Unbelievable Lies*

Andrew T. Little[†]   Sherif Nasser[‡]

July 2, 2018

Abstract

People often tell unbelievable lies. But why bother telling lies which aren't believed? We develop a formal model to address this and related questions, with an emphasis on lies told by politicians. 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 would be optimal. This tension is particularly acute for politicians that care a great deal about perceptions of their performance, who are unable to tell persuasive lies.

*Many thanks to Abraham Aldama, Dimitri Landa, Jay Lyall, Andrew Marshall, Jack Paine, and audiences at APSA 2016, SPSA 2017, APSA 2017, NYU, Rochester, and Stanford for comments and discussion.

[†]Department of Political Science, UC Berkeley.

[‡]SC Johnson College of Business, Cornell University.

Lying pervades social and economic interactions. Parents mislead children to nudge them toward desired actions (and vice versa), job applicants exaggerate on resumes, and once hired employees frequently face incentives to lie about various aspects of their work. Much of advertising is about making a company's products seem more appealing than they truly are.

One domain particularly associated with lying – perhaps even more so recently – is politics. To pick some particularly colorful examples, Syria's Bashar al-Assad claimed to be Syria's premier pharmacist, Saparmurat Niyazov of Turkmenistan boasted that he made an agreement with God stipulating that anyone reading his book the Ruhnama three times would be guaranteed entry to heaven, and Donald Trump had his press secretary declare that his inauguration crowd was the largest ever despite clear photographic evidence to the contrary. These examples are notable in part because the lies seem at best tangentially related to the leaders' political abilities and performance, but of course many lies are about consequential information like the success of the economy, the wisdom of a proposed foreign policy, or the degree of corruption in the government.

Extreme lies can be counterproductive. Take the political context, which we use as our main example. 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 "manipulate" the signal (i.e., distort it upward). The citizen may be 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 (alternatively, "less credulous" or "less trusting"); eventually, she becomes nearly certain that the signal is manipulated. Our first main result shows that, under very general distributional assumptions, the degree of manipulation which maximizes the average belief about the politician performance is always finite. Beyond a certain point, more extreme lies become less persuasive.

When the degree of manipulation is a hidden action taken by the politician, he tells a more extreme lie than would be optimal, even when the cost of manipulation is higher for more extreme lies. 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 several sufficient conditions for this to hold), the politician lies more than he should. Put another way, if the citizen expects the politician to tell the optimal lie, he will find it beneficial to lie a little more.

Next, 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.

Finally, we explore an extension where there are multiple citizens who are heterogeneous in their initial skepticism of the politician. Our main result here is that the politician tells more extreme lies when the audience is mostly comprised of "extremists" who are initially very credulous/trusting or very skeptical. When there are more moderate citizens who are more apt to change their belief about whether the politician is lying, he is more restrained. An implication of this result is that polarization in *trust* can lead to more polarization in beliefs about the politician *performance*.

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. Political lies are a common application, in the context of campaigns (Callander and Wilkie, 2007; Dziuda and Salas, 2017), propaganda (Egorov et al., 2009; Edmond, 2013; Guriev and Treisman, 2015; Horz, 2017), or distorting economic data (Hollyer et al., 2015). 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).

Why lie if the audience is aware that statements might be untruthful and adjusts how they respond accordingly? Regardless of the specific technology or context, information manipulation occurs using standard equilibrium concepts 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. For example, if some senders have aligned incentives with the receiver while others want her to make a "bad" decision, the misaligned types can leverage the receivers belief that they may be a good type to mislead them (e.g., Sobel, 1985). 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, 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, 2016). 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).[2]

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).[3] Unlike these models, where receivers are either fully credulous for fully "Bayesian", we generally focus on "partially credulous" citizens who update their beliefs about whether they are being lied to based on what they hear. Little (2017) uses this updating technology, but focuses on how the presence of credulous citizens affects the behavior of those who know the government is manipulating information. Here we abstract from interactions among citizens to focus more on the decisions of the politician.[4] Most importantly,

[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 (Chen and Xu, 2014; Petrova and Zudenkova, 2015; Guriev and Treisman, 2015; Dziuda and Salas, 2017). Edmond (2013) includes all three dynamics: the government has private information, propaganda is a hidden action which jams the signal observed by citizens.

[2]Though see also Woon (2017) who finds that observers of political lies do become more skeptical of claims rated as more false by neutral observers, as is true in our model.

[3]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 Bueno De Mesquita (2014) also analyze a model where voters do not correctly "filter" a signal of the government performance, albeit with a substantially different technology and purpose.

[4]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

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.[5]

2 The Model

We analyze a game between a sender and a receiver. To fix ideas, we primarily analyze the example where the sender is a politician (pronoun "he") and a receiver is 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 $m \geq 0$ representing 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. Lower values of q correspond to a citizen who is more credulous or trusting; and higher values to being more skeptical about the politician at the outset.

The citizen and politician share a prior belief that θ is drawn from a probability distribution with density $f(\cdot)$. Our most general results only require the following assumptions on this distribution:

Assumption f1. *f is continuous, differentiable, strictly positive on $\mathbb{R}$, and has a finite expectation.*

equilibrium choice. However, this paper does not define or characterize the optimal level of manipulation, which is central to the main results here.

[5] 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 **f**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, as long as the citizen is uncertain about the politician's type in the prior, she will remain uncertain about his 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.

We assume the politician cares about the citizen's posterior expectation of θ. Writing this as a function of s and $\hat{m}$:

$$\hat{\theta} \equiv \mathbb{E}\left[\theta \mid s, \hat{m}\right] = s - (1 - Pr(\omega = 0|s, \hat{m})) \cdot \hat{m}.$$

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:

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

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:

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

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

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 solution concepts where the citizen forms rational expectations (i.e., $\hat{m} = m$), this will drop out in equilibrium. 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.[6] Except when

[6] 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

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

Assumption c1. $c(\cdot)$ *is continuous, twice differentiable, strictly increasing, and convex, with* $c(0) = c'(0) = 0$ *and* $\lim_{m \to \infty} c'(m) = \infty$.

This cost can capture any way manipulation harms the sender *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 the political context, the politician may have to compensate subordinates to manipulate information in his favor, or hire less competent subordinates willing to lie (Zakharov, 2016).

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

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

The citizen and honest politician take no actions, and so we do not need to specify a payoff function for them.[7]

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, $\hat{m}$.

---

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.

[7]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 6.

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 maximizes the manipulative politician payoff, subject to the constraint that the citizen conjecture is correct (i.e., $\hat{m} = m$). 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:[8]

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

In principle, there can be multiple solutions to this optimization problem (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. Further, the non-manipulative politician takes no action. So, when characterizing the Perfect Bayesian Equilibrium to the model, the only requirement is that the manipulation level m^* maximizes the politician payoff when the citizen expects m^*.

Definition The set of equilibrium levels of manipulation are the manipulation levels m such that:

$$M^* = \left\{ m : m \in \arg\max_m \bar{U}_p(m, \hat{m}) \Big|_{\hat{m}=m} \right\} \tag{6}$$

Adopting a similar convention to the optimal manipulation level(s), let $m^* = \max M^*$ (when

[8] We could write this more concisely as $M^{\text{opt}} = \arg\max_m \bar{U}_p(m, m)$, but the formulation in (5) makes the contrast with the equilibrium definition in (6) clearer.

M^* is non-empty, i.e., an equilibrium exists). We discuss the existence and uniqueness of the equilibrium manipulation level in section 5.

Another way to think about the equilibrium level of manipulation is to first define the politician's best response level of manipulation for a given citizen conjecture $\hat{m}$ as

$$m^{br}(\hat{m}) \equiv \arg\max_{m} \bar{U}_p(m, \hat{m}).$$

Since the citizen forms a rational expectation of the equilibrium manipulation, then an equilibrium manipulation level is characterized by $m^* = m^{br}(m^*)$. That is, when the citizen expects the equilibrium manipulation level m^*, the politician's best response is to choose m^*. The difference between optimal and equilibrium levels of manipulation is driven by that fact that (under conditions derived later) $m^{opt} < m^{br}(m^{opt})$. That is, m^{opt} cannot be the equilibrium level of manipulation because when the citizen expects the optimal manipulation level m^{opt}, the politician's best response is to manipulate even more.

Comments on setup. There are several ways one could define the optimal manipulation level, with different treatments of the citizen expectation and what it means to be "credulous". A standard interpretation is to assume that the citizen is correct in thinking the politician only lies with probability q – perhaps due to heterogeneity in the cost of lying – and the optimal manipulation level as we define it corresponds to what is optimal for the manipulative ($\omega = 1$) type.

An alternative interpretation is that all politicians are in fact manipulative, and so credulous citizens are incorrect in thinking the signal might be true. In this case, what we call optimal applies to all politicians, subject to the constraint that citizens are correct in their estimation of how manipulative politicians behave, and are just incorrect in thinking that honest politicians exist.[9]

[9]To see why this restriction is important, suppose the citizen conjecture is 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.

For either interpretation, an important assumption is that the politician sets the manipulation level while being uncertain about his actual performance. Symmetric uncertainty is common in career concerns-style models (following Holmstrom 1999), and greatly simplifies the analysis. The model corresponds well to the type of politicians' (and others') decisions, which affect signals of their performance, and 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.

There are other situations where the politician has some private information about his performance and may report the truth or lie. A natural model here would be to have the politician 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. 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 become a substantially more complicated problems. 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 $q = 1$ or $q = 0$. The former means the citizen knows for certain that the politician manipulates (i.e., is "fully skeptical"). The latter means the citizen knows for sure that the politician does not manipulate (i.e., is "fully credulous" or "fully trusting").

No Credulity. Suppose the citizen knows the politician manipulates. This becomes a standard filtering problem. In particular, $r(\theta, m, \hat{m}) = 0$ for any m (see Equation 2), so the politician payoff (see Equation 4) 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

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

Since an *unexpected* increase in manipulation leads to a one unit increase in the citizen performance assessment, the equilibrium m^* is given by the point where this one unit increase matches the marginal cost. By assumption **c**1, such a solution exists and is unique. As it will recur throughout the analysis (and coincides with a standard career concerns benchmark), we define the manipulation level which solves $c'(m) = 1$ to be m^{CC}. Manipulation is completely ineffective in this case, but the politician still manipulates in equilibrium, ($m^* = m^{\text{CC}} > 0$).

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, $m^{opt}\big|_{q=1} = 0$.

Full Credulity. If the citizen is fully credulous ($q = 0$), then $r(\theta, m, \hat{m})\big|_{q=0} = 1$ for any m, so the expected utility function becomes:

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

Since $\hat{m}$ drops out, the first order condition for both the equilibrium and optimal manipulation levels is $c'(m) = 1$, which is solved by m^{CC}. So, in this case there is no difference between the equilibrium choice and what is optimal. Summarizing:

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

So, when faced with a fully credulous citizen, the politician sets the optimal manipulation level in equilibrium. 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 analysis 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. We start with another instructive special case: when manipulation is free. Setting $c(m) = 0$ and $\hat{m} = m$, 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 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 whenever 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. The marginal benefit from increasing the manipulation level (both actual and expected) is:

$$\text{MB}^{\text{exp}}(m) = \frac{d\overline{\pi}\,(m,m)}{dm} = \overline{r}(m,m) + m\frac{\partial \overline{r}(m,m)}{\partial m}. \tag{7}$$

The superscript in $\text{MB}^{\text{exp}}(m)$ highlights that this is the marginal boost to an *expected* increase in m, which will contrast with an unexpected increase analyzed below. 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 main result is that because of this trade-off, the distortion-maximizing manipulation level is always strictly positive but finite:

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

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. The optimal level of free manipulation is finite: $m_0^{opt} \in (0,\infty)$.

Proof. See the appendix.

Part i states that there is always a return to small degrees of manipulation, i.e., (7) is always positive at $m = 0$. This is because when the manipulation level is small, the endogenous cost to lying more is also small: since the citizen expects manipulative politicians to not change the signal much, increasing the belief that the politician is manipulative has little effect on her beliefs about θ.

Part ii formalizes the notion that as manipulation becomes very extreme, it becomes completely ineffective. An intuition for why the manipulation boost starts decreasing when m is high is that higher m means there is a larger endogenous cost to lying more. If the citizen thinks manipulative politicians change the signal to a great degree, even small changes in the belief about whether he is manipulative can lead to much lower citizen belief about the politician performance.

The manipulation boost eventually goes to zero because even a citizen who is very credulous (but not fully credulous, i.e., q is 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 level of free manipulation is finite, it is trivial that the optimal level of costly manipulation is finite as well:

Proposition 2. *For any prior density $f(\cdot)$ meeting assumption* **f***1,* $m^{opt} \in [0, m_0^{opt})$

Proof. Follows immediately from proposition 1 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 **f**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:

Assumption f2. *$f(\cdot)$ is logarithmically concave*

Assumption f3. *$g''' \leq 0$, where $g(\cdot) \equiv \log f(\cdot)$*

Both assumptions **f**2 and **f**3 are met by many standard distributions such as the normal and extreme value distribution.[10] It is possible for assumption **f**2 to hold but not assumption **f**3, though

[10] It does not hold for Student-t distributions, though instructively, simulations suggest that the results to come hold under this family of distributions as well. At the least, Assumptions **f**2-**f**3 are sufficient but not necessary for subsequent results.

we are unaware of any standard distribution where this is the case.

With this additional structure on f we obtain the following:

Proposition 3. *i. Given assumptions* $\mathtt{f}1$ *and* $\mathtt{f}2$*, 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 $\mathtt{f}1$*,* $\mathtt{f}2$*, and* $\mathtt{f}3$*,* $\overline{\pi}(m, m)$ *is single-peaked; and the unique* m^{opt} *is characterized by:*

$$\mathrm{MB}^{exp}(m^{opt}) = c'(m^{opt}), \tag{8}$$

Proof. See the appendix

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.[11] 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 $\mathtt{f}1$, $\mathtt{f}2$, and $\mathtt{f}3$.

5 Equilibrium Manipulation

We now characterize the equilibrium manipulation level. While the optimal manipulation level was characterized by the point where the boost from an expected increase in m meets the marginal cost, the equilibrium is determined by the point where the boost from an unexpected increase meets the marginal cost. Formally, let:

$$\mathrm{MB}^{\text{unexp}}(m) = \left.\frac{\partial\overline{\pi}(m,\hat{m})}{\partial m}\right|_{\hat{m}=m} = 1 + m\left.\frac{\partial\overline{r}(m,\hat{m})}{\partial m}\right|_{\hat{m}=m} \tag{9}$$

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

The marginal change in the average politician utility for picking a slightly higher manipulation level when the citizen expects m is then $\text{MB}^{\text{unexp}}(m) - c'(m)$. So, the first order condition for an interior manipulation level is an m^* which solves:

$$\text{MB}^{\text{unexp}}(m^*) = c'(m^*). \tag{10}$$

In characterizing the solution(s) to (10), two potential technical challenges arise. First, $\text{MB}^{\text{unexp}}(m)$ can be increasing in m, which can result in multiple solutions to the first order condition. Second, and more problematic, the politician optimization problem is not globally concave, and so the solution(s) may not correspond to a *global* maximizer.

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 diffuse (flat), then the $m \left.\frac{\partial \overline{r}(m,\hat{m})}{\partial m}\right|_{\hat{m}=m}$ term is sufficiently small, which removes the potential for this term to add enough convexity to the objective function. Intuitively, if the prior is very diffuse, 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 $f(\theta) = \lambda^{-1} f_0(\lambda\theta)$. If κ or λ are sufficiently large, then there is a unique equilibrium m^* characterized by (10).*

Proof. See the appendix

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). Either of the following two assumptions is sufficient:

Assumption f4. *λ is sufficiently large that the politician objective function is globally concave for all $\hat{m}$.*

Assumption c2. *κ is sufficiently large that the politician objective function is globally concave for all $\hat{m}$.*

Now we are ready to compare the optimal versus equilibrium manipulation levels. Recall the equilibrium manipulation level is the m which solves $c'(m) = \text{MB}^{\text{unexp}}(m)$ (see equations 9) and the optimal is the level of m which solves $c'(m) = \text{MB}^{\text{exp}}(m)$ (see equation 7). Under assumption f4 or c2, both have a unique solution. Further, since the left-hand sides of these equations are $c'(m)$ (which is increasing), a sufficient condition for the equilibrium manipulation level to higher than optimal is if $\text{MB}^{\text{unexp}}(m) > \text{MB}^{\text{exp}}(m)$ at $m = m^{\text{opt}}$ (or $m = m^*$). Intuitively, if this inequality holds, then the marginal gain to manipulating more than expected when the citizen expects m^{opt} outweighs the marginal cost, and so the best response must be to choose a higher manipulation level.

In general, $\text{MB}^{\text{unexp}}(m) > \text{MB}^{\text{exp}}(m)$ when:

$$1 + m \left.\frac{\partial \overline{r}(m, \hat{m})}{\partial m}\right|_{\hat{m}=m} > \overline{r}(m, m) + m\frac{\partial \overline{r}(m, m)}{\partial m}.$$

Substituting $\frac{\partial \overline{r}(m,m)}{\partial m} = \left.\frac{\partial \overline{r}(m,\hat{m})}{\partial m}\right|_{\hat{m}=m} + \left.\frac{\partial \overline{r}(m,\hat{m})}{\partial \hat{m}}\right|_{\hat{m}=m}$ above and rearranging yields

$$1 - \overline{r}(m, m) > m \cdot \left.\frac{\partial \overline{r}(m, \hat{m})}{\partial \hat{m}}\right|_{\hat{m}=m}. \tag{11}$$

The left-hand side of (11) 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 of (11) 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 11. However, we can prove that any of three unrelated conditions are sufficient for (11) to hold for all m and hence the equilibrium manipulation is higher than optimal:

Proposition 5. *Any of the following conditions is sufficient to ensure that the equilibrium manipulation level is higher than the optimal level ($m^* > m^{opt}$) when $q \in (0, 1]$:*

(1) The cost of manipulation is sufficiently high (κ is large),

(2) the prior distribution is sufficiently diffuse (λ is large), or

(3) q is sufficiently high.

Proof. See the appendix.

Condition 1 is sufficient because, when manipulation is very costly, then both m^* and m^{opt} become very small, so the left-hand side of (11) goes to 0 while the right-hand side remains strictly positive (in fact, it approaches q). Similarly, if the prior is very diffuse, then the $\left.\frac{\partial \overline{r}(m,\hat{m})}{\partial \hat{m}}\right|_{\hat{m}=m}$ term goes to zero, while the left-hand side or (11) approaches q. For condition 3, it is immediate that when $q = 1$ the right-hand side of (11) is equal to 1 while the left-hand side is equal to 0, so the inequality must hold for q sufficiently close to 1 by continuity. (Further, extensive numerical simulations have not uncovered any parameterizations where the result does not hold.)

Discussion 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. Put another way, proposition 5 provides an explanation for a phenomenon that seems common in politics and elsewhere: lies become unbelievable and have a minimal effect on beliefs, even though a more modest 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. Next we present comparative static results which highlight when the difference between equilibrium and optimal manipulation levels are particularly large.

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), the performance prior (f), and the cost function.

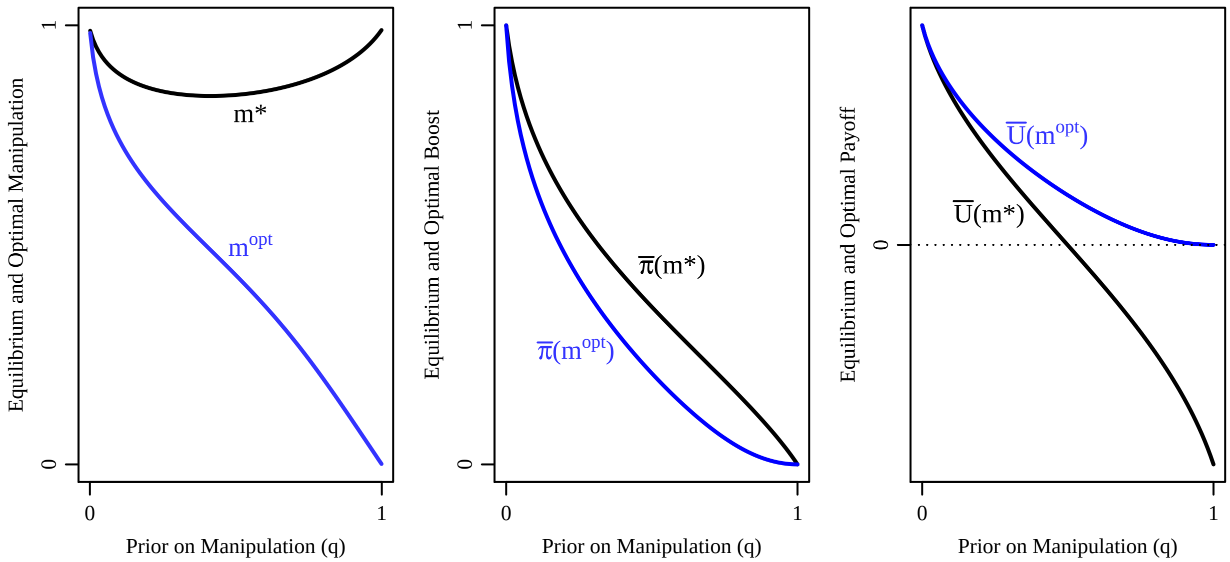
In this section we maintain assumptions $\mathbf{f}1$-$\mathbf{f}3$ and $\mathbf{f}4$ or $\mathbf{c}2$, which ensures (8) and (10) have unique solutions m^{opt} and m^* which are global maximizers. So all comparative statics are obtained by implicitly differentiating these equations.

We first examine how changing q affects the optimal and equilibrium manipulation levels. As discussed in the full and no credulity benchmarks, 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 $q = 1$, there is no endogenous cost to manipulation, so the equilibrium condition becomes $c'(m) = 1$, which is solved my m^{CC}. When q is intermediate, there is always endogenous 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*

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



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 equilibrium manipulation level (black curve) is only decreasing for small q, and for larger q is increasing.

The middle panel shows the manipulation boost when picking the equilibrium and optimal manipulation levels (and the citizen has a correct conjecture about the manipulation level). For these parameters, the boost with optimal manipulation is *lower* than the equilibrium boost. This is because the optimal manipulation level must account for the exogenous cost. To pick a single point, at $q = 0.75$, the optimal manipulation level is around 0.19 and the equilibrium manipulation level is much higher at 0.85. However, the much higher equilibrium manipulation level only translates into a small average manipulation boost of 0.19, while the optimal manipulation boost is 0.06.

Since the politician does a lot more lying without much getting much out of it, he would be better off accepting the small boost of the optimal level.

The right panel compares the utility when choosing the equilibrium and optimal manipulation level. By definition the utility at the optimal level is higher. This difference is particularly stark for higher levels of q. In fact, 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, we add a scale parameter α to how much the politician cares about perceptions of his performance – call this his "neediness". The expected utility function can be written and then normalized as:

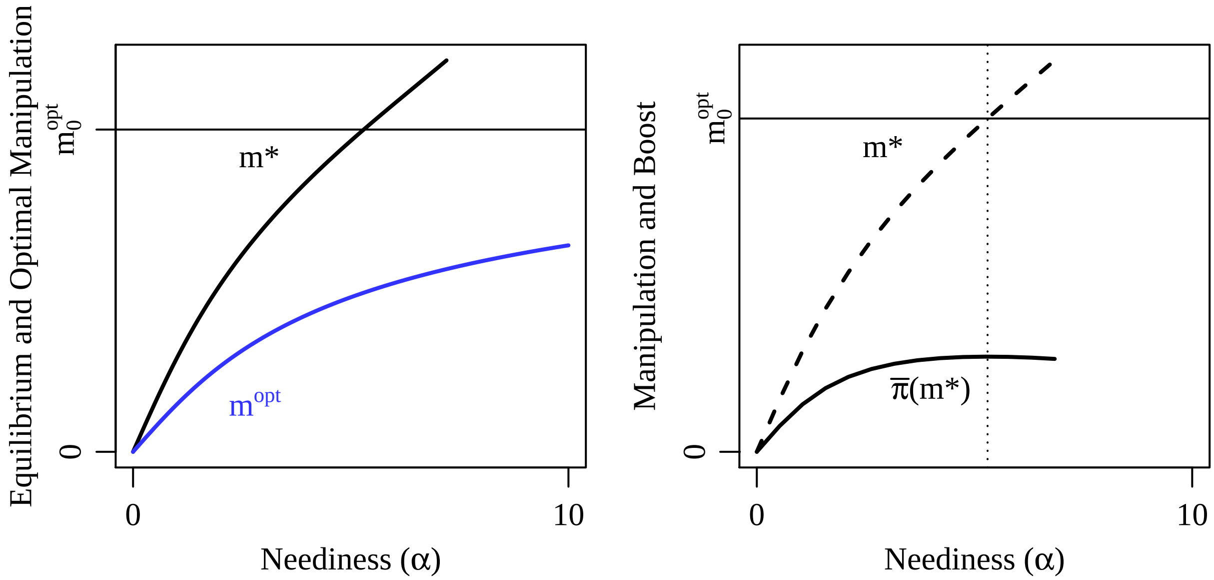
$$U_p(m;\hat{\theta}) = \alpha\hat{\theta} - \kappa c_0(m)$$
$$U_p(m,\hat{\theta})/\alpha = \hat{\theta} - \frac{\kappa}{\alpha}c_0(m)$$

Since maximizing $U_p(\cdot)/\alpha$ with respect to m is the same as maximizing $U_p(\cdot)$ with respect to m, for the purposes of characterizing the equilibrium and optimal manipulation levels changes in κ and α only matter through how they change $\frac{\kappa}{\alpha}$. That is, what matters is the ratio of the costliness of manipulation to how much 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}. So, if the neediness of the politician drives him to lie at a higher level than m_0^{opt}, further increases in m become counterproductive (even setting aside the exogenous cost):

Proposition 7. *Write the politician objective function as $\alpha\hat{\theta} - \kappa c_0(m)$. Where the conditions for*

Figure 2: Equilibrium properties as a function of α



an 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 $m > m_0^{opt}$

Proof. See the appendix

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 citizen's 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 (now a dashed curve) 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 solid curve). However, once m^* gets above m_0^{opt} (the dotted vertical line), the equilibrium manipulation boost (solid curve) bends downwards, if subtlety. So, at a certain point, caring more about being seen as performing well can lead to even more extreme lies that are less effective at changing the belief of the citizen.

Comparative statics on the prior distribution of θ are difficult to pin down analytically. However, we present one suggestive result from a simulation, where f is normally distributed with standard deviation λ (consistent with the scale parameter notation used in proposition 4).[12]

Figure 3 shows how increasing the precision of the prior (λ^{-2}) affects the equilibrium properties for low q (top panels) and high q (bottom panels). In both cases, increasing λ^{-2} decreases the equilibrium and optimal manipulation levels. This decrease always leads to a lower manipulation boost (dashed line in right panels), 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 (solid line in right panels) can go in either direction, as he gets less of a manipulation boost but pays a lower exogenous 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.

This suggests a reason why some politicians choose to allow free or foreign media (particularly in more authoritarian settings where there is heterogeneity in this decision). Doing so presumably gives citizens stronger prior beliefs about the politician performance.[13] Even highly repressive governments may want to allow outside information or free media, if doing so reduces incentives for counterproductive manipulation. Further, the contrast between the top and bottom panels of

[12] 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 $\hat{\theta}$.)

[13] This can be formalized by giving the citizen an additional unmanipulated signal of the politician performance, which has the same effect as increasing the precision of the prior belief.

Figure 3: Equilibrium properties as a function of λ^{-2}, for $q = .3$ (top panels) and $q = .7$ (bottom panels).

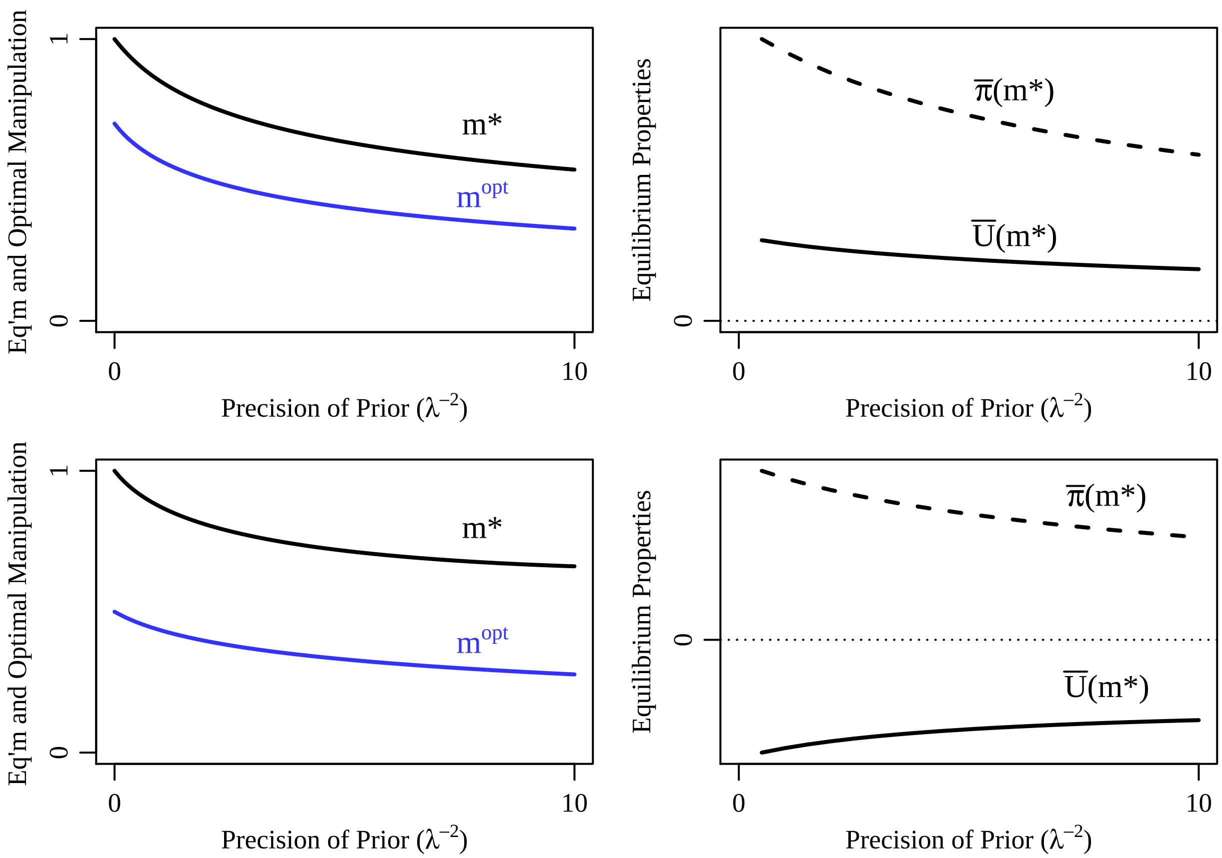


figure 3 suggests that the effect of more precise signals is helpful to the politician when citizens are more skeptical about his honesty.

7 Heterogeneous Audiences and Polarization

In the political context and elsewhere, the audience for manipulated information is frequently not just a single actor. If the audience is homogeneous this poses no issues: we can simple treat the citizen from our main model as a representative citizen and the analysis goes through. However, if the audience is heterogeneous in some manner, its members will react to information in different ways.

There are many ways to specify what differentiates the citizens at the outset. Here we analyze a simple specification which allows us to ask how the polarization of citizens in their initial credulity – or, alternatively, their level of trust in the politician – affects the manipulation choice and the ensuing distribution of beliefs about the politician performance performance. This heterogeneity could be driven by citizens having different prior information about the truthfulness of past statements by the politician. Alternatively, citizens may have an intrinsic desire to be more trusting of politicians they like (or skeptical of politicians they dislike).

Formally, we assume there are multiple citizens who share the same prior about the politician performance (θ), but start with different levels of credulity. There are three types of citizens with different levels of credulity: $0 \leq q_L < q_M < q_H \leq 1$. Write the share of each group as $Pr(q_L) = \eta\psi$, $Pr(q_M) = 1-\psi$, and $Pr(q_H) = (1-\eta)\psi$. We focus on the case where q_L is close to or exactly zero and q_H close to or exactly one. So, $\psi \in [0,\,1]$ measures the combined share of citizens with "extreme" beliefs, and $1-\psi$ represents the fraction of "moderates".

The politician knows the distribution of q.[14] A natural extension of his utility with one citizen is to assume he now cares about the average perception of his performance. The politician utility is then:

$$U_p(\theta, m, \hat{m}) = \sum_{q_i \in \{q_L, q_M, q_H\}} Pr(q_i)\,\hat{\theta}(\theta, m, \hat{m}, q_i) - c(m). \tag{12}$$

The analysis of the citizens' belief formation is identical to the main model; here we add a q_i argument to $\hat{\theta}(\theta, m, \hat{m}, q_i)$ to emphasize this depends on the citizens' initial credulity. The expected utility and hence equilibrium and optimal manipulation levels are defined identically with this new utility function. The analysis of the politician behavior is also similar to the main model, except now in addition to integrating over realizations of θ, the politician needs to account for the reaction

[14] Since citizens only form beliefs about the politician performance, it does matter if they are aware of the credulity levels of others. (See Little (2017) for a model where these higher order beliefs do matter.)

of different types of the citizens.

The result that manipulation becomes ineffective when m gets large (proposition 2) holds in this setting: since the manipulation boost for each individual citizen goes to zero, it goes to zero on average. Similar results hold about the existence of optimal and equilibrium manipulation levels.

To see the main implication of having heterogeneous citizens with this distribution, recall that with one citizen the equilibrium manipulation is highest when q is close to zero or one. This is because the endogenous cost of the citizen becoming more skeptical about the politician as m increases is part of what restrains the politician from lying, and those with intermediate q are most apt to change their beliefs.[15] The heterogeneous analog to this result is that when the population is mostly composed of "extremists" who all start either very skeptical or very trusting of the politician, the equilibrium lie is more extreme. In contrast, when lots of citizens begin moderately skeptical of the politician, the equilibrium lie is lower.

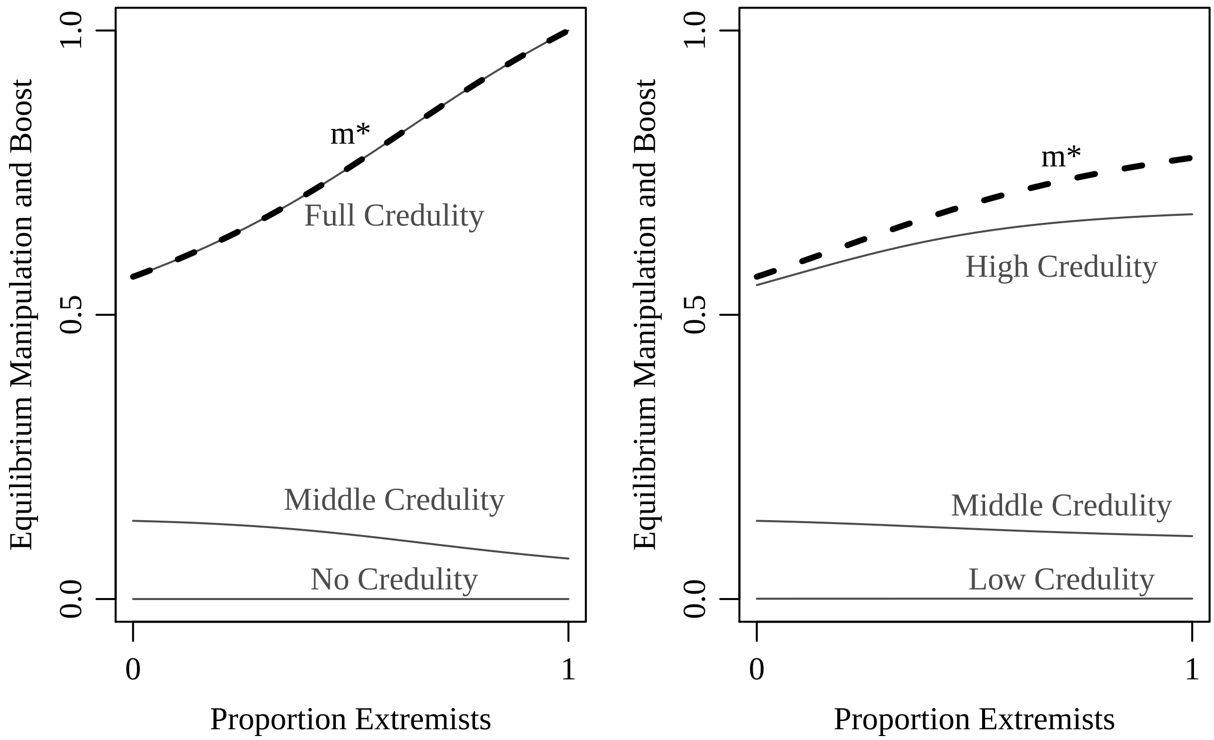
This also has implications for the polarization of beliefs held by citizens at the end. Recall that (by assumption) citizens start with a common prior about the politician performance. However, since they interpret the signal differently, those that begin more skeptical of the politician (higher q) will have a lower posterior belief about the politician performance than those who are more credulous (lower q). Further, these differences are magnified when the equilibrium lie is large. So, having fewer "moderates" in the population in terms of *ex ante* trust leads to more extreme lies an hence more polarization of posterior beliefs about the politician performance.

Figure 4 illustrates. In the left panel, the extremists are fully credulous ($q_L = 0$) or fully convinced the politician is manipulative ($q_H = 1$). In the left panel, those with more extreme beliefs are not fully convinced about the politician type at the outset.

In each panel, the black dashed curve shows that the equilibrium manipulation level is increasing in the proportion of the population which are extremists. Again, this follows from the fact that

[15]To be more precise, those with high low q eventually become convinced the politician is lying too. However, where an equilibrium exists this happens at a very high level of manipulation, beyond the point where the exogenous cost outweighs the benefit to lying more.

Figure 4: Equilibrium Manipulation and Posterior Beliefs as a function of the population composition. In each panel, f is normally distributed with mean 0 and standard deviation $3/10$, $c(m) = m^2$, $q_M = 1/2$ and $q_H = 1 - q_L$. In the left panel, $q_L = 0$, and in the right panel $q_L = 0.001$.



there are fewer moderates to generate the endogenous cost of lying which restrains the politician. Comparing between panels, this effect is even starker as the extremists become more convinced that the politician is either manipulative or not manipulative.

The gray curves show the implications for the final beliefs held by the three groups of citizens. The high credulity citizens end up with a higher manipulation boost – again defined as the difference in the expectation about the politician performance and the truth – as the manipulation level increases, while the low credulity citizens are unmoved no matter what. So, as m^* increases, the posterior beliefs of these two groups polarize.[16]

[16]For this parameterization the middle credulity group moves closer to the low credulity group as m^* increases, though this will not always be the case.

The following proposition formalizes this result:

Proposition 8. *Suppose there are three levels of citizen credulity $q_L < q_M < q_H$, such that $Pr(q_L) = \eta\psi$, $Pr(q_M) = 1 - \psi$, and $Pr(q_H) = (1 - \eta)\psi$. The conditions in Proposition 4 guarantee the existence of a unique equilibrium. Furthermore, if q_L is sufficiently close to 0 and q_H is sufficiently close to 1:*
i. m^* is increasing in ψ, and
ii.$\pi^*(q_L) - \pi^*(q_H)$ is increasing in ψ

Proof. See the appendix

More broadly, this illustration shows how even observing the same "news" can lead people to end up with more polarized beliefs than before. If, for whatever, reason, one group believes a new signal and one does not, what they actually learn will differ. This is particularly true of the signal is "extreme", and in the model here, the signal is extreme precisely when most people are initially very skeptical or trusting.

Similar dynamics can arise when the initial polarization is driven by heterogeneous beliefs about the politician performance. Suppose one group ("supporters") has a prior belief that the politician performance is moderately higher than it really is, and another group ("opponents") has a belief lower than or equal to the truth. Both groups start with a moderate degree of credulity. There is then a manipulated signal which indicates the politician is even better than the supporters' prior. If the signal is not too unbelievable, the supporters will only become marginally more skeptical and further increase their belief about the politician performance. On the other hand, the group with a lower initial prior will become very skeptical of the observed signal and less responsive. So, differing beliefs about the performance lead to polarization in beliefs about whether the politician is manipulative, which can then feed back into polarizing beliefs about performance.

8 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. People 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.

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.

Dziuda, Wioletta, Christian Salas. 2017. Communication with detectable deceit. Manuscript.

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.

Gehlbach, Scott, Konstantin Sonin, Milan Svolik. 2016. Formal models of nondemocratic politics. *Annual Review of Political Science* **19** 565–584.

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. 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.

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. 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.

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. 2016. Office insecurity and electoral manipulation.

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.

Sobel, Joel. 1985. A theory of credibility. *The Review of Economic Studies* 557–573.

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.

Woon, Jonathan. 2017. Political lie detection. Manuscript.

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 1. A more precise statement of part i is that:

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

Note that $r(\theta,0,0) = 1-q$ for all $\theta \;\Rightarrow\; \overline{r}(0,0) = \mathbb{E}_\theta[\overline{r}(\theta,0,0)] = 1-q > 0$. 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 $\lim\limits_{\theta\to-\infty} f(\theta) = \lim\limits_{\theta\to\infty} f(\theta) = 0$, $\int_\theta f'(\theta)d\theta = 0$. So, $\left.\frac{d\overline{\pi}(m,m)}{dm}\right|_{m=0} = 1-q > 0$.

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

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

First observe that since f is a proper density with a finite expectation, $\lim\limits_{x\to\infty} f(x) = 0$ and $\lim\limits_{x\to\infty} xf(x) = 0$, which implies $\lim\limits_{m\to\infty} f(\theta+m) \to 0$ and $\lim\limits_{m\to\infty} mf(\theta+m) \to 0$. So:

$$\lim_{m\to\infty} \pi(\theta,m,m) = \lim_{m\to\infty} mf(\theta+m)\frac{(1-q)}{qf(\theta)+(1-q)f(\theta+m)} = 0. \quad \text{(A.3)}$$

for all θ. Since the desired result is

$$\lim_{m\to\infty}\int_\theta \pi(\theta,m,m)f(\theta)d\theta = 0,$$

what remains to be shown is that we can switch the order of the limit and the integral in this expression, which we show with Lebesgue's Dominated Convergence Theorem (see Royden and

Fitzpatrick (2010), §4.4; and Tao (2011), §1.4). 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 $G^{max}(\theta) \equiv G(\theta, m^{max}(\theta))$. Taken together, Equations A.2 and A.3 imply that

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

Next, represent $\pi(\theta,m,m)$ and $G(\theta,m)$ as two sequences of measurable functions, $\{\pi_m\}$ and $\{G_m\}$ on $\mathbb{R}$. From (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, $|G_m| \le G^{max},\ \forall m$. Moreover, (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

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

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

Proof of Proposition 3. The derivative of $\bar{r}(m)$ is $\frac{d\bar{r}(m,m)}{dm} = \int_\theta \frac{\partial r(\theta,m,m)}{\partial m} f(\theta)d\theta$. Direct substitution from A.1 yields

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

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:

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

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 1). 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 $\left.\frac{\partial^2 \log f(x)}{\partial^2 x}\right|_{x=\theta} - \left.\frac{\partial^2 \log f(x)}{\partial^2 x}\right|_{x=\theta+m}$. 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 $\frac{\partial^3 \log(f(\theta))}{\partial \theta^3} \leq 0$ assumption.
Next, we can write B_0 as follows:

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

The first two terms of the parenthetical are again equal to $\left.\frac{\partial^2 \log f(x)}{\partial^2 x}\right|_{x=\theta} - \left.\frac{\partial^2 \log f(x)}{\partial^2 x}\right|_{x=\theta+m}$, 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, $B \leq 0 \Rightarrow \frac{\partial^2 \log(r(\theta,m,m))}{\partial \theta^2} \leq 0$, 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 $\hat{m}\dfrac{\partial^2\bar{r}\,(m,\hat{m})}{\partial m^2} < \kappa c_0''(m)$. 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.:

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

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 10. From Implicit Function Theorem, the slope of the best response function is

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

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, $\displaystyle\lim_{\lambda\to\infty}\frac{\partial m^{br}(\hat{m})}{\partial \hat{m}} = 0$.

□

Proof of proposition 5 The inequality in (11) holds for a sufficiently small m. And as shown in proposition 7, as $\kappa \to \infty$, both $m^* \to 0$ and $m^{\text{opt}} \to 0$. This proves that condition 1 is sufficient for the result.

For the condition 2, note that for any θ, m, and $\hat{m}$, $r(\theta, m, \hat{m}) \to 1 - q$ as $\lambda \to \infty$. Since r is continuous and differentiable, and pointwise converges to $1 - q$ as $\lambda \to \infty$:

$$\lim_{\lambda\to\infty}\frac{\partial r(\theta,m,\hat{m})}{\partial \hat{m}} = \frac{\partial}{\partial \hat{m}}\lim_{\lambda\to\infty} r(\theta,m,\hat{m}) = \frac{\partial}{\partial \hat{m}}(1-q) = 0.$$

Since $\lim_{\lambda\to\infty} \dfrac{\partial r(\theta, m, \hat{m})}{\partial \hat{m}} = 0$ for all θ, $\lim_{\lambda\to\infty} \dfrac{\partial \overline{r}(m, \hat{m})}{\partial \hat{m}} = 0$ for all m, and so (11) becomes $q > 0$.

For condition 3, inequality in (11) holds for all m at $q = 1$. Further, both sides of the inequality are continuous in q, so it must hold at $m = m^{\text{opt}}$ for some open interval $(\hat{q}, 1)$. $\square$

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 $m\frac{\partial \overline{r}(m,m)}{\partial m} < 1$ 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. $\square$

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

$$\kappa c_0'(m^{opt}) = \alpha m^{opt} \left(\frac{\partial \overline{r}(m, \hat{m})}{\partial m}\bigg|_{\hat{m}=m=m^{opt}} + \frac{\partial \overline{r}(m, \hat{m})}{\partial \hat{m}}\bigg|_{\hat{m}=m=m^{opt}} \right) + \alpha \overline{r}(m, \hat{m} = m)., \quad \text{(A.9)}$$

$$\kappa c_0'(m^*) = \alpha + \alpha m^* \frac{\partial \overline{r}(m, \hat{m})}{\partial m}\bigg|_{\hat{m}=m=m^*} \quad \text{(A.10)}$$

Part i follows from implicitly differentiating these equations

For part ii, the limiting behavior is immediate from the equilibrium conditions. By proposition 1, 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. $\square$

Proof of proposition 8 The expected manipulation boost of a citizen with credulity level $q_i \in \{q_L, q_M, q_H\}$ is $\bar{\pi}(m, \hat{m}, q_i) \equiv m - \hat{m} + \hat{m}\, \overline{r}(m.\hat{m}, q_i)$ (see Equation 3), and $\frac{\partial \bar{\pi}(m,\hat{m},q_i)}{\partial m} = 1 + \hat{m}\, \frac{\partial \overline{r}(m,\hat{m},q_i)}{\partial m}$. The politician's expected payoff function is

$$\bar{U}(m, \hat{m}) \equiv \mathbb{E}[\theta] + \eta\, \psi\, \bar{\pi}\,(m, \hat{m}, q_L) + (1 - \psi)\, \bar{\pi}\,(m, \hat{m}, q_M) + (1 - \eta)\, \psi\, \bar{\pi}\,(m, \hat{m}, q_H) - c(m).$$

Note that the function above is just a weighted average of the payoffs at three different levels of credulity. So analogous conditions of equilibrium existence and uniqueness of Proposition 4 hold (though the actual thresholds in κ and λ required to make the objective function globally concave depend on the credulity distribution). The equilibrium manipulation level, m^* is defined by $\frac{\partial \overline{U}(m,\hat{m})}{\partial m}\Big|_{\hat{m}=m=m^*} = 0$. Let $Y(\psi, m^*) \equiv \frac{\partial \overline{U}(m,\hat{m})}{\partial m}\Big|_{\hat{m}=m=m^*}$. From implicit function theorem

$\frac{\partial m^*}{\partial \psi} = -\frac{\partial Y/\partial \psi}{\partial Y/\partial m^*}$. By Assumption f4, the denominator is negative. Hence $0 < \frac{\partial m^*}{\partial \psi} \Leftrightarrow 0 < \frac{\partial Y}{\partial \psi}$. Rearranged:

$$0 < \frac{\partial m^*}{\partial \psi} \quad \Leftrightarrow \quad \frac{\partial \overline{r}(m, \hat{m}, q_M)}{\partial m}\bigg|_{\hat{m}=m=m^*} < \eta \ \frac{\partial \overline{r}(m, \hat{m}, q_L)}{\partial m}\bigg|_{\hat{m}=m=m^*} + (1-\eta) \ \frac{\partial \overline{r}(m, \hat{m}, q_H)}{\partial m}\bigg|_{\hat{m}=m=m^*}.$$

Note that $\frac{\partial \overline{r}(m,\hat{m},q)}{\partial m}\Big|_{\hat{m}=m} = \int_{-\infty}^{\infty} \frac{(1-q)qf(\theta)(f(\theta)f'(\theta+m)-f'(\theta)f(\theta+m))}{(qf(\theta)+(1-q)f(\theta+m))^2} d\theta$. Since f is log-concave, then f'/f is decreasing and $(f(\theta)f'(\theta+m) - f'(\theta)f(\theta+m)) < 0$, for all $m \in \mathbb{R}^+$. This establishes that $\frac{\partial \overline{r}(m,\hat{m},q_M)}{\partial m}\Big|_{\hat{m}=m} < 0$ for all $q_M \in (0,1)$. Also, $\lim_{q\to 1} \frac{\partial \overline{r}(m,\hat{m},q)}{\partial m}\Big|_{\hat{m}=m} = \lim_{q\to 0} \frac{\partial \overline{r}(m,\hat{m},q)}{\partial m}\Big|_{\hat{m}=m} = 0$. Therefore, the inequality holds when q_L and q_H are sufficiently close to 0 and 1. This proves part i.

Finally, in equilibrium, $\lim_{q\to 1} \pi^*(q) = 0$ and $\lim_{q\to 0} \pi^*(q) = m^*$. Since m^* increases in ψ, then $\pi^*(q_L) - \pi^*(q_H)$ must be increasing in ψ when q_L and q_H are sufficiently close to 0 and 1. □

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 $m'(\theta) > -1$. The posterior belief about ω when anticipating manipulation strategy $\hat{m}(\theta)$ is then 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.

As with the main model, things are simpler when the citizen starts out fully credulous or fully skeptical. 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.

In sum, allowing the politician to know θ does not affect the analysis for the extreme cases where $q = 0$ and $q = 1$, and in this case the equilibrium and optimal behavior do not depend on the revelation of type. For the intermediate case of $q \in (0, 1)$ the optimal and equilibrium strategies may not be constant in θ, though it is not obvious that this undermines the main conclusions of the more tractable version where θ is not known.