is a continuum of new outlets and a one-to-one mapping links them with voters. Thus each voter consumes one elite and one new media outlet. As in the basic model, we think about each voter as representing the potential core audience of the corresponding elite and new media outlet. Because the new media outlets have identical incentives, we only consider equilibria in which they have identical strategies, and treat them as a single decision maker. Like the elite media, the new media observes the politician's common type and sends a message about it to the voter. We assume that the new media is slightly less informed than the elite media: formally, that the message of the new media has a vanishing tremble which in the limit is arbitrarily larger than that of the elite media (but arbitrarily smaller than that of propaganda). We also assume that the message of the new media only reaches the voter with probability α. We will assume that α is not too high, i.e., the new media does not enter too often: this ensures that the politician continues to have incentives to engage in propaganda irrespective of the action of the new media.[15]

It is key to this application that media have audience-seeking preferences. We introduce such preferences for both the elite and new media, and model them by assuming that each elite media outlet wants to maximize the belief of its (single) voter audience that reality is R, while each new media outlet wants to maximize the belief of its (single) voter audience that reality is AR. This is a shortcut that captures competition between the two outlets. If reality is R, then, because the elite media is more informative, the voter should prefer it next period; whereas if reality is AR, then, because the elite media conspires, the voter should prefer the new media next period. Observe that in this setup each outlet is too small to influence the election, but sufficiently large to influence the beliefs of the single voter who constitutes its audience.

We assume that the voter is not aware of the media's audience-seeking preferences. This assumption does not seem essential for our qualitative results, but keeps the analysis simple, and seems to us realistic: audiences often seem to think that the media's preferences for policies and for

[15] One can interpret α as capturing a technological innovation, such as cable television or the internet, that, if successful, enables new media outlets to reach their audiences.

truthfulness are the key determinants of content.[16] Formally, we introduce a new type dimension for the R media outlets, denoted by $\theta_a \in \{0, 1\}$, where zero means that media do not, and one means that media do have audience-seeking preferences. Similarly to the reality types, the voter has an incorrect prior belief about the type distribution: he believes that the probability of $\theta_a = 1$ is zero, while the objective probability is one. We assume that all R media, both elite and new, have the same type θ_a, capturing that in the voter's mind none of them, while in truth all of them have audience-seeking preferences. Because the AR media outlets only exist in the voter's mind, these outlets never care about audiences and do not need the new type dimension.

In this application media outlets' lying cost plays an important role and we thus model it explicitly. We assume that the lying cost of the elite media is higher than that of the new media, which captures the idea that the elite media has a preexisting audience that it would risk losing by manipulating the truth, while the new media does not and is hence less constrained. In the Appendix we provide formal microfoundations in which the elite media has additional informed audience that does not believe government propaganda, knows reality is R, and is thus not contested by the new media. Assuming that the lying cost is proportional to audience size then implies that the lying cost of the elite exceeds that of the new media.

Finally, we assume that the new media outlets are pro-voter, i.e., that they have the same policy preferences as the voter. Because each outlet is small these preferences are inconsequential for behavior, but will rule out the possibility in the voter's mind that the new media are in the conspiracy, because in the AR it is the elite who conspires, and the pro-voter new media is by definition not in the elite.[17]

We introduce a change in notation: because both the elite and the new media send signals about the politician's common type, in this application we denote their messages by s_c^e and s_c^n. The objectives which govern the behavior of the elite and new media can then be written as

[16] Consistent with this assumption, even in the economics literature there has been a debate about the importance of demand- versus supply-side determinants of media content (Gentzkow and Shapiro 2010). We note that we believe a model in which media are audience-seeking with positive probability, and the voter has the correct prior about their audience-seeking, would generate qualitatively similar results to those formulated below.

[17] We believe that allowing the new media to conspire would not change our results, because, being pro-voter, it would not have incentives to mislead the voter.

follows:

$$U_e = 1_{\{\theta_r = R\}} \cdot \left[\phi 1_{\{\theta_a = 1\}} \cdot \mu_v(R|\hat{h}_v^1) + \chi_e 1_{\{s_c^e = \theta_c\}}\right] + 1_{\{\theta_r = AR\}} \cdot (c\tilde{\theta}_c - \lambda\tilde{\theta}_d), \tag{7}$$

$$U_n = \phi 1_{\{\theta_a = 1\}} \cdot \mu_v(AR|\hat{h}_v^1) + \chi_n 1_{\{s_c^n = \theta_c\}}. \tag{8}$$

Consider the objective of the elite media. The first term, active when reality is R, has two parts. The first part captures audience-seeking preferences when $\theta_a = 1$, governed by the voter's posterior belief that reality is R, denoted $\mu_v(R|\hat{h}_v^1)$, and a weight ϕ representing the importance of audience-seeking. The second part captures the elite media's lying cost, denoted χ_e. 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 coordinate and influence elections.[18]

The objective of the new media has two parts, which reflect its audience-seeking preferences when $\theta_a = 1$, and its lying cost χ_n. As mentioned above, we assume that $\chi_e > \chi_n$. Note that for the new media we do not include policy preferences even in the AR. This is for the reason mentioned above that the new media does not conspire even in the AR and thus its policy preferences are irrelevant for its behavior.

The timing of events is the following.

0. The politician's type is realized. The voter observes her divisive type θ_d, the elite and the new media also observes her common type θ_c.

1. The elite sends message $s_c^e \in \{0, 1\}$ the new media sends message s_c^n, and the politician decides on propaganda $p \in \{0, 1\}$. All messages are subject to trembles. The voter always observes propaganda $\hat{p}$ and the elite's message $\hat{s}_c^e$, but only observes the new media's message $\hat{s}_c^n$ with probability α.

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.

[18] Implicit in this formulation 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.

We need some parametric assumptions for our result.

Assumption 6. For the bad pro-voter politician, the benefit of partially hiding her common type in the event in which the new media's message does not reach the voter, is higher then the cost of propaganda:

$$(1-\alpha)E\hat{q}_c cg > f.$$

This assumption strengthens Assumption 1, and makes spreading the alternative reality profitable even if it only influences beliefs in the event in which the new media's message does not reach the voter. This strengthening ensures that spreading propaganda will be profitable for the bad R politician even if the new media reveals her to be bad.

Assumption 7. For the elite media truthtelling dominates audience-seeking, while for the new media audience-seeking dominates truth-telling by a margin

$$\chi_e > \phi > \frac{\chi_n}{q_r}.$$

The first half of the assumption ensures that in equilibrium the elite media reports honestly as in the basic model; the second half that the new media is willing to distort the truth to gain audience. For the latter we need to normalize χ_n by q_r because even in the absence of lying the voter assigns positive probability to the AR, so that the gain to the new media from moving that probability is smaller.

Proposition 3. *Under Assumptions 2 and 6-7, there is a unique pure strategy equilibrium path in which:*

1. *The elite and the politician behave the same way as in the baseline model.*
2. *The objectively existing new media, which has audience-seeker preferences*
   - *Always reports the common type of the pro-voter politician to be good,*
   - *Reports the common type of the pro-elite politician truthfully.*
3. *The imagined new media—which does not have audience-seeking preferences—reports the common type of all politicians truthfully.*

4. The presence of new media amplifies the effect of propaganda and increases the reelection probability of a bad politician relative to the basic model.

The key result is that the new media will report he pro-voter bad politician—who spreads propaganda—to be good. The intuition is the following. Since, in the reality, the elite media is perfectly informative about the common issue, demand for the new media is governed by the strength of the voter's belief in the alternative reality. Thus it is in the best interest of the new media to strengthen the belief in the alternative reality. Reporting the politician good achieves this: since the voter ignores the audience-seeking incentive he perceives the new media to be truthful (except for trembles), and will therefore conclude from the conflicting reports of the elite and the new media that the elite must be conspiring, i.e., that reality is AR. Note, this logic requires that the elite continues to be truthful, but the new media is willing to lie to gain audience, which is ensured by Assumption 7; and that the rest of the equilibrium plays out as in the baseline model, which is ensured by Assumptions 2 and 6 (the latter a strengthening of Assumption 1).

This result helps explain why outlets such as Fox News or conservative talk radio, even though they are not connected to Trump or the Republican establishment, seem to reinforce factually incorrect Trumpian political messages. The result also has two, as yet untested implications. (1) It predicts that government propaganda increases the demand for private propaganda. (2) It predicts that new media such as Fox news affect not only political preferences (DellaVigna and Kaplan 2007) but also beliefs in the alternative reality, i.e., science scepticism, further limiting the adoption of health and climate best practices. The second implication is in line with evidence showing that the consumption of Fox News reduced social distancing during the pandemic (Bursztyn et al. 2020, Simonov et al. 2020).

4 Conclusion

In this paper we built a model of the political economy of alternative realities. In our model a politician can supply an alternative reality which discredits the criticism of the intellectual elite. Key to our approach is to explicitly model an alternative reality that incorporates optimizing actors, and have the voter reason in a Bayesian fashion about, and respond strategically to, imagined behavior in this alternative reality. Requiring that the alternative reality remains consistent both within itself and with available evidence constrains the types of alternative realities that can be spread, as well as the behaviors of the voter, the government, and the media. We have shown that these constraints generate a number of new predictions about politics, media, and adoption of best practices, several of which are consistent with available evidence.

Our analysis leaves several questions unanswered. First, we have not formulated a general theory of strategic behavior under beliefs in an alternative reality: in particular, our equilibrium definitions are specific to our concrete settings. There may be value in formulating a general notion of equilibrium with beliefs in alternative realities.

Second, our approach of explicitly modeling a consistent alternative reality may be useful to understand behavior in other domains. One class of examples may be political ideologies, which often seem to be based on an oversimplified model of the world. Popular belief in the ideology may constrain the politician spreading it: for example, beliefs in the communist ideology may be punctured by the introduction of pro-market reforms and thus lead to loss of support for the political system, a logic which helps explain why the transition to a market economy in Eastern Europe was accompanied by democratization. Another class of examples may be conflict. Misunderstanding the incentives of the counterparty may lead to the breakdown of negotiations and to conflict, and some actors may purposefully engineer such misunderstanding. For example, violence-inciting propaganda often features a false rhetoric of self-defense, and the dehumanization of opponents, as in the case propaganda against the Tutsi in Rwanda (Yanagizawa-Drott 2014). We hope that our conceptual framework can improve the understanding of behavior in such situations.

Third, our approach focuses on the supply of alternative realities but is silent about the demand: why voters are willing to believe in alternative realities. Developing a behavioral-economic theory of the demand side can lead to new predictions about when propaganda is likely to be successful, and what policies can correct beliefs and improve the adoption of best practices.

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.

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.

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.

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.

Che, Yeon-Koo and Konrad Mierendorff, “Optimal Dynamic Allocation of Attention,” *American Economic Review*, August 2019, *109* (8), 2993–3029.

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.

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

Eyster, Erik and Matthew Rabin, "Cursed Equilibrium," *Econometrica*, 2005, *73* (5), 1623–1672.

Fryer, Roland G Jr, Philipp Harms, and Matthew O Jackson, "Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization," *Journal of the European Economic Association*, 08 2018, *17* (5), 1470–1501.

Funk, Cary and Brian Kennedy, "For Earth Day 2020, how Americans see climate change and the environment in 7 charts," Pew Research Center, `https://www.pewresearch.org/fact-tank/2020/04/21/how-americans-see-climate-change-and-the-environment-in-7-charts/`, 2020.

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

___ and ___ , "What Drives Media Slant? Evidence From U.S. Daily Newspapers," *Econometrica*, 2010, *78* (1), 35–71.

___ , 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.

Golub, Benjamin and Matthew O. Jackson, "How Homophily Affects the Speed of Learning and Best-Response Dynamics," *The Quarterly Journal of Economics*, 07 2012, *127* (3), 1287–1338.

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.

Heidhues, Paul, Botond Kőszegi, and Philipp Strack, "Unrealistic Expectations and Misguided Learning," *Econometrica*, 2018, *86* (4), 1159–1214.

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

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

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

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.

Rabin, Matthew and Joel L. Schrag, "First Impressions Matter: A Model of Confirmatory Bias," *The Quarterly Journal of Economics*, 1999, *114* (1), 37–82.

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.

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

A Appendix

A.1 Proof of Proposition 1

Our proof identifies the unique pure strategy equilibrium profile and shows it has the properties highlighted in the proposition. We begin by characterizing the behavior of some actors independently of the whether the incumbent politician is pro-voter or pro-elite, and then proceed to other actors analyzing these two cases separately. The R elite ignores her effect on voters and given the preference for telling the truth always reports honestly. The normal voter updates his beliefs about the politician's common type based on $\hat{s}_c$ and $\hat{p}$. Because of the trembles all realizations of $\hat{s}_c$ and $\hat{p}$ are possible, but because both trembles become very unlikely and $\hat{s}_c$ trembles become arbitrarily more unlikely, in the limit after any observation of messages the normal voter will follow the elite's message:

$$\mu_v[\theta_c = 1|\hat{s}_c, \hat{p}, \theta_d, \theta_m = N] = \hat{s}_c.$$

These beliefs pin down the behavior of the normal voter, in particular, he is more likely to reelect if the politician is good.

Consider the strategy of the R politician. Voter's beliefs about the good R politician are already maximized absent propaganda, so there is no reason for her to engage in propaganda. The bad R politician has two possible strategies: (i) engage in propaganda or (ii) avoid propaganda. We will characterize her choice later.

Case 1: Incumbent politician is pro-voter. Recall from assumption 2 that the AR elite's preference is to remove all types of the pro-voter politician. The AR elite believes that propaganda is ineffective and the voter is normal. Since, as we have seen, the normal voter follows the elite's message, the unique optimal strategy of the AR elite is to always report that the politician is bad.

We next show that in any pure strategy equilibrium the good and bad AR politicians and the bad R politician all find it optimal to use propaganda. There cannot be an equilibrium in which only bad politicians (R, AR, or both) use propaganda, because—given that $q_r > 0$ ensures the persuaded voter has a positive prior of both the bad R and bad AR politician—propaganda would reveal them to be bad, and is hence not worth doing. Therefore, either nobody uses propaganda, or the good AR politician uses propaganda.

If nobody uses it, then switching to propaganda by the bad R politician is profitable. Absent propaganda the voter fully trusts the elite's report and considers the politician bad. Propaganda will be attributed to a tremble but will change the voter's prior, and hence the elite's bad message will result in the voter believing that the politician is bad with probability $\hat{q}_c$, because of a Bayesian updating analogous to equation (10) since such a message can arise in the R for a bad politician and in the AR for either politician. This makes a deviation to propaganda profitable by Assumption 1.

If the good AR politician uses propaganda, then a similar logic establishes that it is not optimal for either the bad R or the bad AR politician to refrain from it. Here too, absent propaganda they would be revealed bad, and propaganda will increase the voter's belief that they are good. The voter's posterior belief will be at least $\hat{q}_c$, because this is the belief that would obtain when propaganda signals the worst possible politician composition due to both the bad R and bad AR politicians (besides the good AR politician) using it. It follows that both AR politicians and the bad R politician must use propaganda. In this profile Assumption 1 ensures that they are all best-responding: on path the politician is believed good with probability $\hat{q}_c$, and a deviation would generate, according to the deviator's understanding of the game, a credible message that she is bad.

The above arguments characterize the behavior of all principals, and the on-path beliefs of both the normal and the persuaded voter. In turn, these beliefs characterize the behavior of the voter since he acts at the last stage of the game and makes decisions where indifference is a zero-probability event. To complete the argument that the proposed path is an equilibrium, we need to consider off-equilibrium information sets. These can only happen at stage 2 of the model. Because of the trembles every history of message profiles is possible; but because propaganda pins down the voter type, the normal voter after propaganda and the persuaded voter absent propaganda never occur in this game, not just on the equilibrium path but in any deviation. Still, our equilibrium definition requires that we specify actions and beliefs at these information sets. Moreover, the behavior of the normal voter after propaganda is payoff relevant because that is the voter the AR elite best responds to. For the normal voter after propaganda, Bayes rule (10) pins down beliefs:

because the voter is normal, he assigns zero probability to the AR types, and his beliefs about the common type follow the elite's message. This also pins down his voting behavior. Consider the persuaded voter absent propaganda. Bayes rule implies that if the elite's message was good, the voter will update that reality is R and form beliefs that follow the elite's message; while if the elite's message was bad, he will observe a history only possible with trembles, will believe that there was a propaganda tremble, assume that the politician sent propaganda, and form beliefs as he does in that subgame. This completes the description of the unique equilibrium.

It is immediate that the unique equilibrium satisfies statements 1 and 2 of the Proposition for Case 1. For statement 3, note that if the bad politician refrains from propaganda the voter will be certain that she is bad; whereas if she engages in propaganda the voter will believe that she is bad with probability $1 - \hat{q}_c < 1$. Thus propaganda increases the reelection probability of the bad pro-voter politician.

Case 2: Incumbent is a pro-elite politician. Consider the AR elite. Because she wants to keep both types of politicians and believes that her audience was not affected by propaganda and forms beliefs based on the signal she sends, her optimal strategy is to always report that the politician is good. Consider the AR politician. By avoiding propaganda she can ensure that the voter thinks she is good, since the AR elite (which the AR politician believes is the elite) always reports her good. This is the best the AR politician can hope for, thus there is no reason to engage in costly propaganda, and the unique best response of both common types is to not engage in propaganda. For the R politician, doing propaganda cannot be part of an equilibrium: the voter observes both propaganda and criticism only for the bad R politician, and will thus conclude from that information that the politician is bad. Thus avoiding propaganda is optimal. Because no principal engages in propaganda, the on-path belief of the voter is to follow the elite's message. This characterizes his behavior too.

We have characterized on-path beliefs and behavior. We now turn to off-equilibrium information sets. Like in Case 1, these are only possible 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 as the normal voter absent propaganda. Since propaganda is off the path, the persuaded voter after propaganda will attribute it to a tremble and form beliefs and behavior just like the persuaded voter absent propaganda. In turn, the persuaded voter absent propaganda, after observing a bad message—since the AR elite always reports the politician is good—learns that he is in R and forms the same beliefs as the normal voter. But after observing a good message, he will form interior beliefs about the reality type as specified by Bayes rule. In either case, 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.

Finally, it is immediate that the unique equilibrium satisfies statements 1 and 2 of the Proposition for Case 2.

A.2 Additional material for competence application

Extending equilibrium definition. To extend the equilibrium to the competence application, we need to introduce some definitions and notation. The type profile in the competence model is $\theta = (\theta_c, \theta_d, \theta_d^0, \theta_k, \theta_r, \theta_m)$. At the beginning of the game all actors have correct priors about the new type dimensions, which we denote by $\mu^0(\theta_k)$, and by $\mu^0(\theta_d|\theta_d^0)$ since the distribution of θ_d depends on the realization of θ_d^0. Analogously to the basic model, we assume that the type of the principals includes all information known to them at the beginning of the game: $\theta_p = \theta_e = (\theta_d^0, \theta_c, \theta_r)$.

Denote the history observed by actor i up to and including stage t by $\hat{h}_i^t$. We allow for private histories since only the politician observes θ_k and only the politician and the elite observe $\hat{\theta}_k$. We denote the change in the private history of actor i between stages $t-1$ and t by $\Delta\hat{h}_i^t$. We encode the new type dimensions in the $\Delta\hat{h}_i^t$: for example, $\Delta\hat{h}_p^2 = (\theta_d, \theta_k)$. We let $\hat{h}^t$ denote at stage t the component of the history observed by all actors, which we call the public history.

We define strategies for stages at which the actor has a move, but for convenience define beliefs and expected utilities for all stages. The perturbed strategy of i—which also incorporates Nature's trembles—at stage t is denoted by $\hat{\sigma}_i^t(.|\theta_i, \hat{h}_i^{t-1})$. The beliefs of i at the end of stage t are denoted by $\mu_i^t(\theta|\theta_i, \hat{h}_i^t)$. For simplicity, when a strategy or a belief does not depend on a particular conditioning variable, such as a component of the type vector θ_i or of the history $\hat{h}_i^t$, we sometimes omit that conditioning variable from the notation.

Belief updating by the principals at stage 1 is as follows. For the politician and the R elite, $\mu_p^1(\theta_m = P|\hat{s}_c, \hat{p}) = \hat{p}$ and $\mu_e^1(\theta_m = P|\theta_r = R, \hat{s}_c, \hat{p}) = \hat{p}$, that is, they correctly believe that $\hat{p}$ determines the mind type of the voter. For the AR elite, $\mu_e^1(\theta_m = P|\theta_r = AR, \hat{s}_c, \hat{p}) = 0$, that is, she does not recognize that the voter's mind type may be altered. The principals' beliefs at the end of stage 1 about all other type components agree with their priors.

Belief updating in all other cases, that is for the voter in stage 1 and for the principals in stages $t \geq 2$, is given by

$$\mu_i^t(\theta|\theta_i, \hat{h}_i^t) = \frac{\mu_i^{t-1}(\theta|\theta_i, \hat{h}_i^{t-1})\hat{\sigma}_{-i}^t(\Delta\hat{h}_i^t|\theta, \hat{h}_i^{t-1})}{\sum_{\theta'} \mu_i^{t-1}(\theta'|\theta_i, \hat{h}_i^{t-1})\hat{\sigma}_{-i}^t(\Delta\hat{h}_i^t|\theta', \hat{h}_i^{t-1})}. \tag{9}$$

This expression has the standard form of Bayesian updating, and is more complicated only because of departures from the standard setting of a multi-stage game with observed actions, not because of our departures from rationality. The complication is the term $\hat{\sigma}_{-i}^t(\Delta\hat{h}_i^t|\theta, \hat{h}_i^{t-1})$, which measures the probability of outcome $\Delta\hat{h}_i^t$ under opponent strategy profile σ_{-i} conditional on type profile θ and private history $\hat{h}_i^{t-1}$ for i. In a standard multi-stage game with observed actions this term would just be opponents' current-stage strategy. In our setting it is modified for three reasons. First, we have trembles. Second, in stage 2 when the politician learns $\Delta\hat{h}_c^2 = (\theta_d, \theta_k)$, the innovation in her private history is coming not from opponents' moves but from Nature. Third, and most important, an actor may need to account for the fact that opponents' behavior is driven by private histories not visible to that actor. This is the case for the voter in stage 4: Because he only observes the competence message $\hat{s}_k$ but not the competence seen by the elite $\hat{\theta}_k$, the $\hat{\sigma}_{-v}^4(\hat{s}_k|\theta, \hat{h}_v^3)$ term must compute the probability of the observed competence message $\hat{s}_k$ taking into account both possible values for the perturbed competence action $\hat{\theta}_k$ observed by the elite. Formally, $\hat{\sigma}_{-v}^4(\hat{s}_k|\theta, \hat{h}_v^3) = \sum_{\hat{\theta}_k=0}^1 \hat{\sigma}_e^4(\hat{s}_k|\hat{\theta}_k, \hat{h}^1, \theta_v)\hat{\sigma}_p^3(\hat{\theta}_k|\theta_d, \theta_k, \hat{h}^1, \theta_p)$. Our notation $\hat{\sigma}_{-i}^t(\Delta\hat{h}_i^t|\theta, \hat{h}_i^{t-1})$ represents all three of these mechanisms.

We now show that the beliefs $\mu_i^t(\theta|\theta_i, \hat{h}_i^t)$ computed by (9) are well-defined at all $(\theta_i, \hat{h}_i^t)$ pairs at which i gets to make a decision. This requires that the denominator be positive for every such $(\theta_i, \hat{h}_i^t)$, that is, under the prior and updating rule of θ_i, private history $\hat{h}_i^t$ must have positive probability. To prove this, consider first the principals: $i = p, e$. For any $\theta_p = \theta_e = (\theta_d^0, \theta_c, \theta_r)$, the trembles in all actions in stages 1-4, and the full support of the distribution of θ_d and θ_k, ensure that all private histories are possible. This is despite the fact that the AR elite never considers $\theta_m = P$ possible, because θ_m is neither in the type nor in the private history of the AR elite. Consider next the voter, $i = v$. The required condition is no longer ensured by the trembles, since with the normal voter a history involving propaganda, or with a persuaded voter a history without propaganda, are impossible. Nevertheless, the condition holds at stage 1 because both the normal and the persuaded voter update from their respective priors, failing to understand that propaganda and their prior should be fully correlated, and thus effectively thinking that the above counterfactual histories are possible. Given this, in subsequent stages the condition holds because trembles make any continuation history a positive-probability event.

Proof of Proposition 2. Our proof characterizes the unique pure strategy equilibrium path and shows that it is supported by an equilibrium. The properties claimed in the Proposition then follow immediately. We begin by characterizing the behavior of some actors independently of the whether the incumbent politician is pro-voter or pro-elite, and then proceed to other actors analyzing these two cases separately. Consider the competence stages after any history. The R elite reports truthfully about competence and the normal voter fully trusts the elite's report. Because the AR elite believes that propaganda was ineffective, she believes that the voter follows her message and by Assumption 3 criticizes if and only if the politician is pro-voter. Both the normal voter (absent propaganda) and the persuaded voter (with propaganda) understand the strategies of the elite, which implies that after every voter history, the elite's message makes both voter types become weakly more informed about the politician's competence type. Moreover, because after a history of no propaganda the voter fully trusts the elite's (truthful) report, every politician type will chose to act competently if he can. Thus the politician may choose to hide her competence only in histories with propaganda.

Case 1: Politician is initially pro-voter. Consider the first stage. The honest R politician understands that by sending propaganda she can partially hide her common type and competence. Since she is good she wants to reveal his common type. And since ϵ is uniform—she is risk neutral—

she is ex-ante indifferent about revealing her competence. Thus avoiding propaganda is her optimal strategy. The good and bad types of the AR politician continue to think that the elite (which they believe is the AR elite) reports both of them bad in the first stage, and reports according to their updated divisive type in the second stage. Since their payoff does not depend on their common type, generically they will choose the same action. It follows that, for generic parameters, there are four possible strategy profiles in the first stage: depending on whether the bad R politician sends propaganda or not, and whether the AR politician (of both types) sends propaganda or not. We now check for each of these profiles whether they can lead to an equilibrium of the two-stage game.

Subcase 1: both send propaganda. This was the unique equilibrium of the original game. Consider a history of observing propaganda. As we have seen, the persuaded voter updates to $\hat{q}_c$ about the politician's common type. The persuaded voter also understands the elite's second-stage behavior, and thus believes that a message of competence can come in two events: (i) in the alternative reality if the politician has flipped to pro-elite, and (ii) in the reality if the politician is bad and competent. We now show that both of these events decrease the value of the politician to the voter. Previously the voter believed the politician was pro-elite with probability ξ and good with probability $\hat{q}_c$. Now in the first event the politician is viewed pro-elite with probability 1, and good with probability q_c because in the alternative reality the elite's message was uninformative (and remains viewed as competent with probability q_k). This change in beliefs is bad for the politician if $(1-\xi)\lambda > (q_c - \hat{q}_c)c$ which follows from Assumption 5 when ξ is small. In the second event the politician is still pro-elite with probability ξ, but is bad with probability 1 and competent with probability 1. This change in beliefs is bad for the politician if $\hat{q}_c c > (1-q_k)k$ which is Assumption 4. It follows that the normal R politician, who understands all this, will choose to avoid a message of competence by acting incompetently. The AR politician also understands this, but believes that the elite is the AR elite whose message she cannot influence. Thus she acts competently if she can.

Now go back to the first stage. Since avoiding propaganda leads to a revelation of competence, a new consequence of propaganda for the R politician is that its (at that stage unknown) competence status is not revealed. But since she is risk neutral this leaves her ex ante utility unchanged. Since propaganda still helps to partially conceal her (at that stage already known) common type, the bad R politician will choose propaganda. For the AR politician the same argument applies. In summary, in subcase 1 with the strategies described here, all actors are mutually best responding on the proposed equilibrium path.

Subcase 2: only the bad R politician sends propaganda. Consider a history of observing propaganda. The voter learns from propaganda that reality is R, the elite is R, and the politician is bad. The R politician understands this and acts competently if she can. This is the same outcome that she would experience absent propaganda, thus in the first stage the bad R politician prefers to avoid propaganda. In subcase 2 we do not have an equilibrium.

Subcase 3: only the AR politicians send propaganda. Consider a history of observing propaganda. The voter learns from propaganda that reality is AR, the elite is AR, and thinks the politician is bad with probability q_c. He believes that the elite's message is based purely on the divisive type. Anticipating this, the R politician prefers to choose propaganda and then act incompetently. This brings the benefit that voters will think she is bad with probability q_c. Her competence is then not revealed in the competence stages but as she is risk neutral it is just as good ex-ante as if it was revealed. Therefore using propaganda is a profitable deviation. In subcase 3 we do not have an equilibrium.

Subcase 4: no politician sends propaganda. After observing propaganda—which the voter attributes to a tremble—a message of incompetence can come in two ways: in the AR if the politician remains pro-voter, or in the R if the politician is incompetent. Then a deviation by the bad R politician to engage in propaganda and act incompetent has the benefit of making the voter think that she is good probability $\hat{q}_c$, at the cost of making him appear somewhat incompetent, which Assumption 4 ensures is worth it. In subcase 4 we do not have an equilibrium.

It follows that only case 1 can lead to an equilibrium, and that case has the properties claimed in the proposition. To close the argument, we need to characterize equilibrium beliefs and behavior at off-equilibrium information sets. These information sets occur in the competence part of the game at stages 3, 4 and 5. At stage 3, besides competence, the politician has four types: good or bad and R or AR. After no propaganda all types act their competence. After propaganda, the proof above characterized the behavior of the bad R type, and both AR types. Thus the only missing piece is the good R politician after propaganda. Her behavior depends on what the elite reported before. If the elite reported her good, then the voter updated to R, so the politician will act her competence. If the elite reported her bad, then the voter remains persuaded, and the politician will act incompetent. At stage 4, after any history, the R elite always reports the politician's realized competence action honestly; and the AR elite always thinks the voter is normal and thus always wants to send a message based on the politician's divisive type. Finally, at stage 5, the voter has two types: normal or persuaded. The normal voter always follows the elite's messages, after any history, including after propaganda. The persuaded voter's behavior after propaganda has been characterized in the proof above. It remains to deal with the persuaded voter absent propaganda. That voter, if the elite's message was good, will update to R and behave accordingly; but if the elite's message was bad, will update to thinking that the politician sent propaganda and update as in that subgame.

Case 2: Politician initially pro-elite. Consider the first stage. The good R politician understands that by sending propaganda she may partially hide her common type and competence. Since she is good and risk-neutral, these changes do not increase her utility. However, propaganda also has the benefit that if her divisive type flips, the AR elite will reveal this, leading the voter to partially update on it. If the probability of the flip ξ is small enough then this effect is dominated by the cost of propaganda. Thus the honest R politician avoids propaganda.

Consider the AR politician. By avoiding propaganda she can ensure that the voter thinks she is good, since the AR elite (which the AR politician believes is the elite) always reports good. Avoiding propaganda leads to a revelation of her competence, but as ϵ is uniform this is not a cost. On the other hand, avoiding propaganda prevents revelation of a flip in his divisive type. This is a cost, but when ξ is small enough it is dominated by the cost of propaganda. Thus the AR politician avoids propaganda.

Consider the bad R politician. Doing propaganda cannot be part of an equilibrium: the voter observes both propaganda and criticism only for the bad R politician, and will thus conclude from that information that the politician is bad. Thus propaganda does not change beliefs and is therefore useless.

We conclude that in the first stage no politician type engages in propaganda. It then follows that in the competence stages all politician types act their competence. Any pure strategy equilibrium must imply this same behavior on the equilibrium path.

It remains to verify that a pure-strategy equilibrium exists. This is more complicated than in Case 1, because we now need to specify beliefs and actions in the competence stages after the zero-probability outcome of propaganda in the first stage. First note that the elite's behavior is pinned down: the R elite always reports the truth while the AR elite reports competence if and only if the politician is pro-elite. We claim that one pure strategy equilibrium is for the AR politician to always act its competence, and for the R politician to pretend incompetent if and only if she is good. We now verify that this is indeed an equilibrium. We note however that for some parameters other pure strategy equilibria may exist.

First note that because propaganda is off the equilibrium path, the voter will attribute it to a tremble and will not update from it about the politician's type. Now consider the good R politician. Pretending incompetence implies that the elite's messages are (good, incompetent). The probability of this message profile in the alternative reality is ξ and thus small. The probability of this message profile in reality R (in the candidate equilibrium) is q_c and thus bounded away from zero. It follows that for ξ low enough the voter puts a large weight on reality being R and the politician being honest, and would believe the politician is competent with probability close to q_k. In contrast, a message profile of (good, competent) would make the voter believe that with large probability reality is AR and that the politician is bad with probability close to $1 - q_c$ and competent with probability close to q_k. Thus acting incompetent is better.

In the case of the bad R politician, the elite's first-stage bad message proves to the voter that reality is R and the politician is bad. Thus the bad R politician cannot do better than acting her competence.

The AR politician believes that she cannot influence the message of the AR elite whom she believes is the elite, and thus she also cannot do better than acting her competence.

Finally consider the voter. Any history of action profiles is possible because of the trembles. Thus the voter of either type can update using Bayes rule. Since this is the last stage, after any history he faces a decision problem which has a solution. He behaves accordingly.

A.3 Additional material for new media application

Microfoundation for different lying costs. We conceptualize the lying cost as a reduced-form representation of reputation concerns about un-modeled future periods. Suppose that the potential audience of each elite and new media outlet pair in fact consists of a continuum of readers. This continuum has measure zero so that they do not matter for the electoral outcome, but they do matter for the profits of the media. Our first key assumption is that a share $1-\eta$ of readers in this continuum are immune to propaganda and always believe that reality is R. Since the elite media is more precise than the new media—and will remain so in future periods—it follows that the new media can never steal these readers and thus ignores them.[19] However, the two media outlets compete for the remaining share of readers η who are gullible to propaganda. Assume that changing the perceived probability of the R reality from 0 to 1 (1 to 0) of a unit mass of readers gives the elite (new) media a utility of ψ. Given that only η share of readers can be influenced, the maximum possible utility from lying is $\eta\psi \equiv \phi$ for both the elite and the new media. Our second key assumption is that the lying cost—a reduced form representation of reputation concerns—is proportional to the volume of potential readers. Therefore, the elite pays the lying cost for all readers but the new media only for the readers (of mass η) who can be grabbed. If the lying cost per unit mass of readers is χ, then $\chi_e = \chi > \eta\chi = \chi_n$.

Equilibrium. We extend our equilibrium concept to the game with the new media. We define the type of the new media to be $\theta_n = (\theta_d, \theta_c, \theta_r, \theta_a)$. Out of these type dimensions, only θ_a affects the new media's payoff function, the other dimensions represent information the new media has at the beginning of the game. The types of the politician, the elite and the voter are unchanged. Prior beliefs about the new type θ_a are as follows. The R principals correctly perceive that $\theta_a = 0$: $\mu_p^0(\theta_a = 0|\theta_r = R) = \mu_e^0(\theta_a = 0|\theta_r = R) = 1$. The AR principals and the voter all falsely believe that $\theta_a = 1$: $\mu_v^0(\theta_a = 1|\theta_v) = \mu_p^0(\theta_a = 1|\theta_r = AR) = \mu_e^0(\theta_a = 1|\theta_r = AR) = 1$. The new media has correct prior beliefs about all types, and all other prior beliefs are as in the basic model.

[19] We continue to assume that the median voter can be persuaded by propaganda. This does not necessarily imply that $\eta > 0.5$, because the elite may overweight people who cannot be persuaded if they have higher advertising value.

Because the politician, the elite, and the new media only move in stage 1, their strategies only depend on their type and are denoted by $\sigma_p(a_p^1|\theta_p)$, $\sigma_e(a_e^1|\theta_e)$, and $\sigma_n(a_n^1|\theta_n)$. The voter moves in stage 2 after observing $\hat{h}_v^1$ which is either $(\hat{s}_c^e, \hat{s}_c^n, \hat{p})$ or $(\hat{s}_c^e, \hat{p})$ depending in whether the message of the new media reaches him; we denote his strategy by $\sigma_v(a_v^2|\theta_v, \hat{h}_v^1)$.

Defining subjective expected utility has the complication that the utility of the media depends on the beliefs of the voter. However, for any system of beliefs subjective expected utility can be defined and leads to the best-response property of equilibrium given those beliefs. Belief consistency does not impose any condition on principals, because they move only at stage 1 where they know only their priors, but it is straightforward to characterize their beliefs at all stages, because they either know or (in the case of the AR principals) think they know all types. If the true type profile after history $\hat{h}^{t-1}$ is $\theta^* = (\theta_d^*, \theta_c^*, \theta_r^*, \theta_m^*, \theta_a^*)$ then the R politician and the R elite believe $\mu_i(\theta^*|\theta_i, \hat{h}^{t-1}) = 1$ while the AR politician believes $\mu_p((\theta_d^*, \theta_c^*, \theta_r^*, \theta_m^*, 1)|\theta_r = AR, \hat{h}^{t-1}) = 1$ and the AR elite believes $\mu_e((\theta_d^*, \theta_c^*, \theta_r^*, N, 1)|\theta_r = AR, \hat{h}^{t-1}) = 1$. The voter also thinks he knows all types except for the politician's common type. Belief consistency for the voter requires that he follows Bayesian updating at the end of stage 1, and is similar to the basic model

$$\mu_v^1(\theta_p|\theta_v, \hat{h}^1) = \frac{\mu_v^0(\theta_p|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{h}^1|\theta_p)}{\sum_{\theta_p'} \mu_v^0(\theta_p'|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{h}^1|\theta_p')}, \tag{10}$$

the main modification being that $\hat{\sigma}_{-v}^1(\hat{h}^1|\theta_p)$ computes the probability of a voter history $\hat{h}^1$ rather than of an action profile, which is necessary because—due to the new media's message not always reaching him—the voter does not always observe the full action profile at stage 1.

Introspection-based equilibrium foundations. Our informal foundation for the equilibrium concept is introspection. However, in this application introspection is more difficult to justify because the voter and the R politician hold contradictory priors about the new media. Thus introspection requires that the voter be aware of the contradictory prior of the R politician and agree to disagree with it. An alternative modeling approach may be to introduce "imagined" types of the good and bad R politician, which, like the voter, perceive the new media to be honest. Then the voter would reason correctly about the imagined types of the R politician, and best respond to that behavior; while actual behavior would be determined by the real R politician who holds correct priors. Under our assumptions that alternative model would generate the same equilibrium behavior, because the imagined types of the R politician find it optimal to behave the same way as the real types. We show this formally in the proof below.

Proof of Proposition 3. Our proof identifies the unique pure strategy equilibrium profile and shows it has the properties highlighted in the proposition. We begin by characterizing the behavior of some actors independently of whether the incumbent politician is pro-voter or pro-elite, and then proceed to other actors analyzing these cases separately. Start with the R elite. Because it is atomistic and cannot influence voters, the honest R elite—which only exists in the normal voter's mind—always reports truthfully. The audience seeker R elite's preference for telling the truth dominates her preference for audience seeking: the benefit of lying is at most ϕ, which is exceeded by the cost of lying χ_e by Assumption 7. Thus she also reports the politician's common type truthfully. The normal voter believes that reality is R and that the elite is honest, and thus—because the new media's tremble is larger than that of the elite–ignores the new media and always follows the elite. The honest new media–who only lives in the minds of the voter and the AR principals—reports the common type truthfully to minimize the lying cost. The good R politician knows that the elite is audience-seeker but understands that it still tells the truth, and thus does not use propaganda because she already gets the best possible outcome absent propaganda.

Case 1: Incumbent politician is pro-voter. Consider the AR elite, and recall that it believes that propaganda is ineffective and the voter is normal. Since, as we have just seen, the normal voter follows the elite's message, the unique optimal strategy of the AR elite—which by Assumption 2 wants the pro-voter politician out—is to always report that the politician is bad.

We next show that the good and bad AR politicians and the bad R politician all use propaganda. Note that there cannot be an equilibrium in which only bad politicians (R, AR, or both) use propaganda, because then propaganda would reveal them to be bad. Therefore, either nobody uses propaganda, or the good AR politician uses propaganda. If nobody uses it, then switching to propaganda by the bad R politician is profitable. Absent propaganda the voter fully trusts the elite's report and considers the politician bad. Propaganda will be attributed to a tremble but will change the voter's prior, and in the event in which the new media cannot speak, will increase the voter's belief that the politician is good to $\hat{q}_c$, because of an analogous Bayesian updating logic to equation (6) in the basic model. Assumption 6 then ensures that propaganda is beneficial. If the good AR politician uses propaganda, then a similar logic establishes that it is not optimal for either the bad R or the bad AR politician to refrain from it. Here too, absent propaganda they would be revealed bad; and propaganda, in the event that the new media cannot speak, will increase the voter's belief that they are good. This posterior belief will be at least $\hat{q}_c$, because this is the belief that would obtain when propaganda signals the worst possible politician composition due to both the bad R and bad AR politicians (besides the good AR politician) using it.

We now turn to the audience-seeker new media and show that she always reports the politician to be good. If the politician is good, then the audience-seeker new media expects no propaganda and a normal voter who never believes in the AR, hence her only concern is to avoid the lying cost and she reports truthfully. If the politician is bad, then, because the bad politician sends propaganda, the voter becomes persuaded. The beliefs of the persuaded voter depend on the message of the new media as follows. (a) If the new media sends a "bad" message, then the persuaded voter believes that the politician is bad because he thinks the new media is honest. At the same time, the persuaded voter believes that the probability of the AR is his prior q_{ar}, because the elite and the honest new media report the bad politician to be bad in both the R and AR realities. (b) If the new media sends a "good" message, then that will convince the persuaded voter—who thinks the new media is honest—that the politician is good and reality is AR. It follows that the gain to the audience-seeker new media from reporting the bad politician good is $\phi(1 - q_{ar})$, while the cost is χ_n, and Assumption 7 implies that she will report the bad politician good. Finally, consider the behavior of the persuaded voter. We have just characterized his beliefs, and, because he acts at the last stage of the game, he is effectively solving a binary decision problem which pins down his behavior. Indifference has zero probability because his preference shock has a smooth distribution.

The argument thus far has uniquely characterized the behavior of all actors. It remains to establish that this pure strategy equilibrium path is in fact an equilibrium, i.e., that no actor has a profitable deviation. We have already done this in the previous paragraph for the audience-seeker new media. Consider the good and bad AR politicians. As we have seen, absent propaganda the normal voter will follow the elite's report and consider them bad. Propaganda, in the event when new media cannot send its message, will increase the belief of the persuaded voter that they are good to $\hat{q}_c$ because of a Bayesian updating analogous to (6), and Assumption 6 thus ensures that there is no profitable deviation from propaganda. Finally consider the bad R politician. As in the case of the AR politician, restraining from propaganda would make the voter believe the elite's report that she is bad. Propaganda, in the event when new media cannot sends its message, will again increase the belief of the persuaded voter that she is good to $\hat{q}_c$. Moreover, if the new media can send a message, then—because it is audience-seeker—it will report the politician good, and the R politician who has the correct priors understands this. It follows that propaganda is even more profitable for the bad R politician than for the AR politicians.

To complete the description of equilibrium we need to characterize behavior at off-equilibrium information sets. These only occur at stage 2 of the model. For the normal voter after propaganda, Bayes rule pins down beliefs: this voter assigns zero probability to the AR types, and his beliefs about the politician's common type follow the elite's message, pinning down his voting behavior. Consider the persuaded voter absent propaganda. If the elite's message is good, the voter will update that reality is R and form beliefs that follow the elite's message regardless of the message of the new media. If the elite's message is bad, the voter will believe that there was a propaganda tremble and updates his beliefs as if he observed propaganda, a bad report from the elite, and the message of the new media, an information set for which we have characterized behavior above. This completes the description of the unique equilibrium.

Model variant with imaginary types for the R politician. We now show that the model variant with imaginary types for the R politician who believe the new media is always honest generates the same equilibrium path. Just like in the original game, in all potential equilibria the R elite tells the truth and the normal voter follows the elite's advice. Consequently, the AR elite criticises if and only if the politician is pro-voter no matter what the politician and the new media do. Assume that the AR politician always uses propaganda, and both imaginary and real R politicians use propaganda if and only if they are bad. This implies that the beliefs of the persuaded voter after each possible signal remain unchanged, since the voter best responds to the new imaginary type which is behaviorally equivalent to the real R politician. We need to show that there is no profitable deviation for any politician type. The good R politicians—either imaginary or real—restrain from propaganda because they expect the elite to praise them, which message is followed by the normal voter and results in the highest possible expected payoff. Similarly, both bad types (imaginary and real) understand that normal voters observing criticism will be certain that the politician is bad. Since the imaginary politician—to whom the persuaded voter best responds—sends propaganda if and only if she is either AR or bad and R, the persuaded voter observing propaganda, elite criticism, and no new media signal believes that the politician is good with $\hat{q}_c$. By Assumption 6, propaganda is worthwhile for both types no matter what they believe about the equilibrium strategy of the new media.

This equilibrium is also unique. To see this first notice that there is no equilibrium such that any good type R politician uses propaganda: good politicians are praised by the elite media which is believed by the normal voters. AR politicians are only in strategic interaction with the imaginary R politicians and not with the real R politicians. Similarly the imaginary R politicians only engage in strategic interaction with the AR politicians and not with the real R politician. As a result, we can rule out all alternative equilibrium strategy profiles of these politician types by only focusing on them. Proof of uniqueness of the strategy profile of AR and imaginary R is similar to the proof in the original game. This is because the real bad R poltician of the original game and the imaginary bad R politician of the new game only differ in their beliefs about the new media but propaganda is a dominant strategy for both without respect of what new media does. Finally, there is no equilibrium where AR politician and imaginary R plays as before but the real bad R politician restrains from propaganda, since using propaganda would make persuaded voter's perception equal to $\hat{q}_c$ if there is no new media signal, which makes propaganda worthwile.

Case 2: Incumbent politician is pro-elite. Consider the AR elite, who believes that propaganda is ineffective and the voter is normal. Since she wants to keep both the good and bad pro-elite politician, her optimal strategy is to always report the politician good. Consider next the AR politician. By avoiding propaganda she can ensure that the voter considers her good, since the AR elite—which the AR politician believes is the elite—always reports good and the normal voter ignores the new media. This is the best the AR politician can hope for, thus there is no reason to engage in costly propaganda, and the unique best response of both the good and bad AR politician is to avoid propaganda. Then, the bad R politician using propaganda cannot be part of an equilibrium, as propaganda would reveal her to be bad. Thus, no politician type uses propaganda. And, since the audience-seeker new media expects no propaganda, she expects to be ignored and thus sends an honest message to minimize the lying cost. The beliefs and behavior of the normal voter on the equilibrium path are then pinned down by the elite's message.

Finally, we characterize behavior at off-equilibrium information sets. Like above, these are only possible 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 will attribute propaganda to a tremble and form beliefs and behavior just as the normal voter absent propaganda. The persuaded voter after propaganda will attribute propaganda to a tremble and behave as if it did not occur. The persuaded voter absent propaganda will update as follows. If the elite's message was bad then, because the AR elite always reports good, he will conclude that reality is R and follow the elite's message regardless of the message of the new media. If the elite's message was good then he will attribute positive probability to the AR (the exact value of which will depend on the message of the new media and can be computed using Bayes rule) and follow the message of the new media. Because this is the last stage of the game, the persuaded voter faces a binary decision problem which that has a solution, and chooses that solution.

New media benefits politician. Having characterized the unique equilibrium, we now show that the presence of the new media weakly increases both the perception of AR and the reelection probability of the bad pro-voter politician. Consider the history in which a bad pro-voter politician uses propaganda and is criticized by the elite. Absent the new media, the voter's posterior that the politician is good is $\hat{q}_c < 1$, and his posterior that reality is AR can be computed analogously to (6) as

$$\mu(\theta_r = AR | \hat{s}_c^e = 0, \hat{p} = 1) = \frac{q_{ar}}{q_{ar} + q_r q_c} < 1. \tag{11}$$

In the presence of the new media, if the message of that media does not reach the voter, then posterior beliefs are the same as above. However, if the message of the new media reaches the voter, then his posterior beliefs become

$$\begin{aligned}\mu(\theta_r = AR|\hat{s}_c^e = 0, \hat{s}_c^n = 1, \hat{p} = 1) = 1,\\ \mu(\theta_c = 1|\hat{s}_c^e = 0, \hat{s}_c^n = 1, \hat{p} = 1) = 1.\end{aligned} \tag{12}$$

This is because the persuaded voter considers the new media honest and thus believes her message on the common type, and infers from the conflicting messages of the elite and the new media that the reality is AR. It follows that new media amplifies beliefs in the alternative reality, improves the perception that the politician is good, and increases the probability that she gets reelected.