DISCUSSION PAPER SERIES

DP17767

THE POLITICAL ECONOMY OF ALTERNATIVE REALITIES

Adam Szeidl and Ferenc Szucs

POLITICAL ECONOMY

ISSN 0265-8003

THE POLITICAL ECONOMY OF ALTERNATIVE REALITIES

Adam Szeidl and Ferenc Szucs

Discussion Paper DP17767
Published 22 December 2022
Submitted 12 December 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

THE POLITICAL ECONOMY OF ALTERNATIVE REALITIES

Abstract

We build a model in which a politician can persuade voters of a coherent alternative reality that serves to discredit the intellectual elite. In the alternative reality, members of the elite conspire, and criticize the politician's competence because 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 criticism. This model makes several predictions. (1) The alternative reality is spread by low-quality politicians and reduces accountability. (2) The alternative reality is only spread in sufficiently divided societies, and the nature of the divisive issue—cultural versus economic—determines whether right-wing or left-wing politicians spread it. (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 demand for non-elite media outlets (e.g., Fox News), which spread misinformation to reinforce beliefs in the alternative reality and sustain that demand. We discuss evidence consistent with these predictions.

JEL Classification: D03, D72, D82, D83

Keywords: N/A

Adam Szeidl - szeidla@ceu.edu
Central European University and CEPR

Ferenc Szucs - ferenc.szucs@su.se
Stockholm Univeristy

The Political Economy of Alternative Realities*

Adam Szeidl
Central European University and CEPR

Ferenc Szucs
Stockholm University

November 25, 2022

Abstract

We build a model in which a politician can persuade voters of a coherent alternative reality that serves to discredit the intellectual elite. In the alternative reality, members of the elite conspire, and criticize the politician's competence because 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 criticism. This model makes several predictions. (1) The alternative reality is spread by low-quality politicians and reduces accountability. (2) The alternative reality is only spread in sufficiently divided societies, and the nature of the divisive issue—cultural versus economic—determines whether right-wing or left-wing politicians spread it. (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 demand for non-elite media outlets (e.g., Fox News), which spread misinformation to reinforce beliefs in the alternative reality and sustain that demand. We discuss evidence consistent with these predictions.

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

JEL codes: D03, D72, D82, D83

*Emails: szeidla@ceu.edu, ferenc.szucs@su.se. We thank Nageeb Ali, Ruben Enikolopov, Matthew Gentzkow, Helios Herrera, Botond Koszegi, Kristof Madarasz, Maria Petrova, Giacomo Ponzetto, Jesse Shapiro, David Stromberg and conference and seminar audiences for comments and discussions, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement number 724501 for funding.

1 Introduction

A majority of Republicans with high science knowledge believe, contrary to the experts' consensus, that human activity does not contribute a great deal to climate change. Similarly, a majority of Republicans believe, contrary 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, that the government and the media 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, 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 a coherent but false alternative reality. Our approach builds on prior work about misinformation in politics, especially Glaeser (2005), Guriev and Treisman (2020), and Eliaz and Spiegler (2020), and contributes with two ideas. First, we explicitly model the actors—conspiring elites—who exist in the (false) alternative reality, and allow the voter to reason about and respond to their behavior, generating strategic interaction between the alternative and the objective reality. Second, we formalize an alternative reality in which members of the intellectual elite conspire, and criticise the competence of a politician because she disagrees with them about a divisive issue such as cultural values. A voter who is persuaded of this alternative reality will distrust the elite's truthful revelation of the politician's competence. We show that the internal coherence of the alternative reality constrains the behavior of many actors, and generates predictions about politics, media, and the non-adoption of best practices that are consistent with evidence.

In Section 2 we present our model. Our basic framework is a principal-agent model in which the incumbent politician and the intellectual elite are the principals and the median voter is the agent. Both principals can send messages to influence the voter's electoral behavior, and the voter then

[1] For beliefs about climate, the 2020 election, and conspiracies, see Funk and Kennedy (2020), Greenberg (2022), and Public Religion Research Institute (2021). More systematically, Stantcheva (2021), Dechezleprêtre, Fabre, Kruse, Planterose, Sanchez Chico and Stantcheva (2022), and Alesina, Ferroni and Stantcheva (2021) document misperceptions along partisan lines about tax and environmental policy and race, and Alesina, Miano and Stantcheva (2020) argue that partisan differences reflect different perceptions of the objective reality.

decides whether to keep or replace the politician. The politician has two payoff-relevant types. (i) A "common" type, over which the voter and the elite have the same preference, and along which the politician can be good or bad. Examples include quality and honesty. The voter does not observe this type dimension. (ii) A "divisive" type, over which the voter and the elite have different preferences, and along which the politician can be pro-voter or pro-elite. Two leading examples 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.[2]

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] Although—due to their current salience—many of our examples will be about cultural division and right-wing alternative realities, we emphasize that the model is equally applicable to left-wing alternative realities, and as we show below, predicts when we should expect one versus the other.

the election outcome. Intuitively, in the alternative reality the "liberal media" criticize Trump's competence not because he is incompetent, but because he is "anti-woke." In turn, the voter, since propaganda persuades him to partially believe the alternative reality, understands this mechanism and distrusts the report of the elite.

A key assumption in this framework is that propaganda can "irrationally" manipulate voters' prior beliefs about the elite. This assumption is consistent with well-identified evidence we discuss below about the impacts of propaganda, and with new suggestive evidence we present in Section 2 that populism is associated with lower trust in science and the media. And the logic of the alternative reality—that elite members act collectively to advance their goals —is consistent with the narrative of many conspiracy theories (Douglas, Uscinski, Sutton, Cichocka, Nefes, Ang and Deravi 2019).

The main results of our model are that (1) propaganda is only ever used if disagreement about the divisive issue between the voter and the elite is sufficiently large, and (2) in that case, the politician sends propaganda if and only if she is pro-voter on the divisive dimension and bad on the common dimension. The intuition for (1) is that the alternative reality in which the elite wants to remove the politician because they disagree can only be plausible if the disagreement is large. The intuition for (2) is that even then, the alternative reality is only plausible if the politician does in fact disagree with the elite, i.e., is pro-voter; and that the politician can only gain from discrediting the elite's truthful report about her if she is bad.

These results have several implications. Most directly, they imply that bad politicians are more likely to use propaganda and 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, here propaganda works by creating distrust in the elite, a mechanism consistent with the suggestive evidence mentioned above that populism is associated with such distrust.

A second implication is predictable variation in propaganda. Our results that propaganda is

only used if (i) disagreement is large, and (ii) by the pro-voter politician, predict both the presence of propaganda and whether it is left-wing or right-wing. In particular, (ii) predicts that when the main divisive issue is cultural values—on which the voter is plausibly to the right of the elite—the pro-voter politician is right-wing and we should observe right-wing populism; whereas when the main divisive is economic redistribution—on which the voter is to the left of the elite—we should observe left-wing populism. We present new cross-country evidence on both (i) and (ii) by showing that a larger cultural disagreement between the high- versus low-educated predicts right-wing (but not left-wing) populism, while a larger economic disagreement between the high- versus low-educated predicts left-wing (but not right-wing) populism. This new evidence suggests that the model captures an empirically important determinant of beliefs in alternative realities.

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 driven by its interests. Thus, propaganda creates distrust in the scientific consensus and leads to the non-adoption of scientific best practices. Consistent with this prediction, we show that Republicans are less likely to believe in climate change or vaccinate against Covid, and Allcott, Boxell, Conway, Gentzkow, Thaler and Yang (2020) show that they are less likely to engage in social distancing. More broadly, the prediction may help explain partisan differences in people's understanding of and reasoning about policies, e.g., concerning taxes or the environment, and about social outcomes such as racial gaps (Stantcheva 2021, Dechezleprêtre et al. 2022, Alesina et al. 2021).

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 dimension will lead the voter to believe that the politician is also part of the conspiracy (formally, that the politician's divisive

type has switched). The politician will then set policy to contradict the elite consensus and thereby maintain the support of the voter. Since addresing major societal challenges, e.g., in the climate and health domains, often require government policy, this prediction highlights a first-order cost of propaganda. Consistent with the prediction, we document that Republican governors were less likely than Democrats to introduce mask mandates or vaccinate publicly.

Our second application is motivated by the salient fact that many non-traditional media outlets, most prominently Fox News, spread alternative realities. This fact is not easily explained by existing theories, which predict that media slant the presentation of facts (Mullainathan and Shleifer 2005, Gentzkow and Shapiro 2006), but not that they present non-truths and alternative realities. Our model provides an explanation based on the idea that the more discredited the elite media, the 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 falsely reporting that the propaganda-spreading politician is good and thereby strengthening beliefs in the alternative reality. This framework makes the new predictions that non-traditional media amplify the effect of propaganda and further reduce trust in scientific best practices. These predictions may help explain the quantitatively large extent of misbeliefs in U.S. society, and the harmful effects of Fox News on social distancing and Covid deaths (Bursztyn, Rao, Roth and Yanagizawa-Drott 2020, Simonov, Sacher, Dubé and Biswas 2020).

Our paper builds on overlapping literatures in political, behavioral, and information economics. Most directly, we build on work studying the supply of misinformation in politics. Foundational contributions include Glaeser (2005) on the supply of hatred and Besley and Prat (2006) on media capture. Ash, Mukand and Rodrik (2021) model the supply of "worldview politics" which alter voters' understanding of how the world works. Closest to our paper, Guriev and Treisman (2020) model "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 political-economic implications. A conceptual framework underlying much of the work on political misinformation

is Bayesian persuasion, formalized by Kamenica and Gentzkow (2011).[3] We depart from that framework 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. Evidence on the supply of misinformation includes studies of 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 assumption that propaganda influences voter beliefs.

Our model of the alternative reality builds on behavioral-economic research on persuasion and narratives. Theories of persuasion include Mullainathan, Schwartzstein and Shleifer (2008), Galperti (2019) and Schwartzstein and Sunderam (2021). Closer to our work, in the political economy domain, Eliaz and Spiegler (2020) and Eliaz, Galperti and Spiegler (2022) study the emergence of competing and false narratives in political equilibrium, Levy, Razin and Young (2022) study political dynamics when one group has a misspecified model, and Gentzkow, Wong and Zhang (2021) study the co-evolution of opinions and trust in news sources. Our conceptual contribution to this work is that the narrative in our model—the alternative reality—involves optimizing agents who strategically interact with the objective reality. This allows us to formalize an alternative reality that can discredit the elite, and derive our political-economic implications.

Our analysis 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) and Besley and Persson (2021). Our contribution to this research is to show how cleavages can be exploited with a strategically-interacting alternative reality that serves to discredit the elite, and the political-economic implications.[4]

Finally, our modelling approach builds on a theoretical work studying learning and interaction under model misspecification, including Berk (1966), Jehiel (2005), Esponda and Pouzo (2016) and Heidhues, Kőszegi and Strack (2018). Our conceptual contribution to this work is to model the decision to create misspecification.

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

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

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.[5] 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 abilities and beliefs, in particular, the AR elite is able to conspire. 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. We now turn to present the formal framework in detail.

Neoclassical setup. It is helpful to start by presenting the neoclassical (non-behavioral) part of the model. There are three classes of actors, the politician p, the intellectual elite e and the voters v. We say classes of actors because both the elite and the voters consist of a unit mass of identical members. We think about the elite as the news media, and assume a one-to-one correspondence between elite members and voters, so that each elite member has exactly one voter as its audience. This assumption ensures that individual elite members cannot affect the election outcome.[6] As we

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

[6] More generally we could allow each elite member to have a zero measure of voters as its audience.

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 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", and 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 about the divisive issue align with that of the voter, while $\theta_d = 0$ means that the politician is pro-elite, i.e., her preferences align with that of the elite. We assume that θ_d is observed by all actors, and that θ_d and θ_c are drawn independently.

After observing the 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.

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 and abilities. 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

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

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 R and the AR principals is $q_r > 0$ and $q_{ar} > 0$ (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 he is not reached by propaganda and persuaded if he is reached by propaganda. We assume that the voter conducts any updating based on the messages he observes from the prior encoded in his mind type. 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.

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.[7] Here $c > 0$ measures the importance of the common issue, and $\lambda > 0$ measures the strength of

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

disagreement on the divisive issue.[8] Thus, the elite derives utility c from a politician who is good on the common issue, but disutility λ from a politician who is pro-voter on the divisive issue. We further assume that each elite member has a small preference for sending a truthful message, thus if otherwise indifferent tells the truth.

The key difference between the R and the AR elite is their ability to coordinate: 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 by 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 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 a common 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

[8] In particular, λ can be thought of as the product of the importance of the divisive issue times the extent of misalignment in the preferences—the difference between the ideal points—of the elite and the voter.

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 all voters behave in the same way and represent them as a single actor.

Trembles. We make the assumption that both the elite's message s_c 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. We discuss real-world examples for the trembles below, and elaborate on their role in the equilibrium analysis after stating the main result. Formally, we assume that 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.

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.

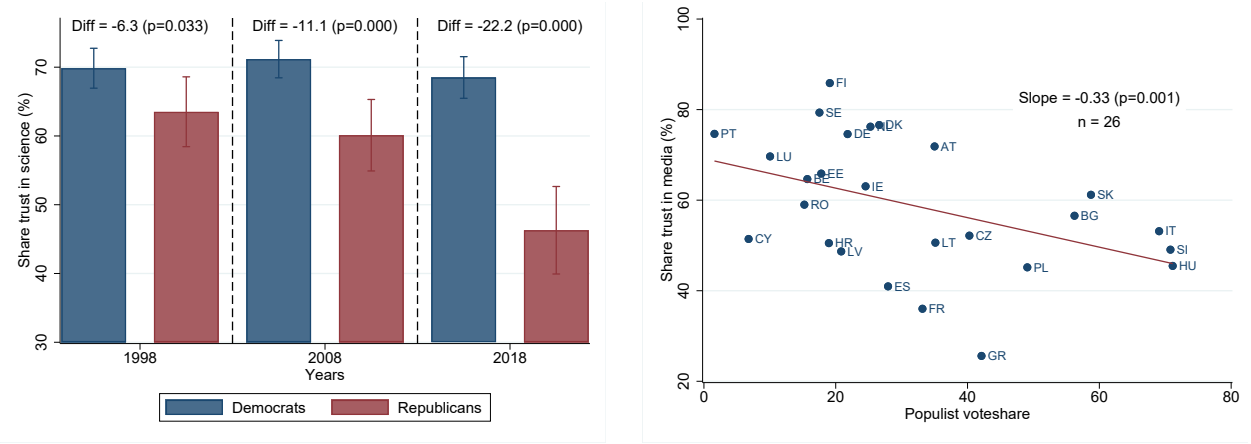
1. The elite sends message $s_c \in \{0, 1\}$ and the politician decides on propaganda $p \in \{0, 1\}$. Both messages are subject to trembles and all actors observe the realized messages $(\hat{s}_c, \hat{p})$. If $\hat{p} = 1$ then the voter's mind type changes to $\theta_m = P$.

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

3. Payoffs realize.

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

Figure 1: Distrust in the intellectual elite



2.2 Equilibrium

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

The key novelty in this definition is the Bayesian updating of the voter. We assume that after stage 1 the posterior of each voter mind type $\theta_m = N, P$ is computed using the prior associated with that mind type. In particular, if the voter is reached by propaganda and becomes persuaded, his posterior is computed from the prior which assigns probability $q_{ar} > 0$ to the alternative reality. This definition allows the voter who is reached by propaganda to make Bayesian inference from the elite message and from propaganda; but the order of updating is that first propaganda changes his prior, and then he makes the inference using the modified prior. Because aside from this novelty our equilibrium concept is essentially standard, we relegate the formal definition to the Appendix.

2.3 Discussion of model assumptions

Departure from rationality. Beliefs in false conspiracy theories plausibly require some departure from rationality (Grimes 2016). Our departure is to assume that propaganda can modify the voter's prior belief and make him assign positive probability to a non-existent alternative reality. The assumption that propaganda affects beliefs is consistent with evidence from different contexts that propaganda affects behavior 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 shows that during the 1998-2018 period of increasingly anti-intellectual Republican party rhetoric, Democrats' trust in science remained largely unchanged while Republicans' trust in science substantially declined. And Figure 1B shows that across 26 European countries the vote share of populist parties is a strong predictor of distrust in the media.[9] Moreover, 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 et al. 2019).

Specifics of the model. Beyond the above departure 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 because it seems plausible that the politician spreading propaganda would want to communicate that she believes in it. As we show below, the AR politician will have a key role in making the alternative reality believable to the voter. Second, we assume that propaganda changes the voter's belief about the elite, but not about whether the politician is good. We do this because changing the belief that the politician is good, absent changing the belief about the elite, would not be effective: the elite's message would immediately correct beliefs.[10] 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 elite

[9] 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.

[10] We note, however, that positive propaganda may be effective in conjunction with changing beliefs about the elite; and that there may be other forms of propaganda, e.g., exploiting fear, which are outside our model.

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 intuition that it is easier to create than to eliminate beliefs in a conspiracy theory.

Equilibrium concept. As is standard in economics, our equilibrium concept assumes that actors know each others' strategies. In our setting with manipulable priors, equilibrium does not seem easily justifiable with learning. However, although a formal foundation is beyond the scope of this work, a plausible informal justification may be based on persuasion and introspection. The propaganda-spreading politician may explain the narrative of the equilibrium to make propaganda persuasive (Shiller 2017, Eliaz and Spiegler 2020). And the voter may fill in any gaps in the politician's narrative by thinking through the motives of the other actors, a process aided by the fact that the equilibrium in our setting will be unique.[11]

Real-world analogues of model components. We highlight some real-world examples for the common and divisive issue and 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, LGBTQ rights, and 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, i.e., 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.

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

2.4 Results

Key to our main result is that propaganda can partially deflect the elite's criticism. In preparation for stating the result, we highlight the logic for how deflection works. Suppose that the politician is pro-voter, and consider the following strategies, which will be part of the equilibrium: (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 the common type honestly while the AR elite criticizes always. Consider the beliefs of the voter after observing propaganda and criticism. If the voter were normal, i.e., assigned zero probability to AR, he would learn that the politician is bad, for two reasons: the R elite's message is truthful, and in R propaganda is only sent by the bad politician. In contrast, the persuaded voter, who assigns probability $q_{ar} > 0$ to the AR, believes that the politician is good with probability

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

Consider the numerator: In reality R both propaganda and criticism imply that the politician is bad, but in reality AR both propaganda and criticism are expected irrespective of whether the politician is good, so the probability of observing both and having a good politician is $q_{ar} q_c$. Consider the denominator: Propaganda and criticism will also arise if the politician is bad, in both R and AR, explaining the term $1 - q_c$. Note that $\hat{q}_c > 0$: the voter updates from propaganda and criticism about the common type only partially, because when reality is AR he expects both messages even for the good politician. It follows that the posterior $\hat{q}_c$ measures the extent to which propaganda deflects criticism, and will therefore play an important role in the analysis.[12]

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

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

Recall that E is the utility from being in power, $\hat{q}_c$ is the expected improvement from propaganda in the voter's belief that the politician is good, c is the benefit of having the good politician and g

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

is the density of the politician's popularity shock. The assumption thus ensures that spreading the conspiracy theory—if it succeeds in changing beliefs—is profitable to the bad politician.

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

We say that the equilibrium is in monotone strategies if whenever an elite type reports the bad politician to be good, she also reports the good politician to be good. We are now ready to state the main result.

Proposition 1. *Suppose Assumption 1 holds.*

1. *If the divisive issue is unimportant, $\lambda < \underline{\lambda}$, then in the unique equilibrium in monotone pure strategies, both in the reality (R) and in the alternative reality (AR):*
   - *The elite reports the common type truthfully,*
   - *No politician sends propaganda.*
2. *If the divisive issue is important, $\lambda > \overline{\lambda}$, then in the unique equilibrium in monotone pure strategies,*
   (a) *In the reality (R):*
      - *The elite reports the common type truthfully,*
      - *The politician sends propaganda if and only if she is pro-voter and bad.*

   (b) *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.*

   (c) *Propaganda increases the reelection probability of the bad pro-voter politician.*

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

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

To understand the inner logic of this equilibrium, it is helpful to flash out how the behavior of the other actors, in both R and AR, supports it. We focus on the interesting case in which $\lambda > \overline{\lambda}$ and the politician is pro-voter. The behavior of the R elite is straightforward: because its members are atomistic and cannot influence the voter, they prefer to report truthfully. Now consider the AR actors (part 2b). The AR elite, as we have seen—since the divisive issue is sufficiently important—wants to remove the pro-voter politician, and therefore always criticizes her competence. Consider next the AR politician. Both the good and the bad type believe that the elite is AR and criticizes their competence. Therefore both choose propaganda to deflect this criticism. Finally, consider the voter. Key to the effectiveness of propaganda is that the voter does not infer from observing it that the politician is bad. This is because propaganda makes the probability of AR positive, and in the AR both the good and the bad politician sends propaganda. This logic, which underlies equation (5), prevents the full revelation of the R politician's bad type.

We next explain how the equilibrium relies on the trembles. The assumption that the elite's

tremble is arbitrarily smaller than the politician's ensures the intuitive step above that the AR elite actually criticizes the politician if she wants her out. Because in equilibrium, absent trembles, propaganda and criticism are perfectly correlated, the AR elite may have an incentive to avoid the lying cost and let the voter update just from observing propaganda. But because her message is arbitrarily less noisy than propaganda, she will have a dominating effect on the voter's updating, which ensures that she will still criticize. A second role of the trembles is to ensure that beliefs are defined after histories off the equilibrium path.

Finally, the intuition for uniqueness—in the interesting case in which the politician is pro-voter—follows through two steps. First, in any pure strategy equilibrium the AR elite always criticizes, because even the persuaded voter assigns positive probability to reality ($q_r > 0$), implying that there is always some to manipulate the voter. Second, any politician who is criticized has a preference for deflecting that criticism, implying that the bad R as well as both AR politicians—who are all criticized—send propaganda.

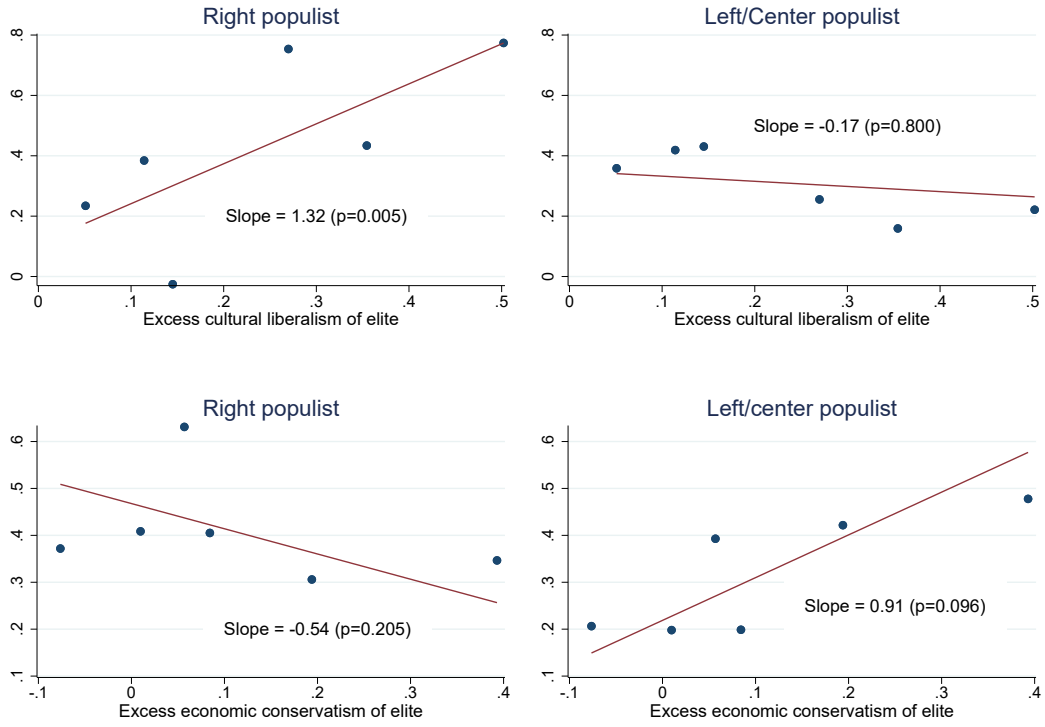
Having characterized the equilibrium, the prediction in part (2c) that propaganda increases the reelection probability of the bad pro-voter politician follows directly. On the equilibrium path, if propaganda fails because of a tremble, the voter remains normal and will correctly interpret the elite's 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 consistent with Guriev and Treismann's (2022) description of informational autocracy, in which autocratic leaders—interpreted as "bad" in our model—stay in power by means of government propaganda. In the Guriev and Treisman (2019) model, propaganda works by improving voters' beliefs about the politician's type. In contrast, here propaganda works by discrediting the elite. In our model, propaganda about the politician's type, absent discrediting, would not work: the

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



elite's truthful message would immediately correct beliefs. Intuitively, discrediting is necessary for "positive propaganda" to be effective. Consistent with the mechanism of discrediting, populism is associated with lower trust in the elite on both sides of the Atlantic (Figure 1).

Propaganda is only used in divided societies by the pro-voter politician. A second implication is predictable variation in propaganda. Our results that propaganda is only used if (i) disagreement is large, and (ii) by the pro-voter politician, predict both the presence of propaganda and whether it is right-wing or left-wing. In particular, (ii) predicts that when the main divisive issue is cultural values, so that the voter is to the right of the elite, the pro-voter politician is right-wing and we should observe right-wing propaganda; but when the main divisive issue is economic redistribution, so that the voter is to the left of the elite, we should observe left-wing propaganda.

We document new cross-country evidence on both (i) and (ii). We use the 7th wave of the World

Values Survey to measure, in 29 democratic countries, the extent of disagreement in the cultural respectively economic domain. Specifically, we measure the excess cultural liberalism of the elite as the gap in attitudes of people with versus without a masters degree on a set of cultural issues: immigration, gender inequality, discrimination of sexual minorities, religion, and national pride. And we measure the excess economic conservatism of the elite as the analogous gap in attitudes about income inequality and state ownership. We then correlate these measures with the presence of populist parties as classified by the Global Party Survey (GPS).[13]

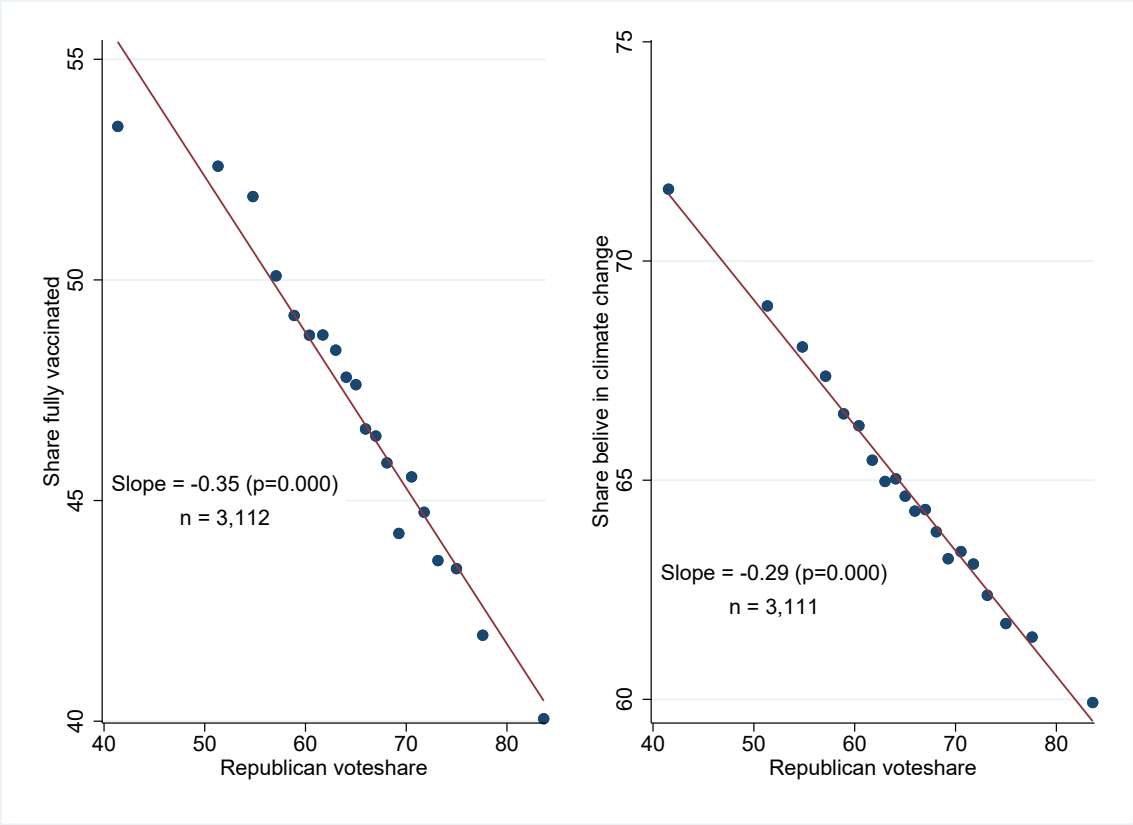
Figure 2 shows the results. The top panels show that cultural disagreement predicts right-wing but not left/center populism; while the bottom panels show that economic disagreement predicts left/center but not right-wing populism. Thus, the panels along the main diagonal support prediction (i) that larger disagreement is associated with more populism; and their comparison with the panels in the off-diagonal support prediction (ii) that the domain of disagreement predicts whether populism is left-wing or right-wing. This new evidence about the emergence and nature of populism suggests that the model succeeds in capturing an empirically important determinant of beliefs in alternative realities.

Moving beyond the Figure, the logic of predictions (i) and (ii) suggests that the growth of beliefs in alternative realities in the U.S. may be driven by growing cultural disagreement, which makes 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 third implication is that once the elite has been discredited, the voter will no longer trust its advice in other non-political domains either. We formalize this point in the Appendix by introducing a new action the voter can take after stage 1, such as vaccinating against Covid or acting in a climate-conscious manner. The action may be good or bad for the voter, and independently, good or bad for the elite. The elite sends a message about whether the action is good for the voter. We show that if the new issue is sufficiently unimportant that it does not change the equilibrium of the base model, then after propaganda the

[13] 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 Movement in Italy or Syriza in Greece.

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



voter will follow the elite's advice too infrequently. Intuitively, because the voter and the elite have imperfectly aligned interests in the new domain, the voter worries that the conspiring elite may be choosing its message to manipulate him, and thus becomes too cautious in following that message.

The prediction that propaganda limits the adoption of best practices outside politics is consistent with the beliefs and behavior of Republicans in the health and climate domains. Figure 3 illustrates this by documenting, across U.S. counties, a strong negative association—controlling for demographics—between the Republican vote share and both vaccination rates and beliefs in human-made climate change.[14] Differences in beliefs seem to generate differences in behavior: Allcott et al. (2020) show that Republicans are less likely to engage in social distancing. And,

[14] 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.

consistent with the mechanism our model highlights, trust in science may be a key factor explaining these differences: Algan, Cohen, Davoine, Foucault and Stantcheva (2021) show that trust in science is a key driver of compliance with Covid-related non-pharmaceutical interventions. More broadly, propaganda-driven distrust in experts may help explain partisan differences in people's understanding of and reasoning about policies, e.g., concerning taxes or the environment, and social outcomes such as racial gaps (Stantcheva 2021, Dechezleprêtre et al. 2022, Alesina et al. 2021).

Discrediting versus censorship. From the perspective of the politician, an alternative to discrediting is to silence the media's criticism using censorship (Guriev and Treisman 2020).[15] Our model offers some insights about this tradeoff. Censorship is known to be expensive, vulnerable to deviations by independent media, and politically costly (Besley and Prat 2006). However, our model suggests that 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 yields the new testable prediction that citizens in autocracies relying more on censorship (e.g., China) should trust the scientific consensus more than those in autocracies relying more on propaganda (e.g., Russia). The tradeoff between discrediting and censorship can change if the media gains access to irrefutable evidence that can puncture the alternative reality: then the politician will have a stronger incentive to use censorship, helping to explain why, following the invasion of Ukraine—which increased the availability of difficult-to-refute evidence—Russia shifted towards state control of the media.

Choice of issue and extent of division. Although outside our formal model, a natural intuition emerging from our framework is that bad politicians have a stronger incentive to become pro-voter on a highly divisive issue. We thus expect bad politicians to (1) focus more on issues on which they are pro-voter (rather than pro-elite), and to (2) increase disagreement (λ) on such issues. These intuitions suggest that effective propaganda, beyond spreading the narrative that elites conspire, should also identify divisive issues and fuel divisions, consistent with descriptions of propaganda (Yanagizawa-Drott 2014, Adena et al. 2015).

[15] See Gehlbach and Sonin (2014) and Shadmehr and Bernhardt (2015) for related theoretical models of censorship.

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 new media outlets which then reinforce 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 the politician's desire to maintain beliefs in the alternative reality constrains her 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.

Framework with government policy. To model this intuition, we extend our basic framework to incorporate government policy. Specifically, we add new stages to the model which require the politician to act competently 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 her competence is realized only after the issue emerges. The politician chooses how to act about the issue: a competent politician can act either competently or incompetently, while an incompetent politician can only act incompetently.[16] If she acts competently, she increases per capita consumption by k. 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 vanishing tremble 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

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

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 suddenly wants to praise her: this is why we view the switch as a metaphor for the politician joining the conspiracy. The switch occurs simultaneously to the competence realization, in both the R and the AR realities, and is only observed by the politician and the elite. We denote the politician's initial divisive type by θ_d^0, and her eventual divisive type by θ_d.

New substantive assumption. In this extended framework we make one new substantive assumption, which allows the persuaded voter to directly "see through" the conspiracy of the AR elite. We assume that the AR elite—who only exists in the voter's mind—thinks that propaganda reaches only a minority of voters and is thus ineffective. Formally, we assume the AR elite believes that the (median) voter's type after propaganda remains normal. This assumption 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. This assumption resonates with conspiracy narratives in which the politician and her followers are aware of the conspiracy, but the conspirators, who take steps to hide their intentions, do not yet realize their awareness.

We need this assumption because the fully opposing incentives, in our stylized model, of the R politician and the AR elite make it difficult for the former to manipulate the criticism of the latter. In any equilibrium, criticism about the new issue either increases or decreases the reelection probability. If it increases then the AR elite has an incentive not to criticize, while if it decreases then the R politician has an incentive not to trigger criticism. To moderate this force, we need a friction that partially decouples the incentives of the two actors. We choose a friction that concerns the AR elite's perception of the voter because it feels realistic in our setting.

Preferences and timing. The behavior of the elite and the voter can now be characterized by the following objectives:

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

where $\tilde{\theta}_k$, $\tilde{\theta_c}$ and $\tilde{\theta}_d$ are the (new) 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 policy preferences, and differs from the basic model because it includes the competence of the elected politician about the new issue. The second term, active when 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.[17] 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 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.

Result. We formally define the equilibrium of this richer model in the Appendix. To state our result, we need a parametric assumption.

Assumption 2. For the voter, a politician who is bad but competent on the new issue is worse than a politician who is good with probability $\hat{q}_c$ but incompetent on the new issue:

$$\hat{q}_c c > k.$$

[17] 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 micro-founded with a continuation game without further reelection concerns.

The assumption implies that the common type is sufficiently important that the politician prefers to (partially) hide her bad type even at the cost of appearing to be incompetent about the new issue. This will ensure that there is no equilibrium in which the bad R politician behaves competently about the new issue.

Finally, we define $\overline{\overline{\lambda}} = \max\{((1-q_c)c + (1-q_k)k)/(1-q_d), (q_c c + q_k k)/q_d\}$. This is a generalization of $\overline{\lambda}$ to the current setting: $\lambda > \overline{\overline{\lambda}}$ ensures the divisive issue is sufficiently important that the elite prefers to remove the pro-voter politician, and keep the pro-elite politician, irrespective of both her common type and competence about the new issue.

Proposition 2. *Under Assumptions 1 and 2, for $\lambda > \overline{\overline{\lambda}}$, for generic parameters and ξ sufficiently low, there is a unique pure strategy equilibrium path in which:*

(a) The first stage of the game unfolds as before.

(b) In the reality (R), absent propaganda, the politician acts competently if she can.

(c) In the reality (R), after propaganda

- *The elite reports about competence truthfully,*
- *The politician acts incompetently.*

(d) In the alternative reality (AR), both absent and after propaganda

- *The elite reports the politician incompetent if and only if the politician is pro-voter,*
- *The politician acts competently if she can.*

(e) After propaganda, elite criticism about competence increases the reelection probability of the bad pro-voter politician.

Part (a) shows that the first stage of the game unfolds as before. Part (b) shows that absent propaganda, behavior is as expected: the politician acts competently whenever her raw competence type allows it.

The key part of the result is (c), which shows that after propaganda, the (bad pro-voter) politician always acts incompetently about the new issue. In equilibrium, this outcome is supported

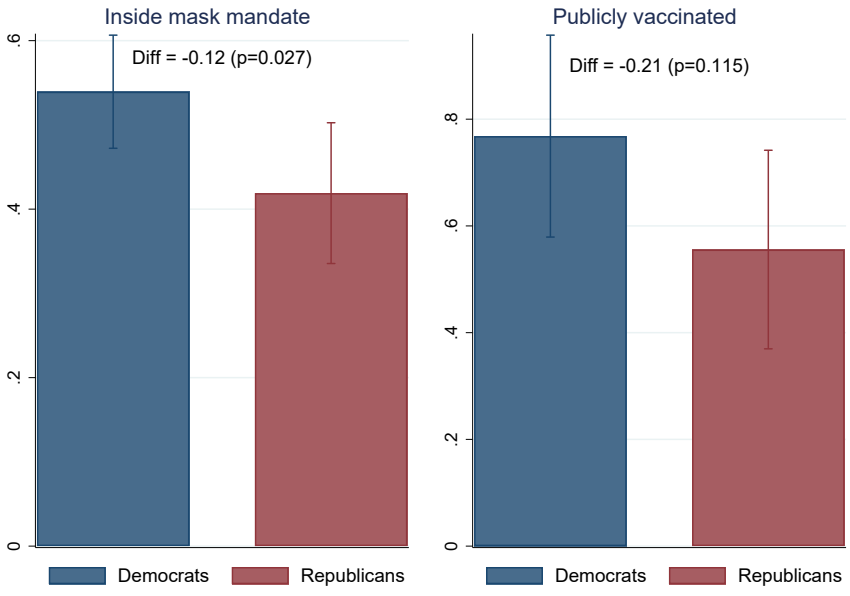
by the behavior in the alternative reality characterized in part (d), namely, that the AR elite criticizes about the new issue if and only if the politician is pro-voter. To see why (c) and (d) constitute an equilibrium, suppose that the (bad pro-voter) politician deviates and acts competently. The elite reports this. The key is to then note that 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, by part (d), it can only come if reality is AR after a type switch to a pro-elite politician. Thus, the deviation would lead the voter to conclude that the politician is pro-elite. Since the voter dislikes pro-elite politicians, to avoid this conclusion the politician acts incompetently.[18] This is how our model formalizes the intuition that praise from the discredited elite is interpreted to mean that the politician is part of the elite conspiracy. Finally, the reason the AR elite criticizes if and only if the politician is pro-voter (i.e., part (d)) is that due to $\lambda > \overline{\overline{\lambda}}$ this reflects her preference, and because the AR elite perceives the voter to be normal and hence manipulable, she always acts on her preference.

To build intuition for uniqueness, focus on the case where the politician is bad and (initially) pro-voter. A first observation is that, because the AR elite believes that the voter is normal and follows her message, in any equilibrium she criticizes about the new issue if and only if the politician is pro-voter. Given this, in any equilibrium, the elite's message that the politician acts competently about the new issue can come in two scenarios: either in reality AR after a type switch (to pro-elite), or in reality R if the politician indeed acts competently. Both of these scenarios decrease the value of the politician to the voter. In the former scenario, as we explained in the previous paragraph, this follows because the voter dislikes the pro-elite politician. In the latter scenario, it follows because then in reality R the politician must be bad (since she was criticized about her common type), and Assumption 2 ensures that being perceived bad is costly relative to any gain in perceived competence. Since acting competently is costly in both scenarios, in any equilibrium the propaganda-spreading politician will act incompetently.

Finally, part (e) says that elite criticism actually helps the propaganda-spreading politician.

[18] There is an additional subtlety, namely that the voter also learns that reality is AR, which implies that he perceives the politician to be less likely to be bad as the elite's first-period criticism is now perceived uninformative. But $\lambda > \overline{\overline{\lambda}}$ ensures that the cost of being perceived pro-elite is higher than the benefit of being perceived as less bad.

Figure 4: Impact on policy



This follows immediately from the previous parts: if the elite, instead of truthful criticism, deviates to praise about the new issue, the voter would conclude that the politician switched to being pro-elite, and would re-elect her with a lower probability. Intuitively, criticism from the "lying New York Times" is a badge of honor for the propaganda-spreading politician.[19]

Implication. The key prediction of this application is that propaganda-spreading politicians choose policies, even in non-political domains, which contradict the expert consensus. It is not just that these politicians ignore expert opinion: they actively set policy to contradict it. Since major societal issues, e.g., in the health and environmental domain, often require government action, this prediction highlights a potentially first-order social cost of propaganda.

Figure 4 presents evidence on this prediction in the Covid context. The left panel shows that across U.S. states and over time, controlling for the severity of the epidemic, Republican governors introduced indoor mask mandates 12 percentage points less often than their Democratic

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

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.[20]

3.2 New media and beliefs in the alternative reality

A salient fact about U.S. media is that several non-traditional outlets, most prominently Fox News, spread and reinforce the false alternative realities propagated by Republican politicians.[21] This fact appears to be unexplained by existing theories. It cannot be easily explained by theories of captured media (Besley and Prat 2006) since there is no evidence that non-traditional outlets are controlled by politicians. Nor can it be easily explained by theories of independent media (Mullainathan and Shleifer 2005, Gentzkow and Shapiro 2006), which predict that media slant the presentation of facts, but not that they present non-truths and alternative realities. In this application we propose an explanation based on the idea that demand for non-traditional media arises because of audiences' distrust in the elite media, implying that it is in the best interest of the non-traditional media to sustain that distrust by reinforcing the alternative reality.

Framework with new media. This application requires that we model the new (non-traditional) media outlets. Paralleling our model of the elite media, we assume that each new media outlet is too small to influence elections: formally, that there is a continuum of new outlets linked by a one-to-one mapping to voters, such that each voter consumes exactly one elite and one new media outlet. Thus, each voter represents the potential core audience of the corresponding elite and new media outlet. Because the new outlets have identical incentives, we only consider equilibria in which they have identical strategies, and treat them as a single decision maker. Like the elite media, the new media observes the politician's common type and sends a message about it to the voter. The new

[20] 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.

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

media is slightly less informed than the elite: formally, the message of the former has a vanishing tremble which becomes arbitrarily larger than that of the latter (but still arbitrarily smaller than that of propaganda). To simplify off-equilibrium belief calculations, we further assume that the elite's tremble is arbitrarily smaller than even the simultaneous occurrence of the new media and the propaganda trembles. We assume that the new media may fail, so that its message only reaches the voter with probability α, where α is not too high (see below in Assumption 3): this ensures that irrespective of the action of the new media the politician will have an incentive for propaganda.[22]

Since our focus is the behavior of the media, we enrich their preferences by (i) incorporating audience-seeking and (ii) explicitly formalizing the lying cost. We model audience-seeking by assuming that each elite media outlet wants to maximize the belief of its audience that reality is R, while each new media outlet wants to maximize the belief that reality is AR. This assumption captures the essence of competition between outlets: If reality is R, then, because the elite media is more informative (has a smaller tremble), the voter should prefer it in the future; whereas if reality is AR, then, because the elite media conspires, the voter should prefer the new media in the future. Turning to lying costs, we now model them explicitly and allow them to be different between the elite media and the new media.

New substantive assumptions. We make two substantive assumptions in this framework. First, we assume that the voter underestimates the strength of the media's audience-seeking preferences, which we model starkly by assuming she is not aware of those preferences. This assumption seems realistic: as discussed by Gentzkow and Shapiro (2010), a prominent view among regulators is that media owners' ideology preference is the key determinant of content, whereas in practice, in the newspaper context, audience preferences are much more important.[23] Formally, we introduce a new type dimension for the R media outlets, $\theta_a \in \{0, 1\}$, where 0 means that media do not, and 1 means that media do have audience-seeking preferences. The voter has the incorrect prior belief that the probability of $\theta_a = 1$ is zero, while the objective probability is one. 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

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

[23] We conjecture that a model in which the voter has the correct prior about the (interior) probability that media are audience-seeking would generate all our results except the on-average amplification effect of the new media.

of them have audience-seeking preferences. Because the AR outlets only exist in the voter's mind, these never care about audiences and do not have the new type dimension.

Second, we assume that lying costs are higher for the elite media than for the new media. This assumption—formally stated as Assumption 4 below—captures that the former has a preexisting audience it could lose by lying, while the latter does not and is hence less constrained.[24]

Finally, we make the natural assumption that the new media outlets are pro-voter, i.e., have the same policy preference as the voter. Because each outlet is small these preferences do not affect behavior, but they rule out the possibility in the voter's mind that the new media are in the conspiracy: it is the elite that potentially conspires, and the pro-voter new media is by definition not in the elite.

Objectives and timing. Since both the elite and the new media send signals about the politician's common type, in this subsection we denote their messages by s_c^e and s_c^n. The behavior of the elite and new media are governed by the following objectives:

$$U_e = 1_{\{\theta_r=R\}} \cdot \left[\phi 1_{\{\theta_a=1\}} \cdot \mu_v(R|\hat{s}_c^e, \hat{s}_c^n, \hat{p}) + \chi_e 1_{\{s_c^e=\theta_c\}}\right] + 1_{\{\theta_r=AR\}} \cdot (c\tilde{\theta_c} - \lambda\tilde{\theta}_d), \tag{6}$$

$$U_n = \phi 1_{\{\theta_a=1\}} \cdot \mu_v(AR|\hat{s}_c^e, \hat{s}_c^n, \hat{p}) + \chi_n 1_{\{s_c^n=\theta_c\}}. \tag{7}$$

Start with the elite. The first term, active when reality is R, has two parts. The first part captures audience-seeking when $\theta_a = 1$, and is governed by the voter's posterior belief that reality is R—denoted $\mu_v(R|\hat{s}_c^e, \hat{s}_c^n, \hat{p})$—and a weight ϕ representing the importance of audience-seeking. The second part captures the 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.[25] 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. Note that for the new media we do not include policy preferences even in the AR. This is for the aforementioned reason that the new media does not conspire thus its policy preferences are irrelevant for behavior.

The timing of events is the following.

[24] In the Appendix we formally develop the model with preexisting audiences that provides microfoundations for this assumption.

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

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 $s_c^n \in \{0,1\}$, 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.

Result. We now state the parametric assumptions needed for our result.

Assumption 3. For the bad pro-voter politician, the benefit of partially hiding her common type, even if only in the event in which the new media's message does not reach the voter, is higher then the cost of propaganda:

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

This assumption strengthens Assumption 1 by making propaganda profitable even if it only influences beliefs when the new media's message does not reach the voter. The assumption ensures that the bad R politician will continue to choose propaganda irrespective of the behavior of the new media.

Assumption 4. For the elite media truth-telling 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 the elite media reports honestly as in the basic model; the second half ensures that the new media is willing to lie 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 the gain to the new media from increasing that probability is smaller.

Proposition 3. *Under Assumptions 3 and 4, if* $\lambda > \overline{\lambda}$*, on the unique pure strategy monotone equilibrium path*

(a) The elite and the politician behave the same way as in Part 2 of Proposition 1.

(b) The objectively existing new media, which has audience-seeking preferences

- *Always reports the pro-voter politician to be good,*
- *Reports the common type of the pro-elite politician truthfully.*

(c) The imagined new media—which does not have audience-seeking preferences—reports the common type of all politicians truthfully.

(d) The presence of new media amplifies the effect of propaganda on misbeliefs and increases the reelection probability of a bad politician.

Part (a) shows that, because the impact of new media is sufficiently small (Assumption 3), the first stage of the game unfolds as before. The key result is in part (b): the new media will report the bad pro-voter politician—who spreads propaganda—to be good. In equilibrium, this behavior is supported by the imagined new media always reporting truthfully (part c). The intuition for (c) is immediate, since the imagined new media does not have audience-seeking preferences. Given this, (b) follows because—to gain audiences—the new media wants to maximize beliefs in the alternative reality. Reporting that the politician is good achieves this: since the voter perceives the new media to be truthful (part (c)), he will conclude from the contradictory reports of the elite and the new media that the elite must be conspiring, i.e., that reality is AR. This logic requires that the elite chooses to be truthful while the new media chooses to lie, which follows because the lying cost of the elite is sufficiently larger than that of the new media (Assumption 4).

Part (d) shows that the presence of new media amplifies the impact of propaganda on misbeliefs. This follows because in the absence of new media, beliefs assign positive probability to reality being R, while in the presence of new media—in the α probability event that they reach the voter—beliefs assign full probability to reality being AR. Note that this prediction relies on the voter underestimating the audience-seeking motive of the new media: with a correctly specified voter

there should not be belief distortion on average. Intuitively, Fox News can strengthen beliefs in the alternative reality because the voter does not fully account for its incentive to build audience.

Implications. The result helps explain our motivating fact that new media like Fox News spread false alternative realities. It also yields two new implications. (1) It predicts that private propaganda amplifies the effect of government propaganda. This prediction may be a quantitatively important reason for the widespread misbeliefs observed in U.S. society today. (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. This prediction is in line with evidence showing that the consumption of Fox News reduced social distancing and increased mortality during the Covid pandemic (Bursztyn et al. 2020, Simonov et al. 2020).

4 Conclusion

In this paper we built a model in which a politician can supply an alternative reality to discredit the criticism of the intellectual elite. Key to our approach is to explicitly model an alternative reality that incorporates optimizing actors, and have the voter reason about and respond strategically to the imagined behavior of these actors. Requiring that the alternative reality is internally consistent and not contradicted by evidence constrains the types of alternative realities that can be spread, and the behavior of the voter, the government, and the media. We have shown that these constraints generate new predictions about politics, media, and adoption of best practices, many of which are consistent with available evidence.

One limitation of our approach is that it is silent about the demand side: why voters are willing to believe in alternative realities. Developing a behavioral-economic theory of the demand side is a promising avenue for research that 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.

Our approach of modeling a coherent and strategic alternative reality may be useful in other domains. One class of examples may be political ideologies which may be represented as oversimplified alternative realities. Widespread beliefs in a political ideology may constrain the politician

spreading that ideology: for example, pro-market reforms may shatter beliefs in the communist ideology and reduce support for the political system, a logic which may 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 amplify conflict, and politicians may purposefully engineer such misunderstanding. Consistent with this intuition, violence-inciting propaganda often features a false rhetoric of self-defense and the dehumanization of opponents (Yanagizawa-Drott 2014). We hope that our conceptual framework can improve the understanding of behavior in such situations.

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.

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

___ , Matteo F Ferroni, and Stefanie Stantcheva, "Perceptions of Racial Gaps, their Causes, and Ways to Reduce Them," Working Paper 29245, National Bureau of Economic Research September 2021.

Algan, Yann, Daniel Cohen, Eva Davoine, Martial Foucault, and Stefanie Stantcheva, "Trust in scientists in times of pandemic: Panel evidence from 12 countries," *Proceedings of the National Academy of Sciences*, 2021, *118* (40), e2108576118.

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

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

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.

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

Dechezleprêtre, Antoine, Adrien Fabre, Tobias Kruse, Bluebery Planterose, Ana Sanchez Chico, and Stefanie Stantcheva, "Fighting Climate Change: International Attitudes Toward Climate Policies," Working Paper 30265, National Bureau of Economic Research July 2022.

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

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

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

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

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

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

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.

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

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

Gehlbach, Scott and Konstantin Sonin, "Government control of the media," *Journal of public Economics*, 2014, *118*, 163–171.

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.

Greenberg, Jon, "Most Republicans still falsely believe Trump's stolen election claims," Politifact, `https://www.politifact.com/article/2022/jun/14/most-republicans-falsely-believe-trumps-stolen-ele/`, 2022.

Grimes, David Robert, "On the Viability of Conspiratorial Beliefs," *PLOS ONE*, 01 2016, *11* (1), 1–17.

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.

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

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

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.

___ , Joshua Schwartzstein, and Andrei Shleifer, "Coarse Thinking and Persuasion*," *The Quarterly Journal of Economics*, 05 2008, *123* (2), 577–619.

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.

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

Shadmehr, Mehdi and Dan Bernhardt, "State censorship," *American Economic Journal: Microeconomics*, 2015, *7* (2), 280–307.

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

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

Stantcheva, Stefanie, "Understanding Tax Policy: How do People Reason?*," *The Quarterly Journal of Economics*, 09 2021, *136* (4), 2309–2369.

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 Additional material for main result and implications in Section 2

Definition of equilibrium. We start with introducing notation. 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 departure from common priors and full rationality. As usual, equilibrium requires that actors best respond and form consistent beliefs. To formulate the best-response condition, we first introduce subjective expected utility. For each actor, at each stage where it moves, its beliefs and the strategy profile generate a probability distribution over final outcomes. This distribution can differ from the objectively correct distribution because the persuaded voter has an incorrect prior about θ. Actor i at stage t uses its subjective probability distribution over outcomes to compute its subjective expected utility, denoted $U_i(\sigma|\hat{h}^t, \theta_i, \mu_i(\theta|\theta_i, \hat{h}^t))$. Then the best-response property of equilibrium is that at each stage t at which i has a move, for all actions σ_i' available to i,

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

Belief consistency does not impose any condition on principals, because they move only at stage 1 where they know only their priors.[26] Belief consistency for the voter requires that he follows Bayesian updating at the end of stage 1:

$$\mu_v^1(\theta_p|\theta_v, \hat{a}^1) = \frac{\mu_v^0(\theta_p|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{a}^1|\theta_p)}{\sum_{\theta'_p} \mu_v^0(\theta'_p|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{a}^1|\theta'_p)} \tag{8}$$

where $\mu_v^0(.|\theta_v)$ is the prior of the voter of type θ_v. This definition accounts for the model's deviation from rationality that the voter's mind type and beliefs may change in stage 1, by computing the posterior for each mind type $\theta_m = N, P$ using the prior associated with that mind type. In particular, if the voter is reached by propaganda and becomes persuaded, (8) computes his posterior from the prior of the persuaded voter $\mu_v^0(.|\theta_d, P)$. Intuitively, because the persuaded voter uses Bayes rule, he infers from the presence of propaganda about the politician's type; but because propaganda also influences his type, this inference is based on the prior modified by propaganda. Implicit in this is that when the voter receives messages $\hat{a}^1 = (\hat{s}_c, \hat{p})$, first propaganda $\hat{p}$ changes his mind type and prior, and then he updates from his new prior based on the information content of $\hat{a}^1$.

Proof of Proposition 1. Our proof identifies the unique pure strategy equilibrium and shows that it has the properties described in the proposition. We begin by characterizing the behavior of some actors independently of the size of λ and of the common type of the politician, and then proceed to other actors analyzing the low/high λ and the pro-voter/pro-elite cases separately.

The R elite, because it cannot coordinate, ignores its effect on the voter and given its preference for truth-telling always sends an honest report s_c. The normal voter updates his beliefs about the politician's common type based on $\hat{s}_c$ and $\hat{p}$. Because $\hat{s}_c$ trembles are arbitrarily more unlikely than $\hat{p}$ trembles, and because absent trembles s_c is truthful, the normal voter's beliefs are fully determined by the elite's message $\hat{s}_c$:

$$\mu_v[\theta_c = 1|\hat{s}_c, \hat{p}, \theta_d, \theta_m = N] = \hat{s}_c. \tag{9}$$

[26] It is straightforward to characterize the beliefs of principals at all stages, because they know all types. If the true type profile after history $\hat{h}^t$ is $\theta^* = (\theta_d^*, \theta_c^*, \theta_r^*, \theta_m^*)$ then the principals believe $\mu_i^t(\theta^*|\theta_i, \hat{h}^t) = 1$.

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: to engage in propaganda or to avoid propaganda. We will characterize her choice later.

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

In this case the AR elite prefers to keep the good politician and remove the bad politician irrespective of their divisive type. This means that both the AR and the R elite have the dominant strategy of reporting the common type truthfully. As a result, the persuaded voter, similarly to the normal voter, follows the elite's message:

$$\mu_v[\theta_c = 1|\hat{s}_c, \hat{p}, \theta_d, \theta_m = P] = \hat{s}_c. \tag{10}$$

Since propaganda is costly but does not discredit the elite, no politician invests in propaganda. Both voter types always believe the elite's message and update accordingly. This completes the description of the unique equilibrium. It is immediate that this equilibrium has the properties claimed in the Proposition.

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

Subcase 1: Incumbent politician is pro-voter.

Existence. We first show that the strategy profile described in the proposition, in which (i) the AR elite always reports the politician bad, (ii) the good and bad AR politician and the bad R politician all use propaganda, constitute an equilibrium.

We begin by characterizing the beliefs of the persuaded voter. Since s_c and p are perfectly correlated, the politician sends propaganda every time she is criticized. The persuaded voter observes these messages with trembles, and because trembles in $\hat{s}_c$ are arbitrarily less likely than in $\hat{p}$, updates based only on the former:

$$\mu_v(\theta_c = 1|\hat{s}_c, \hat{p}, \theta_d = 1, \theta_m = P) = \hat{s}_c + (1 - \hat{s}_c)\hat{q}_c \tag{11}$$

where $\hat{q}_c$ is defined by equation (5). Intuitively, in the proposed equilibrium the elite reports the

politician good if and only if reality is R and the politician is indeed good, implying that a good report ($\hat{s}_c = 1$) can only arise if $\theta_c = 1$. But the elite reports the politician bad ($\hat{s}_c = 0$) both when reality is R and the politician is bad, and when reality is AR irrespective of the politician's type, implying that the persuaded voter follows Bayes rule (8) to update to $\hat{q}_c$ by the calculation underlying equation (5).

We now turn to the behavior of the principals. Since $\lambda > \overline{\lambda}$, the AR elite wants to remove both the good and the bad pro-voter politician. The above equation shows that the persuaded voter's belief is responsive to the elite's message $\hat{s}_c$, implying that the AR elite prefers to report the politician bad (and pay the infinitesimal lying cost) because doing so affects the election outcome.

As to the politician, sending propaganda is optimal for all types who expect the elite to criticize them, that is, the good and bad AR politician and the bad R politician. This follows from Assumption 1, which ensures that partially hiding the common type (securing a belief of $\hat{q}_c$) exceeds the cost of propaganda.

We have confirmed that in the proposed equilibrium the elite and the politician best respond, and we have characterized the beliefs and hence behavior of the voter. To conclude, we clarify that our above arguments also cover off-equilibrium information sets. Such information sets only happen at stage 2, i.e., after the message profile is realized: because propaganda determines the voter type, the normal voter after propaganda and the persuaded voter absent propaganda can never occur in this game. Still, Bayesian updating in our equilibrium definition, (8), specifies beliefs based on how a voter of the given type would update from the information content of the messages he observes. In particular, (9) above specifies the beliefs of the normal voter after propaganda, while (11) specifies the beliefs of the persuaded voter absent propaganda. These beliefs also pin down voting behavior.

Uniqueness. Here we establish that when the politician is pro-voter, no pure strategy equilibrium other than the one above exists. We first show that in every pure strategy equilibrium the AR elite always reports the pro-voter politician bad. Given our focus on monotone strategies, the other possible strategies for the AR elite given the politician is pro-voter are (i) to be truthful; or (ii) to report all politicians good. In both cases, elite criticism would be trusted by the persuaded voter,

creating an incentive to report the pro-voter politician bad.

We next show that there cannot be an equilibrium in which only bad politicians (R, AR, or both) use propaganda. This follows 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 that the politician is bad, and is hence not worth doing.

It follows that in any new pure strategy equilibrium either nobody uses propaganda, or the good AR politician and only one of the bad politician types (either R or AR) use propaganda. We show that 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 (5) 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.

Subcase 2: Incumbent is a pro-elite politician.

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

Start with 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. Consider next the AR elite. Because $\lambda > \overline{\lambda}$ she wants to keep both types of politicians. She expects

no propaganda and a normal voter, and hence finds it optimal to praise the politician. Finally, for the bad R politician, doing propaganda cannot be optimal: since the AR elite always praises, the voter only observes propaganda and criticism if reality is R and the politician is bad. 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.

Next consider off-path information sets. Like in Subcase 1, these only occur in stage 2: we need to deal with the normal voter after propaganda, and—since propaganda is off the equilibrium path—the persuaded voter after any history. The normal voter after propaganda, because it is off the equilibrium path, will attribute propaganda to a tremble and form beliefs and behavior just like the normal voter absent propaganda. Since propaganda is off the path, the persuaded voter after propaganda 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 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.

Uniqueness. We show that there is no other pure strategy equilibrium. If the AR elite follows the strategy of always reporting the politician good, then the above proof also pins down the behavior of all other actors. The AR elite has two other potential monotone strategies: (i) report truthfully, (ii) always report the politician bad. In case (i), both the normal and the persuaded voter will form beliefs about the common type that follow the elite's report. But then, irrespective of the presence of propaganda, the AR elite will prefer to deviate and report the politician good. In case (ii), because $q_r > 0$, both the normal and the persuaded voter update in the direction of the elite's message about the common type. This means that again the AR elite has an incentive to deviate and report the politician good. This concludes Subcase 2.

Statements of the Proposition. It is immediate that the unique equilibrium satisfies statements (a) and (b) in Part 2 of the Proposition. For statement (c), note that if the bad pro-voter 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.

Distrust and the non-adopotion of best practices in other domains. In Section 2.5 we claim that propaganda can hinder the adoption of best practices in other non-political domains. Here we add a new stage to our baseline model to formalize this implication. Suppose that after stage 1 of the base game, the voter needs to take a private action, e.g., whether to vaccinate against Covid, and the elite can provide informed advice about the decision. The cost to voter i of taking the action is $\kappa_i \sim U[\underline{h}, \overline{h}]$, where $\underline{h} < 0$, $\overline{h} > 1$, and $h \equiv (\overline{h} - \underline{h})/2$. The voter knows the cost of the action but is uncertain about the benefit, $v_v \in \{0, 1\}$ where $v_v = 1$ happens with probability ν. Voter i's additional payoff from the action is $U^a_{v,i} = (v_v - \kappa_i) \cdot a_i$ where $a_i = 1$ if i takes the action. The voters' action yields a benefit for the elite as well: $v_e \in \{-\beta, \beta\}$, where $\Pr(v_e = \beta) = \nu$. The parameter β captures the importance of the new issue for the elite. The elite's additional payoff is

$$U^a_e = v_e \cdot \int_0^1 a_i.$$

Observe that the voter benefits from her own action while the elite benefits from the collective action of voters. Crucially, we assume that v_v and v_e are independent. This implies that there is a conflict of interest between the voter and the elite in the new action. The elite can send a message about the value of the action to the voter.

Claim 1. If β is sufficiently small there is a unique pure strategy equilibrium path in which:

1. The first stage unfolds as in the baseline game;
2. Absent propaganda the voter follows the elite's report;
3. After propaganda the voter discounts the elite's report;
4. The voter makes a better decision about the new action in the absence of propaganda.

Proof. Start by assuming that the first stage unfolds as in the base game. Consider the history of no propaganda in the first stage. The voter remains normal. Members of the R elite are truthful

about the new action since they do not internalize their effect on the voter's behavior and dislike lying. The voter, who believes the elite is R, fully updates from the elite's message.

Consider the history of propaganda and elite criticism in the first stage. The voter becomes persuaded and believes the reality is AR with probability $\hat{q}_{ar} = q_{ar}/(q_{ar} + q_r(1 - q_c))$. As we have seen above, the members of the R elite remain truthful. However, the AR elite finds it optimal to manipulate and report the value for the elite v_e instead of the value for the voter v_v. In the alternative reality the elite's signal is completely uninformative since her preferences are uncorrelated with those of the voter. As a result, the voter discounts the elite's message and updates to

$$\mu_v(v_v = 1|\hat{p} = 1, \hat{s}_c = 0, s_v) = \hat{q}_{ar}\nu + (1 - \hat{q}_{ar})s_v.$$

In particular, the voter's belief responds with a weight $1 - \hat{q}_{ar} < 1$ to the elite's message s_v.

Now, consider the history of propaganda and elite praise in the first stage. In this case, the persuaded voter will attribute propaganda to a tremble and learn that the reality is R. In the second stage, the R elite is truthful for the same reasons as before. The voter, who believes that reality is R, follows the elite's advice and forms beliefs

$$\mu_v(v_v = 1|\hat{p} = 1, \hat{s}_c = 1, s_v) = s_v.$$

Consider the first stage. If β is sufficiently small than the payoff of the AR elite is not affected sufficiently by the new action to change its strategy in the first stage. Other players are not affected either. No politician type has an incentive to deviate from her previous strategy as the politician is not affected by the new action. And the R elite, who does not internalize its effect on the whole society does not change her strategy either.

Finally, it is immediate that the voter is worse off in the presence of propaganda, as propaganda effectively garbles an otherwise fully informative message about a payoff-relevant decision.

A.2 Additional material for government policy application in Section 3.1

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

$$\mu_i^t(\theta|\theta_i, \hat{h}_i^t) = \frac{\mu_i^{t-1}(\theta|\theta_i, \hat{h}_i^{t-1})\hat{\sigma}_{-i}^t(\Delta\hat{h}_i^t|\theta, \hat{h}_i^{t-1})}{\sum_{\theta'} \mu_i^{t-1}(\theta'|\theta_i, \hat{h}_i^{t-1})\hat{\sigma}_{-i}^t(\Delta\hat{h}_i^t|\theta', \hat{h}_i^{t-1})}. \tag{12}$$

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 (12) 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 verifies that the proposed path is supported by an equilibrium, and that it is the unique equilibrium path. 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, since she

only focuses on the lying cost, reports truthfully about both the common type and competence. Consequently, the normal voter fully trusts the elite's reports. Because the AR elite believes that propaganda was ineffective, she believes that the voter is normal and follows her message, and by $\lambda > \overline{\overline{\lambda}}$ criticizes—about both the common type and competence—if and only if the politician is pro-voter in the corresponding stage. 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.

Existence. We begin by verifying that on the proposed path there is no profitable deviation after the first stage. Assume that the first stage unfolds as in the baseline model. We show optimality after each possible first stage history.

Consider a history of observing propaganda and criticism in the first stage. Assume the bad R politician always acts incompetently and all other types (good R and good and bad AR) act competently if they can. The persuaded voter understands the elite's and the politician's second-stage behavior and forms the following beliefs (assuming $\xi \approx 0$)

$$\mu(\theta_k = 1|\hat{s}_c = 1, \hat{p} = 1, \hat{s}_k) = q_k,$$
$$\mu(\theta_d = 1|\hat{s}_c = 1, \hat{p} = 1, \hat{s}_k) = 1 - \hat{s}_k,$$
$$\mu(\theta_c = 1|\hat{s}_c = 1, \hat{p} = 1, \hat{s}_k) = \hat{s}_k q_c + (1 - \hat{s}_k)\hat{q}_c.$$

The logic is the following. The voter does not learn anything about competence: both competent and incompetent (bad) R politicians act incompetently, thus the R elite's report is uninformative, and as we have seen above the AR elite's report is uninformative about competence. The voter does learn about the divisive type: he understands that a signal of competence ($\hat{s}_k = 1$) is only possible in the alternative reality if the politician flipped to pro-elite. Finally, the voter also learns about the common type: since a signal of competence implies that reality is AR, the elite must have reported both the good and bad politician as bad, implying that the politician must be good with probability q_c rather than $\hat{q}_c$.

Now consider possible deviations at stages 3 and 4. Since the behavior of the R and AR elites has been pinned down, we need to rule out that the politician has such deviations. The

competence action of the AR politician does not influence the competence report of the AR elite (it only depends on the politician's divisive type), so that politician's unique optimal strategy is to act competently. For the R politician, acting competently generates a competent signal, which, by the above formulas, makes the voter update that the politician pro-elite with probability 1 and good with probability q_c. Because beliefs previously were that the politician is pro-elite with probability ξ and good with probability $\hat{q}_c$, this change is bad for the politician for $\xi \approx 0$ if $\lambda > (q_c - \hat{q}_c)c$ which is implied by $\lambda > \overline{\overline{\lambda}}$. As a result, the R politician will always act incompetently.

Consider a history of observing no propaganda in the first stage. The voter remains normal and follows the elite's signal, so her beliefs are

$$\mu(\theta_k = 1 | \hat{s}_c, \hat{p} = 0, \hat{s}_k) = \hat{s}_k,$$
$$\mu(\theta_d = 1 | \hat{s}_c, \hat{p} = 0, \hat{s}_k) = 1,$$
$$\mu(\theta_c = 1 | \hat{s}_c, \hat{p} = 0, \hat{s}_k) = \hat{s}_c.$$

Given these beliefs the unique optimal strategy of all politician types is to act competently if they can.

Finally, consider a history of observing propaganda but no criticism in the first stage. This is off the equilibrium path, and because the elite's trembles are less likely than those of propaganda, the voter will update that propaganda was a tremble and reality is R, and hence forms the same beliefs as after observing no propaganda. The rest of the game unfolds as after observing no propaganda.

We have confirmed that the proposed path admits no profitable deviations after any first stage history. Now consider the first stage. The good 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. All other politician types expect to be criticized in the first stage. By using propaganda they avoid being detected as bad—increase their perceived probability of being good from 0 to $\hat{q}_c$—at the expense of being unable to signal their potential competence later in the game. Since they are risk neutral, they are indifferent about not being able to signal their competence, and find it optimal to send propaganda.

To close the argument, we need to characterize beliefs and optimal behavior at off-equilibrium information sets. We have seen that after the off-equilibrium profile at stage 1 of propaganda and praise, behavior is identical to that after no propaganda and praise. Further off-equilibrium information sets can occur in the decision stages 3, 4 and 5. At stage 3, besides competence, the politician has four types: good or bad and R or AR. Absent 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.

Uniqueness. We show that in this case, i.e., when the politician is initially pro-voter, the above equilibrium path is generically unique.

Consider the first stage. The behavior of the R and AR elites has been uniquely characterized above. The good R politician understands that by sending propaganda she may be able to 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 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, given these beliefs, the expected payoff of the AR politician does not depend on whether she is good or bad,

generically her behavior will also not depend on whether she is good or bad. 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, and whether the AR politician (of both types) sends propaganda. One of these profiles, when both send propaganda, leads to the equilibrium described above. We now check whether the other three can lead to an equilibrium.

Subcase 1: 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 1 we do not have an equilibrium.

Subcase 2: 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 has 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 this is just as good ex-ante as if it was revealed. Therefore using propaganda is a profitable deviation. In subcase 2 we do not have an equilibrium.

Subcase 3: 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 $\hat{q}_c$. Her competence is not revealed in the competence stages but as she is risk neutral this 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.

It follows that none of these cases can lead to an equilibrium. It remains to check for uniqueness assuming that the first stage unfolds as in the original game. Consider a history of observing propaganda and criticism in the first stage. 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 reporting behavior about competence, 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. Before the elite's report, 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, which (as we noted earlier) is bad for the politician because $(1-\xi)\lambda > (q_c - \hat{q}_c)c$ follows from $\lambda > \overline{\overline{\lambda}}$ if 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 certainly bad for the politician if $\hat{q}_c c > k$, which holds by Assumption 2. It follows that there is no equilibrium in which the normal R politician, who understands all this, will choose to act competently. 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. It follows that after a history of propaganda and criticism the equilibrium path is unique.

Case 2: Politician is initially pro-elite.

Uniqueness. Here we start by establishing the uniqueness of the equilibrium path. Since the behavior of the elite has already been determined, we focus on the politician. 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, potentially allowing 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.

Existence. The above arguments also show that on the proposed path the politician is best responding by avoiding propaganda. However, to establish existence we need to specify off-equilibrium behavior. 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 continues to be 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 where, after propaganda, the AR politician always acts her competence, while the R politician pretends to be 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. Given this, we now check that no politician type has a profitable deviation after propaganda from the proposed profile. Start with 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 good, 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 good with probability

close to 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 the new media application in Section 3.2

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. Assume 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.[27] 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 $\phi \equiv \eta\psi$ for both the elite and the new media. Assume further that the lying cost—a reduced form representation of reputation concerns—is proportional to the mass of potential readers. Then, 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 $\chi_e = \chi > \eta\chi = \chi_n$.

[27] 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.

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 $\theta_a = 1$: $\mu_p^0(\theta_a = 1|\theta_r = R) = \mu_e^0(\theta_a = 1|\theta_r = R) = 1$. The AR principals and the voter all falsely believe that $\theta_a = 0$: $\mu_v^0(\theta_a = 0|\theta_v) = \mu_p^0(\theta_a = 0|\theta_r = AR) = \mu_e^0(\theta_a = 0|\theta_r = AR) = 1$. The new media has correct prior beliefs about all types, and all other prior beliefs are as in the basic model.

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, given any system of beliefs, subjective expected utility can be defined in the usual way 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. Still, it may be useful to characterize their beliefs at all stages, and doing so is straightforward because they 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, elite, and new media believe $\mu_i(\theta^*|\theta_i, \hat{h}^{t-1}) = 1$, while the AR politician, elite, and new media believe $\mu_i((\theta_d^*, \theta_c^*, AR, \theta_m^*, 1)|\theta_r = AR, \hat{h}^{t-1}) = 1$. 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

$$\mu_v^1(\theta_p|\theta_v, \hat{h}_v^1) = \frac{\mu_v^0(\theta_p|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{h}_v^1|\theta_p)}{\sum_{\theta_p'} \mu_v^0(\theta_p'|\theta_v) \cdot \hat{\sigma}_{-v}^1(\hat{h}_v^1|\theta_p')}, \tag{13}$$

the main modification being that $\hat{\sigma}_{-v}^1(\hat{h}^1|\theta_p)$ computes the probability of a voter history $\hat{h}_v^1$ 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.

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 (non-audience-seeking) R elite—which only exists in the voter's mind—always reports truthfully. The audience seeking 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 4. 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 elite's tremble is arbitrarily smaller than even the combined trembles of the politician and the new media—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.

It remains to characterize the behavior of the AR elite and the audience-seeking new media, the bad R politician and the good and bad AR politicians, and the persuaded voter. We do this separately for the cases in which the incumbent politician is pro-voter and pro-elite.

Case 1: Incumbent politician is pro-voter.

Existence.

We show that there is an equilibrium in which the AR elite always criticizes while the audience-seeking new media always praises the politician, and the good and bad AR politicians and the bad R politician all use propaganda. The behavior of all other principals was pinned down above. In the process we also characterize the behavior of the persuaded voter.

We begin by characterizing the beliefs of the voter in the proposed equilibrium. Start with the normal voter. His beliefs about reality are straightforward: he knows $\theta_r = R$. We claim that his beliefs about the politician's common type, irrespective of whether he observes the new media's

message $\hat{s}_c^n$, are given by

$$\mu(\theta_c = 1|\hat{p}, \hat{s}_c^e, \hat{s}_c^n, \theta_m = N) = \hat{s}_c^e.$$

The reason is that the normal voter knows that reality is R and therefore the elite can be fully trusted. This argument works even when the observed profile is off the equilibrium path, for two reasons: the trembles make all paths possible, and elite trembles are arbitrarily less likely than even simultaneous propaganda and new media trembles, making the voter always follow the elite's message. Note that the formula is valid even after propaganda: although in the game we can never have a normal voter after propaganda, Bayesian updating still pins down beliefs.

We now turn to the beliefs of the persuaded voter. First consider the case when he does not observe the message of the new media. We claim that his beliefs about the common type are

$$\mu(\theta_c = 1|\hat{p}, \hat{s}_c^e, \theta_m = P) = \hat{s}_c^e + (1 - \hat{s}_c^e)\hat{q}_c. \tag{14}$$

This follows as in the basic model. When the elite sends praise, the voter knows the politician must be good. When the elite criticizes, the politician may still be good if reality is AR, explaining the $\hat{q}_c$ term. This logic holds even in the absence of propaganda, since the persuaded voter has positive prior on the AR: he will just attribute the lack of propaganda to a tremble.

Continuing with the case in which the persuaded voter does not observe the new media, his beliefs about the AR are given by

$$\mu(\theta_r = AR|\hat{p}, \hat{s}_c^e, \theta_m = P) = (1 - \hat{s}_c^e)\hat{q}_{ar}$$

where $\hat{q}_{ar} = q_{ar}/(q_{ar} + q_r(1 - q_c))$ is the posterior that reality is AR conditional on the voter having prior q_{ar} and observing elite criticism. Intuitively, elite praise implies that reality must be R, while elite criticism can come with both the bad R or with either type of the AR politician and hence implies interior beliefs. Again, the logic holds even in the absence of propaganda which the voter will just attribute to a tremble.

Next consider the case in which the persuaded voter observes the new media. We claim that his beliefs about the politician's common type are given by

$$\mu(\theta_c = 1|\hat{p}, \hat{s}_c^e, \hat{s}_c^n, \theta_m = P) = \max(\hat{s}_c^e, \hat{s}_c^n). \tag{15}$$

To see the logic, consider the possible on-path realizations of the profile $(\hat{p}, \hat{s}_c^e, \hat{s}_c^n)$. When reality is R, these are $(0,1,1)$ if the politician is good and $(1,0,0)$ if the politician is bad, and when reality is AR they are $(1,0,1)$ if the politician is good and $(1,0,0)$ if the politician is bad. Combining across R and AR, profiles $(0,1,1)$ and $(1,0,1)$ emerge if the politician is good and $(1,0,0)$ if the politician is bad. Since the elite's message is more precise than the others even in combination, and since $\hat{s}_c^e = 1$ is only possible if the politician is good, the voter will conclude from that message that the politician is good. When $\hat{s}_c^e = 0$, the voter will attempt to match to the on-path profiles $(1,0,1)$ reflecting that the politician is good and $(1,0,0)$ reflecting that the politician is bad. The value of $\hat{s}_c^n$ will determine which of these matches is closer to the realized profile, and hence the updating.

Finally, if he observes the new media, the persuaded voter's beliefs about the alternative reality are given by

$$\mu(\theta_r = AR|\hat{p}, \hat{s}_c^e, \hat{s}_c^n, \theta_m = P) = (1 - \hat{s}_c^e)(\hat{s}_c^n + (1 - \hat{s}_c^n)q_{ar}). \tag{16}$$

The on-path profiles in R are $(0,1,1)$ and $(1,0,0)$, and in AR are $(1,0,1)$ and $(1,0,0)$. Thus $\hat{s}_c^e = 1$ implies R, explaining the $1 - \hat{s}_c^e$ factor in the expression. When $\hat{s}_c^e = 0$ we consider the profiles separately. $(1,0,1)$ is on-path only in the AR and thus leads to full AR beliefs, consistent with the formula. $(1,0,0)$ is on-path in both the R and the AR and leads to belief q_{ar}: given the new media's message the voter knows the politician is bad, but a bad politician sends propaganda and gets criticized in both R and AR. $(0,0,1)$ is one tremble away from the AR and two trembles away from the R, implying full beliefs in the AR, again consistent with the formula. And $(0,0,0)$ is one tremble away from $(1,0,0)$ which is both in R and AR, implying beliefs q_{ar} as in the formula.

We next verify the optimality of the proposed profile for all actors. Consider the AR elite. Since she believes the politician is AR, she expects propaganda and a persuaded voter. If the new media's signal does not reach the voter, then, similarly to before, (14) implies that by criticizing she can partially convince the voter that pro-voter politician is bad (since $\hat{q}_c < 1$). If the new media's signal—which is expected to be truthful in the voter's mind—does reach the voter, then, by (15), the AR elite either fully convinces the voter that the politician is bad or has no impact. It follows that the AR elite finds it optimal to criticize.

Now consider the good and bad AR politician and the bad R politician. These types all expect

elite criticism. Absent propaganda the voter fully trusts the elite's report, and will conclude that the politician is bad. Propaganda changes the voter's prior about the AR, and in the event in which the new media cannot speak, will increase the voter's belief that the politician is good to $\hat{q}_c > 0$. And if the new media can speak, it cannot decrease the voter's evaluation that the politician is bad further, and by (15) it can increase it if the message is positive. Assumption 3 then ensures that propaganda is beneficial.

We now turn to the audience-seeking new media and show that she always reports that the politician is good. If the politician is good, then the audience-seeking new media expects no propaganda and a normal voter, 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. Equation (16) shows that the beliefs of the persuaded voter about AR depend on the message of the new media: If the new media criticizes, the voter concludes that the probability of AR is q_{ar}, while if the new media praises, the voter concludes that the probability of AR is 1. It follows that the gain to the audience-seeking new media from reporting the bad politician good is $\phi(1 - q_{ar})$. Since the cost χ_n is smaller than this by Assumption 4, the new media will report the bad politician good.

Finally, consider the behavior of the persuaded voter. We have characterized his beliefs above, 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 above argument has characterized the behavior of all actors and confirmed that they are best responding at all decision nodes. We emphasize that this includes off-equilibrium decision nodes as well. Such nodes only occur at stage 2 of the game at which the voter gets to decide, and his beliefs have been characterized for all types and histories: the normal voter always follows the elite, and the beliefs of the persuaded voter were explicitly characterized above.

Uniqueness. We show that there is no equilibrium in which the AR elite, the audience-seeking new media, the AR politician or the bad R politician use a strategy different from above. Since the behavior of the other principals was pinned down above, and the beliefs and behavior of the voter

follow from the behavior of the principals, this will establish uniqueness.

First consider the AR elite. Focusing on monotone strategies, she can either report the common type truthfully, or report all politicians good. In either case, a report that the politician is bad would be credible evidence that the politician is indeed bad. This would create an incentive for the AR elite to report the (pro-voter) politician bad.

Now, consider the good and bad AR politician and the bad R politician. 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 and at least one of the bad politician types (either R or AR) use 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 (5) in the basic model. Assumption 3 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. It follows that all three politician types considered in the paragraph must use propaganda.

Finally, consider the audience-seeking new media. Since voters consider the new media honest, the strategy of the audience-seeking type does not influence the way voters update. Then the same argument used to establish existence shows that her unique optimal strategy is to report all pro-voter politicians good. In particular, if the politician is good, then the audience-seeking new media expects no propaganda and a normal voter, 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 are determined independently of the behavior of the audience-seeking new media, hence our formulas from the existence part apply. Equation (16) shows that the beliefs of the persuaded voter about AR depend on the message of the new media: If the new media criticizes, the voter concludes that the probability of AR is q_{ar}, while if the new media praises, the voter concludes that the probability of AR is 1. It follows that the gain to the audience-seeking new media from reporting the bad politician good is $\phi(1 - q_{ar})$. Since the cost χ_n is smaller than this by Assumption 4, the new media will report the bad politician good.

Case 2: Incumbent politician is pro-elite.

Existence. Assume that the AR elite reports all pro-elite politicians to be good. We show that then no (pro-elite) politician type uses propaganda. Start with the AR politician. Since the AR elite—which the AR politician believes is the elite—always reports her to be good, and since the normal voter ignores the new media, she can ensure that the voter considers her good by avoiding propaganda. 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. It follows that no politician type uses propaganda. Absent propaganda the voter remains normal, therefore reporting all pro-elite politicians to be good is the unique best response of the AR elite. The audience-seeking new media expects to be ignored by the normal voter and thus sends an honest message to minimize the lying cost.

We now characterize the beliefs and behavior of the voter. Because elite trembles are arbitrarily more unlikely then propaganda and new media trembles, he normal voter always follows the elite's message, even after histories that cannot occur on path. 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 voter faces a binary decision problem which that has a solution, and chooses that solution.

Uniqueness. We show that there is no alternative equilibrium. Above we showed that if the AR elite reports all pro-elite politicians good then the equilibrium is uniquely determined. Here we consider the two other monotone strategies of the AR elite: (i) she reports truthfully, or (ii) she reports all types bad. In both cases, the elite's praise of the politician will be fully trusted by voters. This creates an incentive for the AR elite to report all pro-elite politician good.

New media benefits politician. We show that the presence of the new media 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 (5) as

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

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. And if the message of the new media reaches the voter, then his posterior is that the politician is good and reality is AR with certainty. 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.