



## Propaganda and credulity <sup>☆</sup>



Andrew T. Little

Department of Government, Cornell University, United States

### ARTICLE INFO

#### Article history:

Received 13 April 2016

Available online 28 December 2016

#### JEL classification:

D83

#### Keywords:

Political economy

Propaganda

Authoritarian politics

### ABSTRACT

I develop a theory of propaganda which affects mass *behavior* without necessarily affecting mass *beliefs*. A group of citizens observe a signal of their government's performance, which is upwardly inflated by propaganda. Citizens want to support the government if it performs well and if others are supportive (i.e., to *coordinate*). Some citizens are unaware of the propaganda ("credulous"). Because of the coordination motive, the non-credulous still respond to propaganda, and when the coordination motive dominates they perfectly mimic the actions of the credulous. So, all can act as if they believe the government's lies even though most do not. The government benefits from this responsiveness to manipulation since it leads to a more compliant citizenry, but uses more propaganda precisely when citizens are less responsive.

© 2016 Elsevier Inc. All rights reserved.

Politicians lie, and coerce others to lie on their behalf. These lies take many forms, from rewriting the history taught in schools, to preventing the media from reporting on policy failures, to relatively innocuous spinning of the economy's performance in press conferences.

The pervasiveness of political actors lying or manipulating information more generally may seem natural. However, the general acceptance of lying's central role in politics itself poses a puzzle from a game-theoretic perspective: if the "audience" for political information – which could be other elites, party members, or citizens more broadly – know they are being lied to and adjust their beliefs accordingly, why go through the costly and potentially embarrassing effort of distorting information?

Scholars have proposed several explanations for why governments attempt to (and sometimes succeed at) manipulating information to an audience which is strategically sophisticated. For example, manipulation can act as a signal that the government is capable of doing so (Simpser, 2013; Huang, 2015), or render citizens less informed and hence less likely to take risky anti-government action (Gehlbach and Sonin, 2014; Shadmehr and Bernhardt, 2015; Gehlbach and Simpser, 2015; Hollyer et al., 2015). Even if the audience for manipulated information is not fooled, in models with a "career concerns" type information structure (Holmstrom, 1999), governments have to lie to keep up with expectations for and avoid seeming weaker than they truly are (Little, 2012, 2015).<sup>1</sup>

<sup>☆</sup> Many thanks to Avi Acharya, Sarah Bouchat, Alex Debs, Tiberiu Dragu, Oendrilla Dube, Florian Foos, Scott Gehlbach, Florian Hollenbach, Carlo Horz, Navin Kartik, Adam Levine, Peter Lorentzen, Konstantinos Matakos, Tom Pepinsky, Adam Przeworski, Konstantin Sonin, Francesco Squintani, Jeremy Wallace, Jessica Weiss, Galina Zudenkova, audience members at EPSA 2015, APSA 2015, SPSA 2016, MPSA 2016, NYU, Vanderbilt, Harvard, and Wisconsin, several anonymous referees, and the editors of Games and Economic Behavior for comments and discussion. All remaining errors are mine.

E-mail address: [andrew.little@cornell.edu](mailto:andrew.little@cornell.edu).

<sup>1</sup> These mechanisms are not mutually exclusive: for example, a common combination blends private information with Bayesian Persuasion-like dynamics, where those with private information indicating they are less strong manipulate more to "pool" with the stronger types and hide their weakness (Edmond, 2013; Lorentzen, 2014; Petrova and Zudenkova, 2015).

However, a wide variety of empirical results from economics, psychology, and political science indicate that some if not most people *don't* fully adjust for information manipulation even when standard solution concepts indicate that they should. Lab experiments consistently find that subjects tend to believe what others say, even if the sender has transparent incentives to misrepresent their information (Cai and Wang, 2006; Patty and Weber, 2007; Wang et al., 2010). Observational empirical studies of propaganda and information manipulation also tend to find a large influence on the beliefs and behavior of the target audience (e.g., Enikolopov et al., 2011; Yanagizawa-Drott, 2014; Adena et al., 2015).

Motivated by these results, I explore the consequences of a simpler explanation: politicians lie because some people believe them. While those who believe whatever the government tells them are tautologically responsive to propaganda, their presence has powerful effects on the behavior of those who *are* aware that they are being lied to, as well as those doing the lying.

I analyze a model where citizens observe a public signal of the government's performance, using a notion of credulity similar to that in Kartik et al. (2007).<sup>2</sup> The government chooses a (hidden) level of propaganda, which inflates this signal, suggesting the government is performing better than they truly are. I formalize “partially credulous” citizens as those who believe the government may not manipulate information when it does. “Fully credulous” citizens incorrectly believe it is impossible that the government manipulates information. All citizens behave optimally given their beliefs and knowledge about the existence of credulous citizens.

After observing the signal, citizens choose how much to support the government. Following Kuran (1997), two incentives drive this decision: citizens want to support governments that perform well, and want to pick a level of support in line with what others choose.<sup>3</sup> This latter *coordination motive* can drive even those who know the government is lying to behave as if they accept the signal as true because they know the credulous citizens will be influenced by propaganda.

The first main results of the paper are that fully rational and informed citizens are always at least partially responsive to propaganda, and when the coordination motive dominates, the behavior of the informed types approaches that of the credulous types.<sup>4</sup> So, propaganda can successfully influence mass *behavior* even without affecting most citizens' *beliefs*. The second main set of results examine the government's equilibrium choice of propaganda given the citizen behavior. The government wants the citizens' support, but pays an exogenous cost to manipulate. Surprisingly, the government tends to pick a high level of propaganda precisely when it is ineffective. So, when observing people behave as if they believe government propaganda, we can neither conclude that they are truly convinced nor that the government is manipulating their behavior in an optimal fashion.

## 1. The model

The game is played among a government and a continuum of citizens with mass 1.

### 1.1. Information and sequence of moves

There is a fundamental government “performance” (or “strength”) parameter  $\theta \in \mathbb{R}$ . The government and citizens share a common prior belief on  $\theta$  which is distributed according to a density  $f$  with full support on  $\mathbb{R}$ .

The government also has a type  $\omega \in \{0, 1\}$ , which indicates whether they are willing and able to manipulate a signal of their performance. The government learns  $\omega$  (but not  $\theta$ ), and then chooses a manipulation level  $m \geq 0$ .<sup>5</sup>

Next, all citizens observe a public signal of  $\theta$ , given by:

$$s = \theta + \omega m.$$

Proportion  $1 - \epsilon$  of the population also directly observe  $\omega$  (the “informed types”), while  $\epsilon > 0$  do not (the “uninformed”, later “credulous” types). Let the uninformed type prior belief about the probability that the government can manipulate be  $q \in [0, 1]$ . Both  $\epsilon$  and  $q$  are common knowledge: i.e., the uninformed types are aware that some fraction of the population has better information than they do, and the informed types know the probability that the uninformed types assign to the possibility of manipulation.

After observing the signal, citizens simultaneously take an action  $a \in \mathbb{R}$ . The government wants citizen to take “high” actions, so larger values of  $a$  could correspond to speaking positively and publicly about the government, enthusiastically participating in pro-government rallies, working hard as a party member, etc. Equivalently,  $-a$  could represent the degree of anti-government activity taken.

<sup>2</sup> This paper and related work (e.g., Ottaviani and Squintani, 2006; Chen, 2011) study cheap talk models where the receiver sometimes accepts the sender's message at face value, and the audience for (manipulated) information is a single actor or a group of receivers whose actions are independent of each other. In addition to the application to propaganda rather than general communication, the central mechanism driving the results here is interactions among citizens with a motive to coordinate with each other.

<sup>3</sup> Kuran (1997) primarily focuses on small changes in the distribution of preferences can lead to drastic differences in the ability of citizens to coordinate (or not) against the status quo. The model here always exhibits a unique equilibrium, making a sharper prediction that when coordination motives are strong, herding always occurs on the behavior of the least informed citizens.

<sup>4</sup> See Haltiwanger and Waldman (1985) for a similar argument in the context of macroeconomics.

<sup>5</sup> See the online Appendix B for a discussion of what happens if the government knows  $\theta$ .

## 1.2. Payoffs

The government payoff is:

$$u_G(m; \bar{a}) = \bar{a} - c(m),$$

where  $\bar{a}$  is the average action taken by the citizens, and  $c$  is an increasing and convex function capturing the cost of manipulation. To ensure the equilibrium manipulation when  $\omega = 1$  is strictly positive, let  $c'(0) = 0$ .

The government strategy is a manipulation level as a function of  $\omega$ . However, the  $\omega = 0$  type does not change the signal by choosing  $m > 0$  but pays a cost, and hence will always choose  $m = 0$ . So, hereafter let  $m$  represent the amount of manipulation chosen by the  $\omega = 1$  type.

Citizens want to take higher actions when the government performs well and *other* citizens choose high actions. To formalize the citizen incentives, I use a “beauty contest” payoff structure as in employed a recent model of leadership and intra-party policy-making (Dewan and Myatt, 2008); see also Morris and Shin (2002) and Edmond (2015) for models with related payoff structures. The citizen utility as a function of his action ( $a$ ), the true performance ( $\theta$ ), and the average action of others ( $\bar{a}$ ) is:

$$u(a; \theta, \bar{a}) = -\lambda(a - \bar{a})^2 - (1 - \lambda)(a - \theta)^2.$$

The  $(a - \theta)^2$  term captures the fact that citizens want to pick high actions when the government truly is performing well, analogous to what Dewan and Myatt (2008) call “concern for policy” or what Kuran (1997) calls “expressive utility.”

The  $(a - \bar{a})^2$  term implies citizens want to take actions close to the average action of others; analogous to what Dewan and Myatt (2008) call “desire for conformity” and Kuran (1997) calls “reputational utility.” The  $\lambda \in (0, 1)$  parameter weighs the relative importance of the citizens’ two targets. When  $\lambda$  is close to 0, citizens primarily care about matching their action to the true government strength, while when  $\lambda$  is close to 1 citizens care more about matching their action to others. So, I refer to  $\lambda$  as the *coordination motive*.

## 1.3. Solution concept and credulity

In order to capture the fact that some citizens underestimate the government’s ability to manipulation, I characterize the pure strategy Perfect Bayesian Equilibrium (PBE) to the model,<sup>6</sup> and then focus on the case where the government can in fact manipulate. In particular:

**Definition.** An *equilibrium with  $q$ -credulity* is the outcome of the model when citizens play a PBE,  $\omega = 1$ , and the uninformed types’ prior belief is that the probability that the government can manipulate is  $q < 1$ . An *equilibrium with full credulity* is an equilibrium with  $q$ -credulity for  $q = 0$ .<sup>7</sup>

One notable aspect of the solution is that this prior belief need not be “correct.” All that matters is that the uninformed types have a prior belief that manipulation happens with probability  $q$  and it is common knowledge (among all types) that they hold this belief. So, the credulous equilibrium can be interpreted not only as a measure zero event, but a possible outcome when some of the population refuses to accept the possibility that the government is lying to them.

I first characterize the citizen behavior as a function of their conjecture about the government strategy and the observed signal, and then turn to the government’s manipulation choice.

## 2. The citizen stage

For the citizens to be best-responding to the government’s strategy, they must correctly anticipate the (again, pure strategy) level of propaganda when  $\omega = 1$  even if they do not directly observe  $\omega$ . Let  $\hat{m}$  represent the citizen’s conjecture about the degree of manipulation the government will choose when able. Applying Bayes’ rule, the uninformed type belief that the government manipulates as a function of  $s$  and  $\hat{m}$  is:

$$\bar{q}(s, \hat{m}) = \frac{qf(s - \hat{m})}{qf(s - \hat{m}) + (1 - q)f(s)}.$$

<sup>6</sup> The citizen stage always has a unique set of best responses, though as discussed in the online Appendix B it is possible that the regime has no pure strategy.

<sup>7</sup> For  $q \in (0, 1)$ , there is nothing inherently non-standard about the definition of  $q$ -credulity. It is simply what happens when the government can manipulate but some actors are uncertain about whether this is true. The only non-standard aspect of the fully credulous outcome is that it is a measure zero event, as it corresponds to the case where manipulation happens but it is common knowledge that some part of the population believes this is impossible. However, the equilibrium behavior is continuous in  $q$  for all of  $[0, 1]$ , so one way to avoid considering a measure zero event is to treat the outcome with full credulity as the limiting case as  $q \rightarrow 0$ .

Given the quadratic loss payoffs, the optimal action is a weighted average of the expected average action of others and the true performance, where the weight is equal to the coordination motive:

$$BR(I, s) = \lambda \mathbb{E}[\bar{a}|I, s] + (1 - \lambda) \mathbb{E}[\theta|I, s]. \tag{1}$$

Since citizens with the same information always have a unique best response, their equilibrium strategies must be symmetric. Let  $a_I(s, \hat{m})$  be the action taken by a citizen with type  $I$ .

The informed types know that  $\theta = s$  if  $\omega = 0$  and  $\theta = s - m$  if  $\omega = 1$ . The uninformed type's expected belief about  $\theta$  is  $s - \bar{q}(s, \hat{m})\hat{m}$ .

The  $\mathbb{E}[\bar{a}|I, s]$  terms depend linearly on the strategies used by other citizens and belief about the composition of the population. So, setting  $a_I(s, \hat{m}) = \lambda \mathbb{E}[\bar{a}|I, s] + (1 - \lambda) \mathbb{E}[\theta|I, s]$  for  $I \in \{0, 1, \emptyset\}$  gives a system of three linear equations and three unknowns for each  $s$ . This system has a unique solution:

**Proposition 1.** For each signal  $s$  and a fixed common conjecture about the regime's manipulation choice ( $\hat{m}$ ), the citizen stage has a unique set of strategies which are mutual best responses:

$$a_0^*(s, \hat{m}) = s - \frac{\bar{q}(s, \hat{m})\epsilon\lambda}{1 - \lambda(1 - \epsilon)}\hat{m} \tag{2}$$

$$a_1^*(s, \hat{m}) = s - \frac{1 - \lambda(1 - \bar{q}(s, \hat{m})\epsilon)}{1 - \lambda(1 - \epsilon)}\hat{m} \tag{3}$$

$$a_\emptyset^*(s, \hat{m}) = s - \bar{q}(s, \hat{m})\hat{m}. \tag{4}$$

**Proof.** All proofs are in Appendix A.  $\square$

Recall  $s = \theta + m$ , so in the equilibrium with  $q$ -credulity the action taken (as a function of the true government strength) for the credulous type is:

$$a_1^*(\theta + m, \hat{m}) = \theta + m - \bar{q}(\theta + m, \hat{m})\hat{m} = \theta + (m - \hat{m}) + r_\emptyset(\theta, m, \hat{m})\hat{m}$$

where  $r_\emptyset(\theta, m, \hat{m}) = (1 - \bar{q}(\theta + m, \hat{m}))$ . This decomposition highlights two ways manipulation can lead to higher actions among the credulous types. First, the  $m - \hat{m}$  term reflects the fact that if the government manipulates more than they expect (conditional on being willing and able to do so), even the informed citizens will think the regime is more popular than they truly are. This term will drop out when citizens form a correct conjecture in equilibrium. However, the second term ( $r_\emptyset(\theta, m, \hat{m})\hat{m}$ ) reflects the fact that credulous citizens will not fully filter out manipulation because they think there is some chance that the signal was not manipulated. As this term will not drop out, I call  $r_\emptyset(\theta, m, \hat{m})$  the *responsiveness to manipulation*. An analogous calculation gives the responsiveness to manipulation for the informed types is<sup>8</sup>:

$$r_1(\theta, m, \hat{m}) = \frac{\lambda\epsilon(1 - \bar{q}(\theta + m, \hat{m}))}{1 - \lambda(1 - \epsilon)}.$$

The responsiveness to manipulation measures the degree to which citizens behave as if they believe the government's propaganda. When  $r_I = 1$ , a citizen always picks  $a = \theta + m$ , i.e., behaving as if the signal is true or as if he fully believes the government. When  $r_I = 0$ , the type with information  $I$  chooses  $a = \theta + m - \hat{m}$ , meaning they behave as if the regime has manipulated at level  $\hat{m}$ . So, in equilibrium (when  $m = \hat{m}$ ), they choose exactly  $\theta$ .

The responsiveness is a function of the true performance ( $\theta$ ) because the truth affects the signal  $s$ , which affects the credulous types belief about the likelihood of manipulation. So, the more relevant quantity to measure how citizen behavior is affected by manipulation (which will in turn drive the regime decision) is the *average responsiveness to manipulation* over possible realizations of  $\theta$ :

$$\bar{r}_1(m, \hat{m}) = \mathbb{E}_\theta[r_1(\theta, m, \hat{m})] = \frac{\lambda\epsilon(1 - \bar{\bar{q}}(m, \hat{m}))}{1 - \lambda(1 - \epsilon)} \tag{5}$$

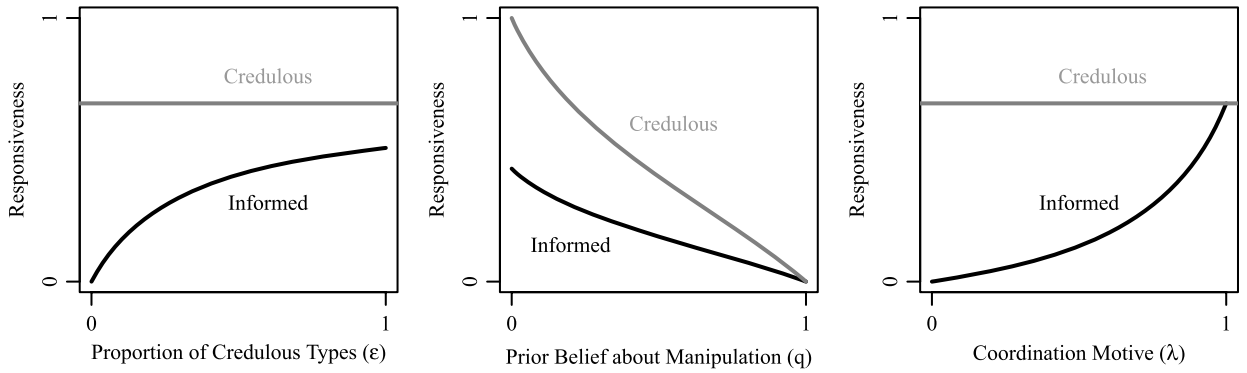
$$\bar{r}_\emptyset(m, \hat{m}) = \mathbb{E}_\theta[r_\emptyset(\theta, m, \hat{m})] = 1 - \bar{\bar{q}}(m, \hat{m}), \tag{6}$$

where  $\bar{\bar{q}}(m, \hat{m}) = \mathbb{E}_\theta[\bar{q}(\theta + m, \hat{m})]$  is the average posterior belief that the government has manipulated among the credulous types.

Fig. 1 shows how the average responsiveness to manipulation for the credulous types (grey curve) and informed types (black curve) varies as a function of the primitives of the model.

The left panel shows the effect of increasing the number of credulous citizens (higher  $\epsilon$ ). The credulous types are always partially responsive, independent of their prevalence, and are more responsive than the informed types. As the fraction of

<sup>8</sup> The notion of responsiveness to manipulation is not meaningful for the informed types who know  $\omega = 0$  as it conditions on  $\omega = 1$ .



**Fig. 1.** Responsiveness to manipulation in the equilibrium with  $q$ -credulity as a function of  $\epsilon$ ,  $q$ ,  $\lambda$ , and  $m$ . The black curve is the informed type responsiveness, and the grey curve is the credulous type responsiveness. In each panel,  $\theta$  is drawn from a standard normal distribution, and where they are not the  $x$ -axis variable, the parameters are set to  $q = 1/5$ ,  $\epsilon = 1/4$ ,  $\lambda = 3/4$ , and  $m = \hat{m} = 1$ .

credulous types increases, the informed types become more responsive to manipulation since the credulous type behavior matters more via the coordination motive. The middle panel shows how the degree of credulity affects the responsiveness to manipulation. Fully credulous types ( $q \rightarrow 0$ ), behave as if they completely believe the government ( $\bar{r}_\theta \rightarrow 1$ ). As the credulous types think it is more likely that the government can manipulate (increasing  $q$ ), they are less responsive to manipulation, as are the informed types since they anticipate the lower responsiveness of the credulous types.

The key result from the citizen interactions is shown in the right panel. When the coordination motive goes to 0, citizens only want to pick an action that matches the state, and hence those who know the government manipulates are unresponsive ( $r_1 = 0$ ). As the coordination motive increases, those who know the government can manipulate become more responsive to manipulation since they care more about picking an action closer to the credulous types. As  $\lambda \rightarrow 1$ , the informed type strategy converges to that of the credulous type. That is, as the citizens only care about coordination, the belief of the credulous type acts as a focal point determining the behavior of everyone.

Formalizing these observations:

**Proposition 2.** For any expected and chosen manipulation level, in an equilibrium with  $q$ -credulity, the informed type responsiveness to manipulation  $\bar{r}_1(m, \hat{m})$  is:

- (i) strictly positive;
- (ii) strictly increasing in  $\lambda$  and  $\epsilon$  and strictly decreasing in  $q$ ; and
- (iii) approaches the responsiveness of the credulous types ( $\bar{r}_1(m, \hat{m}) \rightarrow \bar{r}_\theta(m, \hat{m})$ ) as  $\lambda \rightarrow 1$ .

Part (iii) formalizes the observation in the right panel of Fig. 1: when the coordination motive is high, the informed types mimic the behavior of the credulous types. In other words, the blind lead the sighted. An immediate corollary of this result is that with full credulity ( $q = 0$ , and hence  $\bar{q}(m, \hat{m}) = 0$ ), as  $\lambda \rightarrow 1$ ,  $\bar{r}_1 \rightarrow \bar{r}_\theta = 1$ , i.e., everyone is fully responsive to propaganda.

Why is it that the blind lead the sighted rather than the other way around? In other words, why can't the informed types use their informational advantage to push average actions towards the optimal level where everyone discounts the government's lies? At risk of stretching the analogy, the blind can not follow the sighted precisely because they are blind. That is, if the informed citizens were to "use their sight" and condition their behavior on whether the government is lying, the best response of the uninformed types would be to pick an intermediate action since they do not know whether the signal is manipulated or not. By the coordination motive, this will pull the behavior of both informed types towards the uninformed type action, as they can't force the uninformed types to "see" what they can.

**3. The regime choice**

When  $\omega = 1$ , the average action taken by the citizens (averaging across informed and credulous types) given expected manipulation level  $\hat{m}$  and actual manipulation level  $m$  is:

$$\begin{aligned} \mathbb{E}[\bar{a}|m, \hat{m}] &= \mathbb{E}[\theta] + m - \left( \epsilon \bar{q}(m, \hat{m}) + (1 - \epsilon) \frac{1 - \lambda(1 - \bar{q}(m, \hat{m})\epsilon)}{1 - \lambda(1 - \epsilon)} \right) \hat{m} \\ &= \mathbb{E}[\theta] + m - (1 - \bar{r}(m, \hat{m}))\hat{m}, \end{aligned}$$

where

$$\bar{r}(m; \hat{m}) = \frac{\epsilon}{1 - \lambda(1 - \epsilon)}(1 - \bar{q}(m; \hat{m}))$$

represents the *total responsiveness to manipulation*, averaging across the informed and credulous types. The expected payoff for choosing  $m$  when the citizens expect  $\hat{m}$  is then:

$$\mathbb{E}[u_G(m; \hat{m})] = \mathbb{E}[\theta] + m - (1 - \bar{r}(m, \hat{m}))\hat{m} - c(m).$$

Picking a higher level of manipulation affects the government payoff in three ways. First, it increases the signal seen by the citizens, inducing them all (for a fixed level of responsiveness and expected level of manipulation) to believe the government is performing better. Second, the increase in the signal also changes the posterior beliefs about manipulation for the credulous types. Finally, the government pays a greater cost through the  $c$  function.

An equilibrium  $m^*$  is a level of manipulation such that when the citizens expect  $\hat{m} = m^*$ , it is in fact optimal for the type that can manipulate to choose level  $m^*$ . Taking the derivative of the regime utility and evaluating at  $\hat{m} = m^*$  gives the following the equilibrium condition for the level of manipulation:

$$1 = -m^* \left. \frac{\partial \bar{r}(m, m^*)}{\partial m} \right|_{m=m^*} + c'(m^*). \tag{7}$$

In the fully credulous equilibrium,  $\bar{r}(m, m^*) = \frac{\epsilon}{1-\lambda(1-\epsilon)}$  for all  $m$ , so this expression reduces to  $c'(m^*) = 1$ , which by the assumptions on  $c$  has a unique solution. So, in this case there is a unique equilibrium where  $m^*$  solves  $c'(m^*) = 1$  and the citizens strategies are given by equations (2)–(4) evaluated at  $\hat{m} = m^*$ .

The solution with partial credulity ( $q > 0$ ) is complicated by the fact that the  $\frac{\partial \bar{r}(m, m^*)}{\partial m}$  term does not drop out. So, when committing more propaganda than expected, the government faces an additional cost if this makes the credulous citizens find the propaganda becomes less believable and hence become less responsive, as do the informed citizens because of the coordination motive. The next result shows that for a wide class of distributions, this is in fact true:

**Proposition 3.** *If  $f$  is log-concave, then for any  $\hat{m} > 0$ ,  $\frac{\partial \bar{r}(m, \hat{m})}{\partial m} < 0$ .*

The log-concavity assumption holds for most standard distributions (e.g., normal, exponential, uniform) and is a sufficient but not necessary condition for citizens to be less responsive when  $m$  is higher.

When this effect is large, there can be more than one solution to the equilibrium condition, and the government objective function is not always concave. So, it is possible to have multiple pure strategy equilibria or *no* pure strategy equilibrium (see the Appendix A for an example). However, as long as the increased discounting effect is not too large (or, more precisely, is not too steep in  $m$ ), the model still has a unique pure strategy equilibrium, with the following comparative statics:

**Proposition 4.** (i) *If  $\frac{\epsilon}{1-\lambda(1-\epsilon)}$  is small or  $f$  is sufficiently flat, there is a unique pure strategy PBE to the model for any  $q$ .*

(ii) *In any pure strategy PBE, if  $\frac{\partial \bar{r}(m, m^*)}{\partial m} < 0$  then the equilibrium level of propaganda with partial credulity is decreasing in  $\epsilon$  and  $\lambda$ .*

The comparative statics imply that when citizens are highly responsive to propaganda ( $\epsilon$  and  $\lambda$  are high), the government chooses *less* propaganda in equilibrium (where again  $f$  being log-concave is a sufficient (but not necessary) condition for  $\frac{\partial \bar{r}(m, m^*)}{\partial m} < 0$ ). To see why, first consider the extreme with no credulous citizens. In this case, the equilibrium responsiveness is zero: as in standard career concerns models, the citizens fully filter out manipulation and always choose  $a = \theta$ . However, the citizens would be responsive to unexpected deviations; in particular, the marginal increase the action taken when the regime picks a higher level of manipulation is one.

Now add credulous citizens to the mix. This increases the equilibrium responsiveness to manipulation to something greater than zero, as the credulous citizens are partially fooled in equilibrium and the informed types are dragged along in taking higher actions as well. However, the presence of credulous citizens *decreases* the responsiveness to unexpected increases in propaganda because of the effect more propaganda has on how believable it is. So, contrasting Propositions 2 and 4, increasing  $\epsilon$  and  $\lambda$  increases the equilibrium responsiveness to manipulation, but decreases the responsiveness to off-the-path increases in manipulation by magnifying the effect of making propaganda less believable. Since the choice of manipulation is determined by the latter, the government will end up picking a higher level propaganda precisely when it is less effective.

## 4. Conclusion

### 4.1. Robustness and extensions

The supplemental appendix contains several extensions and further analyses. The core results for the citizen stage hold with a more general signal structure (i.e., under weak restrictions on  $\mathbb{E}[\theta|s, \omega]$ ) and when there are more than two possible manipulation levels. It also explores the implications of allowing the government to be informed about their performance ( $\theta$ ), identifying a special case where the equilibrium behavior is identical to the main model and discussing why the main conclusions are unlikely to change more generally.

The supplemental appendix also contains an extension where all citizens are aware that the government manipulates information, but a fraction are *second-order credulous*, meaning they (incorrectly) believe that some other citizens are credulous. Similar results hold: when coordination motives are strong, the second-order credulous citizens herd on the behavior of what they think the (non-existent) credulous citizens would do, and the fully informed citizens herd on the behavior of the second-order credulous citizens. So, propaganda can influence mass behavior even if it has *no* impact on citizens' beliefs.

#### 4.2. Further discussion

These results have implications for a wide variety of political behavior across regime types. A particularly apt example is Wedeen's description of Syrian's behavior under the first Assad, "people are not required to believe the cult's fictions, and they do not, but they are required to behave *as if* they did" (emphasis in original) (Wedeen, 1999, p. 30). More generally, these results imply that when coordination motives are strong, it is hard to distinguish between genuine belief in the fictions told by politicians and a desire to avoid sticking out by defying the government even when nearly everyone knows they are lying.

When should we expect the coordination motive to be strong? A drawback of the model's abstraction is that it does not explicitly model any political or economic institutions, making it hard to interpret when  $\lambda$  will be high or low in concrete terms. Still, it seems reasonable to speculate that the coordination motive is generally higher in authoritarian countries, or more precisely in countries where not toeing the party line risks retribution. Societies with more norms of deference to authority and conformity (rather than more individualistic norms) may also place a higher weight on speaking and behaving in a similar manner to others, which can lead to the herding behavior described in Proposition 2.

More broadly, the standard game-theoretic assumption that actors form rational expectations about the behavior of others is tremendously powerful and should not be discarded lightly. Assuming a political actor does  $X$  because it fools citizens into doing their bidding can explain any  $X$ , and as a result we should be wary of this line of argumentation. However, in the case of propaganda and censorship, substantial empirical evidence indicates that people can be fooled, and the formal model here demonstrates this has important implications for understanding the manipulation of political information.

Still, the results in the section on the regime behavior indicate that the ability to fool others in this manner is a double-edged sword. Even when citizens are responsive to propaganda despite most not being fooled, the regime may not be able to commit to choose levels of propaganda that are most effective. When propaganda is believable to credulous types, this will generate incentives to make it even more extreme, to the point where it can become too ridiculous to be believed by anyone. Distinguishing between cases where people truly believe versus pretend to believe their government, and when ridiculous lies do or do not change beliefs and behavior may be challenging. This challenge can be seen more clearly and guided by theory, and the modeling done here provides a framework for future work in this vein.

### Appendix A

#### A.1. Proof of Proposition 1

For informed types, the expected average action is:

$$\mathbb{E}[\bar{a}|\omega, s] = \epsilon a_{\emptyset}(s, \hat{m}) + (1 - \epsilon)a_{\omega}(s, \hat{m}),$$

and the expected average action for the uninformed types is:

$$\mathbb{E}[\bar{a}|\emptyset, s] = \epsilon a_{\emptyset}(s, \hat{m}) + (1 - \epsilon)(\bar{q}(s, \hat{m})a_1(s, \hat{m}) + (1 - \bar{q}(s, \hat{m}))a_0(s, \hat{m})).$$

So, the system of equations generated by the best response functions is:

$$a_0(s, \hat{m}) = \lambda(\epsilon a_{\emptyset}(s, \hat{m}) + (1 - \epsilon)a_0(s, \hat{m})) + (1 - \lambda)s$$

$$a_1(s, \hat{m}) = \lambda(\epsilon a_{\emptyset}(s, \hat{m}) + (1 - \epsilon)a_1(s, \hat{m})) + (1 - \lambda)(s - \hat{m})$$

$$a_{\emptyset}(s, \hat{m}) = \lambda(\epsilon a_{\emptyset}(s, \hat{m}) + (1 - \epsilon)(\bar{q}(s, \hat{m})a_1(s, \hat{m}) + (1 - \bar{q}(s, \hat{m}))a_0(s, \hat{m}))) + (1 - \lambda)(s - \bar{q}(s, \hat{m})m),$$

which has a unique solution given by equations (2)–(4).

#### A.2. Proof of part (ii) of Proposition 2

The comparative statics on  $\epsilon$  and  $\lambda$  follow immediately from equations (5) and (6).

For the comparative static on  $q$ , first note that for all  $s$ :

$$\frac{\partial \bar{q}(s, \hat{m})}{\partial q} = \frac{q(1 - q)f(s)f(s - \hat{m})}{(qf(s - \hat{m}) + (1 - q)f(s))^2} > 0.$$

So  $\frac{\partial \bar{q}}{\partial q} > 0$  and therefore  $\frac{\partial \bar{r}_1}{\partial q} < 0$ .

A.3. Proof of Proposition 3

The derivative of  $\bar{r}(m; \hat{m})$  with respect to  $m$  is:

$$\frac{\partial \bar{r}(m; \hat{m})}{\partial m} = \frac{q(1-q)\epsilon}{1-\lambda(1-\epsilon)} \mathbb{E}_\theta \left[ \frac{f(\theta+m-\hat{m})f'(\theta+m) - f(\theta+m)f'(\theta+m-\hat{m})}{(qf(\theta+m-\hat{m}) + (1-q)f(\theta+m))^2} \right].$$

The term inside the expectation is negative if and only if:

$$f(\theta+m-\hat{m})f'(\theta+m) < f(\theta+m)f'(\theta+m-\hat{m}) \iff \frac{f'(\theta+m)}{f(\theta+m)} < \frac{f'(\theta+m-\hat{m})}{f(\theta+m-\hat{m})}. \tag{8}$$

It follows from log concavity that  $\frac{f'(\theta)}{f(\theta)}$  (which is the derivative of  $\log f(\theta)$ ) must be decreasing. So equation (8) holds for all  $\theta$  if  $\hat{m} > 0$ , and hence  $\frac{\partial \bar{r}(m; \hat{m})}{\partial m} < 0$ .

A.4. Proof of Proposition 4

The equilibrium condition can be written  $1 = d(m^*) + c'(m^*)$ , where:

$$d(m^*) = -m^* \left. \frac{\partial \bar{r}(m, m^*)}{\partial m} \right|_{m=m^*} = m^* \frac{\epsilon}{1-\lambda(1-\epsilon)} \left. \frac{\partial \bar{q}(m, m^*)}{\partial m} \right|_{m=m^*}.$$

The conditions in part (i) are to ensure  $\frac{\partial d^*}{\partial m}$  is not too steep, which is sufficient to ensure (1) there is a unique  $m^*$  meeting equation (7), and (2) that the objective function when the citizens expect  $\hat{m} = m^*$  is globally concave. Clearly  $\frac{\partial d^*}{\partial m} \rightarrow 0$  as  $\epsilon \rightarrow 0$ , so  $\epsilon$  sufficiently small ensures the  $d$  function is never too steep. When  $f$  is “flat”,  $\bar{q}(s, m)$  will tend to be flat in  $s$  as  $f(s) \approx f(s - m)$ . For example, if  $f$  is the improper uniform prior, then  $\bar{q}(s, m) = q$ , which implies  $\frac{\partial d^*}{\partial m} = 0$  and hence the equilibrium is unique.

For part (ii), implicitly differentiating the equilibrium condition with respect to  $\epsilon$  gives:

$$\frac{\partial m^*}{\partial \epsilon} = - \frac{\frac{\partial d(m^*)}{\partial \epsilon}}{\frac{\partial d(m^*)}{\partial m^*} + c''(m^*)}$$

the denominator must be negative at  $m^*$ , so the sign is the same as the sign of:

$$\frac{\partial d(m^*)}{\partial \epsilon} = m^* \left. \frac{1-\lambda}{(1-\lambda(1-\epsilon))^2} \frac{\partial \bar{q}(m, m^*)}{\partial m} \right|_{m=m^*}$$

which is positive if and only if  $\left. \frac{\partial \bar{q}(m, m^*)}{\partial m} \right|_{m=m^*}$  is positive, which is true if and only if  $\left. \frac{\partial \bar{r}(m, m^*)}{\partial m} \right|_{m=m^*}$  is negative. The comparative static on  $\lambda$  follows from a similar argument.  $\square$

Appendix B. Supplementary material

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.geb.2016.12.006>.

References

Adena, Maja, Enikolopov, Ruben, Petrova, Maria, Santarosa, Veronica, Zhuravskaya, Ekaterina, 2015. Radio and the rise of the nazis in prewar Germany. *Quart. J. Econ.* 130 (4), 1885–1939.

Cai, Hongbin, Wang, Joseph Tao-Yi, 2006. Overcommunication in strategic information transmission games. *Games Econ. Behav.* 56 (1), 7–36.

Chen, Ying, 2011. Perturbed communication games with honest senders and naive receivers. *J. Econ. Theory* 146 (2), 401–424.

Dewan, Torun, Myatt, David P., 2008. The qualities of leadership: direction, communication, obfuscation. *Amer. Polit. Sci. Rev.* 102 (3), 351–368.

Edmond, Chris, 2013. Information manipulation, coordination, and regime change. *Rev. Econ. Stud.* 80 (4), 1422–1458.

Edmond, Chris, 2015. Information manipulation and social coordination. Manuscript, available at: <http://www.chrisedmond.net/Edmond%20Social%20Coordination%202015.pdf>.

Enikolopov, Ruben, Petrova, Maria, Zhuravskaya, Ekaterina, 2011. Media and political persuasion: evidence from Russia. *Amer. Econ. Rev.* 101 (7), 3253–3285.

Gehlbach, Scott, Simpson, Alberto, 2015. Electoral manipulation as bureaucratic control. *Amer. J. Polit. Sci.* 59 (1), 212–224.

Gehlbach, Scott, Sonin, Konstantin, 2014. Government control of the media. *J. Public Econ.* 118 (0), 163–171.

Haltiwanger, John, Waldman, Michael, 1985. Rational expectations and the limits of rationality: an analysis of heterogeneity. *Amer. Econ. Rev.*, 326–340.

Hollyer, James R., Rosendorff, B. Peter, Vreeland, James Raymond, 2015. Transparency, protest and autocratic instability. *Amer. Polit. Sci. Rev.* 109 (4), 764–784.

Holmstrom, Bengt, 1999. Managerial incentive problems: a dynamic perspective. *Rev. Econ. Stud.* 66 (1), 169–182.

Huang, Haifeng, 2015. Propaganda as signaling. *Comp. Polit.* 47 (4), 419–444.

Kartik, Navin, Ottaviani, Marco, Squintani, Francesco, 2007. Credulity, lies, and costly talk. *J. Econ. Theory* 134 (1), 93–116.

Kuran, Timur, 1997. Private Truths, Public Lies: The Social Consequences of Preference Falsification. Harvard University Press.

Little, Andrew T., 2012. Elections, fraud, and election monitoring in the shadow of revolution. *Quart. J. Polit. Sci.* 7 (3), 249–283.



- Little, Andrew T., 2015. Fraud and monitoring in non-competitive elections. *Polit. Sci. Res. Meth.* 3 (01), 21–41.
- Lorentzen, Peter, 2014. China's strategic censorship. *Amer. J. Polit. Sci.* 58 (2), 402–414.
- Morris, Stephen, Shin, Hyun Song, 2002. Social value of public information. *Amer. Econ. Rev.* 92 (5), 1521–1534.
- Ottaviani, Marco, Squintani, Francesco, 2006. Naive audience and communication bias. *Int. J. Game Theory* 35 (1), 129–150.
- Patty, John W., Weber, Roberto A., 2007. Letting the good times roll: a theory of voter inference and experimental evidence. *Public Choice* 130 (3–4), 293–310.
- Petrova, Maria, Zudenkova, Galina, 2015. Content and Coordination Censorship in Authoritarian Regimes. Manuscript, available at: <http://papers.sioe.org/paper/1406.html>.
- Shadmehr, Mehdi, Bernhardt, Dan, 2015. State censorship. *Amer. Econ. J. Microecon.* 7 (2), 280–307.
- Simpser, Alberto, 2013. *Why Governments and Parties Manipulate Elections: Theory, Practice, and Implications*. Cambridge University Press.
- Wang, Joseph Tao-yi, Spezio, Michael, Camerer, Colin F., 2010. Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender–receiver games. *Amer. Econ. Rev.* 100 (3), 984–1007.
- Wedeen, Lisa, 1999. *Ambiguities of Domination: Politics, Rhetoric, and Symbols in Contemporary Syria*. University of Chicago Press.
- Yanagizawa-Drott, David, 2014. Propaganda and conflict: theory and evidence from the Rwandan genocide. *Quart. J. Econ.* 129 (4), 1947–1994.