Creating Confusion*

Chris Edmond[†] Yang K. Lu[‡]

First draft: December 2017. This draft: November 2018

Abstract

We develop a model in which a politician seeks to prevent people from making informed decisions. The politician can manipulate information at a cost, but cannot commit to an information structure. The receivers are rational and internalize the politician's incentives. In the unique equilibrium of the game, the receivers' beliefs are unbiased but endogenously noisy. We use this model to interpret the rise of social media. We argue that social media simultaneously (i) improves the underlying, intrinsic precision of the receivers' information 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 politician is better off and the receivers are worse off, despite the underlying improvement in their information. But if the costs of manipulation do not fall too much, and if the receivers are also sufficiently well coordinated, the manipulation backfires. In this scenario, the politician would want to invest in commitment devices that prevent them from manipulating information.

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

JEL classifications: C7, D7, D8.

*We particularly thank Kim Sau Chung, Simon Grant, Simon Loertscher, Andy McLennan, Larry Samuelson and Xi Weng for their comments. 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 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 wants to prevent people from making informed decisions. Can the politician achieve this goal even when people are rational and perfectly understand the politician's incentives? Should we be optimistic that new social media technologies will make it easier for people to make informed decisions, making it more difficult for the politician to evade scrutiny? Or will these new technologies simply make it easier for the politician to *create confusion*, frustrating people in their efforts to make informed decisions?[1]

We develop a simple model to answer these questions. We find that the success or failure of the politician's strategy depends crucially on the interaction of two features of the environment: (i) the underlying, intrinsic *precision* of information, and (ii) the *costs* the politician incurs in trying to manipulate information — in propagating "alternative facts," so to speak.[2] We show that the politician's strategy is most successful when the intrinsic precision of the receivers' information is high and yet at the same time the costs the politician incurs in trying to manipulate information are low.

We argue that the rise of new social media technologies changes both of these features of the environment, improving the underlying, intrinsic precision of the receivers' information, but also reducing the politician's costs of manipulation. It turns out that the overall impact of the rise of social media then depends crucially on the size of the reduction in the costs of manipulation. There is a critical threshold such that if the costs of manipulation fall enough, then the politician is made better off and the receivers are worse off, despite the improvement in the underlying, intrinsic precision of their information. But if the costs of manipulation do not fall too much, and if, in addition, the receivers are also *sufficiently well coordinated*, the politician's manipulation *backfires*, 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 single sender who knows the state of the world. There is a continuum of receivers who share a common prior and receive idiosyncratic signals about the state. Each individual receiver wants to take an action that is appropriate for the state, but they may also care about the actions of the other receivers (their actions may be strategic substitutes or strategic complements). The sender seeks to *prevent* the receivers

[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).

[2] In the words of Kellyanne Conway on *Meet the Press* in January 2017. Handy (2017) and Rosen (2018) provide forceful accounts of the relationship between the Trump administration and the US media.

from taking actions that are appropriate for the state. The sender has a technology that allows them to manipulate information by choosing the receivers' signal mean at a cost that is increasing in the distance between the true state and the signal mean. The receivers are rational and internalize the sender'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 receivers' strategies are linear functions of their signals.

Section 3 solves the model and shows that there is a unique (linear) equilibrium. In equilibrium, the receivers' beliefs are unbiased but are made endogenously noisier by the sender's manipulation. In equilibrium, the amount of manipulation and the receivers' responsiveness to their signals are highly sensitive to the costs of manipulation, especially when the receivers' signals are intrinsically precise. When the costs of manipulation are high, an increase in the intrinsic precision of the receivers' signals leads to less manipulation and makes the receivers even more responsive to their signals than they would be if the sender could not manipulate at all. But when the costs of manipulation are low, an increase in the intrinsic precision of signals leads to more manipulation and makes the receivers less responsive to their signals than they would if the sender could not manipulate at all. In the limit where the costs of manipulation are negligible, the sender's manipulation renders the receivers' signals completely uninformative even if the underlying, intrinsic precision of their signals is arbitrarily high.

Section 4 provides conditions under which the sender is made better off or worse off by their ability to manipulate information. These conditions turn out to depend crucially on whether the receivers' actions are *strategic substitutes* or *strategic complements*. We find that the sender gains the most when the costs of manipulation are low and the intrinsic precision of the signals is high. But we also find that the sender's manipulation can backfire if the costs of manipulation are high, the intrinsic precision of the signals is high, and the receivers' actions are *sufficiently strong strategic complements*.

Section 5 shows that the receivers are unambiguously worse off when the sender can manipulate information. And if the costs of manipulation are low, then the receivers become increasingly worse off as their signals become intrinsically more precise, i.e., the scenario where the sender gains the most is also the scenario where the receivers lose the most.

Section 6 uses these results to interpret the rise of social media technologies. We argue that these technologies have (i) facilitated the entry of new media outlets, thereby increasing the intrinsic, potential, precision of available information, but (ii) blur the distinctions between media outlets in a way that makes it easier for "alternative facts" to propagate, thereby reducing the politician's costs of manipulation. We show that there is a threshold around which the equilibrium amount of manipulation is *extremely sensitive* to the costs of manipulation. On one side of this threshold is a "high manipulation" regime and on the other side is a "low manipulation" regime. Near this threshold, small changes in the costs of manipulation cause dramatic changes in the amount of manipulation. The overall 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 the receivers worse off. But if the costs of manipulation can be kept above the threshold, the receivers are better off because of the increase in the intrinsic precision of their information. In this sense, even small changes on the part of social media platforms that make it harder for misinformation to propagate may have large effects.

If the costs of manipulation can be kept above the threshold and in addition the receivers are sufficiently well coordinated, the politician's manipulation backfires. In this case, we would expect the politician to invest in commitment devices (e.g., in a building a reputation for straight talk, or in institutions that provide reliable fact-checking, etc) that would prevent them from manipulating in the first place.

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 has 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 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 — 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 information *receivers*, indexed by $i \in [0, 1]$, and a single informed *sender* attempting to influence their beliefs.

Receivers' payoffs and information. Each receiver wants their action $a_i \in \mathbb{R}$ to be appropriate for the underlying state $\theta \in \mathbb{R}$ (about which they are imperfectly informed). In particular, each receiver chooses a_i to minimize the expected value of the quadratic loss

$$(a_i - \theta)^2 \tag{1}$$

so that each receiver sets their action a_i equal to their expectation of θ.

In forming expectations about θ, the receivers 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 receiver then draws an idiosyncratic signal

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

where the mean y is chosen by the sender, as discussed below, and where the idiosyncratic noise ε_i is IID normal across receivers, independent of θ, with mean zero and precision $\alpha_x > 0$ (i.e., variance $1/\alpha_x$). Based on this information, each receiver sets their action to

$$a_i = \mathbb{E}[\,\theta \,|\, x_i, z\,] \tag{3}$$

To summarize, receivers have one source of information, the prior, that is free of the sender'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 sender's incentives.

Sender's payoffs and information manipulation. The sender knows the value of θ and seeks to *prevent* the receivers from taking actions that are appropriate for the underlying state θ. In particular, the sender obtains a gross benefit

$$\int_0^1 (a_i - \theta)^2 \, di \tag{4}$$

that is increasing in the dispersion of the actions a_i around θ. The sender is endowed with a technology that allows the sender to choose the mean y of the receivers' idiosyncratic signals. In particular, knowing θ, the sender 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* that the sender 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 sender is

$$V = \int_0^1 (a_i - \theta)^2 \, di - c(y - \theta)^2, \qquad c > 0 \tag{5}$$

where the parameter $c > 0$ measures how costly it is for the sender to choose values of y far from θ. The special case $c \to 0$ corresponds to a version of *cheap talk* (i.e., the sender 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 sender will always choose $y = \theta$).

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

Discussion. We interpret the receivers' common prior as representing historically determined beliefs about θ that are taken as given by the sender. We interpret the signals x_i as the receivers collecting new public information about θ that is subject to idiosyncratic noise that represents individual differences in how that public information is interpreted. The sender's choice of $y = \theta + s$ shifts the systematic, public component of the receivers' signals $x_i = \theta + s + \varepsilon_i$ but cannot change the idiosyncratic component ε_i.

We do not allow the sender to *directly* increase the amount of noise in the receivers' signals, i.e., we do not allow the sender to choose α_x. Because the sender is informed, in this alternative setup the sender would be choosing some $\alpha_x(\theta)$, say, and the receivers' Bayesian updating would need to take this dependence on the unknown θ into account. This leads to a more difficult fixed point problem between the sender's chosen $\alpha_x(\theta)$ and the receivers' beliefs that has to be solved numerically.[3] By contrast, our model where the sender distorts the signal mean has a simple analytic solution. Importantly, our model implies that *in equilibrium* the receivers' signals are endogenously noisier despite the fact that the sender cannot directly increase the amount of noise.

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

Before characterizing equilibrium outcomes in the general model with information manipulation, we first review equilibrium outcomes when there is no manipulation. This provides a natural benchmark against which the sender's technology for manipulation can be evaluated.

Equilibrium with no manipulation. Suppose the sender *cannot* manipulate information — i.e., let $c \to \infty$ so that the sender chooses $y = \theta$. This puts us in a standard linear-normal setting where each receiver's posterior expectation of θ is a precision-weighted average of their signal x_i and prior z. In particular, the optimal action of an individual receiver is

$$a(x_i, z) = \mathbb{E}[\,\theta \,|\, x_i, z\,] = \frac{\alpha_x}{\alpha_x + \alpha_z} x_i + \frac{\alpha_z}{\alpha_x + \alpha_z} z. \tag{6}$$

For future reference, let

$$k^*_{nm} := \frac{\alpha}{\alpha + 1}, \qquad \alpha := \frac{\alpha_x}{\alpha_z} > 0 \tag{7}$$

denote the response of each receiver to their signal when there is *no manipulation.* This response coefficient is determined by the relative precision α of the signal to the prior.

[3] Edmond (2013) provides an example of the numerical solution of a closely related fixed point problem.

3 Equilibrium with information manipulation

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

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

$$a(x_i, z) = kx_i + (1-k)z \tag{8}$$

The fact that the receivers' 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 Sender's problem

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

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

$$= (ky + (1-k)z - \theta)^2 + \frac{1}{\alpha_x}k^2 - c(y-\theta)^2 \tag{9}$$

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

$$y(\theta, z) = \frac{c-k}{c-k^2}\theta + \frac{k-k^2}{c-k^2}z \tag{10}$$

where the second-order condition requires

$$c - k^2 \geq 0 \tag{11}$$

Given that receivers use linear strategies, it is optimal for the sender to use a linear strategy.

The coefficients in the sender's strategy sum to one, so we can write

$$y(\theta, z) = (1-\delta)\theta + \delta z \tag{12}$$

where δ depends on the receivers' response coefficient k via

$$\delta(k) := \frac{k-k^2}{c-k^2}, \qquad c - k^2 \geq 0 \tag{13}$$

To interpret the sender's strategy, observe that if, for whatever reason, the sender chooses $\delta(k) = 0$, then the sender is choosing a signal mean y that coincides with the true θ — this corresponds to a situation where the sender chooses not to manipulate information and the receivers' signals x_i are as informative as possible about the true θ (limited only by the exogenous precision, α_x). Alternatively, if the sender chooses $\delta(k) = 1$, then the sender is choosing a signal mean y that coincides with the receivers' prior z — this corresponds to a situation where the receivers' signals x_i provides no additional information about θ.

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

3.2 Receivers' problem

The optimal action for an individual receiver with signal x_i and prior z is still given by $a(x_i, z) = \mathbb{E}[\,\theta \,|\, x_i, z]$. Our task now is to characterize these expectations. If the sender's manipulation strategy is (12), then each individual receiver has two pieces of information: (i) the common prior $z = \theta + \varepsilon_z$, where the aggregate noise ε_z is normal with mean zero and precision α_z, and (ii) the idiosyncratic signal

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

$$= \theta + \delta\varepsilon_z + \varepsilon_i \tag{14}$$

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

$$s_i := \frac{1}{1-\delta}\,(x_i - \delta z) = \theta + \frac{1}{1-\delta}\varepsilon_i \tag{15}$$

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 $y(\theta, z) = \theta$, there is no manipulation from the sender 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 $y(\theta, z) = z$, 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

$$\mathbb{E}[\,\theta \,|\, s_i, z] = \frac{(1-\delta)^2\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}s_i + \frac{\alpha_z}{(1-\delta)^2\alpha_x + \alpha_z}z \tag{16}$$

So in terms of the actual signal x_i they have

$$\mathbb{E}[\,\theta \,|\, x_i, z] = \frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}x_i + \left(1 - \frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}\right)z. \tag{17}$$

Hence indeed the receivers have a strategy of the form

$$a(x_i, z) = kx_i + (1-k)z$$

where the response coefficient k is given by

$$k(\delta) := \frac{(1-\delta)\alpha}{(1-\delta)^2\alpha + 1} \tag{18}$$

and where $\alpha := \alpha_x/\alpha_z$ is the intrinsic precision of the signal relative to the prior.

3.3 Equilibrium determination

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

Now define

$$\mathcal{K}(c) := \{\, k \,:\, 0 \leq k \leq \min[c, 1] \,\}, \qquad c > 0 \tag{19}$$

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 receivers' $k(\delta)$ and the sender's $\delta(k)$.

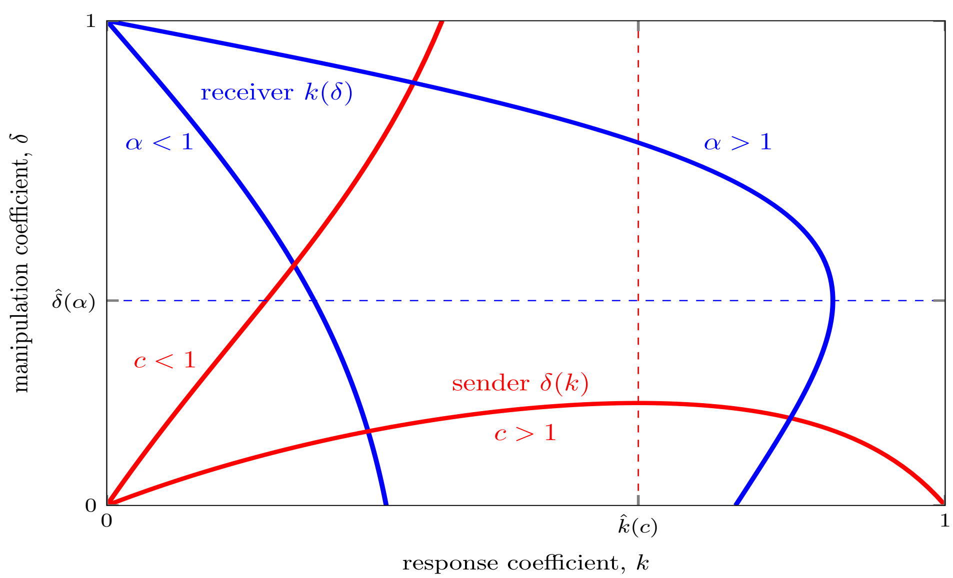

Figure 1: Unique equilibrium.

There is a unique equilibrium, that is, a unique pair k^*, δ^* simultaneously satisfying the receivers' best response $k(\delta)$ and the sender's best response $\delta(k)$. For $\alpha > 1$ there is a critical point $\hat{\delta}(\alpha)$ such that the receivers' $k(\delta)$ is increasing in δ for $\delta < \hat{\delta}(\alpha)$. For $c > 1$ there is a critical point $\hat{k}(c)$ such that the sender'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.

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 sender's equilibrium strategy $y(\theta, z) = (1-\delta^*)\theta + \delta^* z$ and the receivers' equilibrium strategy $a(x_i, z) = k^* x_i + (1-k^*)z$.

When is the receivers' best response non-monotone? As shown in Figure 1, the receivers' best response $k(\delta)$ can be non-monotone in the sender's manipulation δ. Specifically:

LEMMA 1. The receivers' best response $k : [0,1] \to \mathbb{R}_+$ defined by

$$k(\delta) := \frac{(1-\delta)\alpha}{(1-\delta)^2\alpha + 1}, \qquad \delta \in [0,1], \qquad \alpha > 0 \tag{20}$$

is first increasing then decreasing in δ with a single peak at $\delta = \hat{\delta}(\alpha)$ given by

$$\hat{\delta}(\alpha) := \begin{cases} 0 & \text{if } \alpha \leq 1 \\ 1 - 1/\sqrt{\alpha} & \text{if } \alpha > 1 \end{cases} \tag{21}$$

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

Intuitively, we might expect that in the presence of manipulation the receivers are less responsive to their signals x_i than they would be absent manipulation. Lemma 1 tells us that this raw intuition is incomplete. If the intrinsic precision of the signal is relatively low, $\alpha \leq 1$, then the maximum of $k(\delta)$ is obtained at the boundary and so in this special case we get $k(\delta) < k^*_{nm}$, which agrees with the raw intuition. But if the intrinsic precision of the signal is relatively high, $\alpha > 1$, then the maximum of $k(\delta)$ is obtained in the interior so that the receivers will have $k(\delta) < k^*_{nm}$ only when δ is also sufficiently high. That is, if $\alpha > 1$ then the sender will need to choose a relatively large amount of manipulation if they want to achieve a reduction in the receivers' responsiveness, $k(\delta) < k^*_{nm}$.

Why is the receivers' best response non-monotone? Lemma 1 implies that if α is relatively high and the amount of manipulation δ is relatively low, then the receivers 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 (15), and (ii) the effect of δ on the correlation between a receiver's actual 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 (15), 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 receivers' k. If this was the only effect the receivers' response coefficient would be unambiguously decreasing in δ (for any level of α), in line with the raw intuition. But an increase in δ also increases the correlation between a receiver's actual signal x_i and their prior z and the strength of this correlation effect depends on both α and δ. 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 receivers 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 sender's best response non-monotone? Similarly, as shown in Figure 1, the sender's best response $\delta(k)$ can be non-monotone in the receivers' response coefficient k.

LEMMA 2. The sender's best response $\delta : \mathcal{K}(c) \to [0,1]$ defined by

$$\delta(k) := \frac{k - k^2}{c - k^2}, \qquad k \in \mathcal{K}(c), \qquad c > 0 \tag{22}$$

is first increasing then decreasing in k with a single peak at $k = \hat{k}(c)$ given by

$$\hat{k}(c) = \begin{cases} c & \text{if } c < 1 \\ c - \sqrt{c(c-1)} & \text{if } c > 1 \end{cases} \tag{23}$$

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

This lemma says that if $c < 1$ then the maximum of $\delta(k)$ is obtained at the boundary where $k = c$ and the sender's manipulation is $\delta(c) = 1$. Hence in this case, $\delta(k)$ is unambiguously increasing in k. Intuitively, if the cost of manipulation is relatively low, then whenever the receivers respond more to their signals, the sender will choose a higher level of manipulation. But if instead $c > 1$ the maximum is obtained in the interior[4] which means that for high enough k the sender responds by choosing a *lower* level of manipulation $\delta(k)$.

Why is the sender's best response non-monotone? To understand why it can be the case that high values of the receivers' response coefficient k can lead the sender to choose *less* manipulation, we first write the sender's gross payoff as

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

where $A := \int_0^1 a_i \, di$ denotes the population aggregate action. This emphasizes that the sender can be made better off through either shifting the aggregate action A away from θ or through increasing the dispersion of a_i around A. If receivers use the strategy $a_i = kx_i + (1-k)z$ then the gap between a_i and A is $a_i - A = k(x_i - y) = k\varepsilon_i$, proportional to the idiosyncratic

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

noise ε_i. Similarly if the sender uses the strategy $y = (1-\delta)\theta + \delta z$ then the gap between A and θ is $A - \theta = (1 - k(1-\delta))\varepsilon_z$, proportional to the aggregate noise ε_z. Then using $\int_0^1 \varepsilon_i^2 \, di = 1/\alpha_x$ and subtracting off the cost of the manipulation, we can write the sender's objective as:

$$V = \left(B(\delta, k) - C(\delta)\right) \varepsilon_z^2 + \frac{1}{\alpha_x} k^2 \tag{24}$$

where

$$B(\delta, k) := (1 - k(1-\delta))^2, \qquad \text{and} \qquad C(\delta) := c\delta^2. \tag{25}$$

We can now view the sender's problem as being equivalent to choosing $\delta \in [0,1]$ to maximize (24) taking $k \in [0,1]$ as given. Up to constants that do not matter for the optimal policy, $B(\delta, k)$ is the gross benefit from choosing δ and $C(\delta)$ is the cost of it. Whether or not $\delta(k)$ is increasing in k depends on whether δ and k are *strategic complements* or *strategic substitutes* in (24). Since the cost $C(\delta)$ is independent of k, this is determined by the cross-derivative of $B(\delta, k)$ alone. The marginal benefit can be written

$$\frac{\partial B}{\partial \delta} = 2(1 - k(1-\delta))k \tag{26}$$

where, as above, $1 - k(1-\delta)$ measures the *exposure* of the aggregate action A to the aggregate noise ε_z. Then differentiating the marginal benefit with respect to k we can write the cross-derivative as proportional to

$$\underbrace{(1 - k(1-\delta))}_{\text{effectiveness of exposing } A \text{ to } \varepsilon_z} \qquad - \qquad \underbrace{(1-\delta)k}_{\text{change in exposure of } A \text{ to } \varepsilon_z} \tag{27}$$

There are two effects of k on the sender's incentives: (i) if receivers put a higher weight on their private signals, i.e., have a higher k, then a higher δ will be more effective at exposing A to ε_z and this increases the marginal benefit of δ to the sender, but also (ii) a higher k reduces the amount of exposure of A to ε_z and this decreases the marginal benefit of δ. When the first effect dominates, δ and k are *strategic complements* and a higher value of k encourages the sender to also choose a higher δ. But when the second effect dominates, δ and k are *strategic substitutes* and a higher value of k encourages the sender to choose a lower δ.

When k is small (say $k \to 0$), the first effect in (27) dominates regardless of δ. Likewise, when δ is large (say $\delta \to 1$), the first effect in (27) dominates regardless of k. The second effect can dominate, but only through some combination of k being high enough and δ being low enough. More precisely, substituting $\delta(k)$ from (22) into (27) and rearranging, we find that the second effect dominates if and only if $k > \hat{k}(c)$ where $\hat{k}(c)$ is the critical value defined in (23) above. Roughly speaking, for low levels of the cost of manipulation c, the sender will be inclined to choose high levels of δ so that for most values of k the first effect dominates and $\delta(k)$ is increasing. For high levels of c, the sender will be inclined to choose low levels of δ so that the second effect dominates and $\delta(k)$ is decreasing for high enough k.

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 relative precision $\alpha := \alpha_x/\alpha_z > 0$, which measures how responsive individuals would be to their signals absent manipulation, and (ii) the sender's cost of manipulation $c > 0$.

To see how the equilibrium k^* and δ^* vary with α and c, observe from (20) that we can write the receivers' best response as $k(\delta; \alpha)$ independent of the sender's cost c. Likewise, from (22) we can write the sender's best response as $\delta(k; c)$ independent of the relative precision 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 receivers' best response, changes in α shift the receivers' best response $k(\delta; \alpha)$ along an unchanged $\delta(k; c)$ for the sender. Likewise, since c enters only the sender's best response, changes in c shift the sender's best response $\delta(k; c)$ along an unchanged $k(\delta; \alpha)$ for the sender.

Relative precision. Changes in the relative precision $\alpha > 0$ have the following effects:

LEMMA 3.

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

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

$$\alpha < \hat{\alpha}(c) \tag{28}$$

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

For any level of the sender's manipulation δ, a marginal increase in the relative precision α makes the receivers more responsive to their signals so that $k(\delta; \alpha)$ shifts out along the sender's $\delta(k; c)$ thereby increasing k at every δ and hence unambiguously increasing k^*. The effect on the sender's manipulation then hinges on whether or not k and δ are strategic complements *in equilibrium* — i.e., whether the sender's best response $\delta(k; c)$ is increasing or decreasing in k at the equilibrium k^* and δ^*. Then using Lemma 2 we say that k^* and δ^* are strategic complements in equilibrium when α and c are such that $k^*(\alpha, c) < \hat{k}(c)$.

We illustrate this result in Figure 2 which shows the receivers' equilibrium response k^* (left panel) and sender's equilibrium manipulation δ^* (right panel) as functions of the relative precision α for the case of low costs of manipulation $c < 1$ (in blue) and high costs of manipulation $c > 1$ (in red). If $c < 1$ then we know from Lemma 2 that $k \leq c = \hat{k}(c)$ so k and δ are unambiguously strategic complements for the sender and so, as functions of α, the equilibrium k^* and δ^* increase or decrease together, as shown by the blue curves. Alternatively, if $c > 1$, then whether or not k and δ are strategic complements also depends on the level of k^*, which depends on the level of α. If α is low then k^* will also be low so that k^* and δ^* remain strategic complements and therefore increase or decrease together following a change in α. But if α is high enough to make k^* and δ^* strategic substitutes then k^* and δ^* will move in opposite directions following a change in α.

Cost of manipulation. Changes in the cost $c > 0$ have the following effects:

LEMMA 4.

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

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

$$c < \hat{c}(\alpha) \tag{29}$$

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

For any level of the receivers' k, a marginal increase in the cost c decreases the sender's incentive to manipulate so that $\delta(k; c)$ shifts down along the receivers' $k(\delta; \alpha)$ thereby decreasing δ at every k and hence unambiguously decreasing δ^*. The effect on the receivers' k then hinges on whether or not, in equilibrium, the precision effect is larger than the correlation effect — i.e., whether the receivers' $k(\delta; \alpha)$ is increasing or decreasing in δ at the equilibrium k^* and δ^*. Then using Lemma 1 we say that the precision effect dominates in equilibrium when α and c are such that $\delta^*(\alpha, c) > \hat{\delta}(\alpha)$.

We illustrate this result in Figure 3 which shows the receivers' equilibrium response k^* (left panel) and sender's equilibrium manipulation δ^* (right panel) as functions of the cost of manipulation c for the case of low relative precision $\alpha < 1$ (in blue) and high relative precision $\alpha > 1$ (in red). If $\alpha < 1$ then we know from Lemma 1 that the precision effect dominates so that the receivers' $k(\delta; \alpha)$ curve is decreasing and so, as functions of c, the equilibrium k^* and δ^* move in opposite directions following a change in c, as shown by the blue curves. Alternatively, if $\alpha > 1$, then whether the precision or correlation effect dominates also depends on the level of δ^*, which 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.

When is k^* less than k^*_{nm}? Lemma 3 tells us that the receivers' k^* is increasing in the relative precision α. But since the benchmark $k^*_{nm} = \alpha/(\alpha + 1)$ is also increasing in α, this does not yet tell us whether in equilibrium k^* is less or more than k^*_{nm}. In the following lemma, we provide a simply necessary and sufficient condition for $k^* < k^*_{nm}$. In particular:

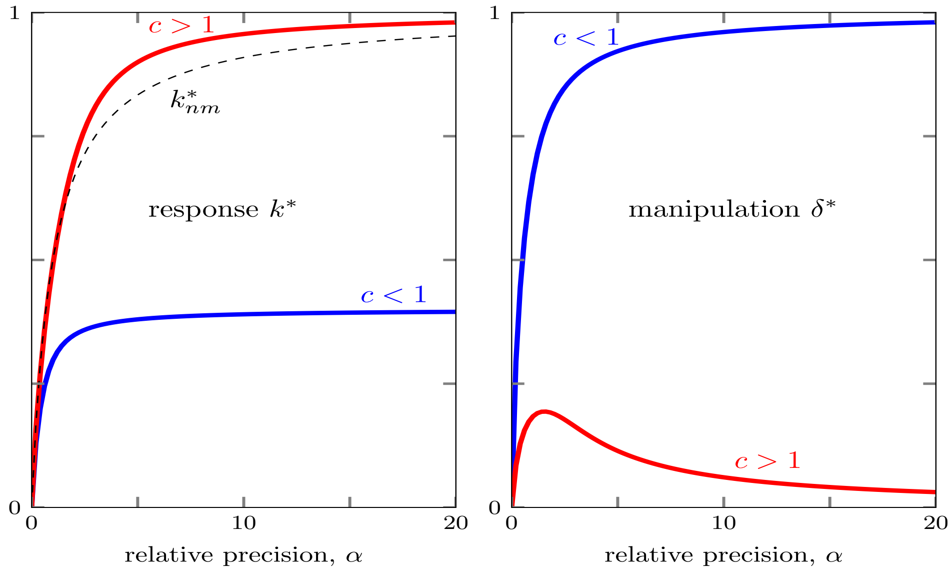

Figure 2: Changes in the relative precision, α.

Receivers' equilibrium response k^* (left panel) and sender's equilibrium manipulation δ^* (right panel) as functions of the relative precision α for various levels of the cost of manipulation c. The receivers' k^* is monotone increasing in α and asymptotes to $\min[c, 1]$ as $\alpha \to \infty$. If $c < 1$ then k^* and δ^* are strategic complements for the sender, 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 k^* and δ^* become strategic substitutes for the sender at which point δ^* starts to decrease and asymptotes to zero as $k^* \to 1$.

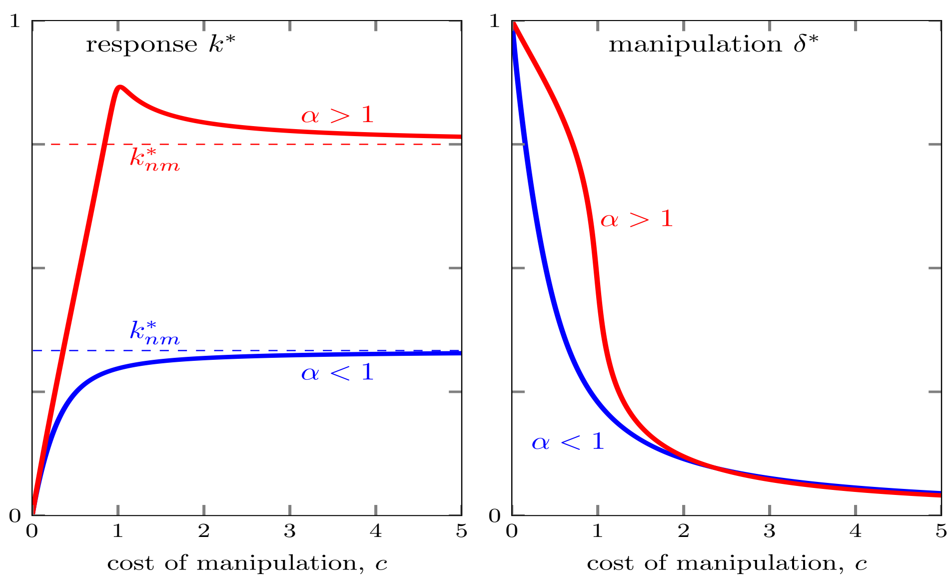

Figure 3: Changes in the cost of manipulation, c.

Receivers' equilibrium response k^* (left panel) and sender's equilibrium manipulation δ^* (right panel) as functions of the sender's cost of manipulation c for various levels of the relative precision α. The sender's δ^* is monotone decreasing in c and asymptotes to zero as $c \to \infty$. If $\alpha < 1$, the precision effect dominates so that as δ^* decreases the receivers' 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 $c \to \infty$.

Lemma 5. Receivers are less responsive to their signals with manipulation

$$k^*(\alpha, c) < k^*_{nm}(\alpha) \qquad \text{if and only if} \qquad c < c^*_{nm}(\alpha) \tag{30}$$

where

$$c^*_{nm}(\alpha) = \begin{cases} \dfrac{\alpha}{\alpha - 1}\left(\dfrac{\alpha}{\alpha+1}\right)^2 & \text{if } \alpha > 1 \\ \\ +\infty & \text{if } \alpha \leq 1 \end{cases} \tag{31}$$

In other words, if $\alpha \leq 1$ then we know that $k^* < k^*_{nm}$ regardless of c but if $\alpha > 1$ then whether or not the receivers' k^* is less than k^*_{nm} does depend on c. The key to this result is Lemma 1 which tells us that in order for $k(\delta) < k^*_{nm}$ it must be the case that the sender has manipulation δ that is sufficiently large, in particular, must have manipulation $\delta > 1 - 1/\alpha$. Hence if α and c are such that $\delta^*(\alpha, c) > 1 - 1/\alpha$ then in equilibrium the sender is indeed choosing enough manipulation that the receivers become less responsive, $k^* < k^*_{nm}$. For $\alpha > 1$, the critical cost $c^*_{nm}(\alpha)$ given in (31) is the unique c that delivers $\delta^*(\alpha, c) = 1 - 1/\alpha$.

Graphically, 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 $\alpha \to \infty$. Thus 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 the cost parameter c is more or less than 1. If $c < 1$ then in this limiting equilibrium the receivers are less responsive to their signals than they would be absent manipulation but if $c > 1$ then they are more responsive.

Fundamental vs. strategic uncertainty. This limiting equilibrium also provides a simple characterization of the size of the change in the receivers' responsiveness $k^*_{nm} - k^*$. In particular, as $\alpha \to \infty$ we have $k^* \to \min[c, 1]$ and $k^*_{nm} \to 1$ so that:

Remark 1. As signals become arbitrarily precise, the change in the receivers' equilibrium responsiveness is

$$\lim_{\alpha \to \infty} \left(k^*_{nm} - k^* \right) = \begin{cases} 1 - c & \text{if } c < 1 \\ \\ 0 & \text{if } c > 1 \end{cases} \tag{32}$$

while the equilibrium amount of manipulation is

$$\lim_{\alpha \to \infty} \delta^* = \begin{cases} 1 & \text{if } c < 1 \\ \\ 0 & \text{if } c > 1 \end{cases} \tag{33}$$

In this limit, there is no *fundamental uncertainty* in the sense that, absent manipulation, all receivers would respond one-for-one to their signals. The limit $\alpha \to \infty$ corresponds to a scenario where the underlying information environment is benign for the receivers. But in the presence of manipulation there remains *strategic uncertainty* even as $\alpha \to \infty$. The consequences of this strategic uncertainty depend on how costly it is for the sender to manipulate.

If the cost is relatively high, $c > 1$, then the limiting equilibrium features no manipulation ($\delta^* = 0$) and receivers that are fully responsive to their signals ($k^* = 1$). But if the cost is relatively low, $c < 1$, then the limiting equilibrium features full manipulation ($\delta^* = 1$) and receivers that are less-than-fully responsive to their signals ($k^* = c < 1$).[5]

Now fix $k^* = \min[c, 1]$ and suppose there is *nearly cheap talk*, $c \to 0$.[6] In this scenario the receivers do not respond to their signals, $k^* \to 0$, they make decisions *based on their prior alone*. This is because, although the *intrinsic precision* of their signals is high, the low cost of manipulation makes k^* and δ^* strategic complements for the sender so that high α also implies high δ^*, thereby making the *equilibrium precision* of their signals low.

4 Sender's welfare

In this section we establish conditions under which the sender is made better off by their ability to manipulate information. But we will also show that this need not always be the case, there are situations where the sender's manipulation *backfires* in the sense that the sender would want to be able to credibly commit to not using their manipulation technology. It turns out that the possibility of manipulation backfiring on the sender depends crucially on there being strategic interactions among the receivers.

4.1 Strategic interactions among receivers

In our benchmark model, the sender's payoffs depend on the actions of the receivers but there are no strategic interactions among the receivers themselves. We now consider a more general version of our model that does allow for such interactions. In particular, we now suppose that the receivers seek to minimize the expected value of the quadratic loss

$$(1-\lambda)(a_i - \theta)^2 + \lambda(a_i - A)^2, \qquad \lambda < 1 \tag{34}$$

where λ governs the strategic interactions among receivers[7] and where $A := \int_0^1 a_i \, di$. Our benchmark model is the special case $\lambda = 0$. If $\lambda < 0$ then each receiver's action a_i and the aggregate action A are *strategic substitutes*. In this case each receiver wants their action a_i to align with θ but also wants to make their a_i stand out from the crowd, differing from A. If $\lambda \in (0, 1)$ then each receiver's action a_i and the aggregate action A are *strategic complements*. In this case, each receiver wants their action a_i to align both with θ and with A.

As we show in the Appendix, it turns out that all of our results so far go through so long as we redefine the composite parameter α to be

$$\alpha := (1-\lambda)\,\frac{\alpha_x}{\alpha_z} > 0. \tag{35}$$

[5]We discuss the knife-edge case $c = 1$ in detail in our supplementary Online Appendix.

[6]The order of limits matters. Here we first take $\alpha \to \infty$ and then take $c \to 0$.

[7]The restriction $\lambda < 1$ ensures that loss function is strictly convex in a_i.

If the receivers' actions are strategic substitutes, $\lambda < 0$, each receiver seeks to differentiate themself and gives more weight to their idiosyncratic signal x_i so that α is higher than the underlying α_x/α_z ratio warrants. But if the receivers' actions are strategic complements, $\lambda \in (0, 1)$, the receivers seek to coordinate with one-another and give more weight to their common prior z so that α is lower than the underlying α_x/α_z ratio warrants.[8]

Because the underlying parameters $\lambda, \alpha_x, \alpha_z$ only matter for the equilibrium and its comparative statics through the composite parameter α we have till now chosen to focus on the special case of $\lambda = 0$ so as to streamline the exposition. But we will now see that the *welfare properties* of the equilibrium do depend on these three underlying parameters separately and in particular depend on the strategic interactions among receivers as measured by λ.

4.2 Sender can benefit from manipulation

Let v^* denote the sender's ex ante expected utility when they can manipulate information, i.e., the expectation of (24) with respect to the prior that θ is normally distributed with mean z and precision α_z. Let v^*_{nm} denote the sender's ex ante expected utility when they can not manipulate information. We say that the sender gains from manipulation if $v^* > v^*_{nm}$. The following result gives sufficient conditions for the sender to gain from manipulation:

PROPOSITION 2. The sender gains from manipulation, $v^* > v^*_{nm}$, if either:

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

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

The key to this proposition is Lemma 5 which says that receivers are relatively less responsive to their signals, $k^* < k^*_{nm}$, if and only if $c < c^*_{nm}(\alpha)$. The proposition then says that there are two basic scenarios in which the sender is better off when they can manipulate information. In the first scenario, the receivers' actions are strategic substitutes, $\lambda < 0$, and the sender gains when in equilibrium the receivers respond relatively more to their idiosyncratic signal x_i than they would absent manipulation, $k^* > k^*_{nm}$. In turn, the receivers are relatively more responsive to their signals when the costs of manipulation are *high enough*, $c > c^*_{nm}(\alpha)$. So, perhaps surprisingly, the sender can be better off when they can manipulate information even if the costs of doing so are high. One might think that, in a signalling game like this, the sender might be "trapped" into taking costly manipulation that merely maintains equilibrium expectations on the part of the receivers while reducing the sender's net payoff. But instead we find that the sender may be able to avoid that kind of trap

[8] In other words, there is no particular relationship between whether the receivers' a_i and A are strategic complements or not and whether the receivers' k^* and the sender's δ^* are strategic complements or not. The latter depends on λ only via the composite parameter α.

when the receivers are averse to coordination. In the second scenario, the receivers' actions are strategic complements, $\lambda > 0$, and the sender gains when in equilibrium the receivers respond relatively less to their idiosyncratic signal x_i than they would absent manipulation, $k^* < k^*_{nm}$, which happens when the costs of manipulation are low enough, $c < c^*_{nm}(\alpha)$. If there are no strategic interactions among receivers, $\lambda = 0$, the sender gains regardless of c.

Decomposition. To understand this result, we decompose the gains from manipulation as

$$v^* - v^*_{nm} = (v(k^*) - v_{nm}(k^*)) + (v_{nm}(k^*) - v_{nm}(k^*_{nm})) \tag{36}$$

where the value functions $v(k)$ and $v_{nm}(k)$ are related by

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

and where, in slight abuse of notation, we now use $V(\delta, k)$ to denote the sender's *ex ante* expected utility if they choose manipulation δ and the receivers have response k. That is

$$V(\delta, k) = \frac{1}{\alpha_z}\left(B(\delta, k) - C(\delta)\right) + \frac{1}{\alpha_x}k^2 \tag{38}$$

As before, $B(\delta, k) = (1 - k(1 - \delta))^2$ and $C(\delta) = c\delta^2$ capture the sender's gross benefits and cost of manipulation. In this notation, $v^* = v(k^*)$ and $v^*_{nm} = v_{nm}(k^*_{nm})$.

Since $v(k) \geq v_{nm}(k)$ for all k, the first term in the decomposition (36) is not negative. So to obtain Proposition 2 it is sufficient that the second term in (36) is positive. Then notice that the second term is a comparison of the function $v_{nm}(k)$ at two different points, k^* and k^*_{nm}. A direct calculation gives

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

This quadratic in k decreases from $v_{ms}(0) = 1/\alpha_z$ till it reaches its global minimum at $k_{min} := \alpha_x/(\alpha_x + \alpha_z)$ and then increases to $v_{ms}(1) = 1/\alpha_x$. Now suppose the receivers' actions are strategic substitutes, $\lambda < 0$. Then $k^*_{nm} > k_{min}$ and so $v_{ms}(k)$ is strictly increasing on $(k^*_{nm}, 1)$ and hence if $k^* > k^*_{nm}$ we know $v_{nm}(k^*) > v_{nm}(k^*_{nm})$ and the second term in (36) is positive. Alternatively, suppose the receivers' actions are strategic complements, $\lambda > 0$. Then $k^*_{nm} < k_{min}$ and so $v_{nm}(k)$ is strictly decreasing on $(0, k^*_{nm})$ and hence if $k^* < k^*_{nm}$ we know $v_{nm}(k^*) > v_{nm}(k^*_{nm})$ and the second term in (36) is again positive. In short, if $\lambda < 0$ the sender gains if receivers are relatively more responsive to their signals, $k^* > k^*_{nm}$, but if $\lambda > 0$ the sender gains if receivers are relatively less responsive to their signals, $k^* < k^*_{nm}$. Then from Lemma 5 we know that $k^* < k^*_{nm}$ if and only if $c < c^*_{nm}(\alpha)$. If $\lambda = 0$ then $k^*_{nm} = k_{min}$ and so $v_{nm}(k^*) \geq v_{nm}(k^*_{nm})$ regardless of k^*.

4.3 But sender's manipulation can backfire

So indeed the sender can benefit from manipulation. But the sender is not always better off. There are situations where the sender's manipulation *backfires* in the sense that the sender would be better off if they could credibly commit to not use their manipulation technology. The following result gives sufficient conditions for this to happen:

PROPOSITION 3.

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

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

Again there are two basic scenarios, depending on whether the receivers' actions are strategic substitutes or complements. In the first scenario, the sender's costs of manipulation are relatively low, $c < 1$, and the receivers' actions are relatively *strong strategic substitutes*, $\lambda < -1/2$. If in addition the intrinsic signal precision α_x is also sufficiently low, $\alpha_x < \underline{\alpha}_x^*$, then the sender's manipulation backfires, $v^* < v_{nm}^*$. Perhaps surprisingly, the sender can be worse off with the manipulation technology even if the costs of using it are very low. One might think that endowing the sender with the ability to manipulate information at low cost would necessarily be to the advantage of the sender, but here we see that this need not be the case. No matter how low the cost of manipulation c is, if the receivers' actions are strong strategic substitutes, $\lambda < -1/2$, the sender will be worse off if the intrinsic signal precision α_x is low enough. Interestingly, however, this outcome does not require the intrinsic signal precision α_x to be *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 sender is worse off by making λ negative enough and the costs of manipulation c low enough. In particular:

REMARK 2. 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 sender's manipulation backfires, $v^* < v_{nm}^*$.

In the second scenario outlined in Proposition 3, the costs of manipulation are relatively high, $c > 1$, and the receivers' actions are relatively *strong strategic complements*, $\lambda > 1/2$. If in addition the intrinsic signal precision α_x is also sufficiently high, $\alpha_x > \overline{\alpha}_x^*$, then the sender's manipulation again backfires, $v^* < v_{nm}^*$. So the sender can indeed be worse off when the costs of manipulation are high, but for this to happen the receivers must have sufficiently strong incentives to coordinate their actions and have signals that are intrinsically quite precise.

To understand Proposition 3, recall the decomposition in (36) above. Since the first term $v(k^*) - v_{nm}(k^*)$ is not negative, to obtain backfiring the second term $v_{nm}(k^*) - v_{nm}(k_{nm}^*)$ must be sufficiently negative. Hence a *necessary condition* for manipulation to backfire when

$\lambda < 0$ is that the costs of manipulation are relatively low, $c < c^*_{nm}(\alpha)$, so that $k^* < k^*_{nm}(\alpha)$ and hence $v_{nm}(k^*) < v_{nm}(k^*_{nm})$ for the reasons given following equation (39) above. Now observe that the function $v_{nm}(k)$ given in (39) 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 $k^* < k^*_{nm}(\alpha)$ 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 sender to be worse off. In turn, $c < c^*_{nm}(\alpha)$ is guaranteed if $c < 1$ and α_x is low enough. Similarly, a necessary condition for manipulation to backfire when $\lambda > 0$ is that the costs of manipulation are relatively high, $c > c^*_{nm}(\alpha)$, so that $k^* > k^*_{nm}(\alpha)$ and hence again $v_{nm}(k^*) < v_{nm}(k^*_{nm})$. As α_x increases, the function $v_{nm}(k)$ behaves more like the decreasing $(1-k)^2$ term so that if $k^* > k^*_{nm}(\alpha)$ 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 sender is again worse off. In turn, $c > c^*_{nm}(\alpha)$ is guaranteed if $c > 1$ and α_x is high enough.

Figure 4 illustrates both gains from manipulation and backfiring in the same figure. The top row shows the sender's gain 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 sender's welfare v^*_{nm} in the absence of manipulation (dashed black). The left column shows the results when the receivers' actions are strong strategic substitutes, $\lambda < -1/2$. The right column shows the results when the receivers' actions are strong strategic complements, $\lambda > 1/2$.

4.4 Sender's payoff levels

A striking feature of Figure 4 is that the sender gains the most from manipulation when c is low and α_x is high, regardless of λ. Although the sender may gain from manipulation, $v^* > v^*_{nm}$, when (say) c is relatively high and $\lambda < 0$, this does not mean that the actual *level* of v^* is high. The levels of v^* and v^*_{nm} are characterized by:

PROPOSITION 4.

(i) For each $\lambda > -1$, the sender's payoff with and without manipulation, v^* and v^*_{nm}, are both strictly decreasing in the intrinsic signal precision α_x.

(ii) Regardless of λ, the sender's payoff has limits

$$\lim_{\alpha_x \to 0^+} v^* = \lim_{\alpha_x \to 0^+} v^*_{nm} = \frac{1}{\alpha_z} \tag{40}$$

$$\lim_{\alpha_x \to \infty} v^* = \max\left[0\,,\ \frac{1-c}{\alpha_z}\right] \qquad \text{and} \qquad \lim_{\alpha_x \to \infty} v^*_{nm} = 0 \tag{41}$$

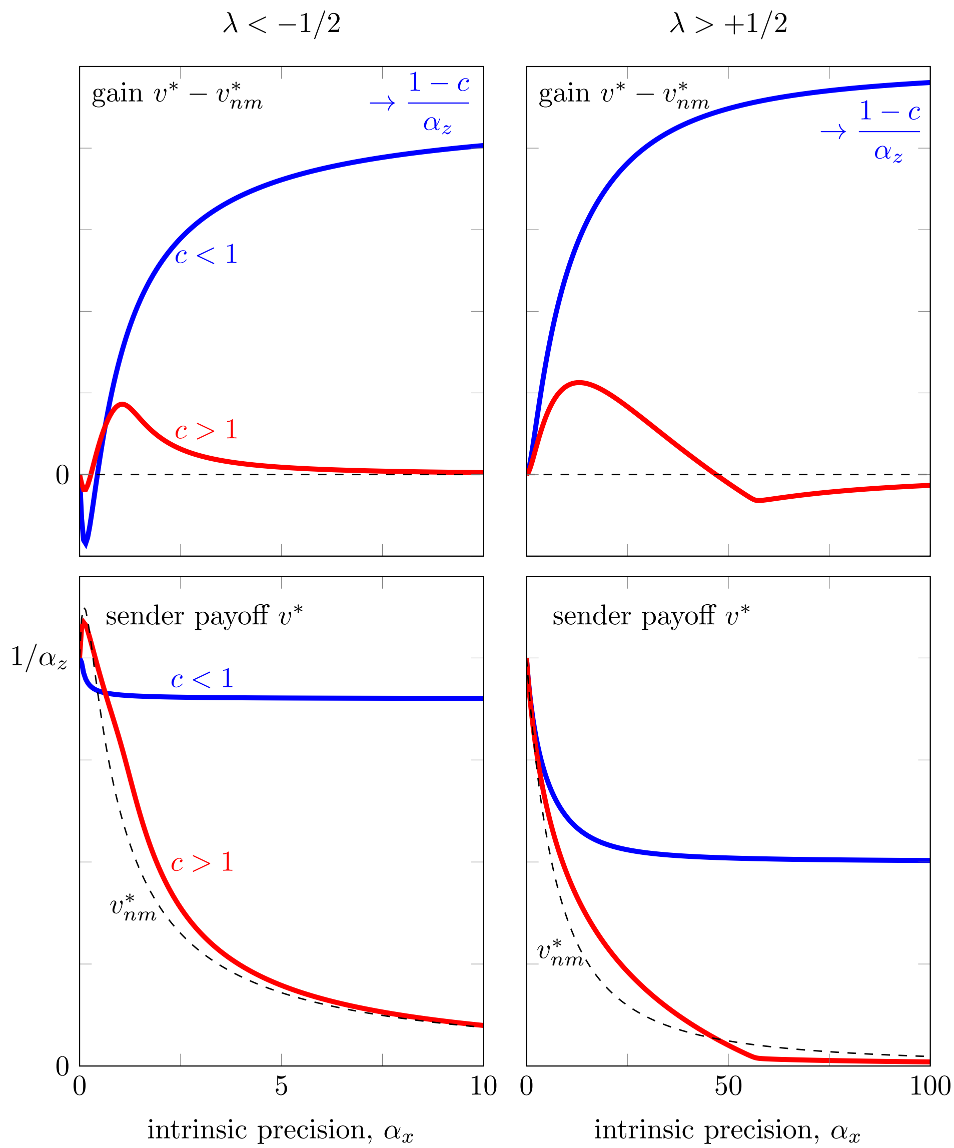

Figure 4: Sender gains most when c is low and α_x is high.

Sender's gain from manipulation $v^* - v^*_{nm}$ (top row) and payoff v^* (bottom row) as functions of α_x for various costs of manipulation c when the receivers' actions are strong strategic substitutes $\lambda < -1/2$ (left column) or strong strategic complements $\lambda > 1/2$ (right column). The sender's payoff absent manipulation v^*_{nm} asymptotes to zero as $\alpha_x \to \infty$. If $c > 1$ the sender'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 sender gains. The sender gains 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.

So long as the receivers' actions are not very strong strategic substitutes, $\lambda > -1$, then in both cases an increase in the intrinsic signal precision α_x reduces the sender's payoff. But the *magnitude* of the reduction in the sender's payoff depends on whether the sender can manipulate information and if so at what cost. In particular, if $c < 1$ then v^* asymptotes to the constant $(1-c)/\alpha_z > 0$ 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.

Direct and indirect effects of α_x. To understand this result, observe that the intrinsic signal precision α_x has both a direct effect on the sender's payoff and an indirect effect through the receivers' response coefficient. The direct effect of increasing α_x is always to reduce the sender's payoff. Likewise, an increase in α_x increases the receivers' response coefficient regardless of whether the sender can manipulate information or not. So the net effect of an increase in α_x will depend on whether the induced increase in the receivers' response coefficient makes the sender worse off or not.

Does an increase in k^* make the sender worse off? To see the effect of changes in the receivers' response coefficient on the sender, write the value function $v(k)$ from (37) above as

$$v(k) := \max_{\delta \in [0,1]} \left[\frac{1}{\alpha_z} \left(B(\delta, k) - C(\delta) \right) + \frac{1}{\alpha_x} k^2 \right] \tag{42}$$

Applying the envelope theorem to calculate $v'(k)$, then evaluating at the equilibrium k^* and collecting terms gives

$$v'(k^*) = -\frac{\lambda}{1-\lambda} \frac{2}{\alpha_x} k^* \tag{43}$$

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

Net effect of α_x. Putting these results together we find that if the receivers' actions are strategic complements, then an increase in α_x unambiguously makes the sender worse off because the direct effect of higher α_x and the indirect effect via higher k^* reinforce each other to reduce the sender's payoff. But if the receivers' actions are strategic substitutes, the net effect is ambiguous because the direct effect of higher α_x and the indirect effect via higher k^* pull in opposite directions. We show in the Appendix that if $\lambda > -1$, i.e., unless the receivers' actions are strong strategic substitutes, the direct effect dominates so that the sender's payoff v^* is strictly decreasing in α_x. Similarly, we find that if $\lambda > -1$ the sender's payoff absent manipulation v^*_{nm} is also strictly decreasing in α_x. The bottom row of Figure 4 shows both the case $\lambda < -1$, so that v^* and v^*_{nm} may not be monotonic in α_x (left panel), and the case $\lambda > -1$, so that both v^* and v^*_{nm} are strictly decreasing in α_x (right panel).

5 Receivers' welfare

We now turn to the implications of the sender's manipulation for the receivers' welfare. Let l^* denote the receivers' ex ante expected loss when the sender can manipulate information, i.e., the expectation of (34) with respect to the prior that θ is normally distributed with mean z and precision α_z. Let l^*_{nm} denote the receivers' ex ante expected loss when the sender can not manipulate information. Our main result here is:

PROPOSITION 5.

(i) The receivers are worse off with manipulation, $l^* \geq l^*_{nm}$, strictly so if $\delta^* > 0$.

(ii) For each $c > 1$, the receivers' loss with manipulation l^* is strictly decreasing in the intrinsic signal precision α_x. For each $c < 1$ there is a critical point α_x^{**} such that l^* is strictly decreasing for $\alpha_x < \alpha_x^{**}$ and strictly increasing for $\alpha_x > \alpha_x^{**}$.

(iii) The receivers' loss without manipulation l^*_{nm} is strictly decreasing in α_x.

(iv) The receivers' loss has limits

$$\lim_{\alpha_x \to 0^+} l^* = \lim_{\alpha_x \to 0^+} l^*_{nm} = \frac{1-\lambda}{\alpha_z} \tag{44}$$

$$\lim_{\alpha_x \to \infty} l^* = \begin{cases} \dfrac{1-\lambda}{\alpha_z} & \text{if } c < 1 \\ 0 & \text{if } c > 1 \end{cases} \qquad \text{and} \qquad \lim_{\alpha_x \to \infty} l^*_{nm} = 0 \tag{45}$$

The first part of this proposition simply says that unambiguously the receivers are worse off when the sender can manipulate information. To understand this result, we first define the receivers' loss function

$$l(\delta) := \min_{k \in [0,1]} L(k, \delta) \tag{46}$$

where $L(k, \delta)$ denotes the receivers' ex ante expected loss if they choose k when the sender has manipulation δ. Collecting terms, it turns out that we can write this

$$L(k,\delta) = \frac{1-\lambda}{\alpha_z}B(\delta,k) + \frac{1}{\alpha_x}k^2 \tag{47}$$

where again $B(\delta,k) = (1-\delta(1-k))^2$ is the sender's gross benefit from manipulation. A straightforward calculation then gives

$$l(\delta) = \frac{(1-\lambda)}{(1-\delta)^2(1-\lambda)\alpha_x + \alpha_z} \tag{48}$$

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

We now turn to the effects of increases in the intrinsic signal precision α_x on the receivers' loss with and without manipulation. Figure 5 illustrates, showing l^* and l^*_{nm} as functions of α_x for various levels of c.

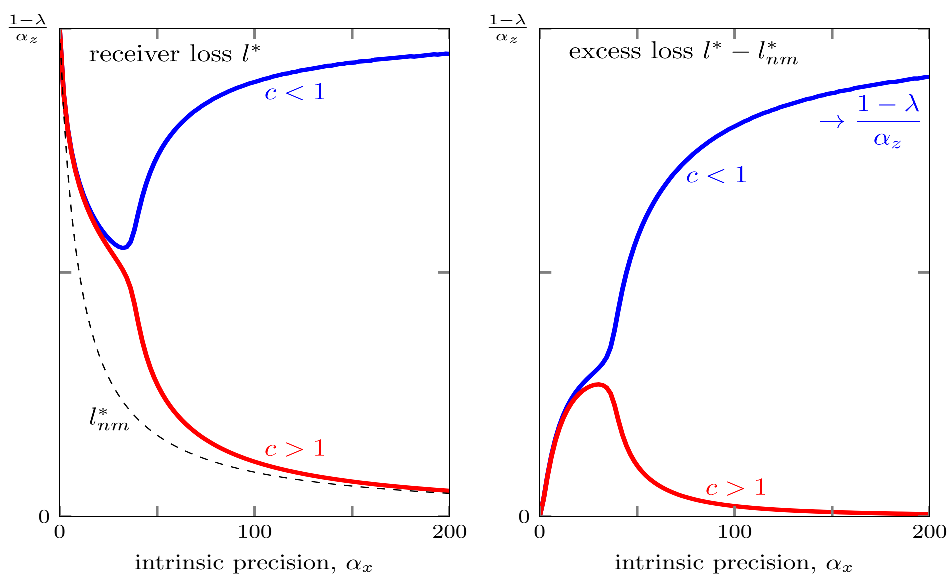

Figure 5: Receivers lose most when α_x high and c low.

Receivers' equilibrium loss l^* (left panel) and excess loss $l^* - l^*_{nm}$ (right panel) as functions of the intrinsic precision α_x for different levels of the cost of manipulation c. If $c > 1$ then l^* is monotonically decreasing in α_x and asymptotes to zero as $\alpha_x \to \infty$. If $c < 1$ then for high enough α_x the receivers' loss l^* is increasing in α_x and asymptotes to $(1-\lambda)/\alpha_z$ as $\alpha_x \to \infty$.

Direct and indirect effects of α_x. Since a higher α_x increases the precision of the receivers' estimates of θ, the direct effect of increasing α_x is to reduce the receivers' loss. When the sender cannot manipulate information, that is the only effect so l^*_{nm} is decreasing in α_x. When the sender can manipulate, there is an additional indirect effect through δ^* with higher values of δ^* increasing the receivers' loss. As shown in Lemma 3, the sender's manipulation can be either increasing or decreasing in the intrinsic precision, depending on whether the receivers' k^* and the sender's δ^* are strategic complements or not. In particular, if the sender's costs of manipulation are high, $c > 1$, then δ^* and k^* are strategic substitutes so that δ^* is decreasing in α_x. In this case, the direct and indirect effects reinforce each other so that unambiguously an increase in α_x decreases the receivers' loss. But if the sender's costs of manipulation are low, $c < 1$, then k^* and δ^* are strategic complements so that increasing α_x increases both k^* and δ^* together. We show in the Appendix that there is then a critical point α_x^{**} such that for $\alpha_x > \alpha_x^{**}$ the indirect effect of α_x via δ^* is strong enough to overcome the direct effect so that on net l^* is increasing in α_x even though l^*_{nm} is decreasing in α_x.

Excess loss from manipulation. When the costs of manipulation are relatively low, $c < 1$, and the intrinsic signal precision is sufficiently high, $\alpha_x > \alpha_x^{**}$, we then know that l^* is increasing in α_x whereas l^*_{nm} is decreasing in α_x. If we define the *excess loss* from manipulation as $l^* - l^*_{nm}$, the excess loss is increasing in α_x for this parameter configuration. Put differently, when c is low and α_x is high then the sender chooses high δ^* and the receivers are not just worse off with manipulation but are becoming increasingly worse off as the intrinsic signal precision α_x increases. As $\alpha_x \to \infty$, the receivers' loss from manipulation asymptotes to $(1-\lambda)/\alpha_z > 0$, i.e., the same loss the receivers would have if $\alpha_x = 0$. By contrast, if $c > 1$ then even though $l^* > l^*_{nm}$ for all α_x the gap between them becomes negligible as $\alpha_x \to \infty$.

6 Social media and political information manipulation

In this section we use our model to interpret the challenges posed by the rise of social media. In Section 6.1 we suppose that the sender is a politician who seeks to prevent the receivers from making informed decisions and we interpret social media as simultaneously increasing the intrinsic precision α_x of the receivers' information and decreasing the politician's costs of manipulation c. The interaction of high intrinsic precision α_x with sufficiently low costs of manipulation c is especially bad news for receivers, and in this sense rationalizes a form of pessimism about the rise of social media. But there is also some good news. We also find that if the intrinsic precision α_x is high then the amount of manipulation by the politician can be extremely sensitive to the cost of manipulation. If the cost of manipulation can be kept above the critical point $c = 1$ then the rise of social media will reduce the amount of manipulation and make the receivers better off. Moreover, if the receivers are *sufficiently well coordinated*, $\lambda > 1/2$, the politician's manipulation will backfire, giving the politician incentives to invest in commitment devices that prevent them undertaking manipulation in the first place. In Section 6.2 we interpret the information manipulation that occurs in our model in terms of common forms of political messaging.

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 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).

6.1 The challenge of social media

As emphasized by 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's AdWords 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, content on these platforms — both from traditional media outlets like the *New York Times* and CNN, as well as from relatively new outlets like *Buzzfeed* and the *Huffington Post*, *Breitbart* and the *Daily Caller* — is largely consumed on mobile devices, in narrow slices, with frequent hot takes, snippets of streaming video and the like, all of which blurs the distinctions between traditional media outlets and their new competitors and blurs the distinctions between reporting and opinion.

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, which leads to a large increase in the number of signals received and hence an increase in the intrinsic quality of information, absent manipulation.[9] But the use of social media technologies also means that media content is consumed in a feed that blurs the distinctions between media

[9] 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.

outlets, making it harder to tell the difference between reliable and unreliable information, and this decreases the politician's costs of propagating "alternative facts". The simultaneous change in α_x and c creates a tension. On the one hand, when information manipulation is sufficiently costly, the increase in the intrinsic signal precision α_x makes the receivers' better off, and in this sense rationalizes a form of optimism about the new media technologies. On the other hand, when the intrinsic signal precision is sufficiently high, the decrease in the costs of manipulation c makes the receivers' worse off and in this sense rationalizes a form of pessimism about the new media technologies.

So which of these effects is stronger? Our model predicts that the interaction of high α_x with low c can be especially *bad news* for the receivers.

Bad news. A high α_x is good for the receivers if the politician cannot manipulate, or if their costs of doing so are sufficiently high. But if the costs of manipulation c become low enough then the high value of α_x stops being good for the receivers and instead starts to harm them. In this sense, the rise of social media technologies that simultaneously make α_x high and make c low can be especially bad for the receivers.

More formally, observe from Proposition 5 part (ii) that if the rise of social media technologies reduces the costs of manipulation c below the critical cost $c = 1$ then for any $\alpha_x > \alpha_x^{**}$, a higher level of α_x amplifies, rather than alleviates, the receivers' losses due to manipulation. This is easiest to see in the limit as $\alpha_x \to \infty$. If $c < 1$ then the receivers' loss $l^* \to (1-\lambda)/\alpha_z$, the same loss the receivers would have even if $\alpha_x = 0$. In this sense, any $c < 1$ is sufficient to completely undo the receivers' hypothetical gains from higher levels of intrinsic signal precision α_x.

To be clear, *any* decrease in c increases the amount of equilibrium manipulation δ^* and so makes the receivers worse off. The question is whether a simultaneous increase in α_x mitigates or amplifies this. The bad news is that if the cost of manipulation is $c < 1$ and the intrinsic precision is $\alpha_x > \alpha_x^{**}$ then any further increase in α_x only amplifies the receivers' loss. In this case, the receivers may actually prefer *less entry* by new media outlets.

But not all is lost. It turns out that a small increase in c can do a lot to curb manipulation.

Good news. We have just seen that a high α_x is good for the receivers if the costs of manipulation c are high but bad for the receivers if the costs of manipulation are low. Flipping this result around, we see that if α_x is high then an increase in the cost of manipulation may be able to make the receivers much better off. The relevant critical cost is $c = 1$. We find that when the cost of manipulation is close to $c = 1$, a small increase from say $c = 1 - \varepsilon$ to $c = 1 + \varepsilon$ can give rise to a large reduction in the equilibrium amount of manipulation δ^* thereby making the receivers much better off. More formally:

Proposition 6.

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

$$\left.\frac{\partial \delta^*}{\partial c}\right|_{c=1} = -\frac{k^*(\alpha,1)}{(1-k^*(\alpha,1))(1+3k^*(\alpha,1))} < 0 \tag{49}$$

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

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

$$\underline{\delta}(\alpha)\,,\,\overline{\delta}(\alpha) = \frac{1}{2}\left(1 \pm \sqrt{1-(4/\alpha)}\right), \qquad \alpha \geq 4 \tag{50}$$

(iii) For any $c > 1$, the sender'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.

To interpret this result, recall from Lemma 4 above that an increase in c reduces the amount of manipulation. This new proposition gives us additional information on the *size* of the reduction in manipulation. In particular, when c is close to the critical cost $c = 1$, 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 $\alpha < 1$ (in blue), $\alpha = 4$ (in pink), and $\alpha > 4$ (in red). 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 that, taking the receivers' k as given, the politician chooses manipulation δ to maximize

$$V(\delta, k) = \frac{1}{\alpha_z}\left(B(\delta, k) - C(\delta)\right) + \frac{1}{\alpha_x}k^2 \tag{51}$$

with benefit $B(\delta, k) = (k\delta + 1 - k)^2$ and cost of manipulation $C(\delta) = c\delta^2$. This is a smooth function of δ and k. But now suppose $\alpha \to \infty$ so that $k \to \min[c, 1]$. Viewed as a function of c, the limit $k = \min[c, 1]$ has a *kink* at $c = 1$. Then as a function of δ and c, the politician's objective $V(\delta, \min[c, 1])$ likewise has a kink at $c = 1$. Because of this kink, the politician's optimal choice of manipulation jumps from $\delta = 1$ for $c > 1$ to $\delta = 0$ for $c < 1$.[10]

[10] Further details can be found in our supplementary Online Appendix.

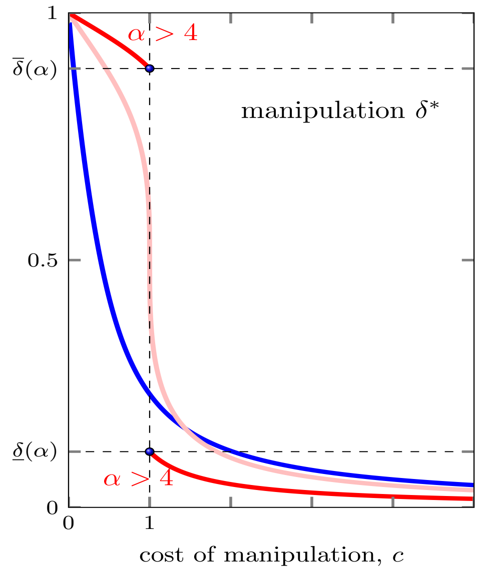

Figure 6: A small increase in c can lead to a large reduction in manipulation δ^*.

Equilibrium manipulation δ^* as a function of c for $\alpha < 1$ (blue), $\alpha = 4$ (pink) and $\alpha > 4$ (red). 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 $c > 1$.

"Regime changes" in the amount of manipulation. The result that small changes in c can lead to large changes in the amount of manipulation (and large changes in the receivers' welfare) is reminiscent of models with multiple equilibria. But because our model features a unique equilibrium, we can precisely identify the parameter changes required to bring about such a "regime change". When α is high enough, i.e., the underlying information environment is otherwise benign, small changes in c near the critical point $c = 1$ can lead to dramatic changes in the amount of manipulation.

In other words, even small changes in the conduct of social media platforms that make it harder for misinformation to propagate may be surprisingly effective. Such changes could in principle come from greater internal efforts to regulate content, to editorialize more in the manner of traditional media outlets, and so on. But they could also come from external regulation or the actions of other institutions.

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 potentially 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 $c = 1$. If we get to $c_1 < 1$, the economy will end up in the high manipulation regime, making the politician better off and everyone else worse off. But if we keep $c_1 > 1$, the economy will end up in the low manipulation regime, making the receivers better off.

Importance of social coordination. So if the costs of manipulation can be kept above the critical point $c = 1$ and α_x is high enough, then the economy can be kept in the low manipulation regime. If, in addition, the receivers are *sufficiently strongly coordinated*, 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, so that it is not just that we are in the low manipulation regime, to the benefit of the receivers, but also that, in this regime, it is in the politician's own self-interest to not manipulate at all. 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 receivers.

Roughly speaking, this suggests that we might see something of a virtuous circle, with policies that mitigate the effects technological change on the ability to propagate misinformation, keeping $c > 1$, in turn leading to greater efforts to create institutions that help keep such manipulation in check. Interestingly, this kind of virtuous circle *cannot* emerge when the receivers strongly prefer to differentiate themselves from one-another. If instead $\lambda < -1/2$ then with $c > 1$ and α_x high enough the politician's manipulation does not backfire and hence the politician faces no incentive to invest in commitment devices to prevent manipulation. In this scenario the receivers get the benefit of being in the low manipulation regime but there are no forces to further curb manipulation.

Of course all this discussion hinges on keeping the costs of manipulation above $c = 1$. Once the costs c fall below that critical point, the regime switches, the politician has no incentive to refrain from manipulation at all, and the receivers end up much worse off.

6.2 Information manipulation and political messaging

We now discuss the politician's information manipulation strategy and how it accords with the kinds of messaging seen in practice.

Creating confusion. Creating confusion is a staple of political messaging, as evidenced by age-old journalistic cliches like "muddying the waters" or "throwing up chaff". But since the 2016 UK Brexit referendum and the 2016 US presidential election, the role of such political messaging has come under greater scrutiny as consumers of media content find themselves on the receiving end of a relentless deluge of spin and "alternative facts". To be clear, the goal of such political messaging is *not* to persuade the receivers of the truth of the alternative facts but instead to prevent the receivers from making informed decisions. As Till (2016) nicely puts it in the Brexit context, "There was more than enough information available to most people for them to make an informed decision but they did not know who or what they could trust". Similarly Pomerantsev (2014) argues that the goal of Russian propaganda under President Putin "is not to persuade anyone, but to keep the viewer hooked and distracted [...] they were trying not so much to convince viewers of any one version of events, but rather to leave them confused, paranoid, and passive" leaving viewers "living in a Kremlin-controlled virtual reality that can no longer be mediated or debated by any appeal to 'truth' ".[11]

Attacking credibility. What can a politician do when they want to change receivers' beliefs but can't directly control the media? A common strategy in this case is to *attack the credibility* of the receivers' information, accusing the media of having hidden biases or other ulterior motives. President Trump cannot directly control the content of articles in the *New York Times*, but he can denounce such articles as "fake news" with the goal of undermining its credibility. Klein (2017), for example, argues that "The Trump administration is creating a baseline expectation among its loyalists that they can't trust anything said by the media" and that "Trump needs to delegitimize the media because he needs to delegitimize facts". Similarly, Tingle (2018) argues that "Rather than regulate and control the shape of the public discourse, Trump lets it career out of control, and provides distractions via Twitter. He does not corral dissidents but tries to destroy their credibility [...] Trump's mastery is not of propaganda but of destruction of the credibility of information". Rosen (2018) argues that for "...Americans who are neither committed supporters nor determined critics of Donald Trump...the campaign to discredit the press works by generating noise and confusion...".

Misdirection and endogenous noise. In our model, the politician creates confusion to prevent the receivers from making informed decisions. They do this through a specific form of *misdirection*. In equilibrium, the systematic part of the receivers' information $y = (1-\delta)\theta+\delta z$ is a mixture of "truth and prejudice" (i.e., the true θ and the prior z). The equilibrium effect of this misdirection is to make the signals x_i endogenously noisier than they would otherwise be. Absent this misdirection, the receivers' 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 receivers' 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. It is as if the receivers' signals are worthless even though the intrinsic information content of their signals may be very high. That is, the underlying information available to the receivers may be good enough to make very informed decisions but in equilibrium they end up making decisions based on their prior alone. In welfare terms, the politician's gain from this endogenous noise is naturally larger when the intrinsic precision α_x is high and the cost of manipulation c is

[11] See Arendt (1973), who concludes that "The ideal subject of totalitarian rule is not the convinced Nazi or the convinced Communist, but people for whom the distinction between fact and fiction (i.e., the reality of experience) and the distinction between true and false (i.e., the standards of thought) no longer exist".

low, because that is when the misdirection leads the receivers to be much less responsive to their signals than is warranted by the intrinsic information content of their signals.

We assume that the politician can only make use of the underlying, fundamental noise in the information structure and cannot *directly* inject noise into x_i. Even with this limitation on the politician's ability to increase the amount of noise, we find that their manipulation may be able to render the receivers' information completely worthless and prevent them from benefitting from the entry of new sources of information. Letting the politician directly inject noise would presumably make their task easier.

7 Conclusions

We develop a model of information manipulation with one sender and many imperfectly informed receivers. The sender engages in costly efforts to prevent receivers from taking actions that are appropriate for the true state of the world. Each individual receiver wants their action to be appropriate for the state but they may also care about their interactions with other receivers. We allow the receivers' actions to be either strategic substitutes or strategic complements, i.e., they may either want to stand out from the crowd or coordinate with the crowd. The receivers are rational and free from "confirmation bias" or any other attributes that would make them easy to deceive. As a consequence, in equilibrium the sender's manipulation does not bias the receivers' beliefs but make them endogenously noisier.

We use this model to understand the effects of new social media technologies when the sender is a politician who wants to prevent receivers from making informed decisions. We interpret the rise of new social media technologies as simultaneously changing two features of the information environment in our model. First, the rise of social media technologies has facilitated the entry of new media outlets, thereby increasing the intrinsic, potential, precision of available information. Second, these technologies also mean that media content is consumed in a feed that blurs the distinctions between media outlets, making it harder to tell the difference between reliable and unreliable information, and this decreases the politician's costs of propagating "alternative facts". Absent manipulation, the first effect would allow the receivers to make more informed decisions and make them better off. But the second effect increases the amount of manipulation and makes the receivers worse off. We find that the interaction of these two effects can be especially bad for the receivers and especially good for the politician. In this scenario, the receivers possess enough information to make informed decisions but their actions reflect little or no actual information.

Of course, one doesn't need this model to arrive at the conclusion that social media technologies may have adverse implications. But we find it striking that severe adverse implications can arise even in a setting that is relatively unfavourable to strategic information manipulation. In particular, the receivers are rational and not easy to deceive. In this sense, we view our results as something of an understatement.

But we also find that the equilibrium amount of manipulation by the politician and the receivers' responsiveness to their information are highly sensitive to the politician's costs of manipulation, especially when the underlying, intrinsic precision of the receivers' information is high. We find that if the costs of manipulation can be maintained above a cutoff, the rise of new social media technologies will work to reduce the politician's manipulation and make the receivers better off. Moreover, if the receivers are sufficiently well coordinated, the rise of new social media technologies can even lead the politician's manipulation to *backfire* entirely, 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 suitable commitment devices that credibly prevent them from manipulating information, e.g., in a personal reputation for straight and coherent talk, in reputable institutions that provide reliable fact-checking, and so on.

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 messages could also just increase the amount of noise facing the information receivers. Second, we assumed that the receivers have identical preferences and prior beliefs. This makes for clear welfare calculations, but partisan differences in preferences and/or prior beliefs seem important especially if one wants a more unified model of political communication and political polarization. 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. Finally, in our model the media is passive and neither captured by the politician nor actively seeking to uncover deception. It would of course be interesting to give the media a more active role.

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 the relative precision $\alpha > 4$. 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 Online Appendix.

Proof of Proposition 1.

An equilibrium is a pair k^*, δ^* simultaneously satisfying the receivers' $k(\delta)$ and the sender'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 sender's best response (13) requires $c \geq k^2$, otherwise the sender 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 sender'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 receivers' $k(\delta)$ 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 sender'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 receivers' $k(\delta)$ 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 (13) into $k(\delta)$ from (18) 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 (13). $\square$

Proof of Lemma 1.

Differentiating the receivers' best response $k(\delta)$ in (20) 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. $\square$

Proof of Lemma 2.

Differentiating the sender's best response $\delta(k)$ in (22) 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 $\delta(1) = 0$. $\square$

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 (20) 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 sender'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 (23). 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 $\hat{\alpha}(c) = +\infty$. $\square$

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 (22) 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 receivers' 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 (21). 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 $\hat{c}(\alpha) = +\infty$. $\square$

Proof of Lemma 5.

Recall that $k^*_{nm}(\alpha) := \alpha/(\alpha+1)$. If $\alpha \le 1$ 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

$$\delta^*_{nm}(\alpha) = \frac{\alpha - 1}{\alpha}, \qquad \alpha > 1 \tag{A19}$$

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

$$c = \frac{\alpha}{\alpha - 1}\left(\frac{\alpha}{\alpha + 1}\right)^2 =: c^*_{nm}(\alpha) \tag{A20}$$

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

$$\delta^*_{nm}(\alpha) = \frac{\alpha - 1}{\alpha} > \hat{\delta}(\alpha) \tag{A21}$$

where $\hat{\delta}(\alpha)$ is the critical point from Lemma 1. Hence $k(\delta; \alpha)$ is decreasing in δ for any $\delta \ge \delta^*_{nm}(\alpha)$. 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 $\delta^*(\alpha, c) > \delta^*_{nm}(\alpha)$. □

B Strategic interactions among receivers

Suppose that the receivers seek to minimize the expected value of the quadratic loss (34). The optimal action of a receiver with signal x_i and prior z is then given by

$$a(x_i, z) = \lambda \mathbb{E}[\, A(\theta, z) \,|\, x_i, z] + (1 - \lambda)\mathbb{E}[\, \theta \,|\, x_i, z] \tag{B1}$$

If other receivers use (8) and the sender uses (12) then the aggregate action is

$$A(\theta, z) = ky(\theta, z) + (1 - k)z = k(1 - \delta)\theta + (1 - k(1 - \delta))z \tag{B2}$$

Collecting terms then gives

$$a(x_i, z) = [1 - \lambda(1 - k(1 - \delta))]\, \mathbb{E}[\, \theta \,|\, x_i, z] + \lambda(1 - k(1 - \delta))z, \tag{B3}$$

(i.e., a weighted average of the posterior and prior expectations). Plugging the expression for the posterior expectation (17) in to (B3) and matching coefficients gives the fixed-point condition

$$k = (\lambda k(1 - \delta) + 1 - \lambda)\, \frac{(1 - \delta)\alpha_x}{(1 - \delta)^2\alpha_x + \alpha_z} \tag{B4}$$

which has the unique solution

$$k(\delta) := \frac{(1 - \delta)\alpha}{(1 - \delta)^2\alpha + 1} \tag{B5}$$

where $\alpha := (1 - \lambda)\alpha_x/\alpha_z$ is the composite parameter redefined in (35).

Notice that the unique solution to the receiver's problem (B5) is given by the same expression as the receivers' best response (18) in the main text. Having strategic interactions among the receivers makes only one difference, we have to redefine the composite parameter α. Once we have redefined α, all the equilibrium properties in Section 2 remain the same even with strategic interactions among the receivers.

C Welfare results

Proof of Proposition 2.

We decompose the sender's gain from manipulation as

$$v^* - v^*_{nm} = (v(k^*) - v_{nm}(k^*)) + (v_{nm}(k^*) - v_{nm}(k^*_{nm})) \tag{C1}$$

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

$$v(k) := \max_{\delta\in[0,1]} V(\delta, k) = \frac{1}{\alpha_z}(1-k)^2\left(\frac{c}{c-k^2}\right) + \frac{1}{\alpha_x}k^2 \tag{C2}$$

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

$$v_{nm}(k) := V(0, k) = \frac{1}{\alpha_z}(1-k)^2 + \frac{1}{\alpha_x}k^2 \leq v(k) \tag{C3}$$

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 (7) and that with strategic interactions among receivers, $\alpha = (1-\lambda)\alpha_x/\alpha_z$. Therefore, when $\lambda < 0$, i.e., the receivers' 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 5 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 (31). Similarly, when $\lambda > 0$, i.e., the receivers' 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 5, 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 Proposition 3.

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

$$v^* = \frac{1}{(1-\lambda)\alpha_x}\left\{k^* - \lambda k^{*2} + \frac{k^{*2}(1-k^*)^2}{c-k^*}\right\} \tag{C4}$$

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

$$v^*_{nm} = \frac{1}{(1-\lambda)\alpha_x}\left\{k^*_{nm} - \lambda k^{*2}_{nm}\right\} \tag{C5}$$

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

$$g(k^*) < f(k^*_{nm}) - f(k^*) \tag{C6}$$

where

$$f(k) := k - \lambda k^2, \qquad g(k) := \frac{k^2(1-k)^2}{c-k} \geq 0 \tag{C7}$$

For part (i) suppose that $\lambda < 0$. We know from Proposition 2 that a necessary condition for the sender's manipulation to backfire is $c < c^*_{nm}(\alpha)$ so that $k < k^*_{nm}$. We can rewrite the inequality in (C6) as

$$\frac{k^{*2}(1-k^*)^2}{c-k^*} < (k^*_{nm} - k)(1 - \lambda(k^*_{nm} + k^*)) \tag{C8}$$

Using the fixed point conditions (A2) for both k^* and k^*_{nm}, we can rewrite the key condition (C8) as

$$\lambda < \frac{1}{c-k^*}\frac{K_1K_2}{K_3K_4} \tag{C9}$$

where

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

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 (C9) is

$$
\frac{1}{c-k^*}\frac{K_1K_2}{K_3K_4} \to \frac{1}{(c-0)}\frac{(-c^2)(c^2)}{(c)(2c^2)} = -\frac{1}{2} \tag{C10}
$$

Hence for any $\lambda < -1/2$ we can find α_x sufficiently close to zero such that (C9) is satisfied and in turn the sender's manipulation backfires, $v^* < v^*_{nm}$.

For part (ii), suppose that $\lambda > 0$. We know from Proposition 2 that the necessary condition for the sender's manipulation to backfire is $c > c^*_{nm}(\alpha)$ so that $k^* > k^*_{nm}$. We can rewrite the inequality in (C6) as

$$
\frac{k^{*2}(1-k^*)^2}{k^* - k^*_{nm}} < (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) \tag{C11}
$$

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

$$
\frac{k^{*2}(1-k^*)}{\frac{k^*_{nm}}{k^*}\left(c\frac{(c-k^*)}{(c-k^{*2})^2}\right) - 1} < (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) \tag{C12}
$$

Observe that if, in addition, $c > 1$ and $\lambda > \frac{1}{2}$, then the RHS of (C12) converges to a strictly positive constant

$$
\lim_{\alpha_x \to \infty} (\lambda(k^*_{nm} + k^*) - 1))(c - k^*) = (\lambda 2 - 1)(c - 1) > 0 \tag{C13}
$$

(since $k^* \to 1$ if $c > 1$). But the LHS of (C12) converges to zero

$$
\lim_{\alpha_x \to \infty} \frac{k^{*2}(1-k^*)}{\frac{k^*_{nm}}{k^*}c\frac{c-k^*}{(c-k^{*2})^2} - 1} = \frac{0^+}{\frac{c}{(c-1)} - 1} = 0^+ \tag{C14}
$$

Therefore, if $c > c^*_{nm}(\alpha)$, $c > 1$ and $\lambda > \frac{1}{2}$ then there exists α^*_x such that for $\alpha_x > \alpha^*_x$ the LHS of (C12) is strictly less than the RHS of (C12) so that the sender's manipulation backfires, $v^* < v^*_{ms}$.

Finally, observe that $c < 1$ is sufficient for $c < c^*_{nm}(\alpha)$ if $1 < c^*_{nm}(\alpha)$. From (31) 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

$$
\underline{\alpha}^*_x < \left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{\alpha_z}{1-\lambda}\right) \tag{C15}
$$

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

$$
\overline{\alpha}^*_x > \left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{\alpha_z}{1-\lambda}\right) \tag{C16}
$$

□

Proof of Proposition 4.

For part (i), differentiating $v_{nm}(k)$ in (C3) evaluated at k^*_{nm} we can write

$$\frac{dv^*_{nm}}{d\alpha_x} = -\frac{(1-\lambda)^3\alpha_x + (1-\lambda)(1+\lambda)\alpha_z}{((1-\lambda)^2\alpha_x + \alpha_z)((1-\lambda)\alpha_x + \alpha_z)} v^*_{nm} \tag{C17}$$

which is negative if

$$\frac{\alpha_x}{\alpha_z} > -\frac{1+\lambda}{(1-\lambda)^2} \tag{C18}$$

If $\lambda > -1$, the RHS of (C18) is negative and hence v^*_{nm} is strictly decreasing in α_x.

The total derivative of v^* with respect to α_x can be written as

$$\frac{dv^*}{d\alpha_x} = v'(k^*)\frac{\partial k^*}{\partial \alpha_x} + \frac{\partial v(k^*;\alpha_x)}{\partial \alpha_x} \tag{C19}$$

where $v'(k^*)$ denotes the derivative of the sender's value function in (C2) evaluated at k^*, that is

$$v'(k^*) = 2\left(\frac{k^*}{\alpha_x} - \frac{R(k^*)}{\alpha_z}\right) = -2\frac{\lambda}{1-\lambda}\left(\frac{k^*}{\alpha_x}\right). \tag{C20}$$

and where $R(k)$ is given in (A2) and we use the equilibrium condition from (A2) that $\frac{1}{1-\lambda}\frac{\alpha_z}{\alpha_x}k^* = R(k^*)$. Then since

$$\frac{\partial v(k^*;\alpha_x)}{\partial \alpha_x} = -\left(\frac{k^*}{\alpha_x}\right)^2 < 0 \tag{C21}$$

we can then write the total derivative (C29) as

$$\frac{dv^*}{d\alpha_x} = -2\frac{\lambda}{1-\lambda}\left(\frac{k^*}{\alpha_x}\right)\frac{\partial k^*}{\partial \alpha_x} - \left(\frac{k^*}{\alpha_x}\right)^2 \tag{C22}$$

which is negative if

$$-2\frac{\lambda}{1-\lambda} < \frac{J_1}{J_2} \tag{C23}$$

where

$$J_1 := (c - k^{*2})^2 - 4k^{*2}(c-k^*)(1-k^*)$$
$$J_2 := (c-k^*)(1-k^*)(c-k^{*2})$$

Observe that as $\alpha_x \to 0$ such that $k^* \to 0$ the ratio $J_1/J_2 \to 1$. The derivative of J_1/J_2 with respect to k^* has the same sign as

$$\frac{\partial J_1}{\partial k^*}J_2 - \frac{\partial J_2}{\partial k^*}J_1 \geq 2\sqrt{c}(1-k^*)(c-k^*)\Big((k^{*2} - 2\sqrt{c}k^* + c)^2 + 4\sqrt{c}k^{*2}(\sqrt{c}-1)^2\Big) \geq 0 \tag{C24}$$

So J_1/J_2 is increasing in k^*. From Lemma 3, we know that k^* is increasing in α and in turn α_x. So J_1/J_2 is increasing in α_x and hence is bounded below by 1. If $\lambda > -1$, the LHS of (C23) is strictly lower than 1. Therefore, the condition (C23) for v^* to be strictly decreasing in α_x must hold.

For part (ii) the limits of v^*_{nm} can be computed directly after evaluating $v_{nm}(k)$ in (C3) at k^*_{nm}, namely

$$\lim_{\alpha_x \to 0^+} v^*_{nm} = \frac{\alpha_z}{\alpha_z^2} = \frac{1}{\alpha_z} \tag{C25}$$

$$\lim_{\alpha_x \to \infty} v^*_{nm} = \lim_{\alpha_x \to +\infty} \frac{(1-\lambda)^2\frac{1}{\alpha_x} + \frac{\alpha_z}{\alpha_x^2}}{(1-\lambda)^2 + 2(1-\lambda)\frac{\alpha_z}{\alpha_x} + \frac{\alpha_z^2}{\alpha_x^2}} = 0 \tag{C26}$$

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

$$\lim_{\alpha_x \to 0^+} v^* = \frac{1}{\alpha_z}(1-0)^2\left(\frac{c}{c-0^2}\right) + \lim_{\alpha_x \to 0^+}\frac{k^{*2}}{\alpha_x} = \frac{1}{\alpha_z} + 0 = \frac{1}{\alpha_z} \tag{C27}$$

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

$$\lim_{\alpha_x \to \infty} v^* = \begin{cases} \frac{1}{\alpha_z}(1-c)^2 \frac{c}{c-c^2} & = \frac{1-c}{\alpha_z} \quad \text{if } c < 1 \\ \frac{1}{\alpha_z}(1-1)^2 \frac{c}{c-1} & = 0 \quad \text{if } c > 1 \end{cases} \tag{C28}$$

□

Proof of Proposition 5.

The proof of part(i) is in the main text. For part (ii), the total derivative of l^* with respect to α_x is

$$\frac{dl^*}{d\alpha_x} = l'(\delta^*)\frac{d\delta^*}{d\alpha_x} + \frac{\partial l(\delta^*;\alpha_x)}{\partial \alpha_x} \tag{C29}$$

Supplementary Lemma 1 in the Online Appendix shows that

$$\frac{dl^*}{d\alpha_x} > 0 \qquad \Leftrightarrow \qquad F(k^*) := k^{*4} - 2k^{*3} + 2ck^* - c^2 > 0 \tag{C30}$$

Supplementary Lemma 2 in the Online Appendix shows that if $c > 1$ then it cannot be the case that $F(k^*) > 0$ and hence the receivers' 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

$$k^*(\alpha_x^{**}, c) = \underline{k} \tag{C31}$$

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 receivers' loss is strictly increasing in α_x.

For part (iii), the derivative of $l^*_{nm} = l(0)$ with respect to α_x can be computed directly as

$$\frac{dl^*_{nm}}{d\alpha_x} = -\frac{(1-\lambda)^2}{((1-\lambda)\alpha_x + \alpha_z)^2} < 0 \tag{C32}$$

For part (iv), the limits of l^*_{nm} can be directly computed from $l(0)$, namely

$$\lim_{\alpha_x \to 0^+} l^*_{ms} = \frac{1-\lambda}{\alpha_z}; \lim_{\alpha_x \to \infty} l^*_{ms} = 0.$$

For the limits of $l^* = l(\delta^*;\alpha_x)$ we repeatedly use that $l(\delta;\alpha_x)$ 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

$$\lim_{\alpha_x \to 0^+} l^* = \frac{(1-\lambda)}{(1-0)^2(1-\lambda)0 + \alpha_z} = \frac{1-\lambda}{\alpha_z} \tag{C33}$$

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

$$\lim_{\alpha_x \to \infty} l^* = \lim_{\alpha_x \to \infty} \frac{(1-\lambda)}{(1-\delta^*)^2(1-\lambda)\alpha_x + \alpha_z} = \frac{1-\lambda}{\alpha_z} \tag{C34}$$

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

$$\lim_{\alpha_x \to \infty} l^* = \lim_{\alpha_x \to \infty} \frac{(1-\lambda)}{(1-\delta^*)^2(1-\lambda)\alpha_x + \alpha_z} = 0 \tag{C35}$$

□

Proof of Proposition 6.

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

$$\left.\frac{\partial \delta^*}{\partial c}\right|_{c=1} = \left(3k^* - \frac{1}{k^*} - 2\right)^{-1} \tag{C36}$$

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 the Online Appendix that when $\alpha > 4$ (i) 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 (ii) 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 sender's best response $\delta(k; c)$ as in (22) is decreasing in c. If $c > 1$, the sender'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 (28) 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.

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.

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.

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.

Handy, Peter, “How Trump Brought the Political Media Class to its Knees,” *Vanity Fair*, October 2017.

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.

Klein, Ezra, "Trump's Real War Isn't with the Media. It's with Facts," *Vox*, January 2017.

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.

Pomerantsev, Peter, "Russia and the Menace of Unreality: How Vladimir Putin is Revolutionizing Information Warfare," *The Atlantic*, September 2014.

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

Rosen, Jay, "Why Trump Is Winning and the Press Is Losing," *New York Review of Books online*, April 15 2018.

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.

Till, Christopher, "Brexit and the Politics of Confusion," *This Is Not a Sociology Blog*, July 2016.

Tingle, Laura, *Follow the Leader: Democracy and the Rise of the Strongman*, Black Inc, September 2018.

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.