Creating Confusion*

Chris Edmond[†] Yang K. Lu[‡]

First draft: December 2017. This draft: February 2020

Abstract

We develop a model in which a politician seeks to discredit the media, to prevent citizens from being well informed. The politician can, at a cost, discredit the underlying sources of information available to media reporters. The reporters are rational and internalize the politician's incentives. In the unique equilibrium of the game, media reports are unbiased but endogenously noisy. We interpret the rise of social media as a shock that simultaneously (i) improves the underlying, intrinsic precision of the information available to reporters, but also (ii) reduces the politician's costs of manipulation. We show that there is a critical threshold such that if the costs of manipulation fall enough, the rise of social media makes the citizens worse off, despite the underlying improvement in the reporters' information. But if the costs of manipulation do not fall too much, the rise of social media reduces the amount of manipulation and makes the citizens better off. If, in addition, the reporters are also sufficiently well coordinated, the manipulation *backfires* on the politician, who would then want to invest in commitment devices that prevent them from manipulating in the first place.

Keywords: persuasion, slant, bias, noise, social media, fake news, alternative facts.

JEL classifications: C7, D7, D8.

*We particularly thank our discussant Wioletta Dziuda for her comments. We have also benefited greatly from conversations with Kim Sau Chung, Simon Grant, Simon Loertscher, Andy McLennan, Larry Samuelson and Xi Weng. We also thank seminar participants at the ANU, City University of Hong Kong, CUHK (Shenzhen), Guanghua Business School at Peking University, Hong Kong Baptist University, Nanyang Technological University, NUS, SMU, University of Melbourne, UNSW, and University of Queensland and participants at the 2019 Comparative Politics and Formal Theory Conference at UC Berkeley for their comments.

[†]University of Melbourne. Email: ⟨cedmond@unimelb.edu.au⟩

[‡]Hong Kong University of Science and Technology. Email: ⟨yanglu@ust.hk⟩

1 Introduction

> "...the campaign to discredit the press works by generating noise and confusion..." (Jay Rosen, "Why Trump Is Winning and the Press Is Losing," *New York Review of Books* online, April 15, 2018)

Consider a politician who seeks to discredit the media, to prevent citizens from being well-informed. Can the politician achieve this goal even when the media is rational and perfectly understands the politician's incentives? Should we be optimistic that new social media technologies will make it more difficult for the politician to discredit inconvenient reporting? Or will these new technologies make it easier for the politician to *create confusion*, frustrating people in their desire to be well-informed.[1]

We develop a simple model to answer these questions. In our model, a collection of reporters each form an assessment of the state of the world using various sources of information. A politician seeks to discredit this reporting by manipulating the sources of information, at a cost. The reporters are rational and internalize the politician's incentives when writing their reports. The reporting is consumed by citizens who value accurate information.

We interpret the rise of social media technologies as a shock that simultaneously: (i) increases the underlying, intrinsic *precision* of the sources of information available to reporters, and (ii) decreases the *costs* the politician incurs in manipulating information. We argue that these new technologies have led to new sources of information, both in the form of new media outlets and in the form of blogging and amateur journalism, thereby increasing the intrinsic precision of information available to reporters, but that these new sources of information are not all subject to the same standard of accountability as traditional journalism and moreover are consumed in a feed that blurs distinctions between sources and that makes it easier for all kinds of news, real and fake, to "go viral," thereby reducing the costs of manipulation.

We find that the rise of social media makes possible a sudden "regime change" in the amount of manipulation: The net effect of a social media shock depend crucially on whether the cost of manipulation can be kept above a critical threshold. In particular, if the underlying precision of information is relatively high and the costs of manipulation fall below this critical threshold, the economy will be in a *high manipulation regime*. In this high manipulation regime, the politician's manipulation successfully prevents the underlying precision of the media's information from passing through to citizens, making them worse off and the politician better off. But if the costs of manipulation can be kept above this critical threshold the economy will be in a *low manipulation regime*. In this low manipulation regime, the

[1] For an overview of the role of social media in the 2016 US presidential election, see Allcott and Gentzkow (2017), Faris et al. (2017) and Guess et al. (2018). In October 2017, representatives of Facebook, Google and Twitter were called to testify before the US Senate on the use of their platforms in spreading fake news, including Russian interference (e.g., Fandos et al., 2017). The role of social media and fake news has also been widely discussed in the context of the 2016 UK Brexit referendum, the 2017 French presidential elections, the 2017 Catalan independence crisis, etc. In November 2017, the European Commission announced its intent to take action to combat the use of social media platforms to spread fake news (e.g., White, 2017).

politician fails to prevent the underlying precision of the media's information from passing through to the citizens, making them better off and the politician worse off. Moreover, we find that if the reporters are *sufficiently well coordinated*, then the politician's manipulation can *backfire*, presenting the politician with incentives to invest in commitment devices that prevent them from manipulating in the first place.

Section 2 outlines the model. There is a politician who knows the state of the world. There is a continuum of reporters who share a common prior and receive idiosyncratic signals about the state. Each individual reporter wants to report the truth, but they may also care about the reports of the other reporters (their actions may be strategic substitutes or strategic complements). The politician seeks to *prevent* the reporters from reporting the truth. The reporting is consumed by citizens who prefer accurate reports in the sense of low mean squared error. The politician has a technology that allows them to manipulate information by choosing the reporters' signal mean at a cost that is increasing in the distance between the true state and the signal mean. The reporters are rational and internalize the politician's incentives. To keep the model tractable, we assume quadratic preferences and normal priors and signal distributions. We study equilibria that are linear in the sense that the reporters' strategies are linear functions of their signals.

Section 3 solves the model and shows that there is a *unique* (linear) equilibrium. In equilibrium, media reports are unbiased but are made endogenously noisier by the politician's manipulation. The equilibrium amount of manipulation is quite sensitive to parameters. In particular, the effects of an increase in the precision of the reporters' signals depends on the level of the costs of manipulation. If the costs of manipulation are high, increasing the precision of the reporters' signals decreases the amount of manipulation and the reporters become more responsive to their signals than they would be if the politician could not manipulate at all. But if the costs of manipulation are low, increasing the precision of the reporters' signals increases the amount of manipulation. In the limit where the costs of manipulation are negligible, the politician's manipulation renders the reporters' signals completely uninformative even if the underlying, intrinsic precision of their signals is arbitrarily high.

Section 4 provides conditions under which the politician is made better off or worse off by their ability to manipulate information. These conditions depend on the *strength* of the strategic interactions among reporters. In particular, we find that the politician's manipulation can backfire if there are sufficiently strong strategic interactions among reporters. But regardless of the strength of strategic interactions among reporters, the politician *gains the most* when the costs of manipulation are low and the intrinsic precision of the signals is high.

Section 5 provides welfare results for the reporters and, perhaps more importantly, for the citizens who consume the reporting. We provide conditions under which the citizens' and reporters' welfare outcomes are aligned or misaligned. The reporters are always made worse off by the politician's manipulation. The citizens are also made worse off by the manipulation unless the reporters' actions are sufficiently strong strategic substitutes, i.e., unless the reporters have an aggressive tendency to differentiate their reports from one-another. Moreover, if the costs of manipulation are low and the intrinsic precision of the signals is high, further increases in the intrinsic precision of the reporters' signals makes both the reporters and citizens increasingly worse off, due to the politician's manipulation. In short, the scenario where the politician gains the most is also the scenario where the reporters and citizens lose the most.

Section 6 contains our results on the rise of social media. We show that if the intrinsic precision of information is high there is a critical threshold for the costs of manipulation. Near this threshold, small changes in the costs of manipulation cause dramatic changes in the amount of manipulation. The net effect of the rise of social media then depends crucially on the size of the reduction in the costs of manipulation. If the costs fall enough, the economy tips into the high manipulation regime, making the politician better off and everyone else worse off. But if the costs of manipulation can be kept above the threshold, the reporters and citizens are made better off. In this sense, even small changes on the part of social media platforms that make it harder for misinformation to propagate may have large welfare effects.

Of course, one doesn't need this model to arrive at the view that social media may have adverse implications. In the aftermath of the 2016 US presidential election, 2016 UK Brexit referendum, etc, such views have become conventional wisdom (e.g., Bennett and Livingston, 2018; Sunstein, 2018). But even quite recently, the conventional wisdom had been the other way round, arguing that social media is a force for good, facilitating online activism and helping to bring about important social and political reforms. These sentiments peaked around 2010, when WikiLeaks was still widely seen as a force for transparency and democratic accountability (e.g., Shafer, 2010) and when hopes for the "Arab Spring" were still high (e.g., Cohen, 2009; Esfandiari, 2010). Similar optimism about the use of social media can still be seen on the part of activists in the #BlackLivesMatter and #MeToo movements (e.g., Codrea-Rado, 2017; Rickford, 2015). In our model, absent manipulation, the rise of social media would allow people to make more informed decisions and make them better off.

Strategic communication with costly talk. Our model is a sender/receiver game with many imperfectly informed receivers whose actions can be strategic substitutes or complements. As in Crawford and Sobel (1982), the preferences of the sender and receivers are not aligned and the sender is informed. But as in Kartik (2009) we have *costly talk*, not cheap talk. By contrast with standard cheap talk models, our model with costly talk features a unique equilibrium. In the limit as the sender's distortion becomes *almost costless*, the unique equilibrium features a kind of *babbling* where the receivers ignore their signals. Our model with costly talk is related to Kartik, Ottaviani and Squintani (2007) and Little (2017) but our receivers are not "credulous" or subject to confirmation bias. If the sender's distortion is so costly that there is no manipulation, and if the receivers' actions are strategic complements, the model reduces to the "beauty contest" game in Morris and Shin (2002).

Bayesian persuasion. In equilibrium, our receivers have unbiased posterior expectations. Despite this, the sender still finds it optimal to send costly distorted messages. This is because of the effects of their messages on other features of the receivers' beliefs, as in the Bayesian persuasion literature following Kamenica and Gentzkow (2011). In particular, the sender can be made better off by the increase in the receivers' posterior variance resulting from the sender's messages. A crucial distinction however is that in Kamenica and Gentzkow (2011), the sender can *commit* to an information structure and this commitment makes the model essentially nonstrategic in that their receiver only needs to solve a single-agent decision problem. Other approaches to information design, such as Bergemann and Morris (2016) also allow the sender to commit. By contrast, our sender *cannot commit* and chooses their message after becoming informed about the state of the world, as in Crawford and Sobel (1982).

Applications to political communication that follow the Bayesian persuasion approach in assuming the sender can commit include Hollyer, Rosendorff and Vreeland (2011), Gehlbach and Sonin (2014), Gehlbach and Simpser (2015) and Rozenas (2016). Our model is more similar to Little (2012, 2015) and Shadmehr and Bernhardt (2015) in that the sender cannot commit and in some cases would find it valuable to commit to *not* manipulate information. Other related work includes Egorov, Guriev and Sonin (2009), Edmond (2013), Lorentzen (2014), Huang (2014), Chen and Xu (2014), and Guriev and Treisman (2015). For overviews of this literature, see Svolik (2012) and Gehlbach, Sonin and Svolik (2016).

Media bias, fake news, and alternative facts. The media bias literature often assumes that receivers *prefer* distorted information[2] — e.g., Mullainathan and Shleifer (2005), Baron (2006), Besley and Prat (2006), Gentzkow and Shapiro (2006), and Bernhardt, Krasa and Polborn (2008). Allcott and Gentzkow (2017) have used this kind of setup to explain how there can be a viable market for "fake news" that coincides with more informative, traditional media (see also Gentzkow, Shapiro and Stone, 2015). To be clear, we view such behavioral biases as very important. Our point is that such biases are *not necessary* for manipulation to be effective. In our model, the sender can still gain from sending costly distorted messages because of the endogenous noise that results from such messages.

Or to put things a bit differently, in our model no one is misled by the politician's "alternative facts" and yet the politician can benefit greatly from the ensuing babble and tumult.

2 Model

There is a unit mass of ex ante identical *reporters*, indexed by $i \in [0, 1]$, that we collectively refer to as the *media*, and a single informed *politician* attempting to influence the media's reports. There is also a large mass of *citizens* who passively consume the media's reports.

[2]For an alternative approach that assumes rational consumers, see Anderson and McLaren (2012).

Reporters. Each individual reporter produces a report $a_i \in \mathbb{R}$ that balances: (i) a desire to accurately report the true underlying state $\theta \in \mathbb{R}$ (about which they are imperfectly informed), with (ii) the way their individual report a_i fits with the average report $A := \int_0^1 a_i \, di$ produced by the entire media. In particular, each reporter chooses a_i to minimize the expected value of the quadratic loss

$$(1-\lambda)(a_i-\theta)^2 + \lambda(a_i - A)^2, \qquad \lambda < 1 \tag{1}$$

where the parameter λ governs the strategic interactions among reporters. If $\lambda = 0$, each reporter cares only about reporting θ and sets their action a_i equal to their expectation of θ. If $\lambda < 0$ then each individuals' report a_i and the average report A are *strategic substitutes*. In this case each individual wants their report a_i to be consistent with θ but also wants to make their a_i stand out from the crowd, differing from A. By contrast if $\lambda \in (0,1)$ then each individual's report a_i and the average report A are *strategic complements*. In this case, each reporter wants their report a_i to be consistent both with θ and with A.

In forming beliefs about θ, the reporters begin with the common prior that θ is distributed normally with mean z and precision $\alpha_z > 0$ (i.e., variance $1/\alpha_z$). Each individual reporter then draws an idiosyncratic signal

$$x_i = y + \varepsilon_i \tag{2}$$

where the mean y is chosen by the politician, as discussed below, and where the idiosyncratic noise ε_i is IID normal across reporters, independent of θ, with mean zero and precision $\alpha_x > 0$ (i.e., variance $1/\alpha_x$).

To summarize, reporters have one source of information, the prior, that is free of the politician's influence and another source of information, the signal x_i, that is not. While the informativeness of the prior is fixed, the informativeness of the signal needs to be determined endogenously in equilibrium in light of the politician's incentives.

Politician. The politician knows the value of θ and seeks to *prevent* the reporters from writing reports that are accurate for the underlying state θ. In particular, the politician obtains a gross benefit

$$\int_0^1 (a_i - \theta)^2 \, di \tag{3}$$

that is increasing in the dispersion of the actions a_i around θ. The politician is endowed with the ability to choose the mean y of the reporters' idiosyncratic signals. In particular, knowing θ, the politician may take a costly action $s \in \mathbb{R}$ to make the signal mean $y = \theta + s$, i.e., the term $s = y - \theta$ can be interpreted as the *slant* or *spin* that the politician is attempting to introduce. This manipulation incurs a quadratic cost $c(y-\theta)^2$, similar to Holmström (1999) and Little (2012, 2015), so that the net payoff to the politician is

$$V = \int_0^1 (a_i - \theta)^2 \, di - c(y-\theta)^2, \qquad c > 0 \tag{4}$$

where the parameter $c > 0$ measures how costly it is for the politician to choose values of y far from θ. The special case $c \to 0$ corresponds to a version of *cheap talk* (i.e., the politician can choose y arbitrarily far from θ without cost). The special case $c \to \infty$ corresponds to a setting without manipulation (i.e., where the politician will always choose $y = \theta$).

Equilibrium. A symmetric *perfect Bayesian equilibrium* of this model consists of individual reporter actions $a(x_i)$ and beliefs, an average action $A(y)$ and the politician's manipulation $y(\theta)$ such that: (i) each reporter rationally takes the manipulation $y(\theta)$ into account when forming their beliefs, (ii) each reporter's action $a(x_i)$ minimizes their expected loss, (iii) the average action $A(y)$ is consistent with the individual actions, and (iv) the politician's $y(\theta)$ maximizes the politician's payoff given the individual actions.

Before solving this model, we briefly explain our interpretation of the setup.

2.1 Discussion

Politician's objective. The politician seeks to prevent the media from accurately reporting the true state θ. The key idea here is that if media reports are sufficiently accurate, they will end up motivating *citizens at large* to do something harmful to the politician's interests — e.g., turning out to vote, or protesting against the politician — and the politician would like to prevent this. For example, suppose that in a subsequent stage of the current game, individual citizens[3] decide whether to turnout against the politician based on an idiosyncratic sample from the set of reports a_i and that citizens will do so only if they obtain a sufficiently precise assessment of θ. The politician then suppresses this turnout if the citizens obtain reports of θ that are imprecise (i.e., the individual reports a_i are scattered around θ). In this sense, the politician is creating a climate of *political disillusionment* amongst the citizens at large, and the set of reporters are the medium through which this occurs.

Note that the politician has no "directional bias" — they are not trying to tilt the reporters' actions towards some ideal point. If the politician had a known directional bias (to the left or right, say), that would be easily extracted by the reporters in forming their beliefs about θ and end up increasing the politician's marginal costs but otherwise leaving the analysis essentially unchanged. We think of our setup as pertaining to the *residual uncertainty* after known biases of the politician have been extracted.

Politician's information manipulation. We think of the politician's manipulation as a systematic policy of *undermining the credibility of the information available to reporters.* This might take the form of spin, seeking to frame coverage in the most favorable possible way, more blatant forms of misdirection and "alternative facts" or outright accusations of

[3] To be clear, we do not model the behavior of these citizens. In Section 5, we evaluate welfare outcomes for these citizens assuming that they prefer reports that are unbiased, precise signals of the state θ.

bias or ulterior motives on the part of reporters' sources. Think of nature drawing θ and the politician being challenged on this — "is it not the case, Mr President, that the truth is θ?"— with the politician then indignantly retorting that the claim of θ is "fake news" and that the real truth is that the state is $y = \theta + s$. In our model, this kind of political spin will not, in equilibrium, lead to any bias in the reports about θ. Instead, as we will see, this kind of political spin will make the reporters' signals x_i *endogenousely* noiser than they otherwise would be. As a result, the politician *may* benefit because the manipulation, in equilibrium, makes the reports less informative and in this way reduces the credibility of the media.

In equilibrium, reporters in our model will be able to justly say that they only report unbiased information and yet at the same time, citizens at large will feel frustrated, getting information that is less informative than it could be, with the politician sometimes able to benefit from the reduced credibility of the media.

Interactions amongst reporters. In the special case $\lambda = 0$ there are no interactions amongst reporters, each reporter cares only about accurately reporting θ and their a_i is simply their (unbiased) expectation of θ. If $\lambda < 0$, the reporters' actions are strategic substitutes. We think of this as capturing, say, competition between journalists who care both about accurately reporting θ but also care about differentiating themselves from rival journalists. If $\lambda > 0$, the reporters' actions are strategic complements. We think of this as capturing, say, social media activists or other users who care both about accurately reporting θ but also care about fitting in with their peers, collecting lots of likes and retweets.

Social media. In our main application, we interpret the rise of social media as a simultaneous increase in the signal precision α_x and a decrease in the cost of manipulation c. What we have in mind is that social media technologies have significantly reduced the costs of collecting, reporting, and disseminating news and information. This leads to the entry of a large number of new online media outlets, blogging, amateur journalism etc, that would, *absent manipulation*, increase the amount of information available to everyone. But these new sources of information are not all subject to the same standards of journalistic ethics and accountability as traditional journalism. Moreover, the new social media technologies also mean that citizens consume much of their media content in a news feed that both blurs distinctions between reliable and unreliable sources of information and also makes it easy for all kinds of news, real and fake, to "go viral" — to be retweeted or shared extremely rapidly. In short, the rise of social media makes it easier for a politician to use spin and misdirection to undermine the credibility of the information reported in the media.

2.2 Equilibrium with no manipulation

Before turning to equilibrium outcomes in the model with information manipulation, we first review equilibrium outcomes when there is *no manipulation*. This provides a natural benchmark against which the politician's ability to manipulate information can be evaluated.

Suppose the politician *cannot* manipulate information — i.e., let $c \to \infty$ so that the politician chooses $y = \theta$. This puts us in a standard linear-normal setting where each reporter's posterior expectation of θ is a precision-weighted average of their signal x_i and prior z, specifically

$$\mathbb{E}[\,\theta \,|\, x_i\,] = \frac{\alpha_x}{\alpha_x + \alpha_z} x_i + \frac{\alpha_z}{\alpha_x + \alpha_z} z. \tag{5}$$

The optimal action of an individual reporter is then

$$a(x_i) = (1-\lambda)\mathbb{E}[\,\theta \,|\, x_i\,] + \lambda \mathbb{E}[\,A(\theta) \,|\, x_i\,]. \tag{6}$$

Equilibrium in linear strategies. We restrict attention to equilibria in which reporters use *linear* strategies of the form

$$a(x_i) = kx_i + hz \tag{7}$$

for some coefficients k, h to be determined in equilibrium. Since $y = \theta$ the corresponding average action is

$$A(\theta) = k\theta + hz. \tag{8}$$

The optimal action of an individual reporter is then

$$a(x_i) = (1 - \lambda + \lambda k)\mathbb{E}[\,\theta \,|\, x_i\,] + \lambda hz. \tag{9}$$

Substituting for $\mathbb{E}[\,\theta \,|\, x_i\,]$ from (5) and matching coefficients with (7), gives the unique solution

$$k^*_{nm} = \frac{(1-\lambda)\alpha_x}{(1-\lambda)\alpha_x + \alpha_z}, \qquad h^*_{nm} = 1 - k^*_{nm}. \tag{10}$$

So, absent manipulation, each reporter has the strategy $a(x_i) = k^*_{nm} x_i + (1 - k^*_{nm})z$. If there are no strategic interactions among reporters, $\lambda = 0$, then each individual's action $a(x_i)$ is simply their posterior expectation (5) with the weight on x_i determined by the α_x/α_z ratio. If the actions are strategic substitutes, $\lambda < 0$, each reporter seeks to differentiate themself and over-weights their idiosyncratic signal relative to the common prior. If the actions are strategic complements, $\lambda > 0$, each reporter seeks to coordinate with their peers and down-weights their idiosyncratic signal relative to the common prior.

For future reference, let

$$k^*_{nm} := \frac{\alpha}{\alpha + 1}, \qquad \alpha := (1-\lambda)\frac{\alpha_x}{\alpha_z} > 0 \tag{11}$$

The composite α, i.e., the relative precision of the signal to the prior α_x/α_z, adjusted by λ to reflect the presence of strategic interactions among reporters, features repeatedly below.

3 Equilibrium with information manipulation

Now suppose the politician *can* manipulate information. In this setting there is a genuine equilibrium fixed-point problem because we need to ensure that the reporters' actions and beliefs and the politician's information manipulation are mutually consistent.

Preliminaries. We again restrict attention to equilibria in which the reporters use linear strategies. We write these as

$$a(x_i) = kx_i + (1-k)z \tag{12}$$

The fact that the reporters' strategies are linear is a genuine restriction. But, as we show in our Supplementary Online Appendix, the fact that the coefficients sum to one is a result and it streamlines the exposition to make use of this result from the start.

3.1 Politician's problem

Given that reporters use linear strategies $a(x_i) = kx_i + (1-k)z$, the politician's problem is to choose $y \in \mathbb{R}$ to maximize

$$V(y) = \int_0^1 \big(k(y+\varepsilon_i) + (1-k)z - \theta \big)^2 \, di - c(y-\theta)^2$$

$$= (ky + (1-k)z - \theta)^2 + \frac{1}{\alpha_x}k^2 - c(y-\theta)^2 \tag{13}$$

Taking the reporters' response coefficient k as given, this is a simple quadratic optimization problem. The solution is

$$y(\theta) = \frac{c-k}{c-k^2}\theta + \frac{k-k^2}{c-k^2}z \tag{14}$$

where the second-order condition requires

$$c - k^2 \geq 0 \tag{15}$$

Given that reporters use linear strategies, it is optimal for the politician to also use a linear strategy. The coefficients in the politician's strategy sum to one, so we can write

$$y(\theta) = (1-\delta)\theta + \delta z \tag{16}$$

where δ depends on the reporters' response coefficient k via

$$\delta(k) := \frac{k-k^2}{c-k^2}, \qquad c - k^2 \geq 0 \tag{17}$$

To interpret the politician's strategy, observe that if, for whatever reason, the politician chooses $\delta(k) = 0$, then the politician is choosing a signal mean y that coincides with the true θ — i.e., the politician chooses not to manipulate information and the reporters' signals

x_i are as informative as possible about the true θ (limited only by the exogenous precision, α_x). Alternatively, if the politician chooses $\delta(k) = 1$, then the politician is choosing a signal mean y that coincides with the reporters' prior z — i.e., the reporters' signals x_i provides no additional information about θ.

In short, the politician's manipulation coefficient $\delta(k)$ summarizes the politician's best response to the reporters' coefficient k. To construct an equilibrium, we need to pair this with the reporters' best response to the politician's manipulation.

3.2 Reporters' problem

To construct the reporters' best response, first observe that the optimal action $a(x_i)$ for an individual reporter with signal x_i continues to satisfy

$$a(x_i) = (1-\lambda)\mathbb{E}[\,\theta\,|\,x_i] + \lambda\mathbb{E}[\,A(\theta)\,|\,x_i].$$

If other reporters use (12) and the politician uses (16) then the aggregate action is

$$A(\theta) = ky(\theta) + (1-k)z = k(1-\delta)\theta + (1-k(1-\delta))z \tag{18}$$

Collecting terms then gives

$$a(x_i) = (1-\lambda(1-k(1-\delta)))\,\mathbb{E}[\,\theta\,|\,x_i] + \lambda(1-k(1-\delta))\,z \tag{19}$$

(i.e., a weighted average of the posterior and prior expectations). To make further progress, we need to determine this individual reporter's posterior expectation $\mathbb{E}[\,\theta\,|\,x_i]$ in the presence of the politician's manipulation strategy (16).

Signal extraction. If the politician's manipulation strategy is (16), then each individual receiver has two pieces of information: (i) the common prior $z = \theta + \varepsilon_z$, where ε_z is normal with mean zero and precision α_z, and (ii) the idiosyncratic signal

$$x_i = y(\theta) + \varepsilon_i = (1-\delta)\theta + \delta z + \varepsilon_i$$

$$= \theta + \delta\varepsilon_z + \varepsilon_i \tag{20}$$

where the ε_i are IID normal with mean zero and precision α_x. The key point is that the politician's manipulation δ makes the signal x_i less correlated with the true θ and more correlated with the prior z. To extract the common component, we construct a *synthetic signal*

$$s_i := \frac{1}{1-\delta}\,(x_i - \delta z) = \theta + \frac{1}{1-\delta}\varepsilon_i \tag{21}$$

The synthetic signal s_i is independent of the prior and normally distributed around the true θ with precision $(1-\delta)^2\alpha_x$. If $\delta = 0$, such that $y(\theta) = \theta$, there is no manipulation from the politician and hence the synthetic signal s_i has precision α_x, i.e., equal to the intrinsic precision of the actual signal x_i. If $\delta = 1$, such that $y(\theta) = z$, the signal x_i is uninformative about θ and the synthetic signal has precision zero.

Conditional on the synthetic signal s_i, an individual receiver has posterior expectation

$$\mathbb{E}[\,\theta\,|\,s_i] = \frac{(1-\delta)^2\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z} s_i + \frac{\alpha_z}{(1-\delta)^2\alpha_x + \alpha_z} z \tag{22}$$

So in terms of the actual signal x_i they have

$$\mathbb{E}[\,\theta\,|\,x_i] = \frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z} x_i + \left(1 - \frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}\right) z \tag{23}$$

Matching coefficients. We can then plug this formula for the posterior expectation back into (19) and match coefficients to get the fixed-point condition

$$k = (1-\lambda(1-k(1-\delta)))\frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}$$

which has the unique solution

$$k(\delta) := \frac{(1-\delta)\alpha}{(1-\delta)^2\alpha + 1} \tag{24}$$

where $\alpha := (1-\lambda)\alpha_x/\alpha_z$ is the composite parameter introduced in (11) above.

In short, $k(\delta)$ gives us the unique value of k determined simultaneously by the reporters in response to the politician's δ. In this sense, we can say that $k(\delta)$ is the reporters' best response to the politician's δ.

Misdirection and endogenous noise. The politician manipulates information through a form of *misdirection*. The systematic part of the reporters' information $y = (1-\delta)\theta + \delta z$ is a mixture of "truth and prejudice" (the true θ and the prior z). The effect of this misdirection is to make the signals x_i *endogenously* noisier than they would otherwise be. Absent this misdirection, the reporters' beliefs would have posterior precision $\alpha_x + \alpha_z$ reflecting the intrinsic information content of their signals and their prior certainty. But with misdirection, the reporters' posterior precision falls to $(1-\delta)^2\alpha_x + \alpha_z$. When there is a lot of misdirection, $\delta \to 1$, the posterior precision falls all the way to α_z, as if the reporters' signals x_i are worthless. In short, the underlying information available to the reporters may be of intrinsically high quality, but with misdirection they can end up writing reports *based on their prior alone*. We now turn to how the endogenous amount of noise is determined in equilibrium.

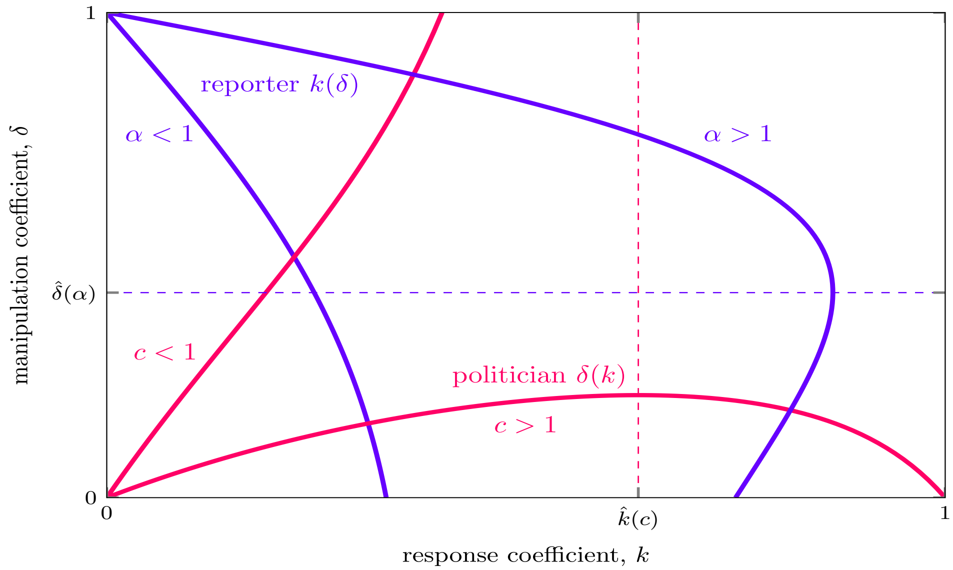

Figure 1: Unique equilibrium.

There is a unique equilibrium, that is, a unique pair k^*, δ^* simultaneously satisfying the reporters' best response $k(\delta)$ and the politician's best response $\delta(k)$. For $\alpha > 1$ there is a critical point $\hat{\delta}(\alpha)$ such that the reporters' $k(\delta)$ is increasing in δ for $\delta < \hat{\delta}(\alpha)$. For $c > 1$ there is a critical point $\hat{k}(c)$ such that the politician's $\delta(k)$ is decreasing in k for $k > \hat{k}(c)$. Note that if $c < 1$ then $k^* \leq c$ and hence k^* cannot be high if c is low.

3.3 Equilibrium determination

To summarize, reporters have strategies of the form $a(x_i) = kx_i + (1-k)z$ where the response coefficient k is a function of the politician's manipulation δ and the politician has a strategy of the form $y(\theta) = (1-\delta)\theta + \delta z$ where the manipulation coefficient δ is a function of the reporters' k. Think of these as two curves, $k(\delta)$ for the reporters and $\delta(k)$ for the politician. Finding equilibria reduces to finding points where these two curves intersect. Let k^* and δ^* denote such equilibrium points.

Now define

$$\mathcal{K}(c) := \{\, k \,:\, 0 \leq k \leq \min[c, 1] \,\}, \qquad c > 0 \tag{25}$$

This is the set of k such that $\delta(k) \in [0, 1]$. The upper bound $k \leq \min[c, 1]$ comes from the fact that if $c \leq 1$ then $\delta(k) \leq 1$ if and only if $k \leq c$. We can now state our first main result:

PROPOSITION 1. There is a unique equilibrium, that is, a unique $k^* \in \mathcal{K}(c)$ and $\delta^* \in [0, 1]$ simultaneously satisfying the reporters' $k(\delta)$ and the politician's $\delta(k)$.

Figure 1 illustrates the result, with k plotted on the horizontal axis and δ plotted on the vertical axis. In general, both these curves are non-monotone but they intersect once, pinning down a unique pair k^*, δ^* from which we can then determine the politician's equilibrium strategy $y(\theta) = (1-\delta^*)\theta + \delta^* z$ and the reporters' equilibrium strategy $a(x_i) = k^* x_i + (1-k^*)z$.

When is the reporters' best response non-monotone?

LEMMA 1. The reporters' best response $k(\delta)$ is first increasing then decreasing in δ with a single peak at $\delta = \hat{\delta}(\alpha)$ given by

$$\hat{\delta}(\alpha) := \begin{cases} 0 & \text{if } \alpha \leq 1 \\ 1 - 1/\sqrt{\alpha} & \text{if } \alpha > 1 \end{cases} \tag{26}$$

with boundary values $k(0) = \alpha/(\alpha+1) =: k^*_{nm}$ and $k(1) = 0$.

Lemma 1 says that if α is relatively high and the amount of manipulation δ is relatively low, then the reporters will in fact be *more responsive* to their signals than they would be in the absence of manipulation. To understand why this can happen, we need to decompose the effect of δ into two parts: (i) the effect of δ on the precision of the synthetic signal s_i in (21), and (ii) the effect of δ on the correlation between reporters' signal x_i and their prior z. We will refer to the former as the "*precision*" effect and to the latter as the "*correlation*" effect. From (21), the synthetic signal precision is $(1-\delta)^2\alpha_x$ and hence is unambiguously decreasing in δ. This reduction in precision acts to decrease the reporters' k. But an increase in δ also increases the correlation between x_i and z. Since z also contains information about the fundamental θ, this increase in correlation acts to increase the reporters' response to their signals x_i. For $\alpha \leq 1$, the precision effect unambiguously dominates so that $k(\delta)$ is strictly decreasing from $k(0) = k^*_{nm}$ to $k(1) = 0$. For $\alpha > 1$, the correlation effect dominates for low levels of δ while the precision effect dominates for high levels of δ so that $k(\delta)$ increases from $k(0) = k^*_{nm}$ to its maximum then decreases to $k(1) = 0$.

That said, the bottom line is that *for high enough* manipulation, it will indeed be the case that the reporters are less responsive to their signals, $k(\delta) < k^*_{nm}$. This hurdle is easy to clear when α is relatively low, but hard to clear when α is relatively high.

When is the politician's best response non-monotone?

LEMMA 2. The politician's best response $\delta(k)$ is first increasing then decreasing in k with a single peak at $k = \hat{k}(c)$ given by

$$\hat{k}(c) = \begin{cases} c & \text{if } c < 1 \\ c - \sqrt{c(c-1)} & \text{if } c > 1 \end{cases} \tag{27}$$

with boundary values $\delta(0) = 0$ and $\delta(c) = 1$ if $c < 1$ and $\delta(1) = 0$ if $c > 1$.

Lemma 2 says that if the cost of manipulation is relatively low, then whenever the reporters respond more to their signals, the politician will choose a higher level of manipulation.[4] But if instead the cost of manipulation is relatively high, then for high enough k the politician responds by choosing a *lower* level of manipulation $\delta(k)$.[5]

[4]If $c < 1$, the maximum of $\delta(k)$ is obtained at the boundary where $k = c$.

[5]If $c > 1$, the critical value $\hat{k}(c) = c - \sqrt{c(c-1)}$ is strictly decreasing in c and hence $\hat{k}(c) < 1$ for $c > 1$.

To understand why it can be the case that higher values of the reporters' response coefficient k can lead the politician to choose *less* manipulation, we first write the politician's gross payoff as

$$\int_0^1 (a_i - \theta)^2 \, di = (A - \theta)^2 + \int_0^1 (a_i - A)^2 \, di$$

In short, the politician can be made better off through increasing the distance between A and θ or through increasing the dispersion of a_i around A. Now observe that if the politician uses the strategy $y = (1 - \delta)\theta + \delta z$ then $A - \theta = (k\delta + 1 - k)(z - \theta)$, proportional to the error in the common prior $z - \theta$. So the first term in the politician's gross payoff is

$$(A - \theta)^2 = (k\delta + 1 - k)^2 (z - \theta)^2$$

Similarly if the reporters use the strategy $a_i = kx_i + (1 - k)z$ then $a_i - A = k(x_i - y) = k\varepsilon_i$, proportional to the idiosyncratic noise ε_i. So the second term in the gross payoff is

$$\int_0^1 (a_i - A)^2 \, di = \frac{1}{\alpha_x} k^2$$

The politician's choice of δ only enters their gross payoff through the distance between A and θ term. The dispersion in a_i around A term is independent of δ.

Subtracting off the cost $c(y - \theta)^2$ and collecting terms gives the politician's objective

$$V = \Big((k\delta + 1 - k)^2 - c\delta^2 \Big) (z - \theta)^2 + \frac{1}{\alpha_x} k^2 \tag{28}$$

We can now view the politician's problem as being equivalent to choosing $\delta \in [0, 1]$ to maximize (28) taking $k \in [0, 1]$ as given. For future reference, let $B(\delta, k) := (k\delta + 1 - k)^2$ denote the benefit the politician obtains from increasing the distance between A and θ and let $C(\delta) := c\delta^2$ denote the associated cost. In this notation, the politician's manipulation is given by $\delta(k) = \text{argmax}_\delta [B(\delta, k) - C(\delta)]$. Since $C(\delta)$ is independent of k, whether or not the best response $\delta(k)$ is increasing or decreasing in k depends on whether the marginal benefit of δ is is increasing or decreasing in k. The marginal benefit of δ is

$$\frac{\partial B}{\partial \delta} = 2(k\delta + 1 - k)k \tag{29}$$

Now recall that $A - \theta = (k\delta + 1 - k)(z - \theta)$ so the term $(k\delta + 1 - k)$ is simply the coefficient on the error in the common prior. There are then two effects of an increase in k on the the marginal benefit of manipulation: (i) an increase in k makes the coefficient $k\delta + 1 - k$ more sensitive to δ, which tends to increase the marginal benefit of manipulation, but also (ii) an increase in k decreases the magnitude of the coefficient $k\delta + 1 - k$, which tends to decrease the marginal benefit of manipulation. When the first effect dominates, a higher k induces the politician to also choose a higher δ. When the second effect dominates, a higher k induces the politician to choose a lower δ.

3.4 Comparative statics

In this section we show how the equilibrium levels of k^* and δ^* vary with the parameters of the model. There are two parameters of interest: (i) the composite parameter $\alpha := (1-\lambda)\alpha_x/\alpha_z > 0$, which measures how responsive reporters would be to their signals absent manipulation, and (ii) the politician's cost of manipulation $c > 0$.

To see how the equilibrium k^* and δ^* vary with α and c, observe from (24) that we can write the reporters' best response as $k(\delta;\alpha)$ independent of the politician's cost c. Likewise, from (17) we can write the politician's best response as $\delta(k;c)$ independent of the composite parameter α. The unique intersection of these curves, as shown in Figure 1, determines the equilibrium coefficients $k^*(\alpha;c)$ and $\delta^*(\alpha;c)$ in terms of these parameters. Since α enters only the reporters' best response, changes in α shift the reporters' best response $k(\delta;\alpha)$ along an unchanged $\delta(k;c)$ for the politician. Likewise, since c enters only the politician's best response, changes in c shift the politician's best response $\delta(k;c)$ along an unchanged $k(\delta;\alpha)$ for the reporters.

Lemma 3. In equilibrium:

(i) The reporters' response $k^*(\alpha,c)$ is strictly increasing in α.

(ii) The politician's manipulation $\delta^*(\alpha,c)$ is strictly increasing in α if and only if

$$\alpha < \hat{\alpha}(c) \tag{30}$$

where $\hat{\alpha}(c)$ is the smallest α such that $k^*(\alpha,c) \geq \hat{k}(c)$.

We illustrate this result in Figure 2 which shows the reporters' equilibrium response k^* (left panel) and politician's equilibrium manipulation δ^* (right panel) as functions of the composite α for the case of low costs of manipulation $c < 1$ and high costs of manipulation $c > 1$. If $c < 1$ then we know from Lemma 2 that $k^* \leq c = \hat{k}(c)$ so that the politician's $\delta(\alpha;c)$ curve is increasing and so, k^* and δ^* unambiguously increase or decrease together. Alternatively, if $c > 1$, then the level of k^* matters, and this depends on the level of α. If α is low then k^* will also be low so that k^* and δ^* remain moving together following a change in α. But if α is high enough to make k^* higher than $\hat{k}(c)$, then k^* and δ^* will move in opposite directions following a change in α.

Lemma 4. In equilibrium:

(i) The politician's manipulation $\delta^*(\alpha,c)$ is strictly decreasing in c.

(ii) The reporters' response $k^*(\alpha,c)$ is strictly increasing in c if and only if

$$c < \hat{c}(\alpha) \tag{31}$$

where $\hat{c}(\alpha)$ is the smallest c such that $\delta^*(\alpha,c) \leq \hat{\delta}(\alpha)$.

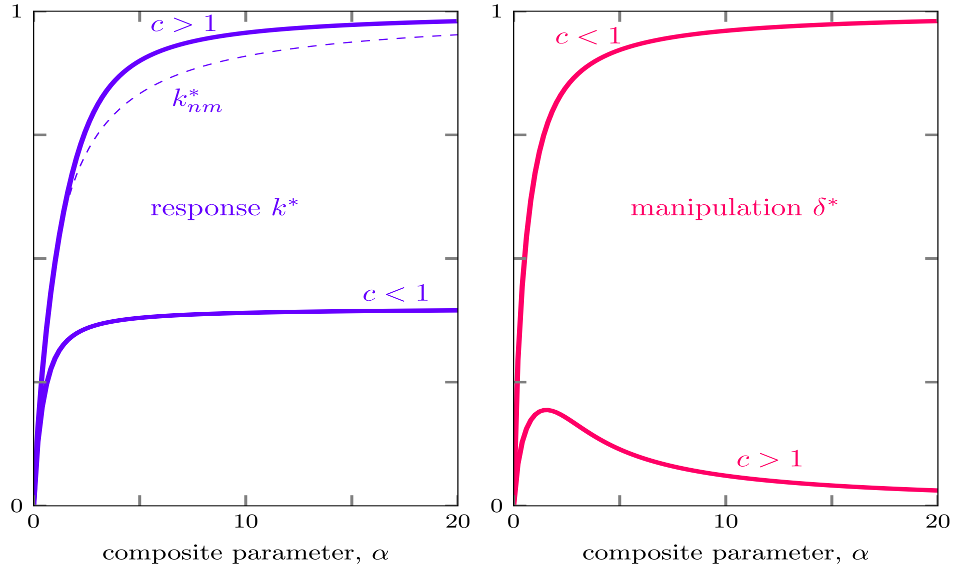

Figure 2: Changes in the composite parameter, α.

Reporters' equilibrium response k^* (left panel) and politician's equilibrium manipulation δ^* (right panel) as functions of the composite α for various levels of the cost of manipulation c. The reporters' k^* is increasing in α and asymptotes to $\min[c, 1]$ as $\alpha \to \infty$. If $c < 1$ then in equilibrium the politician's marginal benefit of manipulation is increasing in k so δ^* increases with k^* as α rises and asymptotes to one as $k^* \to c$. If $c > 1$ then for high enough α we have $k^* > \hat{k}(c)$ so that the politician's marginal benefit of manipulation is decreasing in k so that δ^* starts to decrease and asymptotes to zero as $k^* \to 1$.

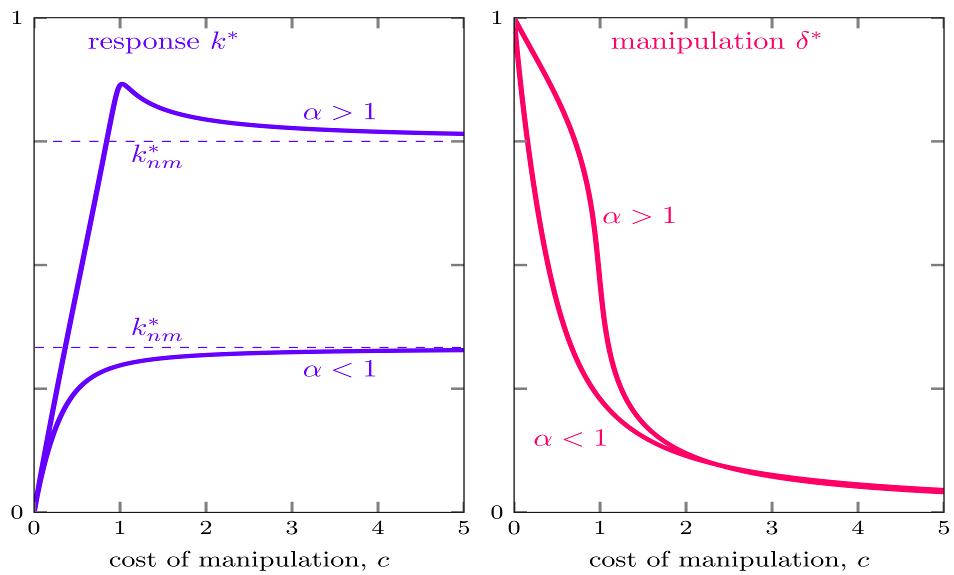

Figure 3: Changes in the cost of manipulation, c.

Reporters' equilibrium response k^* (left panel) and politician's equilibrium manipulation δ^* (right panel) as functions of the politician's cost of manipulation c for various levels of the composite α. The politician's δ^* is decreasing in c and asymptotes to zero as $c \to \infty$. If $\alpha < 1$, the precision effect dominates so that as δ^* decreases the reporters' k^* increases and asymptotes to k^*_{nm} from below as $c \to \infty$. If $\alpha > 1$ then for high enough c we have $\delta^* < \hat{\delta}(\alpha)$ so that the correlation effect begins to dominate at which point k^* starts to decrease and asymptotes to k^*_{nm} from above as $c \to \infty$.

We illustrate this result in Figure 3 which shows the reporters' equilibrium response k^* (left panel) and politician's equilibrium manipulation δ^* (right panel) as functions of the cost of manipulation c for the case of low $\alpha < 1$ and high $\alpha > 1$. If $\alpha < 1$ then we know from Lemma 1 that the precision effect dominates so that the reporters' $k(\delta; \alpha)$ curve is decreasing and so, as functions of c, the equilibrium k^* and δ^* move in opposite directions following a change in c. Alternatively, if $\alpha > 1$, then the level of δ^* matters, and this depends on the level of c. If c is low, then δ^* will be high so the precision effects continues to dominate meaning that k^* and δ^* move in opposite directions following a change in c. But if c is high enough to make δ^* low, then the correlation effect will dominate and k^* and δ^* will move in the same direction following a change in c.

Now that we have a complete understanding of the comparative statics of the model, we can turn to our main interest, the welfare effects of changes in the precision of information and the costs of information manipulation. We begin this analysis in Section 4 with welfare results for the politician. In Section 5 below we then turn to welfare results for the reporters and for the citizens who consume the reporting.

4 Politician's welfare

In this section we provide three results on the welfare effects of changes in the precision of information and the costs of information manipulation for the politician. First, we show that the politician need not benefit from lower costs of information manipulation at the margin. Second, we show that moreover the politician may not benefit from manipulation *at all*. We provide sufficient conditions for which the politician's manipulation *backfires* in the sense that they would want to be able to credibly commit to not use their manipulation technology. Third, we provide sufficient conditions for which the politician does benefit from manipulation, despite the fact that manipulation is costly and has no effect on the reporters' posterior expectations.

4.1 Does the politician benefit when manipulation is cheaper?

Not necessarily. The *direct* effect of a reduction in the cost of manipulation c is to make the politician better off. But there is also an *indirect* effect through the reporter's equilibrium response coefficient k^*, and this can be against the politician's interests. When this indirect effect is strong enough, the net effect is that a reduction in c makes the politician worse off.

To state this result, let v^* denote the politician's ex-ante expected payoff when they can manipulate information, i.e., the expectation of the politician's payoff with respect to the prior that θ is normally distributed with mean z and precision α_z. The following result gives sufficient conditions for v^* to be increasing in the costs of manipulation:

Proposition 2.

(i) For each $\lambda < -1/2$ and $\alpha < \underline{\alpha} < 1$, there exists a cutoff cost c_1^* such that for all $c > c_1^*$ the politician's payoff v^* strictly increases in c.

(ii) For each $\lambda > +1/2$ and $\alpha > \overline{\alpha} > 1$, there exists a cutoff cost c_2^* such that for all $c > c_2^*$ the politician's payoff v^* strictly increases in c.

Supposing c is sufficiently high, there are then two scenarios under which cheaper manipulation can at the margin make the politician worse off. In the first scenario, the reporters' actions are strong strategic substitutes, $\lambda < -1/2$, and the composite parameter α is sufficiently low, $\alpha < \underline{\alpha}$. Equivalently, for given λ, the relative precision of the signal to the prior, α_x/α_z, is sufficiently low. In the second scenario, the reporters' actions are strong strategic complements, $\lambda > +1/2$, and the composite parameter α is sufficiently high, $\alpha > \overline{\alpha}$. Equivalently, for given λ, the relative precision α_x/α_z is sufficiently high.

To understand this result, let $v(k)$ denote the politician's value function with manipulation

$$v(k) := \max_{\delta\in[0,1]} V(\delta, k) \tag{32}$$

where $V(\delta, k)$ denotes the politician's ex-ante expected utility if they choose manipulation δ and the reporters have response coefficient k. That is

$$V(\delta, k) = \frac{1}{\alpha_z}(B(\delta, k) - C(\delta)) + \frac{1}{\alpha_x}k^2 \tag{33}$$

where as before $B(\delta, k) := (k\delta+1-k)^2$ and $C(\delta) := c\delta^2$. Evaluating $V(\delta, k)$ at the politician's best response $\delta(k)$ and collecting terms gives the value function

$$v(k) = V(\delta(k), k) = \frac{1}{\alpha_z}(1-k)^2\left(\frac{c}{c-k^2}\right) + \frac{1}{\alpha_x}k^2 \tag{34}$$

In this notation, $v^* = v(k^*)$. To obtain Proposition 2 we calculate the net effect of c on v^*.

Direct and indirect effects of c. The net effect of c on the politician's v^* can be written

$$\frac{dv^*}{dc} = \underbrace{v'(k^*)\frac{\partial k^*}{\partial c}}_{\text{indirect effect}} + \underbrace{\frac{\partial v(k^*; c)}{\partial c}}_{\text{direct effect}} \tag{35}$$

From the envelope theorem, the direct effect of higher costs of manipulation decreases the politician's payoff v^*. But there is also an indirect effect through the response coefficient k^* that depends on two terms: (i) the politician's marginal value $v'(k)$ evaluated at the equilibrium k^*, and (ii) how k^* varies with c. We discuss each of these two terms next.

Politician's marginal value $v'(k^*)$. The politician's marginal value $v'(k)$ evaluated at the equilibrium k^* is characterized by the following:

Lemma 5. The politician is made better off by a higher equilibrium k^* if and only if $\lambda < 0$. In particular:

$$v'(k^*) = -\frac{\lambda}{1-\lambda}\frac{2}{\alpha_x}k^* \tag{36}$$

In short, the sign of $v'(k^*)$ is the opposite of the sign of λ. If the reporters' actions are strategic substitutes, $\lambda < 0$, then an increase in the response coefficient makes the politician better off, $v'(k^*) > 0$. When individual reporters want to stand out from the crowd, the politician wants the reporters' k^* to be *higher* because this will help scatter their actions a_i. But if instead the reporters' actions are strategic complements, $\lambda > 0$, then an increase in the response coefficient makes the politician worse off, $v'(k^*) < 0$. When individual reporters want to follow the crowd, the politician wants k^* to be *lower* so that in effect the politician can herd the crowd back to their prior z and away from the true θ. Roughly speaking, the politician wants the response coefficient k^* to *amplify* the reporters' interactions, increasing the scatter in a_i when the reporters' actions are strategic substitutes but increasing the herding on a common A when the reporters' actions are strategic complements.

Effect of manipulation cost c on k^* From Lemma 4 above we know that an increase in the costs of manipulation c increases the equilibrium k^* if and only if $c < \hat{c}(\alpha)$. The intuition for this is that an increase in c shifts down the politician's best response $\delta(k; c)$ along an unchanged $k(\delta; \alpha)$ for the reporters. This unambiguously decreases δ^*. Then recall that a decrease in δ^* has two offsetting effects on the information received by reporters: (i) increasing the correlation of their x_i with the true θ, which tends to increase k, but also (ii) decreasing the correlation of their x_i with their prior z (and hence also with θ), which tends to decrease k. It turns out that if $\alpha < 1$ then the first effect always dominates so that an increase in c that decreases δ^* always increases k^*. If $\alpha > 1$ then the first effect dominates if and only if the costs of manipulation are sufficiently low. That is, if $\alpha > 1$, an increase in c increases k^* if and only if $c < \hat{c}(\alpha)$.

To summarize, a higher c can actually make the politician better off. For this to happen, it has to be the case that the indirect effect of c on v^* more than offsets the direct negative effect. This requires that the indirect effect is both positive and large in magnitude. There are two ways for the indirect effect to be positive, either: $\lambda < 0$ and $\alpha < 1$ so that $v'(k^*) > 0$ and k^* is increasing in c, this leads to scenario (i) in Proposition 2; or $\lambda > 0$ and $c > \hat{c}(\alpha)$ so that $v'(k^*) < 0$ and k^* is decreasing in c, this leads to scenario (ii) in Proposition 2. The additional conditions on the parameters in the proposition are sufficient for the indirect effect to be not just positive but also large enough in magnitude.

4.2 Does the politician benefit from manipulation at all?

So the politician can be made better off by marginally higher manipulation costs c. Next we show that moreover there are situations where the politician's manipulation completely *backfires* in the sense that the politician would be better off if they could credibly commit to not use their manipulation technology *at all*, i.e., they would be better off if $c = +\infty$.

Let v^*_{nm} denote the politician's ex-ante expected utility when they cannot manipulate information. We say that the politician's manipulation backfires if $v^* < v^*_{nm}$. The following result gives sufficient conditions for this:

Proposition 3.

(i) For each $\lambda < -1/2$ and $c < 1$, there exists a cutoff signal precision $\underline{\alpha}^*_x$ such that for all $\alpha_x < \underline{\alpha}^*_x$ the politician's manipulation backfires, $v^* < v^*_{nm}$.

(ii) For each $\lambda > +1/2$ and $c > 1$, there exists a cutoff signal precision $\overline{\alpha}^*_x > \underline{\alpha}^*_x$ such that for all $\alpha_x > \overline{\alpha}^*_x$ the politician's manipulation backfires, $v^* < v^*_{nm}$.

Again there are two scenarios depending on whether the reporters' actions are strategic substitutes or complements. But despite this apparent similarity, the first part of Proposition 3 refers to a situation with *low* costs of manipulation whereas both parts of Proposition 2 above refer to situations with relatively *high* costs of manipulation. We now find, perhaps surprisingly, that if the reporters' actions are strong strategic substitutes, $\lambda < -1/2$, the politician can be worse off with the manipulation technology *even if the costs of using it are very low*. One might think that endowing the politician with the ability to manipulate information at low cost would be to their advantage, but here we see that this need not be the case. No matter how low c is, if the reporters' actions are strong strategic substitutes, $\lambda < -1/2$, the politician will be worse off if the signal precision α_x is low enough.[6] Alternatively, if the reporters' actions are strong strategic complements, $\lambda > +1/2$, the politician's manipulation backfires if the costs of manipulation c and the intrinsic signal precision α_x are sufficiently high. This more closely echoes the second part of Proposition 2 above.

We develop intuition for this result in two steps. First, we show that a necessary condition for manipulation to backfire on the politician is for the reporters' response coefficient k^* to move against the politician's interests. Second, we establish conditions on the primitives such that this effect is sufficiently strong.

Necessary condition. Backfiring occurs when the introduction of the manipulation technology causes the reporters' equilibrium response coefficient to change from k^*_{nm} to k^* in a direction that makes the politician worse off. If $\lambda < 0$ the politician prefers higher k^* and

[6]This does not require that the signal precision α_x is *arbitrarily* low. So long as the intrinsic signal precision α_x is less than the prior precision α_z we can find situations where the politician is worse off by making λ negative enough and the costs of manipulation c low enough. In particular, we show in the Supplementary Online Appendix that for any $\alpha_x < \alpha_z$ and for each $\lambda < \underline{\lambda}^*$ where $\underline{\lambda}^*$ is given by

$$\underline{\lambda}^* = -\frac{\alpha_z + \alpha_x}{\alpha_z - \alpha_x} < -1$$

there is a cutoff $\underline{c}^* < 1$ such that for all $c < \underline{c}^*$ the politician's manipulation backfires, $v^* < v^*_{nm}$.

backfiring will occur when reporters are sufficiently less responsive to their signals than they would be absent manipulation, i.e., when k^* is sufficiently smaller than k^*_{nm}. If $\lambda > 0$ the politician prefers lower k^* and backfiring will occur when reporters are sufficiently more responsive to their signals than they would be absent manipulation, i.e., when k^* is sufficiently larger than k^*_{nm}.

To see this, let $v_{nm}(k)$ denote the politician's value function without manipulation

$$v_{nm}(k) := V(0, k) \leq \max_{\delta \in [0,1]} V(\delta, k) =: v(k) \tag{37}$$

where $V(\delta, k)$ again denotes the politician's ex-ante expected payoff with manipulation δ if the reporters' response coefficient is k. In this notation, $v^*_{nm} = v_{nm}(k^*_{nm})$.

We decompose the change in the politician's payoff as

$$v^* - v^*_{nm} = (v(k^*) - v_{nm}(k^*)) + (v_{nm}(k^*) - v_{nm}(k^*_{nm})) \tag{38}$$

Since $v(k) \geq v_{nm}(k)$ for all k, the first term in the decomposition (38) is not negative. So to obtain backfiring the second term $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ must be *sufficiently negative*. Now observe that this second term is a comparison of the function $v_{nm}(k)$ at two different points, k^* and k^*_{nm}, where $v_{nm}(k)$ is given by[7]

$$v_{nm}(k) = \frac{1}{\alpha_z}(1-k)^2 + \frac{1}{\alpha_x}k^2. \tag{39}$$

This quadratic in k decreases from $v_{nm}(0) = 1/\alpha_z$ till it reaches its global minimum at $k_{min} := \alpha_x/(\alpha_x + \alpha_z)$ and then increases to $v_{nm}(1) = 1/\alpha_x$. Now suppose the reporters' actions are strategic substitutes, $\lambda < 0$. Then $k^*_{nm} > k_{min}$ and so $v_{nm}(k)$ is strictly increasing on $(k^*_{nm}, 1)$. So if $\lambda < 0$ a necessary condition for $v_{nm}(k^*) - v_{nm}(k^*_{nm}) < 0$ is that $k^* < k^*_{nm}$. Similarly, if the reporters' actions are strategic complements, $\lambda > 0$, then $k^*_{nm} < k_{min}$ and so $v_{nm}(k)$ is strictly decreasing on $(0, k^*_{nm})$. So if $\lambda > 0$ a necessary condition for $v_{nm}(k^*) - v_{nm}(k^*_{nm}) < 0$ is that $k^* > k^*_{nm}$.

Conditions on the primitives. We now establish conditions on the primitives sufficient to ensure that the gap between $v_{nm}(k^*)$ and $v_{nm}(k^*_{nm})$ is indeed large enough that the politician's manipulation backfires. To do this we use:

Lemma 6. Reporters are less responsive to their signals with manipulation

$$k^*(\alpha, c) < k^*_{nm}(\alpha) \qquad \text{if and only if} \qquad c < c^*_{nm}(\alpha) \tag{40}$$

where

$$c^*_{nm}(\alpha) = \begin{cases} \dfrac{\alpha}{\alpha - 1}\left(\dfrac{\alpha}{\alpha+1}\right)^2 & \text{if } \alpha > 1 \\ +\infty & \text{if } \alpha \leq 1 \end{cases} \tag{41}$$

[7] This expression for $v_{nm}(k)$ can also be obtained as the limit of $v(k)$ from (34) as $c \to \infty$.

In other words, if the composite parameter $\alpha \leq 1$ then we know that $k^* < k^*_{nm}$ regardless of c but if $\alpha > 1$ then the reporters' k^* is less than k^*_{nm} only if c is low enough.[8]

Now observe from (39) that $v_{nm}(k)$ is a linear combination of the terms $(1-k)^2$ and k^2 with the relative importance of the k^2 term being decreasing in α_x. As α_x decreases, the function $v_{nm}(k)$ behaves more like the increasing k^2 term so that if $\lambda < 0$ and $k^* < k^*_{nm}$ then the second term in the decomposition $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ becomes more and more negative, eventually becoming negative enough that the net result is for the politician to be worse off. Similarly, as α_x increases, the function $v_{nm}(k)$ behaves more like the decreasing $(1-k)^2$ term so that if $\lambda > 0$ and $k^* > k^*_{nm}$ the second term in the decomposition $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ becomes more and more negative, eventually becoming negative enough that the net result is that the politician is again worse off.

To summarize, the politician need not benefit from information manipulation and moreover the politician can be made better off when the costs of manipulation are high. Given this, when if at all does the politician actually benefit from information manipulation?

4.3 So when *does* the politician benefit from manipulation?

Although information manipulation can backfire on the politician, there are nonetheless clear situations where the politician benefits from information manipulation. In particular:

PROPOSITION 4. The politician benefits from manipulation, $v^* > v^*_{nm}$, if either:

(i) The reporters' actions are strategic substitutes, $\lambda \leq 0$, and the costs of manipulation are sufficiently high, $c > c^*_{nm}(\alpha)$, or

(ii) The reporters' actions are strategic complements, $\lambda \geq 0$, and the costs of manipulation are sufficiently low, $c < c^*_{nm}(\alpha)$.

These sufficient conditions guarantee that the introduction of the manipulation technology changes the reporters' equilibrium response coefficient from k^*_{nm} to k^* in a direction that benefits the politician, i.e., increasing to $k^* > k^*_{nm}$ if $\lambda < 0$ or decreasing to $k^* < k^*_{nm}$ if $\lambda > 0$. Notice that in the knife-edge special case with no interactions among reporters, $\lambda = 0$, the politician benefits from manipulation regardless of c.

To see this more formally, recall the decomposition (38) above. Since $v(k) \geq v_{nm}(k)$ for all k, the first term is not negative, so for the politician to gain it is sufficient that the second term $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ is positive. When the reporters' actions are strategic substitutes, $\lambda < 0$, $v_{nm}(k)$ is strictly increasing on $(k^*_{nm}, 1)$ and hence $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ is positive if $k^* > k^*_{nm}$. From Lemma 6 we know that $k^* > k^*_{nm}$ if and only if $c > c^*_{nm}(\alpha)$. Similarly, when the reporters' actions are strategic substitutes, $\lambda > 0$, $v_{nm}(k)$ is strictly decreasing on

[8] The function $c^*_{nm}(\alpha)$ is at first steeply decreasing in α, crosses $c^*_{nm}(\alpha) = 1$ and then reaches a minimum before increasing again, approaching $c = 1$ from below as $\alpha \to \infty$. So in the limit as $\alpha \to \infty$, the question of whether or not the equilibrium k^* is less than k^*_{nm} reduces to whether or not c is more or less than 1.

$(0, k^*_{nm})$ and hence $v_{nm}(k^*) - v_{nm}(k^*_{nm})$ is positive if $k^* < k^*_{nm}$, which from Lemma 6 happens if and only if $c < c^*_{nm}(\alpha)$.

When does the politician benefit the most? Figure 4 illustrates both benefits from manipulation and backfiring in the same figure. The top row shows the politician's benefit from manipulation $v^* - v^*_{nm}$ as a function of the intrinsic precision α_x for the case of low costs of manipulation, $c < 1$ (in blue), and the case of high costs of manipulation, $c > 1$ (in red). The bottom row shows the underlying levels v^* for $c < 1$ (in blue) and $c > 1$ (in red) along with the politician's welfare v^*_{nm} in the absence of manipulation (dashed black). The left column shows the results when the reporters' actions are strong strategic substitutes, $\lambda < -1/2$. The right column shows the results when the reporters' actions are strong strategic complements, $\lambda > +1/2$.

A striking feature of Figure 4 is that the politician gains the most from manipulation when c is low and α_x is high, regardless of λ. In particular:

Remark 1. Regardless of λ, the politician's payoff has limits

$$\lim_{\alpha_x \to 0^+} v^* = \lim_{\alpha_x \to 0^+} v^*_{nm} = \frac{1}{\alpha_z} \tag{42}$$

$$\lim_{\alpha_x \to \infty} v^* = \max\left[0\,,\ \frac{1-c}{\alpha_z}\right] \qquad \text{and} \qquad \lim_{\alpha_x \to \infty} v^*_{nm} = 0 \tag{43}$$

The overall effect of an increase in the intrinsic precision α_x on the politician' payoff depends on whether the politician can manipulate information and if so at what cost.[9] In particular, if $c < 1$ then v^* asymptotes to $(1-c)/\alpha_z > 0$ as $\alpha_x \to \infty$ whereas v^*_{nm} asymptotes to zero. If instead $c > 1$, then both v^* and v^*_{nm} asymptote to zero. These asymptotes are independent of λ. In this sense, the politician's asymptotic benefit from manipulation is particularly large when c is low.

These results characterize the welfare effects of changes in the precision of information and the costs of information manipulation *on the politician*. We now turn to the welfare of the reporters who are on the other side of the politician's manipulation and, perhaps more importantly, to the welfare of the citizens at large who consume that reporting.

5 Reporters' and citizens' welfares

In this section we provide two sets of results on the welfare effects of changes in the precision of information and the costs of information manipulation on the reporters and on the citizens at large who consume their reporting. First we show that while manipulation always makes

[9] In our Supplementary Online Appendix we show that for $\lambda > -1$ the politician's payoff is strictly decreasing in α_x and for $\lambda < -1$ the politician's payoff is strictly decreasing in α_x if α_x is high enough.

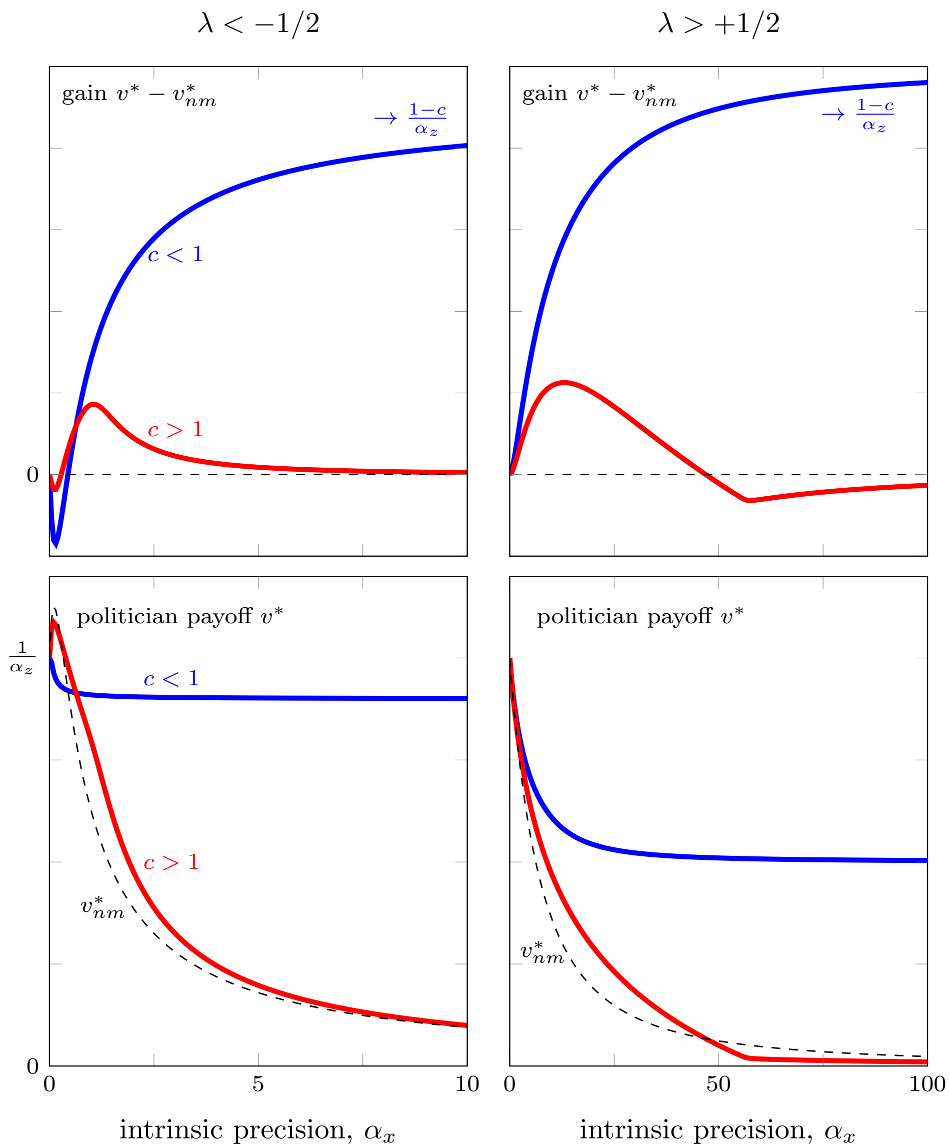

Figure 4: Politician benefits most when c is low and α_x is high.

Politician's benefit from manipulation $v^* - v^*_{nm}$ (top row) and payoff v^* (bottom row) as functions of the intrinsic precision α_x for various costs of manipulation c when the reporters' actions are strong strategic substitutes $\lambda < -1/2$ (left column) or strong strategic complements $\lambda > 1/2$ (right column). The politician's payoff absent manipulation v^*_{nm} asymptotes to zero as $\alpha_x \to \infty$. If $c > 1$ the politician's payoff with manipulation v^* also asymptotes to zero but if $c < 1$ then v^* asymptotes to $(1-c)/\alpha_z > 0$ so that the politician benefits. The politician benefits the most when when c is low and α_x is high. In the left column we use $\lambda < -1$ to highlight that for this parameter setting v^* and v^*_{nm} need not be monotonic in α_x.

the reporters worse off, there is a narrow set of circumstances under which information manipulation, perhaps surprisingly, makes the citizens *better off*. Second we show that, even when manipulation makes both reporters and citizens worse off, an increase in the precision of information need not make either the reporters or the citizens better off at the margin, and, moreover can drive the reporters' and citizens' welfare in opposite directions.

5.1 Are reporters and citizens worse off with manipulation?

We first consider the basic question of whether the politician's manipulation necessarily makes the reporters and citizens worse off. The reporters are the active recipients of the politician's manipulation. The citizens at large are passive consumers of the reporting. The citizens' loss is $\int_0^1 (a_i - \theta)^2 \, di$. As discussed in Section 2.1 above, the key idea here is that when the reports are imprecise about θ (i.e., the reports a_i are scattered around θ), the citizens at large will be demotivated from voting or protesting against the politician.

Let $l_{\mathcal{R}}^*$ and $l_{\mathcal{C}}^*$ denote the reporters' and citizens' ex-ante expected losses evaluated at the equilibrium k^*, δ^* determined in the game between the politician and the reporters. Let $l_{\mathcal{R},nm}^*$ and $l_{\mathcal{C},nm}^*$ denote the reporters' and citizens' ex-ante expected losses when the politician cannot manipulate information. Our main result here is:

PROPOSITION 5.

(i) The reporters are worse off with manipulation, $l_{\mathcal{R}}^* > l_{\mathcal{R},nm}^*$.

(ii) The citizens are worse off with manipulation, $l_{\mathcal{C}}^* > l_{\mathcal{C},nm}^*$, if $\lambda > -1$.

(iii) The citizens are *better off* with manipulation, $l_{\mathcal{C}}^* < l_{\mathcal{C},nm}^*$, if $\lambda < -1$ and $\alpha_x < \widehat{\alpha}_x^{**}$.

So the reporters are always worse off with manipulation. Whether the citizens are worse off or not depends on the strategic interactions among the reporters. If the reporters' actions are not strong strategic substitutes, $\lambda > -1$, the citizens are also unambiguously worse off with manipulation. But if the reporters' actions are strong strategic substitutes, $\lambda < -1$, and if in addition the intrinsic precision of reporters' signals is low enough, $\alpha_x < \widehat{\alpha}_x^{**}$, then, perhaps surprisingly, the citizens are in fact *better off* with manipulation.

Reporters. To understand this result, we first define the reporters' loss function

$$l_{\mathcal{R}}(\delta) := \min_{k \in [0,1]} L_{\mathcal{R}}(k, \delta) \tag{44}$$

where $L_{\mathcal{R}}(k, \delta)$ denotes the reporters' ex-ante expected loss, i.e., the expectation of (1) with respect to the prior that θ is normally distributed with mean z and precision α_z, if they choose k when the politician has manipulation δ. This works out to be

$$L_{\mathcal{R}}(k, \delta) = \frac{1 - \lambda}{\alpha_z} B(\delta, k) + \frac{1}{\alpha_x} k^2 \tag{45}$$

where again $B(\delta, k) := (k\delta + 1 - k)^2$ denotes the politician's benefit from manipulation. Evaluating at the reporters' best response $k(\delta)$ and collecting terms gives

$$l_{\mathcal{R}}(\delta) = L_{\mathcal{R}}(k(\delta), \delta) = \frac{1}{\alpha_x}\left(\frac{k(\delta)}{1-\delta}\right) = \frac{(1-\lambda)}{(1-\delta)^2(1-\lambda)\alpha_x + \alpha_z} \tag{46}$$

The reporters' equilibrium loss is $l_{\mathcal{R}}^* = l_{\mathcal{R}}(\delta^*)$ where δ^* is the politician's equilibrium manipulation. The reporters' loss when the politician cannot manipulate information is then $l_{\mathcal{R},nm}^* = l_{\mathcal{R}}(0)$. Since $l_{\mathcal{R}}(\delta)$ is increasing in δ, we have that the reporters are unambiguously worse off when the politician can manipulate, strictly so whenever $\delta^* > 0$.

Citizens. Recall that the citizens evaluate outcomes according to the loss

$$\int_0^1 (a_i - \theta)^2\, di = (A - \theta)^2 + \int_0^1 (a_i - A)^2\, di \tag{47}$$

So the citizens are at their bliss point if the reporters all produce $a_i = \theta$.

Now let $L_{\mathcal{C}}(k, \delta)$ denote the citizens' ex ante expected loss, i.e., the expectation of (47) with respect to the prior that θ is normally distributed with mean z and precision α_z, if the *reporters* choose k when the politician has manipulation δ. This works out to be

$$L_{\mathcal{C}}(k, \delta) = \frac{1}{\alpha_z} B(\delta, k) + \frac{1}{\alpha_x} k^2 \tag{48}$$

The citizens' equilibrium loss is then $l_{\mathcal{C}}^* = L_{\mathcal{C}}(k^*, \delta^*)$ where k^* is the reporters' equilibrium response and δ^* is the politician's equilibrium manipulation. The citizens' loss when the politician cannot manipulation information is similarly $l_{\mathcal{C},nm}^* = L_{\mathcal{C}}(k_{nm}^*, 0)$.

Wedge between reporters' and citizens' losses. Comparing (48) and (45) we see that

$$L_{\mathcal{C}}(k, \delta) = L_{\mathcal{R}}(k, \delta) + \frac{\lambda}{\alpha_z} B(\delta, k) \tag{49}$$

In the special case $\lambda = 0$, where the reporters care only about accurate reporting with no interactions amongst themselves, the citizens' loss and the reporters' loss *coincide*. More generally, since $B(\delta, k) \geq 0$, the citizens' loss is larger than the reporters' loss whenever $\lambda > 0$ and is less than the reporters's loss whenever $\lambda < 0$.

Intuitively, an incentive to coordinate, $\lambda > 0$, means that individual reporters respond more to their common prior z than its underlying precision warrants. Therefore, from the citizens' point of view, the reporters are excessively responsive to their prior and hence *under-responsive* to the information contained in their signals. For example, if $\lambda \to 1$ the reporters can be quite content when they are producing similar reports, $a_i \approx A$, even if those reports are far from θ and hence very unsatisfactory from the citizens' point of view.

Sufficient condition for citizens' welfare to be aligned with reporters'. Evaluating $L_{\mathcal{R}}(k, \delta)$ at the reporters' best response $k(\delta)$ we can then write

$$l_{\mathcal{C}}(\delta) := L_{\mathcal{C}}(k(\delta), \delta) = l_{\mathcal{R}}(\delta) + \frac{\lambda \alpha_z}{(1-\lambda)^2}\, l_{\mathcal{R}}(\delta)^2 \tag{50}$$

The comparison of the citizens' equilibrium loss with and without manipulation is then reduced to comparing $l^*_{\mathcal{C}} = l_{\mathcal{C}}(\delta^*)$ and $l^*_{\mathcal{C},nm} = l_{\mathcal{C}}(0)$. The effect of manipulation on the citizens is then given by the derivative

$$l'_{\mathcal{C}}(\delta) = l'_{\mathcal{R}}(\delta) \left[1 + \frac{2\lambda \alpha_z}{(1-\lambda)^2}\, l_{\mathcal{R}}(\delta)\right] \tag{51}$$

This expression shows that the effect of δ on the reporters' and the citizens' losses share the same sign when the term in square brackets on the right is positive. If $\lambda > 0$, i.e., if the reporters' actions are strategic complements, this term must be positive. Even if $\lambda < 0$ the term in square brackets can remain positive so long as the reporters' loss $l_{\mathcal{R}}(\delta)$ is not too large in magnitude. Substituting for $l_{\mathcal{R}}(\delta)$ and collecting terms we find that a sufficient condition of this term to be positive is $\lambda > -1$. Since the reporters' loss is unambiguously increasing in δ, we can then conclude that if $\lambda > -1$ the citizens' loss is also increasing in δ and hence the citizens are worse off with manipulation, $l^*_{\mathcal{C}} = l_{\mathcal{C}}(\delta^*) > l_{\mathcal{C}}(0) = l^*_{\mathcal{C},nm}$.

Welfare outcomes need not be aligned. But if the reporters' actions are sufficiently strong strategic substitutes, $\lambda < -1$, the welfare outcomes for the citizens need not be aligned with those of the reporters. Indeed their interests can diverge quite starkly. While the politician's manipulation always makes the reporters worse off, the manipulation can in fact make the citizens *better off* if the intrinsic signal precision is low enough, $\alpha_x < \widehat{\alpha}_x^{**}$.

To understand this, first notice that when $\lambda < -1$, the reporters have a strong incentive to differentiate themselves from one another and their response k to their idiosyncratic signals is, from the citizens' point of view, more than is warranted by the underlying precision of their signals. This is especially problematic for the citizens when the signals are imprecise, i.e., when α_x is very low. By reducing k, the politician's manipulation then "corrects" for this, which makes the citizens better off than they would be absent manipulation.[10]

In this scenario, where the citizens are better off with manipulation, it must also be the case that the manipulation is backfiring on the politician. This is because the politician is trying to increase the dispersion in (47) and because their manipulation is costly. But

[10] The region of the parameter space where the citizens are better off with manipulation is in a sense quite small. The critical point turns out to be

$$\widehat{\alpha}_x^{**} = -\left(\frac{1+\lambda}{(1-\lambda)^2}\right)\alpha_z, \qquad \lambda < -1$$

This is maximized at $\lambda = -3$ for which $\widehat{\alpha}_x^{**} = \alpha_z/8$. Even allowing the value of λ most favorable to this scenario, it only occurs if the intrinsic signal precision α_x is less than one-eighth of the prior precision α_z.

the converse is not true. Manipulation can backfire on the politician even without that manipulation making the citizens better off. In other words, backfiring occurs for a larger set of parameters. From Proposition 5, a necessary condition for citizens to be better of is $\lambda < -1$. By contrast from Proposition 3 above we know that manipulation can backfire on the politician if $\lambda < -1/2$ (i.e., for a larger set of λ) if $c < 1$.[11] And moreover manipulation can backfire on the politician if $\lambda > +1/2$ if $c > 1$. In other words, manipulation can backfire on the politician even in situations where the manipulation also makes the citizens worse off.

We now turn to the effects of an increase in the intrinsic precision α_x.

5.2 Do reporters and citizens benefit from more information?

Not always. In particular, we show that when the politician manipulates information, an increase in α_x that would, *absent manipulation*, make both the reporters and citizens better off, can end up making them worse off instead. Moreover we show that, if the reporters' actions are sufficiently strong strategic substitutes, then changes in α_x can move the reporters' and citizens' welfare in opposite directions.

We begin with the effects of an increase in α_x on the reporters and then use these results to determine the effects of an increase in α_x on the citizens.

Effects of α_x on reporters' loss. Evaluating the reporters' loss (46) at the equilibrium manipulation δ^* and using the comparative statics of δ^* given in Lemma 3 we obtain:

PROPOSITION 6. The reporters' loss $l_{\mathcal{R}}^*$ is strictly decreasing in α_x if and only if $\alpha_x < \alpha_x^{**}$. For $c > 1$ the critical point $\alpha_x^{**} = +\infty$.

So if $c > 1$ the reporters' loss is strictly decreasing in α_x. But if $c < 1$ the reporter's loss is ∪-shaped in α_x, decreasing at first, reaching a minimum at an interior critical point α_x^{**}, then increasing. In short, an increase in α_x does not always make the reporters better off.

To understand this result, first observe that an increase in the intrinsic precision α_x has both a direct effect on the reporters' loss and an indirect effect through the politician's manipulation δ^*. If the politician can not manipulate, then only the direct effect is operative and an increase in α_x reduces the reporters' loss. Now suppose that the politician can manipulate. From Lemma 3 above, we know that the indirect effect of α_x through the politician's δ^* now depends on the magnitude of c. In particular, if the costs of manipulation are relatively high, $c > 1$, then δ^* is decreasing in α_x and so the direct and indirect effects reinforce one another so that the reporters' loss is unambiguously decreasing in α_x. But if the costs of manipulation are relatively low, $c < 1$, then δ^* is decreasing in α_x if and only if

[11] From this we can also conclude that if $\lambda < -1$ and $c < 1$ then the intrinsic precision sufficient to make the citizens better off is lower than the intrinsic precision sufficient to induce backfiring, i.e., $\widehat{\alpha}_x^{**} < \underline{\alpha}_x^{*}$.

α_x is sufficiently low. We show in the Appendix that if $c < 1$ there is a finite critical point α_x^{**} such that the reporters' loss is strictly decreasing in α_x if and only if $\alpha_x < \alpha_x^{**}$. In other words, the direct effect of α_x on the reporters' loss dominates when α_x is low but the indirect effect via the politician's manipulation δ^* dominates when α_x is sufficiently high.

Asymptotic reporters' loss.

REMARK 2. The reporters' equilibrium loss has limits

$$\lim_{\alpha_x \to 0^+} l_{\mathcal{R}}^* = \lim_{\alpha_x \to 0^+} l_{\mathcal{R},nm}^* = \frac{1-\lambda}{\alpha_z} \tag{52}$$

$$\lim_{\alpha_x \to \infty} l_{\mathcal{R}}^* = \begin{cases} \dfrac{1-\lambda}{\alpha_z} & \text{if } c < 1 \\ 0 & \text{if } c > 1 \end{cases} \quad \text{and} \quad \lim_{\alpha_x \to \infty} l_{\mathcal{R},nm}^* = 0 \tag{53}$$

Absent manipulation the reporters' loss is strictly decreasing from $l_{\mathcal{R},nm}^* = (1-\lambda)/\alpha_z$ to 0. If $c > 1$ the loss with manipulation is also strictly decreasing from $l_{\mathcal{R}}^* = (1-\lambda)/\alpha_z$ to 0 so that if $c > 1$ the gap between $l_{\mathcal{R}}^*$ and $l_{\mathcal{R},nm}^*$ becomes negligible in the limit as as $\alpha_x \to \infty$. But if $c < 1$ the loss with manipulation reaches a minimum at α_x^{**} and then starts to increase, returning to $(1-\lambda)/\alpha_z$ in the limit as $\alpha_x \to \infty$, i.e., the same loss the reporters would have if $\alpha_x = 0$. In this scenario, even though the intrinsic precision of their signals is extremely high, the reporters have the same loss as if they had no information other than their prior.

We now use these results to determine the effects of an increase in α_x on the citizens.

Effects of α_x on citizens' loss. Evaluating the expression for the citizens' loss in (50) at the equilibrium manipulation δ^* gives

$$l_{\mathcal{C}}^* = l_{\mathcal{R}}^* + \frac{\lambda \alpha_z}{(1-\lambda)^2} l_{\mathcal{R}}^{*2} \tag{54}$$

Hence the effects of α_x on the citizens' equilibrium loss are given by the total derivative

$$\frac{dl_{\mathcal{C}}^*}{d\alpha_x} = \frac{dl_{\mathcal{R}}^*}{d\alpha_x} \left[1 + \frac{2\lambda \alpha_z}{(1-\lambda)^2} l_{\mathcal{R}}^* \right] \tag{55}$$

This expression is convenient because all the effects of α_x enter $l_{\mathcal{C}}^*$ only through $l_{\mathcal{R}}^*$, which we have just characterized in Proposition 6 above. This gives:

LEMMA 7. The citizens' loss $l_{\mathcal{C}}^*$ and the reporters' loss $l_{\mathcal{R}}^*$ move in the same direction in response to changes in α_x if and only if either (i) $\lambda > -1$, or (ii) $\lambda < -1$ and $\alpha_x \in (\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$. For $c > 1$, $\overline{\alpha}_x^{**} = +\infty$.

From equation (55) the citizens' loss $l_{\mathcal{C}}^*$ and the reporters' loss $l_{\mathcal{R}}^*$ move in the same direction in response to changes in α_x if and only if the term in square brackets, $1+\frac{2\lambda\alpha_z}{(1-\lambda)^2}\,l_{\mathcal{R}}^*$, is positive. Moreover recall from the discussion following Proposition 5 that if $\lambda > -1$ the term in square brackets must be positive so that the citizens' loss and the reporter's loss move in the same direction in response to changes in α_x. If instead $\lambda < -1$ the term in square brackets can still be positive but only if the level of the reporters' loss $l_{\mathcal{R}}^*$ is also sufficiently low. Using Proposition 6, there are then two ways to get a sufficiently low loss for the reporters: either $c > 1$ and α_x is sufficiently high, or $c < 1$ and α_x is neither too high nor too low. More precisely, for each $\lambda < -1$ and each c there is an interval of the form $(\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$ such that if and only if $\alpha_x \in (\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$ the reporters' loss $l_{\mathcal{R}}^*$ is low enough that the citizens' loss and the reporters' loss move in opposite directions in response to changes in α_x. If $c > 1$ then the interval is unbounded above, $\overline{\alpha}_x^{**} = +\infty$, so that the loss functions move in opposite directions when the intrinsic precision is low, $\alpha_x < \underline{\alpha}_x^{**}$, but move in the same direction when the intrinsic precision is high, $\alpha_x > \underline{\alpha}_x^{**}$. Alternatively if $c < 1$ then the interval is bounded above, $\overline{\alpha}_x^{**} < +\infty$, so that the loss functions move in opposite directions when the intrinsic precision is either low, $\alpha_x < \underline{\alpha}_x^{**}$, or high, $\alpha_x > \overline{\alpha}_x^{**}$, but move in the same direction for moderate levels of the intrinsic precision, $\alpha_x \in (\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$.

Combing this lemma with Proposition 6 gives us a complete characterization of the citizens' loss $l_{\mathcal{C}}^*$ in terms of the underlying parameters. In particular:

Proposition 7.

(i) For each $\lambda > -1$ the citizens' loss $l_{\mathcal{C}}^*$ is either

  (a) Strictly decreasing in α_x, if $c > 1$, or

  (b) Strictly decreasing in α_x if and only if $\alpha_x < \alpha_x^{**}$, if $c < 1$.

(ii) For each $\lambda < -1$ the citizens' loss $l_{\mathcal{C}}^*$ is either

  (a) Strictly increasing in α_x if and only if $\alpha_x < \underline{\alpha}_x^{**}$, if $c > 1$, or

  (b) Strictly increasing in α_x for $\alpha_x < \underline{\alpha}_x^{**}$, decreasing in α_x for $\alpha_x \in (\underline{\alpha}_x^{**}, \alpha_x^{**})$, increasing in α_x for $\alpha_x \in (\alpha_x^{**}, \overline{\alpha}_x^{**})$, and decreasing in α_x for $\alpha_x > \overline{\alpha}_x^{**}$, if $c < 1$.

Part (i) here follows because if $\lambda > -1$ the citizens' loss responds to changes in α_x in the same way as the reporters' loss responds to changes in α_x and from Proposition 6 we know that the reporter's loss is strictly decreasing in α_x if $c > 1$ or ∪-shaped in α_x if $c < 1$, so the citizens' loss inherits these properties. Part (ii) follows because if $\lambda < -1$ then the citizens' loss responds in the opposite direction to the reporters' loss if $\alpha_x \notin (\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$. In turn if $c > 1$ we know that the reporters' loss is strictly decreasing in α_x and because the citizens' loss is moving in the opposite direction if and only if $\alpha_x < \underline{\alpha}_x^{**}$ this means that the citizens' loss is *increasing* if and only if $\alpha_x < \underline{\alpha}_x^{**}$. Alternatively if $c < 1$ we know from Proposition 6 that the reporter's loss is decreasing if and only if $\alpha_x < \alpha_x^{**}$. We show in the Appendix that

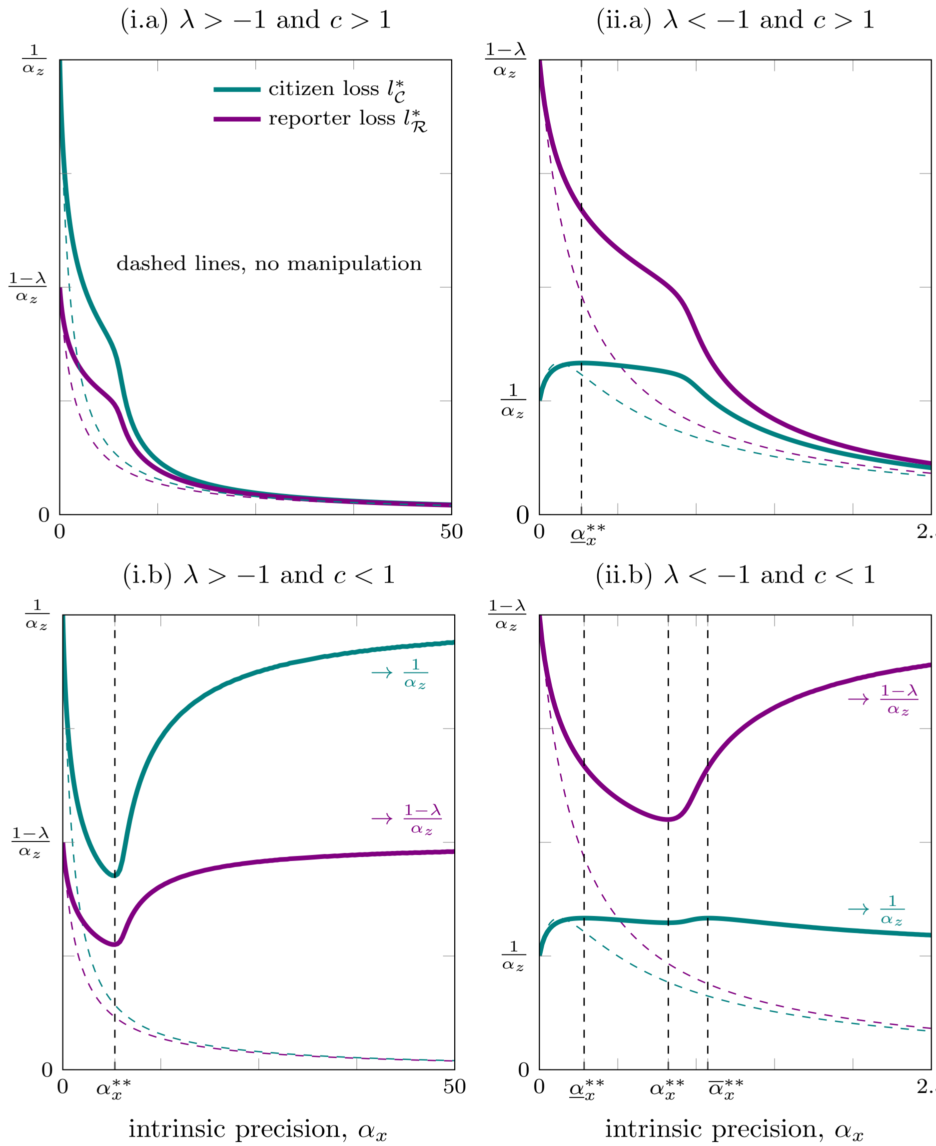

Figure 5: Citizens and reporters lose most when c is low and α_x is high.

Citizens' loss $l_{\mathcal{C}}^*$ and reporters' loss $l_{\mathcal{R}}^*$ as functions of α_x for $c > 1$ (top row) and $c < 1$ (bottom row) and for $\lambda > -1$ (left column) and $\lambda < -1$ (right column). If $\lambda > -1$ both loss functions move in the same direction in response to α_x. If $c > 1$ both loss functions are strictly decreasing (top left). If $c < 1$ both loss functions are $\cup$-shaped with critical point α_x^{**} (bottom left). If $\lambda < -1$ the loss functions move in the same direction only between the critical points $\underline{\alpha}_x^{**}$ and $\overline{\alpha}_x^{**}$ (right column). If $c < 1$ the citizens' loss asymptotes to $1/\alpha_z$ and the reporters' loss asymptotes to $(1-\lambda)/\alpha_z$. For the left column we use $\lambda > 0$ which implies that the reporters' loss is less than the citizens' loss. The colored dashed lines show the corresponding loss functions absent manipulation. If $\lambda < -1$ then for α_x sufficiently small the citizens are better off with manipulation.

the critical points satisfy

$$\underline{\alpha}_x^{**} < \alpha_x^{**} < \overline{\alpha}_x^{**}$$

Since the citizens' loss moves in the same direction as the reporters' loss on the interval $(\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**})$, which includes the reporters' critical point, α_x^{**}, we then know that inside this interval the citizens' and reporters' loss both decrease together for $\alpha_x \in (\underline{\alpha}_x^{**}, \alpha_x^{**})$ and both increase together for $\alpha_x \in (\alpha_x^{**}, \overline{\alpha}_x^{**})$. Outside this interval they move in opposite directions, implying that the citizens' loss is increasing for $\alpha_x < \underline{\alpha}_x^{**}$ but decreasing for $\alpha_x > \overline{\alpha}_x^{**}$. The left panel of Figure 5 illustrates. The left and right columns show the cases $\lambda > -1$ and $\lambda < -1$ respectively. The top and bottom rows show the cases $c > 1$ and $c < 1$ respectively. Each panel shows the loss of the citizens $l_{\mathcal{C}}^*$ and the reporters $l_{\mathcal{R}}^*$ as functions of α_x. The dashed lines demarcate the critical points α_x^{**} and $\underline{\alpha}_x^{**}, \overline{\alpha}_x^{**}$.

To summarize, there are two basic scenarios, depending on the magnitude of λ. In the first scenario, the reporters' actions are not strong strategic substitutes, $\lambda > -1$, and in response to marginal changes in α_x the citizens' loss $l_{\mathcal{C}}^*$ and the reporters' loss $l_{\mathcal{R}}^*$ always move in the same direction. In this first scenario, the welfare outcomes for the citizens and the reporters are essentially aligned, differing only in terms of the levels implied by the sign of λ. In the second scenario, the reporters' actions are strong strategic substitutes, $\lambda < -1$, and in response to marginal changes in α_x the citizens' loss $l_{\mathcal{C}}^*$ and the reporters' loss $l_{\mathcal{R}}^*$ move in opposite directions for certain values of α_x. In other words, just as the citizens may be better off if the politician can manipulate, there are also clear circumstances where the citizens' loss can be increasing in the intrinsic precision α_x, i.e., that a marginal increase in α_x makes the citizens worse off. This happens for example if $\lambda < -1$ and α_x is sufficiently low. The intuition for this is similar to the intuition for why manipulation can make the citizens better off, as discussed above. If $\lambda < -1$, the reporters' strong incentive to differentiate their reports from one another makes the reporters, from the citizens point of view, excessively responsive to their idiosyncratic signals. When the signals are imprecise, i.e., for low levels of α_x, a marginal improvement in α_x that increases the reporters' responsiveness to their signals only exacerbates this distortion and thereby makes the citizens worse off.

Asymptotic citizens' loss. Notice from Proposition 7 that for $c < 1$ and high levels of α_x, the citizens' loss $l_{\mathcal{C}}^*$ is increasing in α_x if $\lambda > -1$ but is decreasing in α_x if $\lambda < -1$. This does not mean that the citizens' loss $l_{\mathcal{C}}^*$ asymptotes to different levels depending on whether $\lambda > -1$ or not. In fact, we can show:

Remark 3. Regardless of λ, the citizens' equilibrium loss has limits

$$\lim_{\alpha_x \to 0^+} l^*_{\mathcal{C}} = \lim_{\alpha_x \to 0^+} l^*_{\mathcal{C},nm} = \frac{1}{\alpha_z} \tag{56}$$

$$\lim_{\alpha_x \to \infty} l^*_{\mathcal{C}} = \begin{cases} \dfrac{1}{\alpha_z} & \text{if } c < 1 \\ 0 & \text{if } c > 1 \end{cases} \qquad \text{and} \qquad \lim_{\alpha_x \to \infty} l^*_{\mathcal{C},nm} = 0 \tag{57}$$

As with the reporters' loss, the limit of the citizens' loss as $\alpha_x \to \infty$ is sensitive to the costs of manipulation c. If $c < 1$, as $\alpha_x \to \infty$ the citizens's loss $l^*_{\mathcal{C}}$ asymptotes to the same loss $1/\alpha_z$ the citizens would have if $\alpha_x = 0$. If $c > 1$, the citizens's loss $l^*_{\mathcal{C}}$ asymptotes to zero, the same limit of the citizens' loss without manipulation, $l^*_{\mathcal{C},nm}$.

We now turn to a specific interpretation of changes in α_x and c.

6 Social media

In this section we use our model to interpret the challenges posed by the rise of social media. We argue that the rise of social media makes possible a sudden "regime change" in the amount of manipulation. As a result, whether the social media revolution benefits the citizens depends crucially on whether the cost of manipulation can be kept above a critical level.[12] If this can be achieved, the rise of social media will decrease manipulation and make the citizens better off. Moreover, if in addition, the reporters are *sufficiently well coordinated*, the politician's manipulation will backfire, giving the politician incentives to invest in commitment devices that prevent them undertaking manipulation in the first place. But if instead the cost of manipulation falls below the critical level, then the rise of social media will greatly increase manipulation. As a consequence, the reporters' actions will increasingly reflect their prior alone and the citizens will be increasingly worse off.

Pessimism and optimism about new media technologies. The strategic use of information manipulation, whether it be blatant propaganda or more subtle forms of misdirection and obfuscation, is a timeless feature of human communication. The role that new technologies play in either facilitating or impeding this information manipulation is widely debated and optimism or pessimism on this issue seems to fluctuate as new technologies develop. For example, in the postwar era a pessimistic view emphasized the close connections between mass media technologies like print media, radio and cinema and the immersive propaganda

[12] To be clear, we identify *social welfare* with the welfare of the citizens, not with the reporters. The key idea here is that the reporters act as information intermediaries that collect and disseminate information to the citizens. We take the view that what matters in the end is how informed the citizens are about the fundamental which in turn depends on how accurate the reporters' actions (reports) are about the fundamental, i.e., the loss function of the citizens defined in equation (47) above.

of totalitarian regimes (e.g., Friedrich and Brzezinski, 1965; Arendt, 1973; Zeman, 1973). But in the 1990s and 2000s, a more optimistic view stressed the potential benefits of the internet and other, relatively more decentralized methods of communication, in undermining attempts to control information. This optimism seems to have reached its zenith during the "Arab Spring" protests against autocratic regimes in Tunisia, Egypt, Libya and elsewhere beginning in 2010. But increasingly the dominance of social media like Facebook and Twitter has led to renewed pessimism (e.g., Morozov, 2011). In particular, the apparent role of such platforms in facilitating the spread of misleading information during major political events, like the 2016 UK Brexit referendum and the 2016 US presidential election, has led to newly intense scrutiny of social media technologies (e.g., Faris et al., 2017).

The challenge of social media As emphasized by Bruns and Highfield (2012) and Allcott and Gentzkow (2017), social media technologies have two features that are particularly relevant. First, they have low barriers to entry and it has become increasingly easy to commercialize social media content through tools like Google and Facebook advertising. This has lead to a proliferation of new entrants that have been able to establish a viable market for their content. Second, social media technologies have significantly reduced the costs of collecting, reporting and disseminating information, and have thus lead to a rapidly expanded role of blogging and amateur journalism in the media industry. As emphasized by Fielder (2009) and Ward (2011), these new sources of information are not all subject to the same standards of accountability as traditional journalism. Moreover, the new social media technologies also mean that citizens consume much of their media content in a feed that both blurs distinctions between reliable and unreliable sources of information and also makes it easy for all kinds of news, real and fake, to "go viral" — to be rapidly retweeted or shared.

In the context of our model, we view these two features of social media as simultaneously (i) *increasing* the underlying, intrinsic signal precision α_x, but (ii) *decreasing* the costs of manipulation c. Social media technologies facilitate the entry of new media outlets and amateur journalists, which leads to a large increase in the news and information collected and disseminated. Absent manipulation, this would mean more signals and hence an increase in the intrinsic quality of information.[13] But the entry of low-accountability media outlets and amateur journalism and the technological ease with which stories can go viral, diffusing rapidly in the population, makes it easier for a politician to use spin and misdirection to undermine the credibility of the information reported in the media.

In short, the simultaneous change in α_x and c creates a tension. We now turn to analyze the net effect of this tension.

[13] Recall that the signals are $x_i = y + \varepsilon_i$ with ε_i representing idiosyncratic differences in how the common component y is interpreted. If the intrinsic signal precision α_x is high, there is in fact not much scope for different individuals to interpret the common y differently. In this sense, a high α_x corresponds to a high-quality information environment, absent manipulation.

6.1 "Regime changes" in the amount of manipulation.

Recall from Lemma 4 that an increase in c reduces the amount of information manipulation in equilibrium. The following proposition gives us additional information on the *size* of the reduction in manipulation.

PROPOSITION 8.

(i) For each $\alpha \leq 4$, the politician's equilibrium manipulation $\delta^*(\alpha, c)$ is smoothly decreasing in c with

$$\left.\frac{\partial \delta^*}{\partial c}\right|_{c=1} = -\frac{k^*(\alpha,1)}{(1-k^*(\alpha,1))(1+3k^*(\alpha,1))} < 0 \tag{58}$$

This derivative is strictly decreasing in α and approaches $-\infty$ as $\alpha \to 4$.

(ii) For each $\alpha > 4$, the politician's manipulation jumps discontinuously from $\overline{\delta}(\alpha)$ as $c \to 1^-$ to $\underline{\delta}(\alpha)$ as $c \to 1^+$ where

$$\underline{\delta}(\alpha)\,,\,\overline{\delta}(\alpha) = \frac{1}{2}\left(1 \pm \sqrt{1-(4/\alpha)}\right), \qquad \alpha \geq 4 \tag{59}$$

(iii) For any $c > 1$, the politician's equilibrium manipulation $\delta^*(\alpha, c)$ is bounded above by $1/2$ and can be made arbitrarily close to zero by making α large enough.

In particular, when c is close to the critical point $c = 1$, there will be an especially *large* reduction in manipulation when the composite parameter $\alpha = (1-\lambda)\alpha_x/\alpha_z$ is high, e.g., when the intrinsic signal precision α_x is high. This large reduction in manipulation close to $c = 1$ is most stark when $\alpha > 4$. In this case, a small increase from $c = 1-\varepsilon$ to $c = 1+\varepsilon$ will cause the amount of manipulation to *jump* from $\overline{\delta}(\alpha) > 1/2$ down to $\underline{\delta}(\alpha) < 1/2$. In the limit as $\alpha \to \infty$ we have $\overline{\delta}(\alpha) \to 1$ and $\underline{\delta}(\alpha) \to 0$ so that the manipulation jumps from $\delta^* = 1$ (full manipulation) to $\delta^* = 0$ (no manipulation). We illustrate this in Figure 6 which shows the equilibrium manipulation δ^* as a function of c for $\alpha < 1$ (lighter), $\alpha = 4$, and $\alpha > 4$ (darker). For $\alpha < 1$ the manipulation is smoothly decreasing in c with a mild slope at $c = 1$. For $\alpha = 4$ the derivative at $c = 1$ is very steep. For $\alpha > 4$ the manipulation jumps from $\overline{\delta}(\alpha) > 1/2$ to $\underline{\delta}(\alpha) < 1/2$ at $c = 1$.

Intuition for large changes in manipulation near $c = 1$. To understand this result, recall from (28) that the politician's optimal manipulation can be written

$$\delta(k) = \underset{\delta \in [0,1]}{\operatorname{argmax}} \left[\, B(\delta, k) - C(\delta)\,\right] \tag{60}$$

where $B(\delta, k)$ denotes the politician's benefit from manipulation, which is increasing in the distance between the reporters' average action A and the true θ, and where likewise $C(\delta)$ is the cost of manipulation, which is increasing in the distance between the manipulated average signal y and the true θ, with coefficient c. As the intrinsic precision α_x increases, the reporters become more responsive to their signals, i.e., k increases, so that the reporters'

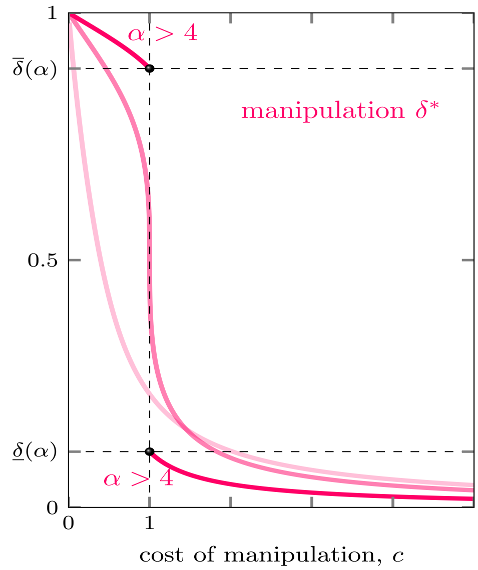

Figure 6: A small increase in c can lead to a large reduction in manipulation δ^*.

Equilibrium manipulation δ^* as a function of c for $\alpha < 1$ (lighter), $\alpha = 4$ and $\alpha > 4$ (darker). For $\alpha \leq 4$, the manipulation δ^* is continuous in c. But for $\alpha > 4$ the manipulation jumps discontinuously at $c = 1$. In the limit as $\alpha \to \infty$ the boundaries $\underline{\delta}(\alpha) \to 0^+$ and $\overline{\delta}(\alpha) \to 1^+$ so that the manipulation jumps by the maximum amount, from $\delta^* = 0$ if $c < 1$ to $\delta^* = 1$ if $c > 1$.

average action A becomes close to the average signal, $A \to y$. In short, the politician's benefit from manipulation is increasingly similar to their cost of manipulation, differing only by the magnitude of c. Small changes in c near $c = 1$ can thus lead to large changes in the amount of manipulation when α_x is high.

More formally, recall that the benefit is $B(\delta, k) := (k\delta + 1 - k)^2$ and the cost is $C(\delta) := c\delta^2$. Now suppose that $k \to 1$ so that the benefit is approximately $B(\delta, 1) = \delta^2$. Then we have $B(\delta, 1) - C(\delta) = (1 - c)\delta^2$ so that the optimal manipulation is extremely sensitive to small changes in c around $c = 1$. A small increase from $c = 1 - \varepsilon$ to $c = 1 + \varepsilon$ tips the politician's objective from increasing in δ to decreasing in δ, so that the optimal amount of manipulation jumps from $\delta = 1$ for $c < 1$ to $\delta = 0$ for $c > 1$.[14]

6.2 Two kinds of social media revolutions.

We interpret the rise of social media technologies as simultaneously increasing the intrinsic signal precision α_x (and hence increasing α) and at the same time decreasing c. Given this, the preceding discussion implies that there are really two kinds of social media revolutions, with quite different implications. To be concrete, suppose that initially the economy has relatively high costs of manipulation $c_0 > 1$ and that following the social media revolution these costs are $c_1 < c_0$. And suppose that the social media revolution makes α_x high. Then the key consideration is whether the decrease in c is large enough to push the costs of manipulation below the critical point $c = 1$.

[14]Further details can be found in our Supplementary Online Appendix.

High manipulation regime. If we get to $c_1 < 1$ when the intrinsic precision α_x is high, then the economy will end up in a *high manipulation regime* where the amount of manipulation δ^* jumps up and where increases in the intrinsic precision α_x only further increase the amount of manipulation δ^*, as in Lemma 3. If in addition $\lambda > -1$, so that the reporters' actions are not strong strategic substitutes, increases in α_x will amplify the citizens' loss, as in Proposition 7. Indeed, as shown in Remark 3, in the limit as $\alpha_x \to \infty$, the citizens' loss asymptotes to $1/\alpha_z$, i.e., the same loss the citizens would have if $\alpha_x = 0$. In this sense, the citizens lose all the potential benefits from a high intrinsic precision α_x.

Moreover, from Proposition 7 and Remark 3 we know that when the reporters' actions are *strong strategic substitutes*, $\lambda < -1$, the citizens' loss decreases in α_x but converges to $1/\alpha_z$, again the same loss the citizens would have if $\alpha_x = 0$. This means if $\lambda < -1$ and $c < 1$, the citizens with any level of α_x, no matter how high α_x is, are even worse off than when $\alpha_x = 0$. In other words, when the costs of manipulation are low and the reporters strongly prefer to differentiate their reports from each other, the citizens would prefer the reporters not to respond to their signals at all.

Low manipulation regime. But if the cost of manipulation does not fall that much, if we keep $c_1 > 1$, then the economy will end up in a *low manipulation regime*, where the amount of manipulation δ^* is bounded above by $1/2$ and where increases in the intrinsic precision α_x further decrease the amount of manipulation δ^*, as in Lemma 3 and Proposition 8. As a consequence, the citizens' loss decreases in α_x, as in Proposition 7. Indeed, as shown in Remark 3, in the limit as $\alpha_x \to \infty$, the citizens' loss asymptotes to zero, the same limit as the citizens' loss without manipulation. In this sense, keeping $c_1 > 1$ is sufficient to ensure the citizens benefit from higher levels of intrinsic precision α_x.

If, in addition, the reporters' actions are *strong strategic complements*, this may bring about other forces to further curb manipulation. In particular, if $\lambda > 1/2$ then with $c > 1$ and α_x high enough the politician's manipulation will backfire, as in Proposition 3. So here it is not just that we are in the low manipulation regime, which directly benefits citizens, but also that, in this regime, it is in the politician's own self-interest to not manipulate. In this scenario, the politician has an incentive to invest in suitable commitment devices (e.g., in a personal reputation for straight and coherent talk, or in reputable institutions that provide reliable fact-checking) that would prevent them from manipulating in the first place. Such commitment devices would be welfare-enhancing for both the politician and the citizens.

In short, even relatively small changes in the conduct of social media platforms that make it harder to manipulate information may be surprisingly effective. Such changes could come from greater internal efforts to regulate social media content, better technologies that help distinguish reliable information sources from less reliable ones, more rigorous scrutiny of politicians' speeches and interviews, etc. Even when the politician cannot commit to not manipulate, such increases in c would directly reduce the politician's incentive to engage in manipulation. But if in addition the reporters are well coordinated, we would expect to see politicians invest in commitment devices that help keep their own manipulation in check.

7 Conclusions

We argue that even small changes in the behavior of social media platforms that make it harder for misinformation to spread may play an important role in ensuring that society benefits from the rapid pace of change driven by social media technologies. We arrive at this conclusion by developing a model of information manipulation with one politician and many imperfectly informed reporters. The politician seeks to prevent reporters from accurately reporting the true state of the world, by manipulating the sources of reporters' information. The reporters are rational and internalize the politician's incentives when writing their reports. The reporting is consumed by citizens who value accurate information.

We interpret the rise of social media technologies as a shock that simultaneously changes two features of the information environment. First, these new technologies have led to new sources of information, both in the form of new media outlets and in the form of blogging and amateur journalism, thereby increasing the underlying, intrinsic precision of the sources of information available to reporters. Second, these new sources of information are not all subject to the same standards of accountability as traditional journalism and moreover are consumed in a feed that blurs distinctions between sources and that makes it easier for all kinds of news, real and fake, to go viral, thereby reducing the costs of manipulation.

We find that in the unique equilibrium of our model, the amount of information manipulation is very sensitive to the politician's costs of manipulation, especially when the underlying, intrinsic precision of the reporters' information is high. A *social media shock* that increases the intrinsic precision of the reporters' information at the same time as it decreases the costs of the politician's manipulation can have starkly different implications for equilibrium outcomes and social welfare, depending on the details. In particular, if these social media technologies reduce the costs of manipulation below a critical threshold, the economy will end up in a *high manipulation regime*, where increases in the intrinsic precision of reporters' information only further increase the amount of manipulation. As a result, the citizens are made increasingly worse off by the rise of social media technologies. But if the costs of manipulation can be maintained above this critical threshold, the economy will be in a *low manipulation regime*, where the rise of social media technologies helps reduce the politician's manipulation and makes the citizens better off. Moreover, if in addition the reporters are sufficiently coordinated, then in this low manipulation regime the rise of social media can lead the politician's manipulation to backfire, making the politician worse off than they would be if they could not manipulate at all. In this scenario, a politician would seek to invest in commitment devices that credibly prevent them from manipulating information, e.g., in a reputation for straight talk, in institutions that promote accountability, etc.

In keeping the model simple, we have abstracted from a number of important issues. First, in political contexts *competition* between multiple senders seems like an important consideration that, at least in principle, could mitigate some of the effects outlined here. But perhaps not – after all, competing distorted message could also just increase the amount of noise facing the information receivers. Second, we assume that the reporters have identical preferences and prior beliefs. This makes for clear welfare calculations, but partisan differences in preference and/or prior beliefs seem important especially if one wants a more unified model of political communication and political polarization. Finally, it would also be valuable to assess in what ways confirmation biases or other behavioral attributes interact with the endogenous noise mechanism that we emphasize.

Appendix

A Equilibrium results

Caveat on equilibrium results. As explained in our Supplementary Online Appendix, in the knife-edge case that $c = 1$ exactly there are *two* equilibria if $\alpha > 4$. This knife-edge case is essentially negligible in the sense that for any c arbitrarily close to 1 there is a unique (linear) equilibrium for any $\alpha > 0$, but formally this means we should handle the case $c = 1$ separately. The proofs of the equilibrium results and welfare results below should be understood to pertain to any generic $c \neq 1$ but to streamline the exposition we have chosen not to keep listing the $c \neq 1$ exception. For example, we report various derivatives of equilibrium outcomes with respect to c without always noting that these derivatives may not exist at $c = 1$. These derivatives should of course be read in terms of left-hand or right-hand derivatives as $c \to 1^-$ or $c \to 1^+$ as the case may be. We report the knife-edge case $c = 1$ separately in our Supplementary Online Appendix.

Proof of Proposition 1.

An equilibrium is a pair k^*, δ^* simultaneously satisfying the reporters' $k(\delta)$ and the politician's $\delta(k)$. We first show that in any equilibrium, $k^* \in \mathcal{K}(c) := \{k : 0 \leq k \leq \min(c, 1)\}$ and $\delta^* \in [0, 1]$. We then show there is a unique such equilibrium.

Recall that the politician's best response (17) requires $c \geq k^2$, otherwise the politician is at a corner with $\delta(k) = 0$. We thus focus on $k \in [-\sqrt{c}, +\sqrt{c}]$ and we distinguish two cases, depending on the magnitude of c.

(i) If $c \geq 1$, then $1 \leq \sqrt{c} \leq c$. From the politician's $\delta(k)$, we have $\delta(k) < 0$ if $k < 0$ and $\delta(k) < 0$ if $k \in (1, \sqrt{c}\,]$ but $\delta(k) \in [0, 1]$ if $k \in [0, 1]$. From the reporters' $k(\delta)$ we have $k(\delta) < 0$ if $\delta > 1$ and $k(\delta) < 1$ if $\delta < 0$. The only possible crossing points must be in the unit square with $k^* \in [0, 1]$ and $\delta^* \in [0, 1]$.

(ii) If $c \in (0, 1)$, then $0 < c < \sqrt{c} < 1$. From the politician's $\delta(k)$ we have $\delta(k) < 0$ if $k < 0$ and $\delta(k) > 1$ if $k \in (c, \sqrt{c}\,]$ but $\delta(k) \in [0, 1]$ if $k \in [0, c]$. From the reporters' $k(\delta)$ we have $k(\delta) < 0$ if $\delta > 1$ and $k(\delta) < 1$ if $\delta < 0$. The only possible crossing points must be in a *subset* of the unit square with $k^* \in [0, c]$ and $\delta^* \in [0, 1]$.

Plugging the expression for $\delta(k)$ from (17) into $k(\delta)$ from (24) and simplifying, we can write the equilibrium problem as finding $k^* \in \mathcal{K}(c)$ that satisfies

$$L(k) = R(k) \tag{A1}$$

where

$$L(k) := \frac{1}{\alpha} k, \qquad R(k) := c \frac{(c - k)(1 - k)}{(c - k^2)^2} \tag{A2}$$

and

$$R'(k) = c\left(\frac{1}{c-k^2}\right)^3 P(k), \qquad P(k) := \left(2k^3 - 3k^2 - 3ck^2 + 6ck - c^2 - c\right) \tag{A3}$$

Recall that $k \in \mathcal{K}(c)$ implies $c - k^2 \geq 0$. The sign of $R'(k)$ is thus the same as the sign of the polynomial $P(k)$. Computing the maximum of $P(k)$ over $k \in \mathcal{K}(c)$ gives

$$\overline{P}(c) := \max_{k \in \mathcal{K}(c)} P(k) = (2c - c^2 - 1)\max(c, 1) \leq 0 \tag{A4}$$

with equality only in the knife-edge case $c = 1$. We can then conclude $R'(k) \leq 0$ for all $k \in \mathcal{K}(c)$.

Observe that $L'(k) = 1/\alpha > 0$ so that the function $H(k) := L(k) - R(k)$ is strictly increasing from $H(0) = -1$ to $H(\min(c,1)) = \min(c,1)/\alpha > 0$ and hence there is a unique $k^* \in [0, \min(c,1)]$ such that $H(k^*) = 0$ or $L(k^*) = R(k^*)$. We can then recover the unique $\delta^* = \delta(k^*) \in [0,1]$ from (17). □

Proof of Lemma 1.

Differentiating the reporters' best response $k(\delta)$ in (24) with respect to δ gives

$$k'(\delta) = \alpha\frac{(1-\delta)^2\alpha - 1}{((1-\delta)^2\alpha + 1)^2}, \qquad \delta \in [0,1], \qquad \alpha > 0 \tag{A5}$$

Hence

$$k'(\delta) > 0 \qquad \Leftrightarrow \qquad \delta < 1 - 1/\sqrt{\alpha} \tag{A6}$$

If $\alpha \leq 1$ then $1 - 1/\sqrt{\alpha} \leq 0$ and $k(\delta)$ is decreasing for all $\delta \in [0,1]$. If $\alpha > 1$ then $1 - 1/\sqrt{\alpha} \in (0,1)$ and $k(\delta)$ is first increasing and then decreasing in δ. Hence $\hat{\delta}(\alpha) := \max[\,0, 1 - 1/\sqrt{\alpha}\,]$ is the critical point. Plugging in $\delta = 0$ and $\delta = 1$ gives the boundary values $k(0) = \alpha/(\alpha+1)$ and $k(1) = 0$ respectively. □

Proof of Lemma 2.

Differentiating the politician's best response $\delta(k)$ in (17) with respect to k gives

$$\delta'(k) = \left(\frac{1}{c-k^2}\right)^2 \left(k^2 - 2ck + c\right), \qquad k \in \mathcal{K}(c), \qquad c > 0 \tag{A7}$$

Hence

$$\delta'(k) > 0 \qquad \Leftrightarrow \qquad k^2 - 2ck + c > 0 \tag{A8}$$

If $c < 1$, then $k^2 - 2ck + c > 0$ for all $k \in [0,c]$ and $\delta(k)$ is increasing for all $k \in [0,c]$. If $c > 1$, then $k^2 - 2ck + c > 0$ if and only if $k < c - \sqrt{c(c-1)} < 1$. Hence $\hat{k}(c)$ as defined in the lemma is the critical point in both cases. Plugging in $k = 0$ gives $\delta(0) = 0$ for any c. If $c \leq 1$ then plugging in $k = c$ gives $\delta(c) = 1$. If $c > 1$ (so that $k = 1$ is admissible) then plugging in $k = 1$ gives $\delta(1) = 0$. □

Proof of Lemma 3.

In equilibrium we have $k^* = k(\delta^*; \alpha)$ and $\delta^* = \delta(k^*; c)$ which determine the functions $k^*(\alpha, c)$ and $\delta^*(\alpha, c)$. For part (i), applying the implicit function theorem gives

$$\frac{\partial k^*}{\partial \alpha} = \left(\frac{1}{1 - k'(\delta^*)\delta'(k^*)}\right)\frac{\partial k(\delta^*; \alpha)}{\partial \alpha} \tag{A9}$$

where, in slight abuse of notation, $k'(\delta^*)$ and $\delta'(k^*)$ denote the derivatives of the best response functions evaluated at equilibrium. Now observe from (24) that

$$\frac{\partial k(\delta; \alpha)}{\partial \alpha} = \frac{1-\delta}{((1-\delta)^2\alpha + 1)^2} \in [0,1], \qquad \delta \in [0,1], \qquad \alpha > 0 \tag{A10}$$

We will now show that at equilibrium the product $k'(\delta^*)\delta'(k^*)$ is nonpositive. To do this, first evaluate $k'(\delta)$ at the equilibrium k^*, δ^* to obtain

$$k'(\delta^*) = \left(\frac{k^*}{c - k^*}\right)\left(2ck^* - k^{*\,2} - c\right) \tag{A11}$$

Then evaluate $\delta'(k)$ from (A7) at k^* to get

$$k'(\delta^*)\delta'(k^*) = -\left(\frac{k^*}{c - k^*}\right)\left(\frac{k^{*2} - 2ck^* + c}{c - k^{*2}}\right)^2 \leq 0 \tag{A12}$$

Hence $k^*(\alpha, c)$ is strictly increasing in α. For part (ii) we use the politician's best response to calculate

$$\frac{\partial \delta^*}{\partial \alpha} = \delta'(k^*)\frac{\partial k^*}{\partial \alpha} \tag{A13}$$

From Lemma 2 we know that $\delta'(k) > 0$ if and only if $k < \hat{k}(c)$ where $\hat{k}(c)$ is defined in (27). Hence

$$\frac{\partial \delta^*}{\partial \alpha} > 0 \qquad \Leftrightarrow \qquad k^*(\alpha, c) < \hat{k}(c) \tag{A14}$$

For any $c > 0$, the critical $\hat{\alpha}(c)$ is found using the result from part (i) that $k^*(\alpha, c)$ is strictly increasing in α to find the smallest α such that $k^*(\alpha, c) \geq \hat{k}(c)$. If there is no such value, i.e., if $c < 1$, we set $\hat{\alpha}(c) = +\infty$. □

Proof of Lemma 4.

In equilibrium we have $k^* = k(\delta^*; \alpha)$ and $\delta^* = \delta(k^*; c)$ which determine the functions $k^*(\alpha, c)$ and $\delta^*(\alpha, c)$. For part (i), applying the implicit function theorem gives

$$\frac{\partial \delta^*}{\partial c} = \left(\frac{1}{1 - k'(\delta^*)\delta'(k^*)}\right)\frac{\partial \delta(k^*; c)}{\partial c} \tag{A15}$$

We already know from (A12) that $k'(\delta^*)\delta'(k^*) \leq 0$. And from (17) observe that

$$\frac{\partial \delta(k; c)}{\partial c} = -\frac{k - k^2}{(c - k^2)^2} < 0 \tag{A16}$$

Hence $\delta^*(\alpha, c)$ is strictly decreasing in c. For part (ii) we use the reporters' best response to calculate

$$\frac{\partial k^*}{\partial c} = k'(\delta^*)\frac{\partial \delta^*}{\partial c} \tag{A17}$$

From Lemma 1 we know that $k'(\delta) < 0$ if and only if $\delta > \hat{\delta}(\alpha)$ where $\hat{\delta}(\alpha)$ is defined in (26). Hence

$$\frac{\partial k^*}{\partial c} > 0 \qquad \Leftrightarrow \qquad \delta^*(\alpha, c) > \hat{\delta}(\alpha) \tag{A18}$$

For any $\alpha > 0$ the critical $\hat{c}(\alpha)$ is found using the result from part (i) that $\delta^*(\alpha, c)$ is strictly decreasing in c to find the smallest c such that $\delta^*(\alpha, c) \leq \hat{\delta}(\alpha)$. If there is no such value, i.e., if $\alpha < 1$, we set $\hat{c}(\alpha) = +\infty$. □

B Politician's welfare

Proof of Lemma 5.

The derivative of the politician's value function from (34) evaluated at the equilibrium k^* can be written

$$v'(k^*) = 2\left(\frac{k^*}{\alpha_x} - \frac{R(k^*)}{\alpha_z}\right) = -2\frac{\lambda}{1 - \lambda}\left(\frac{k^*}{\alpha_x}\right) \tag{B1}$$

where we use the equilibrium condition from (A2) above to write $R(k^*) = \frac{1}{1-\lambda}\frac{\alpha_z}{\alpha_x}k^*$. □

Proof of Proposition 2.

The total derivative of the equilibrium v^* with respect to the parameter c is

$$\frac{dv^*}{dc} = \underbrace{v'(k^*)\frac{\partial k^*}{\partial c}}_{\text{indirect effect}} + \underbrace{\frac{\partial v(k^*;c)}{\partial c}}_{\text{direct effect}} \tag{B2}$$

where $v'(k^*)$ is obtained from (E29) above. Taking the partial derivative of the politician's value function in (34) with respect to c gives

$$\frac{\partial v(k^*;c)}{\partial c} = -\frac{(1-k^*)^2}{\alpha_z}\frac{k^{*2}}{(c-k^{*2})^2} \tag{B3}$$

From the proofs of Lemma 3 and Lemma 4, we can then write

$$\frac{\partial k^*}{\partial c} = \frac{k^*}{c-k^*}\left(2ck^* - k^{*2} - c\right)\frac{1}{1+\frac{k^*}{c-k^*}\left(\frac{2ck^*-k^{*2}-c}{c-k^{*2}}\right)^2}\frac{-k^*(1-k^*)}{(c-k^{*2})^2} \tag{B4}$$

Hence v^* is strictly increasing in c if and only if

$$-\frac{\lambda}{1-\lambda}\frac{2}{\alpha_x}k^*\frac{\partial k^*}{\partial c} > \frac{(1-k^*)^2}{\alpha_z}\frac{k^{*2}}{(c-k^{*2})^2} \tag{B5}$$

Since the RHS of (B5) is positive, for the inequality (B5) to hold it must be the case that the LHS of (B5) is positive too, which requires either (i) $\partial k^*/\partial c > 0$ and $\lambda < 0$, or (ii) $\partial k^*/\partial c < 0$ and $\lambda > 0$. These conditions lead to part (i) and part (ii) of the proposition.

For part (i) suppose that condition (i) holds, i.e., $\partial k^*/\partial c > 0$ and $\lambda < 0$. From (A17) and (A11), we know $\partial k^*/\partial c > 0$ if and only if $k'(\delta^*) < 0$ which in turn requires $2ck^* - k^{*2} - c < 0$. The inequality (B5) can then be written as

$$-2\lambda > \alpha\frac{J_1}{J_2} \tag{B6}$$

where

$$J_1 = c^3 - c^3k - 6c^2k^2 + 10c^2k^3 + 4ck^3 - 9ck^4 - 4c^2k^4 + 5ck^5 \tag{B7}$$
$$J_2 = c^3k - 2c^3k^2 - c^2k^3 - 4c^2k^4 - ck^5 - 2ck^6 + k^7 \tag{B8}$$

From Lemma 4, we know that if $\alpha < 1$ then $\partial k^*/\partial c > 0$ with the limit $k^* = \alpha/(1+\alpha)$ as $c \to \infty$. Therefore

$$\lim_{c\to\infty}\alpha\frac{J_1}{J_2} = \lim_{c\to\infty}\alpha\frac{1-k}{k-2k^2} = \frac{1}{1-2\frac{\alpha}{\alpha+1}} \tag{B9}$$

This limit increases in α, starting from 1 when $\alpha = 0$ and diverging to $+\infty$ as $\alpha \to 1$. Therefore if $-2\lambda > 1$, i.e., for each $\lambda < -1/2$, there is a critical point $\underline{\alpha} < 1$ such that

$$-2\lambda > \lim_{c\to\infty}\alpha\frac{J_1}{J_2} = \frac{1}{1-2\frac{\alpha}{\alpha+1}} \qquad \text{for } \alpha < \underline{\alpha} < 1 \tag{B10}$$

Hence for $\lambda < -1/2$ and $\alpha < \underline{\alpha} < 1$ there exists a cutoff c_1^* such that v^* is strictly increasing in c for $c > c_1^*$.

For part (ii), suppose that condition (ii) holds, i.e., $\partial k^*/\partial c < 0$ and $\lambda > 0$. From (A17) and (A11), we know $\partial k^*/\partial c < 0$ if and only if $k'(\delta^*) > 0$ which in turn requires $2ck^* - k^{*2} - c > 0$. The inequality (B5) can be written as

$$2\lambda > -\alpha\frac{J_1}{J_2} \tag{B11}$$

with J_1 and J_2 as defined above. From Lemma 4, we know that if $\alpha > 1$ then $\partial k^*/\partial c < 0$ if $c > \widehat{c}(\alpha)$. So we now have

$$\lim_{c\to\infty} -\alpha\frac{J_1}{J_2} = \lim_{c\to\infty} -\alpha\frac{1-k}{k-2k^2} = \frac{1}{2\frac{\alpha}{\alpha+1}-1} \tag{B12}$$

This limit decreases in α, diverging to $+\infty$ as $\alpha \to 1$ and converging to 1 as $\alpha \to \infty$. Therefore if $2\lambda > 1$, i.e., for each $\lambda > +1/2$, there is a critical point $\overline{\alpha} > 1$ such that

$$2\lambda > \lim_{c\to\infty} -\alpha\frac{J_1}{J_2} = \frac{1}{2\frac{\alpha}{\alpha+1}-1} \qquad \text{for } \alpha > \overline{\alpha} > 1 \tag{B13}$$

Hence for $\lambda > 1/2$ and $\alpha > \overline{\alpha} > 1$ there exists a cutoff c_2^* such that v^* is strictly increasing in c for $c > c_2^*$. □

Proof of Lemma 6.

Recall that $k^*_{nm}(\alpha) := \alpha/(\alpha+1)$. If $\alpha \leq 1$ then any $c < +\infty$ implies $\delta^*(\alpha, c) > 0$ and hence $k^*(\alpha, c) < k^*_{nm}(\alpha)$. With $\alpha > 1$ we find combinations of (α, c) that give $k^*(\alpha, c) = k^*_{nm}(\alpha)$. To do so, first determine the equilibrium δ^* that equates $k(\delta; \alpha)$ and $k^*_{nm}(\alpha)$, namely

$$\delta^*_{nm}(\alpha) = \frac{\alpha - 1}{\alpha}, \qquad \alpha > 1 \tag{B14}$$

Then solve for c that equates $\delta(k^*_{nm}(\alpha); c)$ and $\delta^*_{nm}(\alpha)$, namely

$$c = \frac{\alpha}{\alpha - 1}\left(\frac{\alpha}{\alpha + 1}\right)^2 =: c^*_{nm}(\alpha) \tag{B15}$$

(with $c^*_{nm}(\alpha) = +\infty$ for $\alpha \leq 1$). We now show that $k^*(\alpha, c) < k^*_{nm}(\alpha)$ iff $c < c^*_{nm}(\alpha)$. Observe that

$$\delta^*_{nm}(\alpha) = \frac{\alpha - 1}{\alpha} > \hat{\delta}(\alpha) \tag{B16}$$

where $\hat{\delta}(\alpha)$ is the critical point from Lemma 1. Hence $k(\delta; \alpha)$ is decreasing in δ for any $\delta \geq \delta^*_{nm}(\alpha)$. Now observe that $k(\delta^*_{nm}(\alpha); \alpha) = k^*_{nm}(\alpha)$ so that $k^*(\alpha, c) < k^*_{nm}(\alpha)$ iff $\delta^*(\alpha, c) > \delta^*_{nm}(\alpha)$. From Lemma 4 we know that $\delta^*(\alpha, c)$ is strictly decreasing in c hence any $c < c^*_{nm}(\alpha)$ is equivalent to $\delta^*(\alpha, c) > \delta^*_{nm}(\alpha)$. □

Proof of Proposition 3.

Using the fixed point condition (A2) with the redefined $\alpha = (1 - \lambda)\alpha_x/\alpha_z$, we can write

$$v^* = \frac{1}{(1 - \lambda)\alpha_x}\left\{k^* - \lambda k^{*2} + \frac{k^{*2}(1 - k^*)^2}{c - k^*}\right\} \tag{B17}$$

Using the analogous condition for k^*_{nm}, we can write

$$v^*_{nm} = \frac{1}{(1 - \lambda)\alpha_x}\left\{k^*_{nm} - \lambda k^{*2}_{nm}\right\} \tag{B18}$$

Hence the politician's manipulation backfires, $v^* < v^*_{nm}$, if and only if

$$g(k^*) < f(k^*_{nm}) - f(k^*) \tag{B19}$$

where

$$f(k) := k - \lambda k^2, \qquad g(k) := \frac{k^2(1 - k)^2}{c - k} \geq 0 \tag{B20}$$

For part (i) suppose that $\lambda < 0$. We know from (38) and (39) that a necessary condition for the politician's manipulation to backfire is $k < k^*_{nm}$. We can rewrite the inequality in (B19) as

$$\frac{k^{*2}(1 - k^*)^2}{c - k^*} < (k^*_{nm} - k)(1 - \lambda(k^*_{nm} + k^*)) \tag{B21}$$

Using the fixed point conditions (A2) for both k^* and k^*_{nm}, we can rewrite the key condition (B21) as

$$\lambda < \frac{1}{c - k^*}\frac{K_1 K_2}{K_3 K_4} \tag{B22}$$

where

$$K_1 := 4ck^{*2} - c^2 - k^{*3} - 2ck^{*3} - k^{*4} + k^{*5}$$
$$K_2 := c(c - k^*)(1 - k^*) + k^*(c - k^{*2})^2 > 0$$
$$K_3 := k^{*3} - 2ck^* + c > 0$$
$$K_4 := (1 + k^*)(c - k^{*2})^2 + c(c - k^*)(1 - k^*) > 0$$

Now consider taking $\alpha_x \to 0$ for fixed $\lambda < 0$ such that $k^* \to 0$. We then have the following limits

$$K_1 \to -c^2, \quad K_2 \to +c^2, \quad K_3 \to c, \quad K_4 \to 2c^2$$

So in the limit the RHS of (B22) is

$$\frac{1}{c-k^*}\frac{K_1K_2}{K_3K_4} \to \frac{1}{(c-0)}\frac{(-c^2)(c^2)}{(c)(2c^2)} = -\frac{1}{2} \tag{B23}$$

Hence for any $\lambda < -1/2$ we can find α_x sufficiently close to zero such that (B22) is satisfied and in turn the politician's manipulation backfires, $v^* < v^*_{nm}$.

For part (ii), suppose that $\lambda > 0$. We know from (38) and (39) that the necessary condition for the politician's manipulation to backfire is $k^* > k^*_{nm}$. We can rewrite the inequality in (B19) as

$$\frac{k^{*2}(1-k^*)^2}{k^* - k^*_{nm}} < (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) \tag{B24}$$

Using the fixed point conditions (A2) for both k^* and k^*_{nm}, we can rewrite this key condition as

$$\frac{k^{*2}(1-k^*)}{\frac{k^*_{nm}}{k^*}\left(c\frac{(c-k^*)}{(c-k^{*2})^2}\right) - 1} < (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) \tag{B25}$$

Observe that if, in addition, $c > 1$ and $\lambda > \frac{1}{2}$, then the RHS of (B25) converges to a strictly positive constant

$$\lim_{\alpha_x \to \infty} (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) = (\lambda 2 - 1)(c - 1) > 0 \tag{B26}$$

(since $k^* \to 1$ if $c > 1$). But the LHS of (B25) converges to zero

$$\lim_{\alpha_x \to \infty} \frac{k^{*2}(1-k^*)}{\frac{k^*_{nm}}{k^*}c\frac{c-k^*}{(c-k^{*2})^2} - 1} = \frac{0^+}{\frac{c}{(c-1)} - 1} = 0^+ \tag{B27}$$

Therefore, if $k^* > k^*_{nm}$, $c > 1$ and $\lambda > \frac{1}{2}$ then there exists α^*_x such that for $\alpha_x > \alpha^*_x$ the LHS of (B25) is strictly less than the RHS of (B25) so that the politician's manipulation backfires, $v^* < v^*_{ms}$.

Finally, we know from Lemma 6 that $k^* < k^*_{nm}$ if and only if $c < c^*_{nm}(\alpha)$. Also observe that $c < 1$ is sufficient for $c < c^*_{nm}(\alpha)$ if $1 < c^*_{nm}(\alpha)$. From (41) we have $1 < c^*_{nm}(\alpha)$ if $\alpha < 1$, or if $\alpha > 1$ and $\alpha < (1+\sqrt{5})/2$. Since $\alpha = (1-\lambda)\alpha_x/\alpha_z$, the critical point $\underline{\alpha}^*_x$ must be

$$\underline{\alpha}^*_x < \left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{\alpha_z}{1-\lambda}\right) \tag{B28}$$

Likewise, $c > 1$ is sufficient condition for $c > c^*_{nm}(\alpha)$ if $1 > c^*_{nm}(\alpha)$, and we need $\alpha > \alpha_2 = (1+\sqrt{5})/2$ to ensure that $1 > c^*_{nm}(\alpha)$. Since $\alpha = (1-\lambda)\alpha_x/\alpha_z$, the critical point $\overline{\alpha}^*_x$ must be

$$\overline{\alpha}^*_x > \left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{\alpha_z}{1-\lambda}\right) \tag{B29}$$

□

Proof of Proposition 4.

We decompose the politician's gain from manipulation as

$$v^* - v^*_{nm} = (v(k^*) - v_{nm}(k^*)) + (v_{nm}(k^*) - v_{nm}(k^*_{nm})) \tag{B30}$$

where $v(k)$ is the politician's value function with manipulation

$$v(k) := \max_{\delta \in [0,1]} V(\delta, k) = \frac{1}{\alpha_z}(1-k)^2\left(\frac{c}{c-k^2}\right) + \frac{1}{\alpha_x}k^2 \tag{B31}$$

and $v_{nm}(k)$ is the politician's value function without manipulation:

$$v_{nm}(k) := V(0,k) = \frac{1}{\alpha_z}(1-k)^2 + \frac{1}{\alpha_x}k^2 \leq v(k) \tag{B32}$$

First observe that $v(k^*) \geq v_{nm}(k^*)$ for any k^* hence for $v^* > v^*_{nm}$ it suffices that $v_{nm}(k^*) > v_{nm}(k_{nm})$. Then observe that $v'_{nm}(k) < 0$ for all k in the interval $(0, \frac{\alpha_x}{\alpha_x+\alpha_z})$, and $v'_{nm}(k) > 0$ for all k in the interval $(\frac{\alpha_x}{\alpha_x+\alpha_z}, 1)$. Recall that $k^*_{nm} = \alpha/(\alpha+1)$ from (11) and that with strategic interactions among reporters, $\alpha = (1-\lambda)\alpha_x/\alpha_z$. Therefore, when $\lambda < 0$, i.e., the reporters' actions are strategic substitutes, $k^*_{nm} > \frac{\alpha_x}{\alpha_x+\alpha_z}$ and hence if $k^* > k^*_{nm}$ then $v_{nm}(k^*) > v_{nm}(k_{nm})$. From Lemma 6 we know that a necessary and sufficient condition for $k^* > k^*_{nm}$ is that the cost of manipulation be $c > c^*_{nm}(\alpha)$ where $c^*_{nm}(\alpha)$ is the critical cost given in (41). Similarly, when $\lambda > 0$, i.e., the reporters' actions are strategic complements, $k^*_{nm} < \frac{\alpha_x}{\alpha_x+\alpha_z}$ so that if $k^* < k^*_{nm}$ then $v_{nm}(k^*) > v_{nm}(k_{nm})$. Again from Lemma 6, the necessary and sufficient condition for $k^* < k^*_{nm}$ is $c < c^*_{nm}(\alpha)$. When $\lambda = 0$, $k^*_{nm} = \frac{\alpha_x}{\alpha_x+\alpha_z}$, the minimizer of $v_{nm}(k)$. Hence, $v_{nm}(k^*) \geq v_{nm}(k^*_{nm})$ for any k^* and strictly so if $k^* \neq k^*_{nm}$. □

Proof of Remark 1.

The limits of v^*_{nm} can be computed directly after evaluating $v_{nm}(k)$ in (B32) at k^*_{nm}, namely

$$\lim_{\alpha_x \to 0^+} v^*_{nm} = \frac{\alpha_z}{\alpha_z^2} = \frac{1}{\alpha_z} \tag{B33}$$

$$\lim_{\alpha_x \to \infty} v^*_{nm} = \lim_{\alpha_x \to +\infty} \frac{(1-\lambda)^2 \frac{1}{\alpha_x} + \frac{\alpha_z}{\alpha_x^2}}{(1-\lambda)^2 + 2(1-\lambda)\frac{\alpha_z}{\alpha_x} + \frac{\alpha_z^2}{\alpha_x^2}} = 0 \tag{B34}$$

For the limits of $v^* = v(k^*; \alpha_x)$ we repeatedly use that $v(k; \alpha_x)$ is continuous in k and that k^* is continuous in α_x. In the limit as $\alpha_x \to 0^+$ we have $k^* \to 0^+$ so that

$$\lim_{\alpha_x \to 0^+} v^* = \frac{1}{\alpha_z}(1-0)^2 \left(\frac{c}{c-0^2}\right) + \lim_{\alpha_x \to 0^+} \frac{k^{*2}}{\alpha_x} = \frac{1}{\alpha_z} + 0 = \frac{1}{\alpha_z} \tag{B35}$$

where we have used L'Hôpital's rule and (A9) and (A10) to calculate that

$$\lim_{\alpha_x \to 0^+} \frac{k^{*2}}{\alpha_x} = \lim_{\alpha_x \to 0^+} 2k^* \frac{dk^*}{d\alpha_x} = \lim_{\alpha_x \to 0^+} 2k^* \left(\frac{1}{1-k'(\delta^*)\delta'(k^*)}\right) \frac{(1-\delta^*)}{((1-\delta^*)^2\alpha+1)^2}(1-\lambda)\frac{1}{\alpha_z} = 0$$

where the limit follows because $\delta^* \in [0,1]$ for all α_x and $k^* \to 0$ and hence from (A12) $k'(\delta^*)\delta'(k^*) \to 0$ as $\alpha_x \to 0^+$. At the other extreme, in the limit as $\alpha_x \to \infty$ we have $k^* \to \min(c, 1)$ so that

$$\lim_{\alpha_x \to \infty} v^* = \begin{cases} \frac{1}{\alpha_z}(1-c)^2 \frac{c}{c-c^2} & = \frac{1-c}{\alpha_z} & \text{if } c < 1 \\ \frac{1}{\alpha_z}(1-1)^2 \frac{c}{c-1} & = 0 & \text{if } c > 1 \end{cases} \tag{B36}$$

□

C Reporters' and citizens' welfare

Proof of Proposition 5.

The proof of part(i) is in the main text. For part (ii) use $l^*_{\mathcal{C}} = l_{\mathcal{C}}(\delta^*)$ and and equation (50) to write

$$l^*_{\mathcal{C}} = l_{\mathcal{R}}(\delta^*) + \frac{\lambda\alpha_z}{(1-\lambda)^2}\, l_{\mathcal{R}}(\delta^*)^2 \tag{C1}$$

and likewise

$$l^*_{\mathcal{C},nm} = l_{\mathcal{R}}(0) + \frac{\lambda\alpha_z}{(1-\lambda)^2}\, l_{\mathcal{R}}(0)^2 \tag{C2}$$

Differencing these expressions we can write

$$l^*_{\mathcal{C}} - l^*_{\mathcal{C},nm} = \Big(l_{\mathcal{R}}(\delta^*) - l_{\mathcal{R}}(0)\Big) \times \left[1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(l_{\mathcal{R}}(\delta^*) + l_{\mathcal{R}}(0)\Big)\right] \tag{C3}$$

Now write the term in square brackets on the LHS as $\Delta(\delta^*)$ where $\Delta(\delta)$ is the function

$$\Delta(\delta) := 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(l_{\mathcal{R}}(\delta) + l_{\mathcal{R}}(0)\Big) \tag{C4}$$

From part (i) we know $l_{\mathcal{R}}(\delta^*) > l_{\mathcal{R}}(0)$ so $l^*_{\mathcal{C}} > l^*_{\mathcal{C},nm}$ if and only if $\Delta(\delta^*) > 0$. Since $\alpha_z > 0$ and $l_{\mathcal{R}}(\delta^*) > l_{\mathcal{R}}(0) > 0$, a sufficient condition for $\Delta(\delta^*) > 0$ is that $\lambda > 0$. To prove part (ii) we need to show that any $\lambda > -1$ is also sufficient. To see this, observe that since $l_{\mathcal{R}}(\delta)$ is strictly increasing in δ, for $\lambda < 0$ we also know that $\Delta(\delta)$ is strictly decreasing in δ which in turn implies $\Delta(\delta) \geq \Delta(1)$. Hence if $\lambda < 0$ a sufficient condition for $\Delta(\delta^*) > 0$ is that $\Delta(1) > 0$. Calculating $\Delta(1)$ gives

$$\begin{aligned}\Delta(1) &= 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(l_{\mathcal{R}}(1) + l_{\mathcal{R}}(0)\Big) \\ &= 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(\frac{1-\lambda}{\alpha_z} + \frac{1-\lambda}{(1-\lambda)\alpha_x + \alpha_z}\Big)\end{aligned}$$

where the second equality follows from the expression for $l_{\mathcal{R}}(\delta)$ in equation (46) evaluated at $\delta = 1$ and $\delta = 0$. Simplifying further

$$\Delta(1) = 1 + \frac{\lambda}{1-\lambda}\Big(1 + \frac{1}{1+\alpha}\Big) \tag{C5}$$

where $\alpha := (1-\lambda)\alpha_x/\alpha_z > 0$. So for $\lambda < 0$ a sufficient condition for $\Delta(1) > 0$ and hence $\Delta(\delta^*) > 0$ is

$$\frac{\lambda}{1-\lambda}\Big(1 + \frac{1}{1+\alpha}\Big) > -1 \tag{C6}$$

or equivalently

$$1 + \alpha > -\lambda \tag{C7}$$

Since $\alpha > 0$ a sufficient condition for this is $\lambda > -1$. To summarize, any $\lambda > -1$ is sufficient for $\Delta(\delta^*) > 0$ and hence sufficient for $l^*_{\mathcal{C}} > l^*_{\mathcal{C},nm}$. For part (iii) we then know that $\lambda < -1$ is necessary for $l^*_{\mathcal{C}} < l^*_{\mathcal{C},nm}$. Recall that $\Delta(\delta)$ is strictly decreasing in δ, i.e., $\Delta(\delta) \leq \Delta(0)$, for $\lambda < 0$. Hence for $\lambda < -1$ a sufficient condition for $\Delta(\delta^*) < 0$ is that $\Delta(0) < 0$. Calculating $\Delta(0)$ gives

$$\begin{aligned}\Delta(0) &= 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(l_{\mathcal{R}}(0) + l_{\mathcal{R}}(0)\Big) \\ &= 1 + \frac{\lambda\alpha_z}{(1-\lambda)^2}\Big(\frac{2(1-\lambda)}{(1-\lambda)\alpha_x + \alpha_z}\Big) \\ &= 1 + \frac{\lambda}{1-\lambda}\Big(\frac{2}{1+\alpha}\Big)\end{aligned}$$

So for $\lambda < -1$ a sufficient condition for $\Delta(0) < 0$ and hence $\Delta(\delta^*) < 0$ is

$$\alpha < \frac{\lambda+1}{\lambda-1} = -\left(\frac{1+\lambda}{1-\lambda}\right), \qquad \lambda < -1 \tag{C8}$$

Since $\alpha := (1-\lambda)\alpha_x/\alpha_z > 0$ we rewrite this as

$$\alpha_x < \widehat{\alpha}^{**}_x := -\left(\frac{1+\lambda}{(1-\lambda)^2}\right)\alpha_z, \qquad \lambda < -1 \tag{C9}$$

To summarize, for each $\lambda < -1$ there is a critical point $\widehat{\alpha}^{**}_x$ such that together $\lambda < -1$ and $\alpha_x < \widehat{\alpha}^{**}_x$ are sufficient for $\Delta(\delta^*) < \Delta(0) < 0$ and hence sufficient for $l^*_{\mathcal{C}} < l^*_{\mathcal{C},nm}$. □

Proof of Proposition 6.

The total derivative of $l_{\mathcal{R}}^*$ with respect to α_x is

$$\frac{dl_{\mathcal{R}}^*}{d\alpha_x} = l_{\mathcal{R}}'(\delta^*)\frac{d\delta^*}{d\alpha_x} + \frac{\partial l_{\mathcal{R}}(\delta^*;\alpha_x)}{\partial \alpha_x} \tag{C10}$$

Supplementary Lemma 1 in the Supplementary Online Appendix shows that

$$\frac{dl_{\mathcal{R}}^*}{d\alpha_x} > 0 \qquad \Leftrightarrow \qquad F(k^*) := k^{*4} - 2k^{*3} + 2ck^* - c^2 > 0 \tag{C11}$$

Supplementary Lemma 2 in the Supplementary Online Appendix shows that if $c > 1$ then it cannot be the case that $F(k^*) > 0$ and hence the reporters' loss is unambiguously decreasing in α_x. If $c < 1$ then there is an interval $(\underline{k}, \overline{k})$ with $0 < \underline{k} < \overline{k} < 1$ such that $F(k) > 0$ for $k \in (\underline{k}, \overline{k})$ and $F(k) \leq 0$ otherwise. Moreover, the cutoffs are on either side of c so that $0 < \underline{k} < c < \overline{k} < 1$.

Since $k^*(\alpha_x, c)$ is strictly increasing in α_x from 0 to $\min(c, 1)$, for any fixed $c < 1$ there is then a critical point α_x^{**} solving

$$k^*(\alpha_x^{**}, c) = \underline{k} \tag{C12}$$

such that for any $\alpha_x > \alpha_x^{**}$ we have $k^*(\alpha_x, c) \in (\underline{k}(c), c)$ so that $F(k^*) > 0$ and hence for $\alpha_x > \alpha_x^{**}$ the reporters' loss is strictly increasing in α_x.

Proof of Remark 2.

The limits of $l_{\mathcal{R},nm}^*$ can be directly computed from $l_{\mathcal{R}}(0)$, namely

$$\lim_{\alpha_x \to 0^+} l_{\mathcal{R},nm}^* = \frac{1-\lambda}{\alpha_z}, \qquad \text{and} \qquad \lim_{\alpha_x \to \infty} l_{\mathcal{R},nm}^* = 0.$$

For the limits of $l_{\mathcal{R}}^* = l_{\mathcal{R}}(\delta^*; \alpha_x)$ we repeatedly use that $l_{\mathcal{R}}(\delta; \alpha_x)$ is continuous in δ and that δ^* is continuous in α_x. In the limit as $\alpha_x \to 0^+$ we have $\delta^* \to 0$ so that

$$\lim_{\alpha_x \to 0^+} l_{\mathcal{R}}^* = \frac{(1-\lambda)}{(1-0)^2(1-\lambda)0 + \alpha_z} = \frac{1-\lambda}{\alpha_z} \tag{C13}$$

At the other extreme, as $\alpha_x \to \infty$ we have $\delta^* \to 1$ if $c < 1$ so that

$$\lim_{\alpha_x \to \infty} l_{\mathcal{R}}^* = \lim_{\alpha_x \to \infty} \frac{(1-\lambda)}{(1-\delta^*)^2(1-\lambda)\alpha_x + \alpha_z} = \frac{1-\lambda}{\alpha_z} \tag{C14}$$

where we have used L'Hôpital's rule to calculate that

$$\lim_{\alpha_x \to \infty} (1-\delta^*)^2 \alpha_x = \lim_{\alpha_x \to \infty} 2(1-\delta^*)\frac{d\delta^*}{d\alpha_x}\alpha_x^2 = \lim_{\alpha_x \to \infty} 2(1-\delta^*)\delta'(k^*)\left(\frac{\frac{\alpha_z}{(1-\lambda)}k^*}{\frac{\alpha_z}{(1-\lambda)\alpha_x} - R'(k^*)}\frac{1}{\alpha_x^2}\right)\alpha_x^2 = 0$$

(since if $c < 1$ we have $\delta^* \to 1$ and $k^* \to c$ and from (A7) and (A3) we have $\delta'(c) = -R'(c) = 1/(c - c^2)$). Alternatively, if $c > 1$ then $\delta^* \to 0$ as $\alpha_x \to \infty$ so that we simply have

$$\lim_{\alpha_x \to \infty} l_{\mathcal{R}}^* = \lim_{\alpha_x \to \infty} \frac{(1-\lambda)}{(1-\delta^*)^2(1-\lambda)\alpha_x + \alpha_z} = 0 \tag{C15}$$

□

Proof of Lemma 7.

From equation (55) in the main text we see that the derivative of $l_{\mathcal{C}}^*$ with respect to α_x and the derivative of $l_{\mathcal{R}}^*$ with respect to α_x have the same sign if and only if

$$1 + \frac{2\lambda\alpha_z}{(1-\lambda)^2} l_{\mathcal{R}}^* > 0 \tag{C16}$$

Write this key term $T(\delta^*) > 0$ where

$$T(\delta) := 1 + \frac{2\lambda\alpha_z}{(1-\lambda)^2}\, l_{\mathcal{R}}(\delta) \tag{C17}$$

Clearly $\lambda \geq 0$ suffices for $T(\delta^*) > 0$. When $\lambda < 0$, $T(\delta^*) > 0$ if and only if

$$l^*_{\mathcal{R}} < -\frac{(1-\lambda)^2}{2\lambda\alpha_z} := l_{crit} \tag{C18}$$

From Proposition 6 and Remark 2 we know that the maximum of $l^*_{\mathcal{R}}$ is $l^*_{max} = (1-\lambda)/\alpha_z$. If $l^*_{max} < l_{crit}$, i.e., if $\lambda \in (-1, 0)$, the inequality (C18) holds and therefore $T(\delta^*) > 0$. Alternatively, if $l^*_{max} > l_{crit}$, i.e., if $\lambda \in (-\infty, -1)$, there exists a subset of α_x such that the inequality (C18) does not hold and in turn $T(\delta^*) < 0$.

We now determine the set of α_x such that (C18) does not hold, conditional on $\lambda < -1$. For any $c > 1$ we know from Proposition 6 and Remark 2 that $l^*_{\mathcal{R}}$ is strictly decreasing in α_x with $\lim_{\alpha_x \to 0^+} l^*_{\mathcal{R}} = l^*_{max}$ and $\lim_{\alpha_x \to \infty} l^*_{\mathcal{R}} = 0$. Hence for each $\lambda < -1$ and $c > 1$ there is a unique critical point $\underline{\alpha}^{**}_x > 0$ such that $T(\delta^*) > 0$ if and only if $\alpha_x > \underline{\alpha}^{**}_x$. Similarly, for any $c < 1$ we again know from Proposition 6 and Remark 2 that $l^*_{\mathcal{R}}$ is strictly decreasing in α_x if and only if $\alpha_x < \alpha^{**}_x$ and $\lim_{\alpha_x \to 0^+} l^*_{\mathcal{R}} = \lim_{\alpha_x \to \infty} l^*_{\mathcal{R}} = l^*_{max}$. Let l^*_{min} denote the reporter's loss at the $\alpha_x = \alpha^{**}_x$ that achieves the minimum. For any $c < 1$ and any fixed loss $l \in (l^*_{min}, l^*_{max})$ there are two critical points $\underline{\alpha}_x(l) < \alpha^{**}_x < \overline{\alpha}_x(l)$ such that $l^*_{\mathcal{R}} < l$ if and only if $\alpha_x \in (\underline{\alpha}_x(l), \overline{\alpha}_x(l))$. Then for each $\lambda < -1$ and $c < 1$ there are two possibilities, either $l_{crit} \in (l^*_{min}, l^*_{max})$ or $l_{crit} \leq l^*_{min}$. For the interior cases $l_{crit} \in (l^*_{min}, l^*_{max})$ we define the critical points by $\underline{\alpha}^{**}_x := \underline{\alpha}_x(l_{crit})$ and $\overline{\alpha}^{**}_x := \overline{\alpha}_x(l_{crit})$. For the boundary case $l_{crit} \leq l^*_{min}$ we define the critical points by $\underline{\alpha}^{**}_x = \overline{\alpha}^{**}_x = +\infty$. Given these critical points, we have $T(\delta^*) > 0$ if and only if $\alpha_x \in (\underline{\alpha}^{**}_x, \overline{\alpha}^{**}_x)$. □

Proof of Proposition 7.

From Lemma 7 the citizens' loss $l^*_{\mathcal{C}}$ and the reporters' loss $l^*_{\mathcal{R}}$ move in the same direction in response to α_x if and only if either: $\lambda > -1$, or $\lambda < -1$ and $\alpha_x \in (\underline{\alpha}^{**}_x, \overline{\alpha}^{**}_x)$ with $\overline{\alpha}^{**}_x = +\infty$ if $c > 1$. Moreover from Proposition 6 we know that $l^*_{\mathcal{R}}$ is strictly decreasing in α_x if $c > 1$.

For part (i) fix $\lambda > -1$. For case (a), if $c > 1$ we know $l^*_{\mathcal{R}}$ is strictly decreasing and hence $l^*_{\mathcal{C}}$ is also strictly decreasing in α_x. Similarly for case (b), if $c < 1$ we know $l^*_{\mathcal{R}}$ is strictly decreasing if and only if $\alpha_x < \underline{\alpha}^{**}_x$ and hence $l^*_{\mathcal{C}}$ is also strictly decreasing if and only if $\alpha_x < \underline{\alpha}^{**}_x$.

For part (ii) fix $\lambda < -1$. For case (a), if $c > 1$ we know $l^*_{\mathcal{R}}$ is strictly decreasing and that $l^*_{\mathcal{C}}$ moves in the same direction if $\alpha_x \in (\underline{\alpha}^{**}_x, \infty)$ hence $l^*_{\mathcal{C}}$ is strictly increasing for $\alpha_x \in (0, \underline{\alpha}^{**}_x)$ and strictly decreasing for $\alpha_x \in (\underline{\alpha}^{**}_x, \infty)$. Similarly for case (b), if $c < 1$ we know $l^*_{\mathcal{R}}$ is strictly decreasing if and only if $\alpha_x < \underline{\alpha}^{**}_x$ and hence $l^*_{\mathcal{C}}$ is strictly increasing for $\alpha_x \in (0, \underline{\alpha}^{**}_x)$, strictly decreasing for $\alpha_x \in (\underline{\alpha}^{**}_x, \alpha^{**}_x)$, strictly increasing again for $\alpha_x \in (\alpha^{**}_x, \overline{\alpha}^{**}_x)$ and finally strictly decreasing again for $\alpha_x \in (\overline{\alpha}^{**}_x, \infty)$. □

Proof of Remark 3.

Since the citizens' loss $l^*_{\mathcal{C}}$ is the reporters' loss $l^*_{\mathcal{R}}$ evaluated at the special case $\lambda = 0$ these limits follow from evaluating their counterparts in Remark 2 at $\lambda = 0$. □

D Social media

Proof of Proposition 8.

For part (i), we use expressions (A12), (A15) and (A16) to rewrite the derivative as

$$\left.\frac{\partial \delta^*}{\partial c}\right|_{c=1} = \left(3k^* - \frac{1}{k^*} - 2\right)^{-1} \tag{D1}$$

This is decreasing in k^* and approaches $-\infty$ as $k^* \to 1$. From Lemma 3 we know that k^* is increasing in α so that the derivative above is decreasing in α. Moreover, we show in the Online Appendix, that $k^* = 1$ at $\alpha = 4, c = 1$. Thus the derivative above approaches $-\infty$ as $\alpha \to 4$.

For part (ii), we show in our Supplementary Online Appendix that when $\alpha > 4$ each equilibrium with $c < 1$ has $\delta^* > \overline{\delta}(\alpha)$ with the limit equal to $\overline{\delta}(\alpha)$ as c approaches to 1 from below, and each equilibrium with $c > 1$ has $\delta^* < \underline{\delta}(\alpha)$ with the limit equal to $\underline{\delta}(\alpha)$ as c approaches to 1 from above.

For part (iii), observe that the politician's best response $\delta(k; c)$ as in (17) is decreasing in c. If $c > 1$, the politician's best response is thus bounded above by $\delta(k; 1) = k/(1 + k)$, which in turn is bounded above by $1/2$ for all $k < 1$. Lemma 2 implies that if $c > 1$ then $\delta(k; c)$ peaks at $\hat{k}(c) < 1$. Therefore, the equilibrium $\delta^* = \delta(k^*; c)$ must be bounded above by $1/2$. Lemma 2 also implies that if $c > 1$ then $\delta(k; c)$ is decreasing in k for $k > \hat{k}(c)$. Hence, for any $c > 1$, there exists a finite $\hat{\alpha}(c)$ defined in (30) such that if $\alpha > \hat{\alpha}(c)$ then the equilibrium k^*, δ^* moves along the decreasing part of $\delta(k; c)$ with $\delta^* \to 0$ as $\alpha \to \infty$. □

References

Allcott, Hunt and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 2017, *31* (2), 211–236.

Anderson, Simon P. and John McLaren, "Media Mergers and Media Bias with Rational Consumers," *Journal of the European Economic Association*, August 2012, *10* (4), 831–859.

Arendt, Hannah, *The Origins of Totalitarianism*, revised ed., André Deutsch, 1973.

Baron, David P., "Perisistent Media Bias," *Journal of Public Economics*, 2006, *90* (1-2), 1–36.

Bennett, W. Lance and Steven Livingston, "The Disinformation Order: Disruptive Communication and the Decline of Democratic Institutions," *European Journal of Communication*, April 2018, *33* (2), 122–139.

Bergemann, Dirk and Stephen Morris, "Information Design, Bayes Persuasion and Bayes Correlated Equilibrium," *American Economic Review*, 2016, *106* (5), 586–591.

Bernhardt, Dan, Stefan Krasa, and Mattias Polborn, "Political Polarization and the Electoral Effects of Media Bias," *Journal of Public Economics*, 2008, *92*, 1092–1104.

Besley, Timothy and Andrea Prat, "Handcuffs for the Grabbing Hand? The Role of the Media in Political Accountability," *American Economic Review*, 2006, *96* (3), 720–736.

Bruns, Axel and Tim Highfield, "Blogs, Twitter, and Breaking News : The Produsage of Citizen Journalism," in Rebecca Ann Lind, ed., *Produsing Theory in a Digital World : The Intersection of Audiences and Production in Contemporary Theory*, Peter Lang Publishing, New York 2012, pp. 15–32.

Chen, Jidong and Yiqing Xu, "Information Manipulation and Reform in Authoritarian Regimes," 2014. working paper.

Codrea-Rado, Anna, "#MeToo Floods Social Media With Stories of Harassment and Assault," *New York Times*, October 16 2017.

Cohen, Noam, "Twitter on the Barricades: Six Lessons Learned," *New York Times*, June 20 2009.

Crawford, Vincent P. and Joel Sobel, "Strategic Information Transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

Edmond, Chris, "Information Manipulation, Coordination, and Regime Change," *Review of Economic Studies*, October 2013, *80* (4), 1422–1458.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin, "Why Resource-Poor Dsictators Allow Freer Media: A Theory and Evidence from Panel Data," *American Political Science Review*, 2009, *103* (4), 645–668.

Esfandiari, Golnaz, "The Twitter Devolution," *Foreign Policy*, June 2010.

Fandos, Nicholas, Cecilia Kang, and Mike Isaac, "House Intelligence Committee Releases Incendiary Russian Social Media Ads," *New York Times*, November 2017.

Faris, Rob, Hal Roberts, Bruce Etling, Nikki Bourassa, Ethan Zuckerman, and Yochai Benkler, "Partisanship, Propaganda, and Disinformation: Online Media and the 2016 US Presidential Election," Technical Report, Berkman Klein Center for Internet and Society at Harvard University, August 2017.

Fielder, Tom, "Crisis Alert: Barack Obama Meets a Citizen Journalist," in Stuart Allan and Einar Thorsen, eds., *Citizen Journalism – Global Perspectives*, Peter Lang Publishing, New York 2009, pp. 209–221.

Friedrich, Carl J. and Zbigniew K. Brzezinski, *Totalitarian Dictatorship and Autocracy*, second ed., Harvard University Press, 1965.

Gehlbach, Scott and Alberto Simpser, "Electoral Manipulation as Bureaucratic Control," *American Journal of Political Science*, 2015, *59* (1), 212–224.

_ and Konstantin Sonin, "Government Control of the Media," *Journal of Public Economics*, 2014, *118*, 163–171.

_ , _ , and Milan Svolik, "Formal Models of Nondemocratic Politics," *Annual Review of Political Science*, 2016, *19*, 565–584.

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

_ , _ , and Daniel F. Stone, "Media Bias in the Market Place: Theory," in Simon P. Anderson, Joel Waldfogel, and David Stromberg, eds., *Handbook of Media Economics*, 2015.

Guess, Andrew, Brendan Nyhan, and Jason Reifler, "Selective Exposure to Misinformation: Evidence from the Consumption of Fake News During the 2016 US Presidential Campaign," January 2018. Princeton University working paper.

Guriev, Sergei M. and Daniel Treisman, "How Modern Dictators Survive: Cooptation, Censorship, Propaganda, and Repression," 2015. working paper.

Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland, "Democracy and Transparency," *Journal of Politics*, 2011, *73* (4), 1191–1205.

Holmström, Bengt, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, 1999, *66* (1), 169–182.

Huang, Haifeng, "Propaganda as Signaling," *Comparative Politics*, forthcoming 2014.

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

Kartik, Navin, "Strategic Communication with Lying Costs," *Review of Economic Studies*, 2009, *76* (4), 1359–1395.

_ , Marco Ottaviani, and Francesco Squintani, "Credulity, Lies, and Costly Talk," *Journal of Economic Theory*, 2007, *134* (1), 93–116.

Little, Andrew T., "Elections, Fraud, and Election Monitoring in the Shadow of Revolution," *Quarterly Journal of Political Science*, 2012, *7* (3), 249–283.

_ , "Fraud and Monitoring in Noncompetitive Elections," *Political Science Research and Methods*, 2015, *3* (1), 21–41.

_ , "Propaganda and Credulity," *Games and Economic Behavior*, 2017, *102*, 224–232.

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

Morozov, Evgeny, *The Net Delusion*, Allen Lane, 2011.

Morris, Stephen and Hyun Song Shin, "Social Value of Public Information," *American Economic Review*, December 2002, *92* (5), 1521–1534.

Mullainathan, Sendhil and Andrei Shleifer, "The Market for News," *American Economic Review*, 2005, *95* (4), 1031–1053.

Rickford, Russell, "Black Lives Matter: Toward a Modern Practice of Mass Struggle," *New Labor Forum*, December 2015, *25* (1), 34–42.

Rozenas, Arturas, "Office Insecurity and Electoral Manipulation," *Journal of Politics*, 2016, *78* (1), 232–248.

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

Shafer, Jack, "Why I Love WikiLeaks: For Restoring Distrust in Our Most Important Institutions," *Slate*, November 30 2010.

Sunstein, Cass R., *#Republic: Divided Democracy in the Age of Social Media*, Princeton University Press, 2018.

Svolik, Milan, *The Politics of Authoritarian Rule*, Cambridge University Press, 2012.

Ward, Stephen J., *Ethics and the Media: An Introduction*, Cambridge University Press, 2011.

White, Aoife, "Tide of Fake News is 'Almost Overwhelming,' EU Warns," *Bloomberg*, November 2017.

Zeman, Z.A.B., *Nazi Propaganda*, second ed., Oxford University Press, 1973.