Unbelievable Lies*

Andrew T. Little† Sherif Nasser‡

August 31, 2017

Abstract

Politicians tell unbelievable lies. But why bother telling lies which aren't believed? We develop a formal model to address this and related questions. The "optimal lie" to manipulate the beliefs of an initially credulous citizen (i.e., one who believes the politician might tell the truth) is never too extreme. However, if lying is free, politicians can not restrain themselves to tell the most effective lies: since the optimal lie is partially believed, they are always tempted to exaggerate more. Even if lying is costly, politicians generally tell more extreme lies than they would like to *ex ante* commit to. This tension is particularly acute for politicians that care a great deal about perceptions of their performance, who are unable to tell persuasive lies.

*Some material in this paper previously circulated in unpublished drafts of Little (2017). Many thanks to Abraham Aldama, Dimitri Landa, Andrew Marshall, Jack Paine, and audiences at APSA 2016, SPSA 2017, NYU, Rochester, and Stanford for comments and discussion.

†Department of Political Science, UC Berkeley.

‡SC Johnson College of Business, Cornell University.

Politicians frequently tell lies so extreme that no reasonable observer could believe them. Some of these lies are at best tangentially related to their political abilities or performance, such as Bashar Assad's claim to be Syria's premier pharmacist, or Saparmurat Niyazov of Turkmenistan boasting that he made an agreement with God stipulating that anyone reading his book the Ruhnama three times would be guaranteed entry to heaven. However, many lies are about political consequential information like the performance of the economy, the wisdom of a proposed foreign policy, or the degree of corruption in the government.

Extreme lies can be counterproductive. If the goal is to persuade citizens or elites that the economy is growing at a healthy clip, reporting that GDP doubled over the last year is more likely to convince the audience that the speaker is making things up. Would it not be more effective to report a number only modestly higher than the truth, which has a higher chance of being believed? If so, why are extreme lies so common?

We develop a theory of these "unbelievable lies." A citizen observes a signal of a politician's performance. The politician would like the citizen to believe performance is high. To that end, he can increase ("manipulate") the signal. The citizen is initially "credulous", in the sense that she believes it is possible the politician can manipulate the signal but is not sure of this fact. Extremely high signals make the citizen more skeptical; eventually, she becomes nearly certain that the signal is manipulated. Our first main result shows that, under weak distributional assumptions, the degree of manipulation which maximizes the average belief about the politician is always finite. Beyond a certain point, more extreme lies become less persuasive.

Next, we show that when the exact degree of manipulation is a costly and hidden action taken by the politician, he tells a more extreme lie than would be optimal. This is because the equilibrium lie is determined by the point where the benefit to an *unexpected* increase in manipulation equals the marginal cost. However, the optimal lie is characterized by the point where an *expected* increase in manipulation equals the marginal cost. So, as long as an unexpected increase in manipulation changes the citizen beliefs more than an expected increase (and we provide a sufficient condition for this to hold), the politician lies more than they would like to.

Finally, we explore how the parameters of the model affect the optimal and equilibrium manipulation levels. Our most provocative result is about the behavior of politicians that care a lot about appearing popular (or face little exogenous cost to lying). As politicians become extremely "needy", the optimal manipulation approaches the level which maximizes the distortion of citizens beliefs. However, the equilibrium level increases without bound (if an equilibrium exists). Combining these observations, caring too much about appearing effective makes politicians unable to achieve this goal. We also show that the politician equilibrium payoff can be increasing in the precision of the prior belief about his performance, suggesting a benefit to allowing free or foreign media.

1 Literature

Much of the literature on strategic communication studies the degree to which informed parties truthfully convey their information to receivers, and so is to some degree about lying. A smaller literature studies political lies in particular, in the context of campaign promises (Callander and Wilkie, 2007), propaganda (Egorov et al., 2009; Edmond, 2013; Guriev and Treisman, 2015; Horz, 2017), or distorting economic data (Hollyer et al., 2015, 2014). While lying politicians are universal across time and space, recent political economy work on authoritarian politics has placed a particularly strong focus on information manipulation (see Gehlbach et al. 2016 for recent review).

Regardless of the specific technology, information manipulation occurs using standard equilibrium concepts with a strategic audience due to a combination of four factors (models of communication where the audience is *not* fully strategic are discussed below). First, following Crawford and Sobel (1982), informed actors may selectively release information to manipulate the beliefs of a receiver. Second, the sender may have private information not only about the state of the world but their ability or need to manipulate. So, a high degree of manipulation can serve as a costly signal that a politician is able to do so easily (Simpser, 2013; Huang, 2014). Third, some kinds of manipulation must be unobserved by the audience it aims to influence. This can lead to "career concerns" (Holmstrom, 1999) dynamics where manipulation occurs even though the audience correctly adjusts for it, as manipulating less than expected would make the sender look weak (Little, 2012, 2015). Fourth, information manipulation can serve as what Gentzkow and Kamenica (2011) call "Bayesian persuasion," and others in more closely related models call "signal jamming" (Edmond, 2013; Rozenas, 2015). In these models, manipulation is beneficial even if it doesn't make the audience think the sender is better *on average*, but because it rearranges the distribution of posterior beliefs in a favorable manner (Chen and Xu, 2014; Gehlbach and Sonin, 2014; Shadmehr and Bernhardt, 2015; Gehlbach and Simpser, 2015; Guriev and Treisman, 2015; Hollyer et al., 2015).[1]

This work assumes the audience for manipulated information is fully aware of the fact they are being lied to. While it is clearly valuable to devise theories of lying and information manipulation with a fully strategic audience, substantial empirical evidence indicates that some if not most people *don't* fully adjust for information manipulation in various contexts. For example, lab experiments consistently find that subjects tend to believe what others say, even if the sender has transparent incentives to lie (Cai and Wang, 2006; Patty and Weber, 2007; Dickson, 2010; Wang et al., 2010).

To formalize this, we employ a notion of credulity similar to that in Kartik et al. (2007), who study a cheap talk game where the receiver sometimes accepts the sender's message at face value (see also Ottaviani and Squintani 2006; Chen 2011).[2] More specifically, our model builds on Little

[1]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 (Lorentzen, 2014; Chen and Xu, 2014; Petrova and Zudenkova, 2015; Guriev and Treisman, 2015). Edmond (2013) includes all three dynamics: the government has private information, propaganda is a hidden action which jams the signal observed by citizens.

[2]See also Horz (2017), where a receiver chooses whether to become skeptical about what a sender tells him, which leads to analogous tradeoff where more distorted messages are less apt to be accepted. Ashworth and De Mesquita (2014) also analyze a model where voters do no correctly "filter" a signal of the government performance, albeit with a substantially different technology and purpose.

(2017), which focuses more on how the presence of credulous citizens affects the behavior of those who know for sure the government is manipulating information. Here we abstract from interactions among citizens to focus more on the decisions of the politician.[3] Most importantly, our paper is unique in the focus on the optimal level of manipulation in the presence of credulous citizens, and how this compares to equilibrium choices.[4]

2 The Model

We analyze a game between a politician (he) and a citizen (she). The citizen wants to learn the politician's true performance, θ, which neither actor directly observes. The citizen learns about θ from a signal s. We aim to capture a scenario where the signal is distorted by the politician, but the citizen is "credulous" in the sense that she thinks there is some chance the signal is unmanipulated.

To formalize this, suppose the politician can be one of two types: manipulative or truthful. A manipulative politician can upwardly distort the signal, with $m \geq 0$ denoting the level of manipulation. A truthful politician does not manipulate the signal. Hence, the signal can be expressed as $s = \theta + \omega\, m$, where $\omega \in \{0, 1\}$ is an indicator variable that takes the value of 0 if the politician is truthful, or 1 if he is manipulative.

The citizen does not know the politician's type. Her prior on ω is that it takes the value of 1 or 0 with respective probabilities $q \in (0, 1)$ and $1 - q$, independent of θ. The common knowledge prior on θ is that it is drawn from a probability distribution with density $f(\cdot)$. Our most general results only require the following assumptions on this distribution:

Assumption 1. *f is continuous, differentiable, strictly positive on $\mathbb{R}$, and has a finite expectation*

[3]In particular, Little (2017) contains comparative static results where some exogenous parameters increase the responsiveness of citizens to manipulation (measured in a similar way to this paper) but decrease the government's equilibrium choice. However, this paper does not define or characterize the optimal level of manipulation, which is central to the main results here.

[4]Using a different information structure and manipulation technology, Shadmehr and Bernhardt (2015) show that a government would increase their payoff if they could commit to censor slightly less than their equilibrium choice.

To assess the politician performance, the citizen must form a belief about the probability that the signal is manipulated and have a conjecture about the level of manipulation conditional on it occurring. Let $\hat{m}$ denote the citizen's conjecture about the level of manipulation. So upon observing signal, s, she updates her posterior probability that the signal is truthful (i.e., not manipulated) according to Bayes' rule:

$$Pr(\omega = 0|s, \hat{m}) = \frac{Pr(\omega = 0, s)}{Pr(s)} = \frac{(1-q)f(s)}{qf(s-\hat{m}) + (1-q)f(s)}. \tag{1}$$

By assumption 1, as long as $q \in (0,1)$, $Pr(\omega = 0|s, \hat{m}) \in (0,1)$, $\forall (s, \hat{m}) \in \mathbb{R} \times \mathbb{R}^+$. That is, unless the citizen is certain about the politician's type in the prior, she will not be certain about the politician's type in the posterior. Of course, she may come to believe that the politician being honest is more or less likely, and this inference will depend on the observed signal.

The average of the citizen's posterior belief about θ as a function of s and $\hat{m}$ is:

$$\begin{aligned}\mathbb{E}_\omega[\theta \mid s, \hat{m}] =& Pr(\omega = 0|s, \hat{m})s + (1 - Pr(\omega = 0|s, \hat{m}))(s - \hat{m}) \\ =& s - (1 - Pr(\omega = 0|s, \hat{m})) \cdot \hat{m}.\end{aligned}$$

A key quantity in our analysis is how much the citizen belief diverges from the truth when the politician is indeed manipulative. When this is the case (i.e., $\omega = 1$), the beliefs about the probability the politician is truthful can be expressed by substituting $\theta + m$ for s:

$$Pr(\omega = 0|s = \theta + m, \hat{m}) = \frac{(1-q)f(\theta + m)}{qf(\theta + m - \hat{m}) + (1-q)f(\theta + m)} \equiv r(\theta, m, \hat{m}) \tag{2}$$

Following Little (2017) (and for reasons which will become apparent), we refer to $r(\theta, m, \hat{m})$ as the *responsiveness to manipulation*. The average belief about the politician performance when he is indeed manipulative can now be written:

$$\hat{\theta}(\theta, m, \hat{m}) \equiv \mathbb{E}_{\omega}\left[\theta \mid s = \theta + m, \hat{m}\right] = \theta + \pi(\theta, m, \hat{m}),$$

$$\text{where} \qquad \pi(\theta, m, \hat{m}) \equiv \underbrace{m - \hat{m}}_{\text{filtering}} + \underbrace{\hat{m} \cdot r(\theta, m, \hat{m}).}_{\text{uncertainty about } \omega} \tag{3}$$

Call $\pi(\theta, m, \hat{m})$ the *manipulation boost* that a manipulative politician achieves. The expression highlights how the citizen's belief about θ can be manipulated in two senses. First, as in a standard "career concerns" model, if the politician manipulates more than expected ($m > \hat{m}$), then the citizen would filter less than the true level of manipulation *even if she is certain that* $\omega = 1$. Using standard equilibrium concepts where the citizen forms rational expectations (i.e., $\hat{m} = m$), this will drop out. The second source of manipulation is the $\hat{m} \cdot r(\theta, m, \hat{m})$ term, which is the degree to which the citizen is "fooled" since she is not certain whether or not the politician is manipulative. When the citizen forms rational expectations, this is equal to the true manipulation level times the probability that the citizen believes the politician is truthful. We refer to this probability as the responsiveness to manipulation because when $r(\cdot) = 0$, she is not fooled by manipulation at all through this channel (and will form a correct belief about θ in equilibrium), but when $r(\cdot) = 1$, her belief about the performance is $\theta + m$, so in a sense her views are perfectly manipulated.

The manipulative politician payoff as a function of his true performance and manipulation choice is:

$$U_p(\theta, m, \hat{m}) = \hat{\theta}(\theta, m, \hat{m}) - c(m)$$

The first term captures the value the politician places on the citizen's posterior belief; and the second captures the exogenous cost of manipulation incurred by the politician.[5] Except when

[5] The assumption about types of politicians can be easily reduced to cost. That is, the truthful type has a prohibitively high cost that makes it optimal for him to set $m = 0$, while the manipulative type has lower cost making it optimal for him to set $m > 0$. As such, q and $1 - q$ can be interpreted as the relative sizes of each group/type of politicians.

analyzing the special case where manipulation is free (i.e., $c(m) = 0$), we assume the following:

Assumption 2. $c(\cdot)$ *is continuous, differentiable, strictly increasing, and convex, with* $c'(0) = 0$ *and* $\lim_{m\to\infty} c'(m) = \infty$.

This cost can capture any way manipulation harms the politician *except* the loss of credibility from the fact higher signals are less believable, which will arise endogenously. At the simplest level, this can represent a (psychological) "lying cost" (Kartik et al., 2007; Kartik, 2009). In addition, the politician may have to compensate subordinates to manipulate information in his favor, or hire less competent subordinates willing to lie (Zakharov, 2016).

Let $\bar{r}(m, \hat{m}) \equiv \mathbb{E}_\theta [r(\theta, m, \hat{m})]$ denote the *expected responsiveness to manipulation* and $\bar{\pi}(m, \hat{m}) \equiv \mathbb{E}_\theta [\pi(\theta, m, \hat{m})]$ denote the *expected manipulation boost.* The politician's expected payoff averages over the possible realizations of θ:

$$\bar{U}_p(m, \hat{m}) \equiv \mathbb{E}_\theta [U_p(\theta, m, \hat{m})] = \mathbb{E}[\theta] + \bar{\pi}(m, \hat{m}) = \mathbb{E}[\theta] + m - \hat{m} + \hat{m} \cdot \bar{r}(m, \hat{m}). \quad (4)$$

The citizen and honest politician take no actions, and so we do not need to specify a payoff function for them.[6]

To summarize, the sequence of stages in the game are:

Stage 0: Nature chooses θ and ω

Stage 1: The politician observes ω. If $\omega = 1$, he chooses his manipulation level, m; otherwise, he takes no action.

Stage 2: The citizen forms her conjecture of the manipulation level, $\hat{m}$.

[6]It is trivial to embed the manipulative politician behavior in a model where the citizen an honest politician take actions. For the citizen, we could assign a utility function which is maximized by taking an action equal to the average of her posterior belief about θ, and replace the citizen belief with this action in the politician payoff. For the honest politician, we could assume that his manipulative action does not impact the signal, and since it is costly he will choose $m = 0$. See also footnote 5.

Stage 3: The citizen observes the signal, $s = \theta + \omega m$, and forms her posterior beliefs.

To formalize what would be the best manipulation level for a politician interacting with a credulous but rational citizen, we define the optimal manipulation level to be the choice of m that, subject to the constraint that the citizen conjecture is correct (i.e., $\hat{m} = m$), maximizes the manipulative politician payoff. Put another way, to limit the ways in which the citizen can be fooled by manipulation, we suppose she knows how much manipulation occurs *conditional on the politician being a* $\omega = 1$ *type*, but is uncertain as to whether $\omega = 1$ or $\omega = 0$.

Definition The set of optimal levels of manipulation is:

$$M^{\text{opt}} = \arg\max_{m} \bar{U}_p(m, m) \tag{5}$$

In principle there can be multiple solutions to this optimization problem (e.g., if $\bar{U}_p$ has two peaks at exactly the same height). Let $m^{\text{opt}} = \max M^{\text{opt}}$ be "the" optimal level using a tie-breaking rule of selecting the largest maximizer. Later we impose an assumption which will ensure M^{opt} has a unique solution, and hence the tie-breaking rule is irrelevant.

The citizen takes no action and her beliefs are already built into the manipulative politician payoff. So, a standard equilibrium condition for the model only requires that the manipulative politician choice is a best response given the citizen beliefs:

Definition The set of equilibrium levels of manipulation are the manipulation levels m such that when the citizen expects manipulation level m, it is in fact optimal to choose m:

$$M^* = \{m : m = \arg\max_{m} \bar{U}_p(m, \hat{m} = m)\} \tag{6}$$

Adopting a similar convention to the optimal manipulation level(s), let $m^* = \max M^*$ (when M^* is non-empty, i.e., an equilibrium exists). We discuss the existence and uniqueness of the equilibrium manipulation level in detail in section 5.

Comments on setup. Our definition of the equilibrium manipulation level is standard, corresponding to what the choice would be when using a standard solution concept like Perfect Bayesian Equilibrium. There are several ways one could define the optimal manipulation level, with different treatments of the citizen expectation. For example, keeping the citizen conjecture fixed at any $\hat{m}$, the posterior belief about the politician performance is at least $\theta + m - \hat{m}$, and so is unbounded as $m \to \infty$. So, if the citizen's conjecture about the manipulation choice is unrelated to what the manipulative politician actually does, there is no limit to how distorted her belief can become. In a sense, the citizen can be fooled in two ways: having an incorrect conjecture about how manipulative politicians behave, and being uncertain about whether they face a manipulative politician. Our definitions only allow for manipulation through the second channel, which limits the difference between the equilibrium (where citizen conjectures must be correct) and optimal cases.

An unsatisfying aspect of this setup for some applications is the assumption that the politician has no private information about his performance. Symmetric uncertainty is common in career concerns-style models (following Holmstrom 1999), and greatly simplifies the analysis. And some decisions made by politicians (and others) which affect signals of their performance are plausibly made without private information. For example, political leaders decide how to structure and who to hire in statistical agencies or their communication office with the aim of affecting distortion of later-realized performance indicators.

Still, an admittedly more natural setup to capture situations where the politician knows his performance and may report the truth or lie would be to have him observe θ (or at least a noisy signal of θ) and then to choose the signal s. Mapping this to our formalization requires a trivial change and a consequential one. The trivial change is to reinterpret the cost as being increasing in the distance between s and θ. I.e., writing m as $s - \theta$, the cost is now $c(s - \theta)$.

More consequential is that the politician strategy is not a single manipulation level m, but a function mapping θ to the manipulation level $m(\theta)$. If we constrain this function to not depend on θ – i.e., assume the politician must lie the same amount regardless of his performance – then the analysis is unchanged. However without this constraint, characterizing and comparing the optimal and equilibrium choices requires a substantially more complicated analysis. We present some of this analysis in the appendix, which indicates that the main conclusions we draw from the more tractable formulation are unlikely to change.

3 Full and No Credulity

As benchmarks, we first discuss the special cases where the citizen knows for certain that the politician manipulates ($q = 1$) or is fully credulous ($q = 0$).

Suppose the citizen knows the politician manipulates. This becomes a standard career concerns filtering problem. In particular, $r(\theta, m, \hat{m}) = 0$ for any m, so the politician payoff for picking manipulation level m when the citizen expects $\hat{m}$ is:

$$\bar{U}(m, \hat{m})\big|_{q=1} = \mathbb{E}[\theta] + m - \hat{m} - c(m)$$

Regardless of $\hat{m}$, the first order condition for an equilibrium manipulation level (see Equations 4 and 6) becomes

$$1 = c'(m^*).$$

Since an *unexpected* increasing in manipulation leads to a one unit increase in the citizen performance assessment, the equilibrium level is given by the point where this one unit increase matches the marginal cost. As it will recur throughout the analysis (and coincides with a standard career concerns benchmark), we refer to this manipulation level as $m^{\text{CC}} > 0$. (By assumption 2, such a solution exists and is unique.)

Now consider the optimal manipulation level. At $q = 1$, the maximand in (5) becomes $\bar{U}(m, m)\big|_{q=1} = \mathbb{E}[\theta] - c(m)$. In this case, the citizen knows the politician is manipulating and forms a correct conjecture about the behavior of manipulative politicians, so she always correctly learns θ. Since c is increasing and manipulation confers no benefit, $m^{opt}|_{q=1} = 0$.

On the other extreme, if the citizen is fully credulous, then $r(\theta, m, \hat{m})|_{q=0} = 1$ for any m, so the expected utility function becomes:

$$\bar{U}(m, \hat{m})\Big|_{q=0} = \mathbb{E}[\theta] + m - c(m)$$

Since the $\hat{m}$, drops out, the first order condition for both the equilibrium and optimal manipulation levels is $1 = c'(m)$, which is solved by m^{CC}. Summarizing:

Remark 1. *For any cost function meeting assumption 2, for* $q = 1$, $m^{opt} = 0$ *and* $m^* = m^{CC}$. *When* $q = 0$, $m^{opt} = m^* = m^{CC}$.

So, when faced with a fully credulous citizen, the politician *is* able to pick the optimal manipulation level. When faced with a citizen who knows the politician is manipulating, he ends up manipulating even though it is completely ineffective.

The remainder of the paper considers the more interesting case where q is intermediate, and hence the degree to which the citizen believes the politician depends on the signal she observes. In general, we find that any deviation from the full credulity case will make the politician want to restrain himself from telling extreme lies because they become less believable. However, the incentives to lie more than expected make it hard for the politician to manipulate beliefs effectively.

4 Optimal Manipulation

We now analyze the optimal manipulation level when the citizen is not certain about the politician type at the outset: $q \in (0, 1)$.

Free manipulation. To start, consider the special case where manipulation is free. Dropping the cost and setting $\hat{m} = m$, the political payoff becomes $\theta + \overline{\pi}(m, m)$. (In general, we will write

$H(\cdots, m, m)$ to denote $H(\cdots, m, \hat{m} = m)$, for any function H.) So, the optimal manipulation level with no exogenous cost is that which maximizes $\overline{\pi}(m, m)$. Call this m_0^{opt} (again, using a tie-breaking rule of selecting the largest optimizer in the knife-edged case where this is necessary).

In general, there is a tradeoff where more manipulation makes the politician look more effective for a fixed responsiveness to manipulation, but can also decrease this responsiveness. A marginal increase in the manipulation level (both actual and expected) is:

$$\frac{d\overline{\pi}(m, m)}{dm} = \overline{r}(m, m) + m\frac{\partial \overline{r}(m, m)}{\partial m} \tag{7}$$

As it will always be negative under assumptions specified below, we refer to the second term – i.e., the decrease in the manipulation boost due to decreasing responsiveness – as the *endogenous cost of manipulation*.

Our first result is that because of this trade-off, the distortion-maximizing manipulation level is always strictly positive but finite:

Proposition 2. *For any prior density $f(\cdot)$ meeting assumption 1 and $q \in (0, 1)$*

i. For sufficiently small m, $\overline{\pi}(m, m)$ is increasing in m.

ii. At the limit, the expected manipulation boost goes to zero: $\lim_{m\to\infty} \overline{\pi}(m, m) = 0$.

iii. $m_0^{opt} \in (0, \infty)$

Proof. See the appendix.

Part i states that there is always a return to small degrees of manipulation, i.e., equation 7 is always positive at $m = 0$.[7]

Part ii formalizes the notion that as manipulation becomes very extreme, it becomes completely ineffective. This is because even a citizen who is very credulous (but not fully credulous, i.e., q

[7]This follows from the fact that the derivative of the average responsiveness at $m = 0$ also zero, and hence $\frac{d\overline{\pi}(m,m)}{dm}\Big|_{m=0} = \overline{r}(0, 0) = 1 - q$

small but strictly greater than zero) will become nearly certain that the politician is manipulative. Formally, $\overline{r}(m,m) \to 0$. The technical challenge is to show that this convergence happens "fast enough" that $m\,\overline{r}(m,m) \to 0$ as well, i.e., that $\overline{r}(m,m)$ converges to zero faster than $1/m$. The proof shows that a sufficient condition for sufficiently fast convergence is if f has a finite expectation.

Part iii immediately follows from the first two: since the manipulation boost is zero when $m = 0$ and approaches zero again as m gets arbitrarily large, it must have a finite maximizer.

Costly Manipulation. Since we have already shown the optimal manipulation level when there is no cost is finite, it is trivial that the optimal level of manipulation when it is costly is smaller than this and hence finite as well:

Proposition 3. *For any prior density $f(\cdot)$ meeting assumption 1, $m^{opt} \in [0, m_0^{opt})$*

Proof. Follows immediately from proposition 2 and $c' > 0$. □

To make comparisons to the equilibrium manipulation level and facilitate comparative statics, it will be useful to show that the politician objective function is single-peaked with a unique solution. Unfortunately, assumption 1 is not sufficient to ensure this: e.g., if f is multimodal the objective function can be multimodal as well. We present results with two further assumptions on f with additional restrictions:

Assumption 3. *$f(\cdot)$ is logarithmically concave*

Assumption 4. *$g''' \leq 0$, where $g(\cdot) \equiv \log f(\cdot)$*

Both assumptions 3 and 4 are met by many standard distributions such as the normal and extreme value distribution.[8] It is possible for assumption 3 to hold but not assumption 4, though we are unaware of any standard distribution where this is the case.

[8] It does not hold for Student-t distributions, though instructively, simulation suggest that the results to come seem to hold under this family of distributions as well. At the least, Assumptions 3-4 sufficient but not necessary for subsequent results.

With this additional structure on f we obtain the following:

Lemma 1. *i. Given assumptions 1 and 3, the average responsiveness to manipulation is strictly decreasing in the level of manipulation, i.e.,* $\frac{d\overline{r}(m,m)}{dm} < 0$.

ii. Given assumptions 1, 3, and 4, $\overline{\pi}(m,m)$ *is single-peaked; and the unique* m^{opt} *is characterized by:*

$$\left.\frac{d\overline{\pi}(m,m)}{dm}\right|_{m=m^{opt}} = c'(m^{opt}), \tag{8}$$

Part i gives a condition for the average responsiveness to be decreasing in the level of manipulation, which implies there is always a tradeoff in manipulating more.[9] Part ii implies this tradeoff makes the objective function single-peaked, which facilitates more straightforward comparisons to the equilibrium level and comparative statics.

For the remainder of the paper we assume that f meets assumptions 1, 3, and 4.

5 Equilibrium Manipulation

We now characterize the equilibrium manipulation level. The first order condition for an interior manipulation level is an m^* which solves:

$$c'(m^*) = 1 + m^* \left.\frac{\partial\overline{r}(m,\hat{m})}{\partial m}\right|_{\hat{m}=m^*}. \tag{9}$$

Two potential technical challenges arise. First, the right hand side of equation 9 can be increasing in m^*, which can result in multiple solutions to the first order condition. Second, and more problematic, since the politician optimization problem is not globally concave, the solution(s) may not correspond to a *global* maximizer.

[9]This will not necessarily hold for all m if, for example, f is bimodal.

There are two sufficient conditions for there to be a unique solution to this equation which is in fact a global maximizer of the politician utility. First, if the cost function is sufficiently convex, then the objective function will be globally concave, and the solution unique. Second, if the prior is sufficiently "flat", then the $m \left. \frac{\partial \overline{r}(m,\hat{m})}{\partial m} \right|_{\hat{m}=m}$ term is always small, which removes the potential for this term to add enough convexity to the objective function. Intuitively, if the prior is very flat, then the endogenous cost to manipulating more is generally not too high, and, in particular, does not change rapidly. Formally:

Proposition 4. *Write the cost function as $\kappa c_0(m)$ for some baseline cost function $c_0(m)$ and $\kappa > 0$, and add a scale parameter to the prior such that $f(\theta) = \lambda^{-1} f_0(\lambda\theta)$. Then if κ or λ are sufficiently large, then there is a unique equilibrium m^* characterized by equation 9.*

For the remainder of the formal analysis, we focus on the case where a unique equilibrium exists (though our illustrations will show examples where no equilibrium exists for part of the parameter space):

Assumption 5. *Either the condition on κ or the condition on λ identified in proposition 4 is met.*

Comparing the optimal versus equilibrium manipulation level requires comparing the solutions to equations 8 and 9. The left-hand sides of these equations are the same, so the equilibrium manipulation level is higher than optimal if and only if at the optimal manipulation level:

$$\begin{aligned} m^* > m^{opt} \qquad & \Leftrightarrow \qquad 1 + m \left. \frac{\partial \overline{r}(m,\hat{m})}{\partial m} \right|_{\hat{m}=m} > \overline{r}(m,m) + m \left(\left. \frac{\partial \overline{r}(m,\hat{m})}{\partial m} \right|_{\hat{m}=m} + \left. \frac{\partial \overline{r}(m,\hat{m})}{\partial \hat{m}} \right|_{\hat{m}=m} \right) \\ & \Leftrightarrow \qquad 1 - \overline{r}(m,m) > m \cdot \left. \frac{\partial \overline{r}(m,\hat{m})}{\partial \hat{m}} \right|_{\hat{m}=m} . \qquad (10) \end{aligned}$$

The left-hand side of equation 10 reflects the difference between an unexpected and expected increase in m for a fixed level of responsiveness. An unexpected increase in m increases the signal by one unit. An expected increase is partially filtered, leading to an $\overline{r}(m,m)$ unit increase in the citizen belief.

The right-hand side is the difference between how an expected and unexpected deviation changes the responsiveness to manipulation. An increase in the expected manipulation level (starting at a point where expectations are correct) can be decomposed into the sum of the effect of increasing the citizen expectation about manipulation and the increase in the actual manipulation level. So, this difference is equal to the effect of increasing the expected manipulation level but **not** the actual manipulation level.

Unfortunately the $\left.\frac{\partial \overline{r}(m,\hat{m})}{\partial \hat{m}}\right|_{\hat{m}=m}$ term can be positive or negative, and is hard to compare to the left-hand side of equation 9. Still, recall that we have already demonstrated that at $q = 1$, $m^{\text{opt}} = 0 < m^{\text{CC}} = m^*$. Cast in terms of equation 10, as $q \to 1$, $\overline{r}(m, m) \to 0$ for any m. So, the left-hand side of equation 10 goes to 1 while the right-hand side goes to zero, leaving $1 > 0$. So, by continuity we have a straightforward condition for the equilibrium manipulation to be higher than optimal:

Proposition 5. *There exists a $\hat{q} \in [0, 1)$ such that if $q > \hat{q}$, the equilibrium manipulation level is higher than optimal ($m^* > m^{opt}$).*

Proof. As shown above, equation 10 holds at $q = 1$. Further, both sides of the inequality are continuous in q, so it must hold for some open interval $(\hat{q}, 1)$. □

A good lie has to be big enough to meaningfully distort the truth, but not so large as to become too obvious. And whenever the distortion is not too obvious, the speaker has an incentive to tell a larger lie. Proposition 5 formalizes this idea, showing general conditions under which politicians lie more than is optimal. The condition stated requires that the audience be sufficiently skeptical about the politician (high q), though in all parameterizations we have examined any level of skepticism beyond uncritically accepting what the politician says ($q > 0$) is sufficient.[10]

Proposition 5 provides an explanation for a phenomenon that seems common in politics and elsewhere: lies become unbelievable and have no effect on beliefs, even though a more modest

[10]Recall that when $q = 0$, $m^{\text{opt}} = m^{\text{CC}} = m^*$.

lie would be effective. If so, a credulous citizenry may be less useful for politicians than it might seem.

Still, the difference between what is optimal and the equilibrium manipulation may be small, or at least small enough that the politician still benefits from the presence of credulous citizens. Our final technical analysis is a set of comparative static results which show this is often not true.

6 Comparative Statics

The parameters of the model which allow for comparative statics are the priors on the probability that the politician is manipulative (q) and the performance (f), and the cost function.

In this section we maintain assumptions 1-5 which ensure equations 8 and 9 have unique solutions m^{opt} and m^* which are global maximizers. So all comparative statics are obtained by implicitly differentiating these equations.

We first examine how changing q affects the optimal and equilibrium manipulation levels. As discussed in the full and no credulity benchmark analysis, the optimal manipulation level is $m^{\text{opt}} = m^{CC}$ when $q = 0$ and $m^{\text{opt}} = 0$ when $q = 1$. More generally, when the citizen is less credulous (higher q), the optimal level of manipulation goes down.

The equilibrium level of manipulation is non-monotone. The intuition behind this is easiest to see by first recalling that when $q = 0$ or $q = 1$, the equilibrium condition becomes $c'(m) = 1$, which is solved my m^{CC}. When q is intermediate, there is always some indirect cost to lying more, leading to less manipulation than in either extreme. Summarizing:

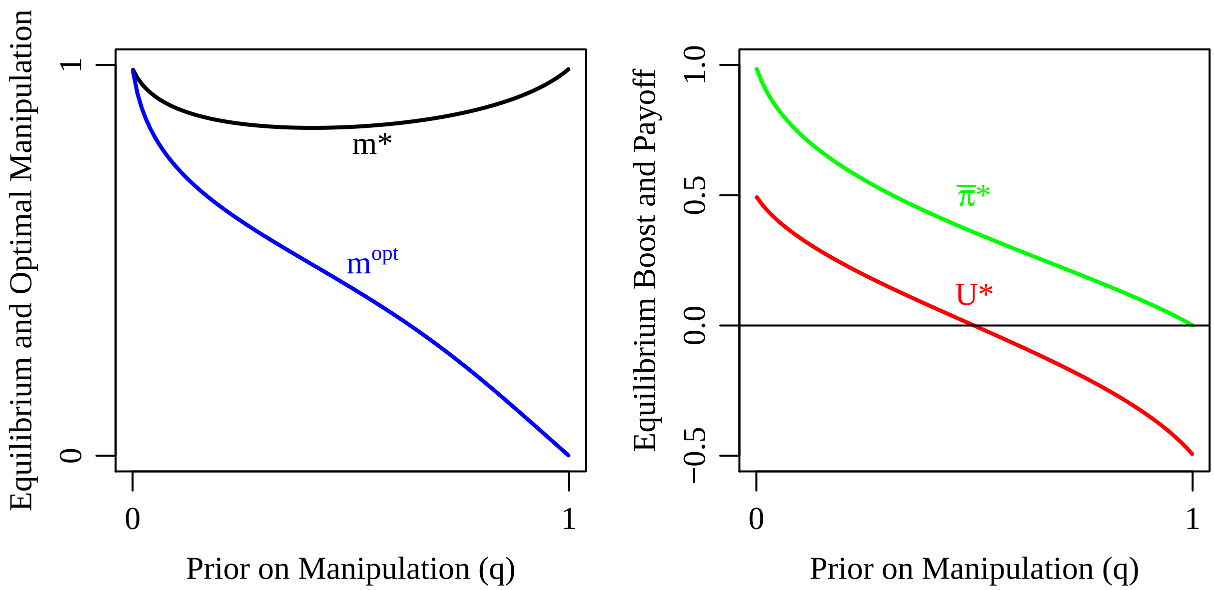
Proposition 6. *i. The optimal level of manipulation is decreasing in q, and*

ii. The equilibrium level of manipulation is decreasing for small q and increasing for large q

Proof. See the appendix

Figure 1 illustrates this result. The left panel shows that the optimal level of manipulation (blue curve) starts at m^{CC} as $q \to 0$, and is monotone decreasing to 0 as $q \to 1$. However, the

Figure 1: Equilibrium properties as a function of q, with f a standard normal and $c(m) = m^2$



equilibrium manipulation level (black curve) is only decreasing for small q, and for larger q is increasing. The right panel shows the implications for other equilibrium properties. The green curve shows the equilibrium manipulation boost ($\overline{\pi}(m^*, m^*)$, and the red curve the manipulative politician's average payoff ($\hat{U}_p(m^*, m^*)$). Both are decreasing in q: so as the citizen becomes more credulous, her beliefs are manipulated less in equilibrium and the politician is worse off. Interestingly, for $q \gtrsim 1/2$, the politician utility is less than 0, which is the mean of the prior and hence what his average payoff would be if he never manipulated and the citizen learned his type. So, for moderately high q, the politician is able to partially manipulate beliefs about this performance, but the cost to doing so outweighs the (equilibrium) benefits.

To examine how making manipulation more or less costly affects the outcomes, we return to the notation of proposition 4 and write the cost function as $c(m) = \kappa c_0(m)$. Further, note that if we add a scale parameter α to how much the politician cares about perceptions of his performance ("neediness"), the expected utility function can be written and then normalized as:

$$U_p(m;\hat{\theta}) = \alpha\hat{\theta} - c(m)$$
$$U_p(m,\hat{\theta})/\alpha = \hat{\theta} - c(m)/\alpha$$

So, a downward scaling of the cost function (i.e., decreasing κ) can capture changes not only in the cost but also the degree to which the politician cares about perceptions of his performance.

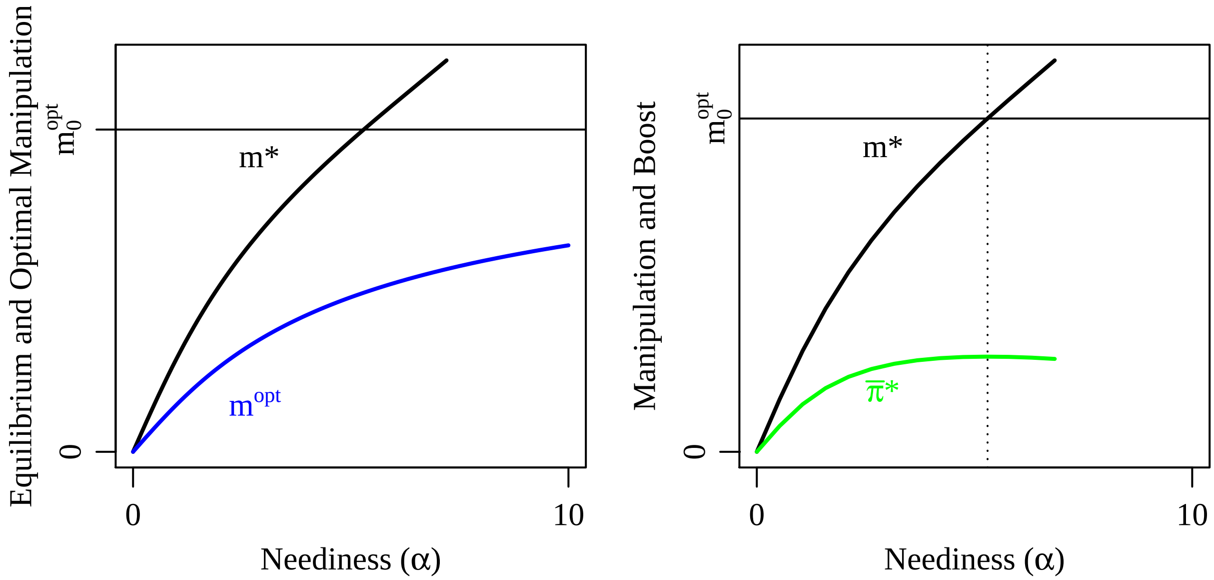
Not surprisingly, needier politicians have higher equilibrium and optimal manipulation levels. However, there is an upper bound on how high the optimal level can get, which is exactly the level which leads to the biggest average manipulation boost, i.e., m_0^{opt}:

Proposition 7. *Write the politician objective function as $\alpha\hat{\theta} - \kappa c_0(m)$. Where the conditions for a pure strategy equilibrium are met:*
i. The optimal and equilibrium manipulation levels are increasing in α and decreasing in κ,
ii. as $\alpha \to \infty$ or $\kappa \to 0$, $m^* \to \infty$ and $m^{opt} \to m_0^{opt}$, and
iii. the equilibrium manipulation boost is increasing in α (and decreasing in k) if $m^* < m_0^{opt}$, and is decreasing in α (and increasing in k) if $m > m_0^{opt}$

Figure 2 illustrates this result with respect to α. The left panel shows that both the optimal and equilibrium manipulation levels are increasing in how much the politician cares about the citizens' belief. However, note that when α is sufficiently high, the m^* curve stops as an equilibrium no longer exists. (This is because increasing α makes the objective function "less concave" by scaling down the effective cost parameter.)

The right panel illustrates part iii of proposition 7. For small α, the equilibrium manipulation level is less than m_0^{opt}, and so increasing the neediness of the politician leads to more manipulation and a higher equilibrium manipulation boost (the green curve). However, once m^* gets above m_0^{opt} (the dotted vertical line), the equilibrium manipulation boost bends downwards. So, at a certain

Figure 2: Equilibrium properties as a function of α



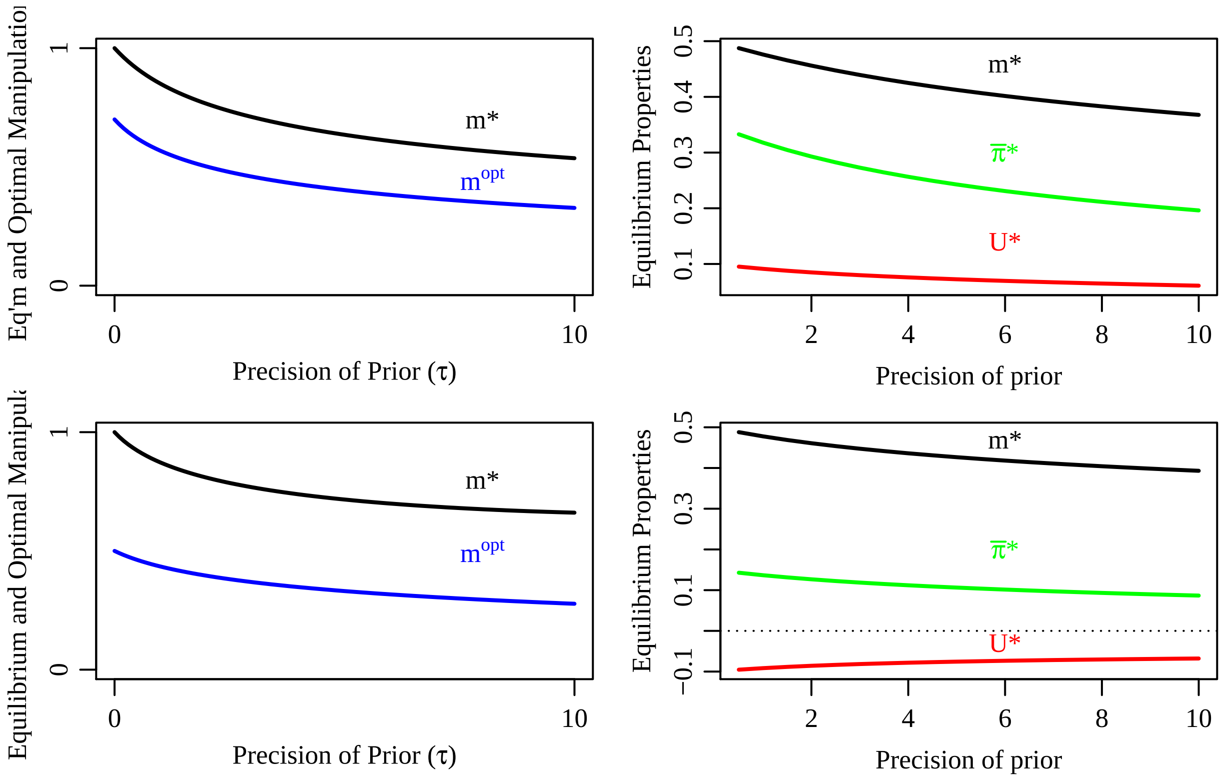
point, caring more about being seen as performing well can lead to even more extreme lies that are less effective at distorting the belief of the citizen.

Comparative statics on the prior distribution of θ prove difficulty analytically. However, we present one suggestive result from a simulation, where f is normally distributed with mean μ and precision τ.

The mean of the prior has no effect on the equilibrium or optimal manipulation level. This follows from the fact that the politician payoff is linear in the belief about his performance. So, the returns to increasing that belief do not depend on whether he is generally popular or unpopular. (This would not be the case if, for example, the politician payoff was strictly concave in $\hat{\theta}$.)

Figure 3 shows how increasing the precision of the prior affects the equilibrium properties for low q (top panels) and high q (bottom panels). In both cases, increasing τ decreases the equilibrium and optimal manipulation levels. This decrease always leads to a lower manipulation boost, as the manipulation goes down and the citizen has a easier time distinguish between clean and manipulated signals. However, the effect on the manipulative politician utility can go in either direction,

Figure 3: Equilibrium properties as a function of τ, for $q = .3$ (top panels) and $q = .7$ (bottom panels).



as he gets less of a manipulation boost but pays a lower cost. In the top right panel, the citizen is more credulous, and the latter effect dominates: adding more information in the prior lowers the politician payoff. However, in the bottom panel, where the citizen is less credulous at the outset, the cost effect dominates. So, the politician payoff is increasing in the precision of the prior.

7 Discussion

Most explanations of outlandish lying try to figure out how such lies can be effective, or assumes the speaker is pathological or irrational. We provide a theory where neither is true. Politicians tell lies that do not effectively manipulate beliefs precisely because they are rational, and can not restrain themselves to tell the kind of moderate lies which would be believed.

Several extensions could provide additional insight into the relationship between information manipulation, government survival, and allowing outside information (e.g., foreign media, free press).

The results suggest that the ability to manipulate information easily may backfire since it exacerbates the difference between optimal and equilibrium manipulation. So, governments that can manipulate easily may distort information to a greater degree even though this harms them. This would seem to contradict the fact that many long-lived autocracies (e.g., the Kim dynasty in North Korea) are the most extreme manipulators of information. However, this is consistent with a model where regimes that have more discretionary resources more generally spend more on information manipulation as well as technologies that actually increase their chances of survival (e.g., transfers to elites, repression, public goods). So, we can observe a positive correlation between government survival and information manipulation even though in a sense information manipulation is harming the regime.

Another policy decision the model speaks to is the degree to allow free or foreign media. One way this can be incorporated is to give the citizen an additional unmanipulated signal of the politician performance. Preliminary analysis indicates this has the same effect as increasing the precision of the prior belief, which figure 3 shows can be helpful to the regime. This suggests a reason why even highly repressive governments may want to allow outside information or free media, which can lower incentives for counterproductive manipulation. Further, the contrast between the top and bottom panels of figure 3 suggests that the effect of more precise signals is helpful to the politician when citizens are more skeptical about his honesty.

The model could also be extended to a dynamic setting. In addition to learning about the politician performance over time, the citizen will also update his beliefs about his honesty. The fact that manipulating more today makes citizens more skeptical tomorrow may be a force which restrains politicians from lying too much. Still, as long as the returns to unexpected manipulation are higher than those to expected manipulation, the politician will lie too much, potentially "wasting" his credibility quickly even if this leads to a skeptical audience in the future.

References

Ashworth, Scott, Ethan Bueno De Mesquita. 2014. Is voter competence good for voters?: Information, rationality, and democratic performance. *American Political Science Review* **108**(3) 565–587.

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

Callander, Steven, Simon Wilkie. 2007. Lies, damned lies, and political campaigns. *Games and Economic Behavior* **60**(2) 262–286.

Chen, Jidong, Yiqing Xu. 2014. Information manipulation and reform in authoritarian regimes. Mansucript, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2487437.

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

Crawford, Vincent P, Joel Sobel. 1982. Strategic information transmission. *Econometrica: Journal of the Econometric Society* **50**(6) 1431–1451.

Dickson, Eric S. 2010. Leadership, followership, and beliefs about the world: An experiment. Manuscript, available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.114.7512&rep=rep1&type=pdf.

Edmond, Chris. 2013. Information manipulation, coordination, and regime change. *The Review of Economic Studies* **80**(4) 1422–1458.

Egorov, Georgy, Sergei Guriev, Konstantin Sonin. 2009. Why resource-poor dictators allow freer media: a theory and evidence from panel data. *American Political Science Review* **103**(4) 645–668.

Gehlbach, Scott, Alberto Simpser. 2015. Electoral manipulation as bureaucratic control. *American Journal of Political Science* **59**(1) 212–224.

Gehlbach, Scott, Konstantin Sonin. 2014. Government control of the media. *Journal of Public Economics* **118**(0) 163 – 171. doi:http://dx.doi.org/10.1016/j.jpubeco.2014.06.004.

Gehlbach, Scott, Konstantin Sonin, Milan Svolik. 2016. Formal models of nondemocratic politics.

Gentzkow, Matthew, Emir Kamenica. 2011. Bayesian persuasion. *American Economic Review* **101**(6) 2590–2615.

Guriev, Sergei M, Daniel Treisman. 2015. How modern dictators survive: Cooptation, censorship, propaganda, and repression. CEPR Discussion Paper No. DP10454.

Hollyer, James R, B Peter Rosendorff, James Raymond Vreeland. 2014. Why do autocrats disclose? APSA 2014 Annual Meeting Paper.

Hollyer, James R, B Peter Rosendorff, James Raymond Vreeland. 2015. Transparency, protest, and autocratic instability.

Holmstrom, Bengt. 1999. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies* **66**(1) 169–182.

Horz, Carlo M. 2017. Propaganda, cognitive constraints, and instruments of political survival. MPSA 2017 Annual Meeting Paper.

Huang, Haifeng. 2014. Propaganda as signaling. Forthcoming, Comparative Politics.

Kartik, Navin. 2009. Strategic communication with lying costs. *The Review of Economic Studies* **76**(4) 1359–1395.

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

Little, Andrew T. 2012. Elections, fraud, and election monitoring in the shadow of revolution. *Quarterly Journal of Political Science* **7**(3) 249–283.

Little, Andrew T. 2015. Fraud and monitoring in non-competitive elections. *Political Science Research and Methods* **3**(01) 21–41.

Little, Andrew T. 2017. Propaganda and credulity. *Games and Economic Behavior* **102** 224–232.

Lorentzen, Peter. 2014. China's strategic censorship. *American Journal of Political Science* **58**(2) 402–414.

Ottaviani, Marco, Francesco Squintani. 2006. Naive audience and communication bias. *International Journal of Game Theory* **35**(1) 129–150.

Patty, John W, Roberto A Weber. 2007. Letting the good times roll: A theory of voter inference and experimental evidence. *Public Choice* **130**(3-4) 293–310.

Petrova, Maria, Galina Zudenkova. 2015. Content and coordination censorship in authoritarian regimes. Manuscript, available at http://papers.sioe.org/paper/1406.html.

Royden, H. L., P. M. Fitzpatrick. 2010. *Real Analysis*. 4th ed. Pearson.

Rozenas, Arturas. 2015. Office insecurity and electoral manipulation. Forthcoming, Journal of Politics.

Saumard, Adrien, Jon A. Wellner. 2014. Log-concavity and strong log-concavity: A review. *Statist. Surv.* **8** 45–114.

Shadmehr, Mehdi, Dan Bernhardt. 2015. State censorship. *American Economic Journal: Microeconomics* **7**(2) 280–307. doi:10.1257/mic.20130221.

Simpser, Alberto. 2013. *Why Governments and Parties Manipulate Elections: Theory, Practice, and Implications*. Cambridge University Press.

Tao, Terrence. 2011. *An Introduction to Measure Theory*. American Mathematical Society, Graduate Studies in Mathematics Series, Vol. 125.

Wang, Joseph Tao-yi, Michael Spezio, Colin F. Camerer. 2010. Pinocchio's pupil: Using eye-tracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review* **100**(3) 984–1007. doi:10.1257/aer.100.3.984.

Zakharov, Alexei V. 2016. The loyalty-competence trade-off in dictatorships and outside options for subordinates. *The Journal of Politics* **78**(2) 457–466.

Appendix A Proofs

Proof of Proposition 2. A more precise statement of part i is that:

$$\left.\frac{d\overline{\pi}(m,m)}{dm}\right|_{m=0} = \overline{r}(0,0) + 0 \cdot \left.\frac{d\overline{r}(m,m)}{dm}\right|_{m=0} = 1 - q > 0.$$

Note that $r(\theta,0,0) = 1-q$, for all $\theta \Rightarrow \overline{r}(0,0) = \mathbb{E}_\theta[\overline{r}(\theta,0,0)] = 1-q$. So as long as $\left.\frac{d\overline{r}(m,m)}{dm}\right|_{m=0}$ is finite, the second term drops out, completing the proof. Differentiating $r(\theta,m,m)$ w.r.t. m yields

$$\frac{\partial r(\theta,m,m)}{\partial m} = \frac{q(1-q)f(\theta)f'(\theta+m)}{((1-q)f(\theta+m)+qf(\theta))^2} \Rightarrow \left.\frac{\partial r(\theta,m,m)}{\partial m}\right|_{m=0} = \frac{q(1-q)f'(\theta)}{f(\theta)}. \quad \text{(A.1)}$$

Thus, $\left.\frac{d\overline{r}(m,m)}{dm}\right|_{m=0} = \mathbb{E}_\theta\left[\left.\frac{\partial r(m,\theta)}{\partial m}\right|_{m=0}\right] = q(1-q)\int_\theta f'(\theta)d\theta$. Since $\lim_{\theta\to-\infty} f(\theta) = \lim_{\theta\to\infty} f(\theta) = 0$. So, $\left.\frac{d\overline{\pi}(m,m)}{dm}\right|_{m=0} = 1-q$.

For part ii, recall that $\pi(\theta,m,m) \equiv \frac{m\,(1-q)f(\theta+m)}{qf(\theta)+(1-q)f(\theta+m)}$, and

$$\pi(\theta,0,0) = 0, \qquad \text{and} \qquad \left.\frac{\partial\pi(\theta,m,m)}{\partial m}\right|_{m=0} = 1-q > 0. \quad \text{(A.2)}$$

First observe that since f is a proper density with a finite expectation, $\lim_{x\to\infty} xf(x) = 0$; which implies

$$\lim_{m\to\infty} \pi(\theta, m, m) = 0. \tag{A.3}$$

Define $G(\theta, m) \equiv \pi(\theta, m, m)\, f(\theta)$; $m^{max}(\theta) \equiv \arg\max_m \pi(\theta, m, m) = \arg\max_m G(\theta, m)$; and $G^{max}(\theta) \equiv G\left(\theta, m^{max}(\theta)\right)$. Taken together, Equations A.2 and A.3 imply that

$$m^{max}(\theta) \in (0, \infty),\ \forall \theta \in \mathbb{R} \qquad \Longrightarrow \qquad G^{max}(\theta) \in (0, \infty),\ \forall \theta \in \mathbb{R}. \tag{A.4}$$

The remainder of the proof relies on Lebesgue's Dominated Convergence Theorem (see Royden and Fitzpatrick (2010), §4.4; and Tao (2011), §1.4); thus, it is useful to represent $\pi(\theta, m, m)$ and $G(\theta, m)$ as two sequences of measurable functions, $\{\pi_m\}$ and $\{G_m\}$ on $\mathbb{R}$. From Equation A.3, it follows that $\int_{\mathbb{R}} \lim_{m\to\infty} G_m = 0$. Note that G^{max} dominates $\{G_m\}$ on $\mathbb{R}$, because, by definition, $|G_m| \leq G^{max}$, $\forall m$. Moreover, Equation A.4 establishes that G^{max} is (Lebesgue) integrable over $\mathbb{R}$. As such, Lebesgue's Dominated Convergence Theorem applies to $\{G_m\}$ such that

$$\lim_{m\to\infty} \bar{\pi}(m) = \lim_{m\to\infty} \int_{\mathbb{R}} G_m = \int_{\mathbb{R}} \lim_{m\to\infty} G_m = 0.$$

completing part ii. Part iii follows directly from parts i and ii. □

Proof of Lemma 1. The derivative of $\overline{r}(m)$ is $\frac{d\overline{r}(m,m)}{dm} = \int_\theta \frac{\partial r(\theta,m,m)}{\partial m} f(\theta) d\theta$. Direct substitution from A.1 yields

$$\frac{d\overline{r}(m,m)}{dm} = \int_\theta w(\theta, m) f'(\theta + m) d\theta, \quad \text{where} \quad w(\theta, m) \equiv \frac{(1-q)qf(\theta)^2}{(qf(\theta) + (1-q)f(\theta+m))^2}.$$

From inspection w is strictly positive, but $f'(\theta + m)$ can be positive or negative. In particular, since f is a log-concave density with full support on $\mathbb{R}$, there must exist a θ^* such that f' is positive for all $\theta < \theta^*$ and negative for all $\theta > \theta^*$. The idea of the proof is to show that more weight via $w(m, \theta)$ is placed on the negative part of f'. Taking the derivative of $w(m, \theta)$ w.r.t. θ:

$$\frac{\partial w}{\partial \theta} = \frac{2(1-q)^2 q f(\theta)(f(\theta) f'(m+\theta) - f(m+\theta) f'(\theta))}{(qf(\theta) + (1-q) f(m+\theta))^3},$$

which is positive if and only if

$$f(\theta) f'(m+\theta) > f(m+\theta) f'(\theta) \iff \frac{f'(m+\theta)}{f(m+\theta)} > \frac{f'(\theta)}{f(\theta)}.$$

A sufficient condition to ensure the above holds is if $\frac{f'(\theta)}{f(\theta)}$ is decreasing, or

$$\frac{f(\theta)f''(\theta)-f'(\theta)^2}{f(\theta)^2}>0,$$

which is true by log-concavity. Hence, w is strictly increasing in θ, which allows us to establish the following inequality:

$$\begin{aligned}\frac{d\bar{r}(m,m)}{dm} &= \int_{\theta=-\infty}^{\theta^*-m} w(\theta,m)f'(\theta+m)d\theta + \int_{\theta=\theta^*-m}^{\infty} w(\theta,m)f'(\theta+m)d\theta \\ &< w(\theta^*-m,m)\underbrace{\left(\int_{\theta=-\infty}^{\theta^*-m} f'(\theta+m)d\theta + \int_{\theta=\theta^*-m}^{\infty} f'(\theta+m)d\theta\right)}_{=\int_\theta f'(\theta+m)d\theta=0} = 0\end{aligned}$$

Which completes part i.

The proof of part ii proceeds along the following steps:

a. Let $g(\cdot)\equiv\log f(\cdot)$, then $g'''\leq 0$ is a sufficient condition to guarantee that $r(\theta,m,m)$ is log concave in both m and θ.

b. Log-concavity is preserved by marginalization (see Saumard and Wellner (2014), §3.1.3). That is, if $r(\theta,m,m)$ is log concave in both m and θ, then $\bar{r}(m,m)=\int_{-\infty}^{\infty} r(\theta,m,m)\,f(\theta)\,d\theta$ must be log concave in m.

c. Log-concavity is preserved by products (see Saumard and Wellner (2014), §3.1.2). That is, if $\bar{r}(m,m)$ is log concave in m, then $\overline{\pi}(m,m)\equiv m\,\bar{r}(m,m)$ must also be log concave in m.

d. $\overline{\pi}(m,m)$ is a continuous function and $m^{opt}\equiv\arg\max_m \overline{\pi}(m,m)\in(0,\infty)$ (see part iii of Proposition 2). Therefore, if $\overline{\pi}(m,m)$ is log concave, then m^{opt} must be unique and defined by the F.O.C.

Proving step (a) completes the proof as the remaining steps directly follow. We start by assuming that $\frac{\partial^3\log(f(\theta))}{\partial\theta^3}\leq 0$ and then show below that $\frac{\partial^2\log(r(\theta,m,m))}{\partial m^2}\leq 0$ and $\frac{\partial^2\log(r(\theta,m,m))}{\partial\theta^2}\leq 0$ must follow. Differentiating twice w.r.t. m yields

$$\frac{\partial^2\log(r(\theta,m,m))}{\partial m^2}=\frac{-q(1-q)f(\theta)f(\theta+m)A}{f(\theta+m)^2(qf(\theta)+(1-q)f(\theta+m))^2}, \tag{A.5}$$

where $A\equiv(2-qf(\theta))f'(m+\theta)^2-(1-qf(\theta))f(m+\theta)f''(m+\theta)$.

Log-concavity of f implies $f'(m+\theta)^2 > f(m+\theta)f''(m+\theta)$, and $(2-qf(\theta)) > (1-qf(\theta))$, so $A > 0$ and hence $\frac{\partial^2 \log(r(\theta,m,m))}{\partial m^2} < 0$. Differentiating twice w.r.t. θ yields

$$\frac{\partial^2 \log(r(\theta, m, m))}{\partial \theta^2} = \frac{qB}{f(\theta+m)^2((1-q)f(\theta+m)+qf(\theta))^2}, \text{ where} \tag{A.6}$$

$$B \equiv f(\theta)f(\theta+m)\left(qf(\theta)f''(\theta+m) - 2(1-q)f'(\theta+m)^2\right) - (1-q)f''(\theta)f(\theta+m)^3 - qf(\theta)^2 f'(\theta+m)^2 \\ + f(\theta+m)^2\left(f(\theta)\left((1-q)f''(\theta+m) - qf''(\theta)\right) + 2(1-q)f'(\theta)f'(\theta+m) + qf'(\theta)^2\right).$$

Let B_1 and B_0 denote B evaluated at $q = 1$ and at $q = 0$. Simplifying these expressions gives:

$$B_1 = f(\theta+m)^2\left(f'(\theta)^2 - f(\theta)f''(\theta)\right) - f(\theta)^2\left(f'(\theta+m)^2 - f(\theta+m)f''(\theta+m)\right), \text{and}$$
$$B_0 = f(\theta+m)\left(2f'(\theta)f(\theta+m)f'(\theta+m) - f''(\theta)f(\theta+m)^2 - f(\theta)\left(2f'(\theta+m)^2 - f(\theta+m)f''(\theta+m)\right)\right).$$

Since B is linear in q, then, taken together, $B_1 \leq 0$ and $B_0 \leq 0$ imply that $B \leq 0$. Starting with B_1, we can sign this by writing it as:

$$B_1 = f(\theta)^2 f(\theta+m)^2\left(\frac{f'(\theta)^2 - f''(\theta)}{f(\theta)^2} - \frac{f'(\theta+m)^2 - f''(\theta+m)}{f(\theta+m)^2}\right),$$

i.e., a strictly positive term times $\frac{\partial^2 \log f(x)}{\partial^2 x}\Big|_{x=\theta} - \frac{\partial^2 \log f(x)}{\partial^2 x}\Big|_{x=\theta+m}$. So if $\frac{\partial^2 \log f(\theta)}{\partial^2 \theta}$ is decreasing in θ, then B_1 is negative, which is ensured by the $\dfrac{\partial^3 \log(f(\theta))}{\partial \theta^3} \leq 0$ assumption.
Next, we can write B_0 as follows:

$$B_0 = f(\theta)f(\theta+m)^3\left(\frac{f'(\theta)^2 - f(\theta)f''(\theta)}{f(\theta)^2} - \frac{f'(\theta+m)^2 - f(\theta+m)f''(\theta+m)}{f(\theta+m)^2} \right. \\ \left. - \frac{(f'(\theta)f(\theta+m) - f(\theta)f'(\theta+m))^2}{f(\theta)^2 f(\theta+m)^2}\right).$$

The first two terms of the parenthetical are again equal to $\frac{\partial^2 \log f(x)}{\partial^2 x}\Big|_{x=\theta} - \frac{\partial^2 \log f(x)}{\partial^2 x}\Big|_{x=\theta+m}$, which is less than or equal to zero. The third term in the parenthetical is also negative by log-concavity of f. So $B_1 \leq 0$.

Finally, from Equation A.6, $B \leq 0 \Rightarrow \dfrac{\partial^2 \log(r(\theta, m, m))}{\partial \theta^2} \leq 0$, i.e., $r(\theta, m, m)$ is log concave in θ. □

Proof of Proposition 4 For a given citizen conjecture, $\hat{m}$, the politician best response is:

$$m^{br}(\hat{m}) = \arg\max_m \bar{U}_p(m, \hat{m}) = \arg\max_m \mathbb{E}[\theta] + \overline{\pi}(m, \hat{m}) - c(m).$$

The objective function is strictly concave in m if $\hat{m}\frac{\partial^2 \bar{r}(m, \hat{m})}{\partial m^2} < \kappa c_0''(m)$. The right-hand side increases without bound in κ,and so when this parameter is sufficiently large the inequality holds.

As $\lambda \to \infty$, $r(\theta, m, \hat{m}) = 1 - q$ for any θ, m, and $\hat{m}$. Further, r is continuous in all arguments, and so:

$$\lim_{\lambda \to \infty} \hat{m}\frac{\partial^2 \bar{r}(m, \hat{m})}{\partial m^2} = \hat{m}\frac{\partial^2 \lim_{\lambda \to \infty} \bar{r}(m, \hat{m})}{\partial m^2} = 0$$

So, for sufficiently large κ or λ, the politician's objective function is strictly concave and his *unique* best response function, $m^{br}(\hat{m})$, follows from the following F.O.C.:

$$\left.\frac{\partial \bar{U}_p(m, \hat{m})}{\partial m}\right|_{m=m^{br}(\hat{m})} = 1 + \hat{m}\left.\frac{\partial \bar{r}(m, \hat{m})}{\partial m}\right|_{m=m^{br}(\hat{m})} - \kappa c_0'(m^{br}(\hat{m})) = 0. \tag{A.7}$$

Since the citizen rationally expects the true level of manipulation, then we must have $\hat{m} = m$ in any pure strategy equilibrium. In other words, a pure strategy PBE will exist only if the FOC in (A.7) crosses the 45 degree line (i.e., $m = \hat{m}$). From (A.7), $c'(m^{br}(0)) = 1 \Rightarrow m^{br}(0) > 0$; thus, if $m^{br}(\hat{m})$ never increases at a faster rate than 1, then it must cross the $m = \hat{m}$ only once and the unique equilibrium manipulation level is defined by the F.O.C. in Equation 9. From Implicit Function Theorem, the slope of the best response function is

$$\frac{\partial m^{br}(\hat{m})}{\partial \hat{m}} = \left.\frac{\left(\frac{\partial \bar{r}(m,\hat{m})}{\partial m} + \hat{m}\frac{\partial^2 \bar{r}(m,\hat{m})}{\partial m \partial \hat{m}}\right)}{\left(\kappa c_0''(m) - \hat{m}\frac{\partial^2 \bar{r}(m,\hat{m})}{\partial m^2}\right)}\right|_{m=m^{br}(\hat{m})}. \tag{A.8}$$

By the same arguments as above, if κ or λ are sufficiently large, the denominator in (A.8) is positive and the numerator goes to zero; hence, $\lim_{\lambda \to \infty} \frac{\partial m^{br}(\hat{m})}{\partial \hat{m}} = 0$.

□

Proof of proposition 6 Part i follows from implicitly differentiating the first order condition for the optimal manipulation level.

For part ii, it is shown in the main text that the equilibrium level is m^{CC} for both $q = 0$ and $q = 1$. Further, since $m\frac{\partial \bar{r}(m,m)}{\partial m} < 1$ for $q \in (0, 1)$, the right hand side of the equilibrium condition is strictly less than 1 for this range of q. so, $m^* < m^{CC}$ for $q \in (0, 1)$. Further, the equilibrium manipulation level is continuous in q. So, it must be decreasing for q close to 0 and increasing for q close to 1. □

Proof of proposition 7 The first order conditions for the equilibrium and optimal manipulation levels given this transformation are now:

$$kc_0'(m^{opt}) = \alpha m^{opt} \left(\frac{\partial \overline{r}(m, \hat{m})}{\partial m}\bigg|_{\hat{m}=m=m^{opt}} + \frac{\partial \overline{r}(m, \hat{m})}{\partial \hat{m}}\bigg|_{\hat{m}=m=m^{opt}} \right) + \alpha \overline{r}(m, \hat{m} = m)., \quad \text{(A.9)}$$

$$kc_0'(m^*) = \alpha + \alpha m^* \frac{\partial \overline{r}(m, \hat{m})}{\partial m}\bigg|_{\hat{m}=m=m^*} \quad \text{(A.10)}$$

Part i follows from implicitly differentiating these equations

For part ii, the limiting behavior is immediate from the equilibrium conditions. By proposition 2, the manipulation boost is increasing for $m < m_0^{\text{opt}}$ and decreasing for $m > m_0^{\text{opt}}$ and the equilibrium choice is increasing (and with range $\mathbb{R}_+$) in m, which gives the second claim. □

Politician (Partially) Informed about Performance

If the politician knows θ, then their strategy is a mapping from θ and ω to a manipulation level. Write the manipulation level $m(\theta)$.

The equilibrium signal when $\omega = 1$ as a function of θ is then:

$$s_1(\theta) = \theta + m(\theta)$$

As long as m is bounded – a reasonable presumption with a convex cost function where c' increases without bound, s_1 will have full support on $\mathbb{R}$. Since the signal distribution when $\omega = 0$ also has full support on $\mathbb{R}$.

To simplify, suppose $s_1(\theta)$ is continuous and monotone, which is guaranteed if $m(\theta)$ is continuous with $m'(\theta) > -1$. This implies s_0 and s_1 both admit a density and the posterior belief about ω when anticipating manipulation strategy $\hat{m}(\theta)$ is given by:

$$Pr(\omega = 0 | s, \hat{m}(\theta)) = \frac{(1-q)f(s)}{qf(\theta^{-1}(s)) + (1-q)f(s)}$$

where $\theta^{-1}(s; \hat{m})$ is the (unique) solution to $s = \theta + \hat{m}(\theta)$. The optimal manipulation level is the function that solves:

$$\arg\max_{m(\theta)} \mathbb{E}_\theta[\theta + r(m(\theta), m(\theta), \theta)m(\theta) - c(m(\theta))]$$

where:

$$r(m, m(\theta), \theta) = \frac{(1-q)f(\theta)}{(1-q)f(\theta) + qf(\theta^{-1}(\theta + m; m(\theta)))}$$

This is a hard functional analysis problem.

The equilibrium condition is that for all θ:

$$m(\theta) \in \arg\max_{m} \theta + m - (1 - r(m, m(\theta), \theta))m - c(m)$$

The first order condition at each θ is then:

$$c'(m) = 1 + r(m, m(\theta), \theta) + m\frac{\partial r}{\partial m}$$

where:

$$\frac{\partial r}{\partial m} = \frac{(1-q)f(\theta)qf'(\theta^{-1}(q + m, m(\theta)))\frac{\partial \theta_{-1}}{\partial m}}{(1-q)f(\theta) + qf(\theta^{-1}(\theta + m; m(\theta)))^2}$$

This is hard differential equation.

Full and No Credulity With full credulity $(q = 0)$, the objective function for the optimal manipulation becomes

$$\arg\max_{m(\theta)} \mathbb{E}_\theta[\theta + m(\theta) - c(m(\theta))]$$

which is maximized by $m(\theta) = m^{CC}$ for all m. The same holds for the equilibrium choice.

With no credulity, the objective function for optimal manipulation becomes:

$$\arg\max_{m(\theta)} \mathbb{E}_\theta[\theta - c(m(\theta))]$$

which is clearly maximized by $m(\theta) = 0$ for all θ.

The equilibrium choice becomes:

$$m(\theta) \in \arg\max_{m} \theta + m - c(m)$$

which is solved by $m(\theta) = m^{CC}$ for all θ.