Creating Confusion: Information Manipulation and Imperfect Coordination*

Chris Edmond[†] Yang K. Lu[‡]

December 2017

Abstract

We develop a sender/receiver game with many imperfectly informed and imperfectly coordinated receivers. The sender can manipulate the receivers' information at a cost, but cannot commit to an information structure. The sender tries to prevent the receivers from taking actions that align with the state of the world. This game has a unique equilibrium. In that equilibrium, the receivers' beliefs are unbiased but the sender can still gain from the endogenous noise that is a byproduct of their manipulation. The sender gains most when the receivers have intrinsically very precise information but the sender's costs of manipulation are low. In this situation, the receivers possess information that is a very reliable guide to the truth but their actions do not reflect it. In particular, when the costs of manipulation are negligible the sender manipulates a lot and the receivers become unresponsive to their signals even though their signals are intrinsically very precise. By contrast, if the costs of manipulation were high and the receivers were sufficiently well-coordinated, the manipulation would backfire on the sender. In this situation, the sender would want to commit to not manipulate information. We give a political interpretation.

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

JEL classifications: C7, D7, D8.

*We particularly thank Kim Sau Chung, Simon Grant, Andy McLennan and Larry Samuelson for their comments. We also thank seminar participants at the ANU, City University of Hong Kong, CUHK (Shenzhen), 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

Consider a media establishment consisting of many journalists. Each individual journalist would like to write reports that represent their best assessment of some contentious facts, but, collectively, these journalists are subject to some groupthink in that they also like their reports to fit in with each other. Confronting the media is a politician who seeks to prevent them from accurately reporting the facts. Can the politician achieve this goal? Even when the journalists are perfectly rational and understand the politician's incentives completely? Or will the politician's strategy backfire leaving everyone, including the politician, worse off? Should we be optimistic that new social media technologies will make it more difficult for a politician to stop the facts getting out? Or will these new social media technologies make it too easy to spread "fake news" — creating confusion, and leaving consumers of news no better informed?[1]

We develop a simple model to answer these questions. We find that the success or failure of the politician's media strategy depends crucially on the interaction of two features of the environment: (i) the intrinsic, potential, *precision* of the media's information, and (ii) the *costs* the politician incurs in trying to achieve slanted media coverage — i.e., in propagating "alternative facts".[2] We find that the politician's media strategy is most successful when the intrinsic precision of the media's information is high and yet at the same time the cost the politician incurs in trying to achieve slanted coverage is low. In this situation, individual journalists possess information that is highly precise but the reports they write convey little or no actual information. New media technologies that simultaneously improve the precision of the journalists' information and yet at the same time make it easy to spread fake news end up making the politician better off and everybody else worse off.

Section 2 outlines the model. There is a single politician who knows the state of the world. There are many imperfectly coordinated journalists with imperfect information about the state of the world. The journalists try to write reports that align both with their individual beliefs about the true state of the world and with the reports of their fellow journalists. The politician tries to prevent the journalists from accurately reporting the facts. In particular, the politician can distort the journalists' signals about the state at a cost that is increasing in the amount of slant they want to introduce. The journalists are perfectly rational and internalize the politician's incentives when forming their beliefs. To keep the model tractable, we assume quadratic preferences and normal priors and signal distributions. We

[1] For an overview of the role of social media in the 2016 US presidential election, see Allcott and Gentzkow (2017) and Faris et al. (2017). 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 on leaving the European Union, 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] As per Kellyanne Conway's use of the phrase on *Meet the Press* in January 2017. Handy (2017) provides a forceful account of the relationship between the Trump administration and the US media.

study equilibria that are linear in the sense that journalists' strategies are linear functions of their signals. Section 3 solves this model when the politician cannot manipulate information. This serves as a natural benchmark for the remaining analysis.

Section 4 solves the model when the politician *can* manipulate information and shows that there is a unique (linear) equilibrium. In equilibrium, individual journalists have unbiased beliefs and write unbiased reports. But these reports are *noisier* than they would be absent manipulation. We find that if the journalists' information is intrinsically precise but the politician's costs of manipulation are low, then there will be a sizeable gap between the information content of the journalists' reports absent manipulation and the equilibrium information content of their reports in the presence of manipulation. Indeed, if the costs of manipulation are negligible then the politician's manipulation renders the journalists' reports completely uninformative even if the journalists' information is arbitrarily precise.

Section 5 provides a welfare analysis. In the absence of manipulation, more precise information would make the politician worse off and make the journalists (and consumers who value informative journalism) better off. We provide four welfare results in the presence of manipulation. First, when the journalists' information is intrinsically precise but the politician's costs of manipulation are low then the politician indeed benefits from their manipulation. A strong form of this is that if the politician's costs of manipulation are negligible we find that the politician can completely insulate themself from the consequences of intrinsically precise information. Second, the manipulation can *backfire* on the politician. This happens when the journalists' information is intrinsically precise, the politician's costs of manipulation are high, and the journalists are sufficiently *well coordinated*. Third, the manipulation makes the journalists unambiguously worse off. Fourth, when the costs of manipulation are low the journalists become even worse off as their information becomes more precise.

Section 6 discusses our results. We interpret the rise of new social media technologies as simultaneously increasing the intrinsic precision of information and reducing the politician's costs of manipulation. New social media technologies like Facebook, Google, Twitter and YouTube have facilitated the entry of new media outlets — thereby increasing the intrinsic (or potential) precision of information. But content on these platforms is usually consumed in a feed that blurs the distinctions between media outlets in a way that makes it easier for "fake news" and "alternative facts" to propagate — thereby reducing the costs of information manipulation. Our model implies that simultaneously increasing the intrinsic precision of information but decreasing the cost of manipulation tends to work to the advantage of politicians with an interest in preventing accurate reporting and to the disadvantage of everyone else. Section 7 offers some concluding remarks. All proofs are in the Appendix.

Strategic communication with costly talk. Our model is a sender/receiver game with many imperfectly coordinated receivers. As in Crawford and Sobel (1982), the preferences of the sender and receivers are not aligned and the sender is informed about the state of the world. But we have *costly talk*, not cheap talk. As in Kartik (2009), our sender can send distorted messages at a cost. By contrast with standard cheap talk models, our model with costly talk features a unique equilibrium. In the limit as the sender's distortion become *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 any kind of confirmation bias. At the other extreme, if the sender's distortion is infinitely costly then there is no manipulation and our receivers play a coordination game as 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. The receivers take actions $a_i \in \mathbb{R}$ that balance two considerations: (i) each receiver wants their action to align with the true underlying state $\theta \in \mathbb{R}$ (about which they are imperfectly informed), but also (ii) each receiver wants their action to align with the population aggregate action $A := \int_0^1 a_i \, di$. In particular, as in Morris and Shin (2002), each individual chooses a_i to minimize the expected value of the quadratic loss

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

where the parameter $\lambda \in [0, 1)$ measures the relative strength of the *coordination motive*. If $\lambda = 0$, each individual receiver will set their a_i equal to their expectation of θ without regard to the actions (or beliefs) of other receivers. Alternatively, if $\lambda \to 1$, each individual receiver will set their a_i equal to their expectation of the aggregate action A and this will involve determining the expectations of their fellow receivers.

In the special case where θ is known with *certainty*, the optimal action is

$$a_i = \lambda A + (1 - \lambda)\theta \tag{2}$$

Under certainty, in equilibrium each individual takes action $a_i = A = \theta$ and thus in this special case there is no tension between aligning with θ and aligning with the rest of the population — i.e., in this special case, the population is perfectly coordinated on θ.

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

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

that is increasing in the dispersion of the actions a_i around θ. The sender is endowed with a technology that, in principle, gives it some influence over the receivers' beliefs. Knowing θ, the sender may take a costly action $s \in \mathbb{R}$ in an attempt to induce the receivers into believing the state is actually $y = \theta + s$, 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

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

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 $y = \theta$).

Receivers' information. The receivers know the sender's objective but are imperfectly informed about θ. The receivers begin with a common prior for θ. In particular, their prior belief is 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

$$x_i = y + \varepsilon_i = \theta + s + \varepsilon_i \tag{5}$$

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 $1/\alpha_x$). In other words, 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.

Equilibrium. A symmetric *perfect Bayesian equilibrium* of this model consists of individual receiver actions $a(x_i, z)$ and beliefs, an aggregate action $A(y, z)$ 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, (iii) the aggregate action $A(y, z)$ is consistent with the individual actions, and (iv) the sender's $y(\theta, z)$ maximizes the sender's payoff given the individual actions.

Discussion. 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.

3 Equilibrium without information manipulation

Suppose the sender *cannot* manipulate information — i.e., let $c \to \infty$ so that the sender always chooses $y = \theta$. Then this model essentially reduces to the setting of Morris and Shin (2002). The optimal action of a receiver with signal x_i and prior z satisfies, in general,

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

In this linear-normal setup, each receiver's posterior beliefs are also normal. In particular, with $y = \theta$ a receiver with signal x_i and prior z has posterior expectation

$$\mathbb{E}[\, \theta \,|\, x_i, z \,] = \frac{\alpha_x}{\alpha_x + \alpha_z} x_i + \frac{\alpha_z}{\alpha_x + \alpha_z} z \tag{7}$$

and all receivers have posterior precision $\alpha_x + \alpha_z$.

Linear strategies. As in Morris and Shin (2002), we restrict attention to equilibria in which the receivers use *linear* strategies of the form

$$a(x_i, z) = kx_i + hz \tag{8}$$

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

$$A(\theta, z) = k\theta + hz \tag{9}$$

Equilibrium in linear strategies. If other receivers use linear strategies, then from (6) and (15), the optimal action of any individual receiver is

$$\begin{aligned} a(x_i, z) &= \lambda \left(k\mathbb{E}[\,\theta\,|\,x_i, z] + hz\right) + (1-\lambda)\mathbb{E}[\,\theta\,|\,x_i, z] \\ &= (\lambda k + 1 - \lambda)\,\frac{\alpha_x}{\alpha_x + \alpha_z} x_i + \left((\lambda k + 1 - \lambda)\,\frac{\alpha_z}{\alpha_x + \alpha_z} + \lambda h\right) z \end{aligned} \tag{10}$$

Equating coefficients between (8) and (10) then allows us to solve for the equilibrium responses to the signal x_i and prior z. In particular, equating the sums of the coefficients

$$k + h = (\lambda k + 1 - \lambda) + \lambda h$$

so that $k + h = 1$. In short, actions simply have the form $a(x_i, z) = kx_i + (1-k)z$. Moreover the coefficient k on the signal x_i must satisfy the fixed-point condition

$$k = (\lambda k + 1 - \lambda)\,\frac{\alpha_x}{\alpha_x + \alpha_z} \tag{11}$$

which has the unique solution

$$k^*_{ms} = \frac{(1-\lambda)\alpha_x}{(1-\lambda)\alpha_x + \alpha_z} \in (0, 1) \tag{12}$$

To interpret this formula, first observe that in the absence of the coordination motive ($\lambda = 0$) each individual sets their action $a(x_i, z) = \mathbb{E}[\,\theta\,|\,x_i, z]$, i.e., they set their action equal to their expectation of the underlying state. Thus in this case we have $k^*_{ms} = \alpha_x/(\alpha_x + \alpha_z)$, the same weight as given to their signal x_i in the posterior expectation (7), which depends only on the precision of the signal relative to that of the prior, α_x/α_z. But in the presence of the coordination motive, $\lambda \in (0, 1)$, the weight on the signal x_i gets reduced from α_x to $(1-\lambda)\alpha_x$ with the weight on the prior mean z correspondingly increased. In the presence of the coordination motive, individual receivers are seeking to predict each others' actions and hence their common prior gets more weight than its precision alone warrants. If the coordination motive is very strong ($\lambda \to 1$), each receiver completely ignores their individual signal even if it is much more precise than the prior.

For future reference, let

$$k^*_{ms} = \frac{\alpha}{\alpha + 1}, \qquad \alpha := (1 - \lambda)\frac{\alpha_x}{\alpha_z} > 0 \tag{13}$$

The composite parameter $\alpha > 0$ features repeatedly below. For brevity, we will refer to α as measure of the *relative precision* of the signal x_i to the prior z, though it also incorporates the effects of the coordination motive via λ. The key thing is that the magnitude of α determines how much receivers would respond to their signals if there is no manipulation.

4 Equilibrium with information manipulation

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

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

$$a(x_i, z) = kx_i + (1 - k)z \tag{14}$$

with coefficients summing to one. We show in the Appendix that this is without loss of generality — in any such equilibrium, the coefficients in the receivers' strategy will always sum to one — and it streamlines the exposition to make use of this result from the start. As a function of the sender's y, the corresponding aggregate action is then

$$A(y, z) = ky + (1 - k)z \tag{15}$$

4.1 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

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

$$= (ky + (1 - k)z - \theta)^2 + \frac{1}{\alpha_x}k^2 - c(y - \theta)^2 \tag{16}$$

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

$$y(\theta, z) = \frac{c - k}{c - k^2}\theta + \frac{k - k^2}{c - k^2}z \tag{17}$$

where the second-order condition requires

$$c - k^2 \geq 0 \tag{18}$$

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

$$y(\theta, z) = (1 - \delta)\theta + \delta z \tag{19}$$

where δ depends on the receivers' response coefficient k via

$$\delta(k) := \frac{k - k^2}{c - k^2}, \qquad c - k^2 \geq 0 \tag{20}$$

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.

4.2 Receivers' problem

To construct the receivers' best response, first observe that the optimal action $a(x_i, z)$ for an individual receiver with signal x_i and prior z continues to satisfy

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

If other receivers use (14) and the sender uses (19) then the aggregate action is

$$A(\theta, z) = ky(\theta, z) + (1 - k)z = k(1 - \delta)\theta + (1 - k(1 - \delta))z \tag{21}$$

Collecting terms then gives

$$a(x_i, z) = [1 - \lambda(1 - k(1 - \delta))]\, \mathbb{E}[\, \theta \,|\, x_i, z] + \lambda(1 - k(1 - \delta))z \tag{22}$$

(i.e., a weighted average of the posterior and prior expectations). To make further progress, we need to determine this individual receiver's posterior expectation $\mathbb{E}[\, \theta \,|\, x_i, z]$ in the presence of the sender's manipulation strategy (19).

Signal extraction. If the sender's manipulation strategy is (19), then each individual receiver has two pieces of information: (i) the common prior $z = \theta + \varepsilon_z$, where ε_z is normal with mean zero and precision α_z, and (ii) the idiosyncratic signal

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

$$= \theta + \delta\varepsilon_z + \varepsilon_i \tag{23}$$

where the ε_i are IID normal with mean zero and precision α_x. In short, the signal x_i is correlated with the prior via the common component ε_z. To extract the common component, we construct a *synthetic signal*

$$s_i := \frac{1}{1-\delta}(x_i - \delta z) = \theta + \frac{1}{1-\delta}\varepsilon_i \tag{24}$$

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

$$\mathbb{E}[\,\theta \mid s_i, z] = \frac{(1-\delta)^2\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}s_i + \frac{\alpha_z}{(1-\delta)^2\alpha_x + \alpha_z}z \tag{25}$$

So in terms of the actual signal x_i they have

$$\mathbb{E}[\,\theta \mid x_i, z] = \frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}x_i + \left(1 - \frac{(1-\delta)\alpha_x}{(1-\delta)^2\alpha_x + \alpha_z}\right)z \tag{26}$$

Matching coefficients. We can then plug this formula for the posterior expectation back into (22) and match coefficients to get the fixed-point condition

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

which has the unique solution

$$k(\delta) := \frac{(1-\delta)\alpha}{(1-\delta)^2\alpha + 1} \tag{27}$$

where $\alpha := (1-\lambda)\alpha_x/\alpha_z$ is the composite parameter introduced in (13) above.

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

4.3 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.

A simplifying result. The receivers' response coefficient $k(\delta)$ given in (27) admits any $\delta \in \mathbb{R}$ as an argument and in principle can take on both negative values and values larger than one. Similarly, the sender's manipulation coefficient $\delta(k)$ given in (20) admits any $k \in [-\sqrt{c}, +\sqrt{c}\,]$ as an argument and in principle can take on both negative values and values larger than one. It turns out that these cases can be ignored. In particular:

Lemma 1. In any equilibrium, the receivers' $k^* \in [0,1]$ and the sender's $\delta^* \in [0,1]$.

In other words, the only possible crossing points of $k(\delta)$ and $\delta(k)$ are in the unit square. We make use of this result to simplify much of the exposition below.

Receivers' best response. We then have:

Lemma 2. The receivers' best response $k : [0,1] \to \mathbb{R}_+$ defined by

$$k(\delta) := \frac{(1-\delta)\alpha}{(1-\delta)^2\alpha + 1}, \qquad \delta \in [0,1], \qquad \alpha > 0 \tag{28}$$

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

$$\hat{\delta}(\alpha) := \begin{cases} 0 & \text{if } \alpha \leq 1 \\ 1 - 1/\sqrt{\alpha} & \text{if } \alpha > 1 \end{cases} \tag{29}$$

with boundary values $k(0) = \alpha/(\alpha+1) =: k^*_{ms}$ 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 2 tells us that this raw intuition is incomplete. If $\alpha \leq 1$ the maximum of $k(\delta)$ is obtained at the boundary and so in this special case we get $k(\delta) < k^*_{ms}$, which agrees with the raw intuition. But if $\alpha > 1$ the maximum of $k(\delta)$ is obtained in the interior so that the receivers will have $k(\delta) < k^*_{ms}$ only when δ is 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, $k(\delta) < k^*_{ms}$.

Why might receivers be more responsive to their signals? Lemma 2 implies that if α is relatively high and δ is relatively low, then the receivers will 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 (24), 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 (24), 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 $k(0) = k^*_{ms}$ 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 $k(0) = k^*_{ms}$ 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, $k(\delta) < k^*_{ms}$. This hurdle is easy to clear when α is relatively low, but hard to clear when α is relatively high.

Sender's best response. To characterize the sender's best response, recall that the sender's second-order condition (18) requires that the receiver's response coefficient k is in $[-\sqrt{c}, +\sqrt{c}]$. Any other k leads the sender to choose the corner solution $\delta(k) = 0$. Moreover, Lemma 1 tells us that in any equilibrium we must have $k \in [0,1]$ and $\delta(k) \in [0,1]$ so that we can further restrict attention to the interval

$$\mathcal{K}(c) := \{\, k \,:\, 0 \leq k \leq \min[c,1] \,\}, \qquad c > 0 \tag{30}$$

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 $k \leq c$. With this notation in hand, we have:

LEMMA 3. The sender's best response $\delta : \mathcal{K}(c) \to [0,1]$ defined by

$$\delta(k) := \frac{k-k^2}{c-k^2}, \qquad k \in \mathcal{K}(c), \qquad c > 0 \tag{31}$$

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

$$\hat{k}(c) = \begin{cases} c & \text{if } c \leq 1 \\ c - \sqrt{c(c-1)} & \text{if } c > 1 \end{cases} \tag{32}$$

with boundary values $\delta(0) = 0$ and $\delta(\min[c,1]) = 1$.

This lemma says that if $c \leq 1$ 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[3] which means that for high enough k the sender responds by choosing a *lower* level of manipulation $\delta(k)$.

Strategic complements and substitutes. 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, it's instructive to revisit the sender's problem in light of these strategies. Using $y = (1-\delta)\theta + \delta z$ the sender's objective function (4) can be rewritten

$$V(\delta; k) = \left(B(\delta; k) - C(\delta)\right)(z-\theta)^2 + \frac{1}{\alpha_x}k^2 \tag{33}$$

where

$$B(\delta; k) := (k\delta + 1 - k)^2, \qquad \text{and} \qquad C(\delta) := c\delta^2 \tag{34}$$

and we can view the sender's problem as being equivalent to choosing $\delta \in [0,1]$ to maximize (33) 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 (33). 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

$$\frac{\partial B}{\partial \delta} = 2(k\delta + 1 - k)k \tag{35}$$

where $k\delta + 1 - k = 1 - (1-\delta)k$ measures the *exposure* of the aggregate action A to the "aggregate noise" $\varepsilon_z = z - \theta$, as in (21) above. Then differentiating the marginal benefit with respect to k we can write the cross-derivative as proportional to

$$\underbrace{(k\delta + 1 - k)}_{\text{effectiveness of exposing } A \text{ to } \varepsilon_z} \qquad - \qquad \underbrace{(1-\delta)k}_{\text{change in exposure of } A \text{ to } \varepsilon_z} \tag{36}$$

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.

[3] 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 $c > 1$.

When k is small (say $k \to 0$), the first effect in (36) dominates regardless of δ. Likewise, when δ is large (say $\delta \to 1$), the first effect in (36) 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 (31) into (36) and rearranging, we find that the second effect dominates if and only if

$$k > \hat{k}(c) \tag{37}$$

where $\hat{k}(c)$ is the critical value defined in (32) 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 will dominate and $\delta(k)$ is decreasing for high enough k.

4.4 Equilibrium uniqueness

With the receivers' best response $k(\delta)$ and the sender's best response $\delta(k)$ in hand, the first main result of the paper is:

Proposition 1. There is a unique equilibrium, that is, a unique pair k^*, δ^* simultaneously satisfying the receivers' $k(\delta)$ and the sender's $\delta(k)$.

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

In equilibrium the sender is engaged in active information manipulation and the receivers are making use of their signals despite that active manipulation. But this still leaves open important questions. How large is the equilibrium amount of manipulation δ^*? Is it large enough to induce the receivers to be less responsive to their signals, $k^* < k^*_{ms}$? To answer these questions, we need to figure out how the equilibrium levels of k^* and δ^* vary with the underlying parameters of the model.

4.5 Comparative statics

Our goal now is to show how the equilibrium levels of the receivers' response k^* and the sender's manipulation δ^* vary with the underlying parameters of the model. There are two parameters of interest, (i) the relative precision $\alpha := (1-\lambda)\alpha_x/\alpha_z > 0$, 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 (28) that we can write the receivers' best response as $k(\delta; \alpha)$ independent of the sender's cost c. Likewise, from (31) we can write the sender's best response as $\delta(k; c)$ independent of the relative

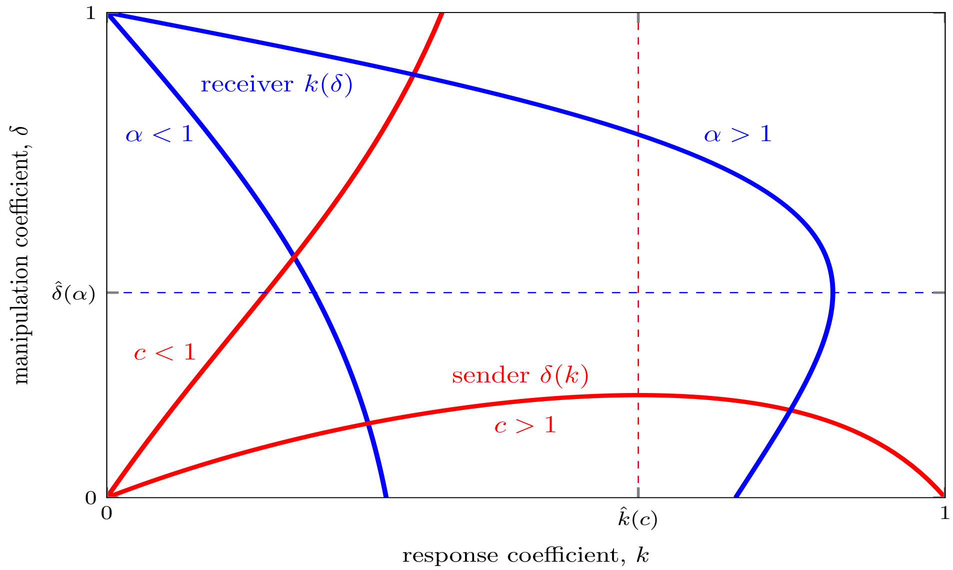

Figure 1: Unique equilibrium.

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 $\delta < \hat{\delta}$. 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.

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. With these observations in hand, we can derive the main comparative statics of the model.

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

LEMMA 4.

(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

$$\alpha < \hat{\alpha}(c) \tag{38}$$

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

For any given level of the sender's manipulation $\delta \in (0, 1)$, 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 $\delta(k; c)$

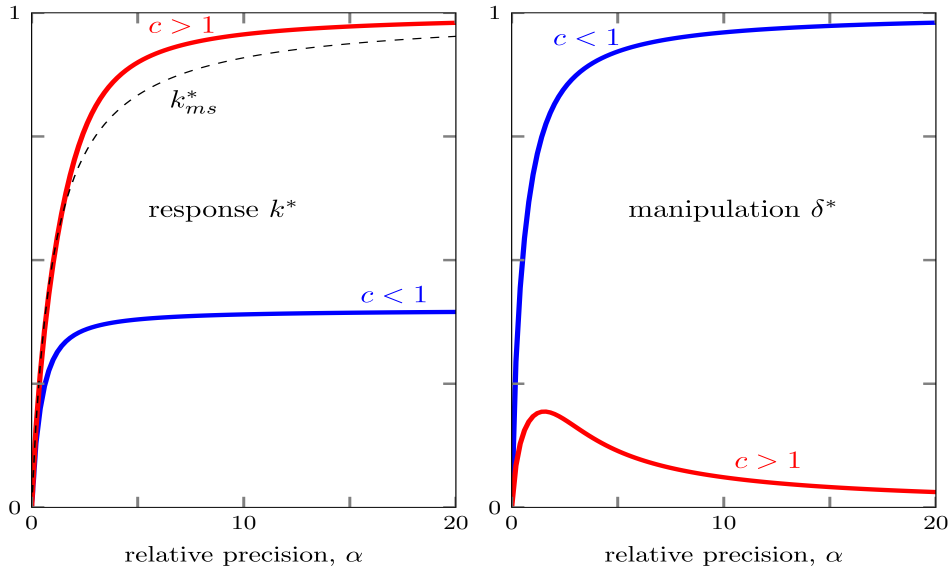

Figure 2: Changes in the relative precision, α.

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 $k^* \to 1$.

is increasing or decreasing in k at the equilibrium k^* and δ^*. Then using Lemma 3 we say that k^* and δ^* are strategic complements in equilibrium when α and c are such that

$$k^*(\alpha, c) < \hat{k}(c) \tag{39}$$

In terms of primitives, if $c < 1$ then from Lemma 3 we know that $k \leq c = \hat{k}(c)$ so it is surely the case that k and δ are strategic complements for the sender and so, as functions of α, the equilibrium k^* and δ^* increase or decrease together. 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. Figure 2 illustrates.

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

Lemma 5.

(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

$$c < \hat{c}(\alpha) \tag{40}$$

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

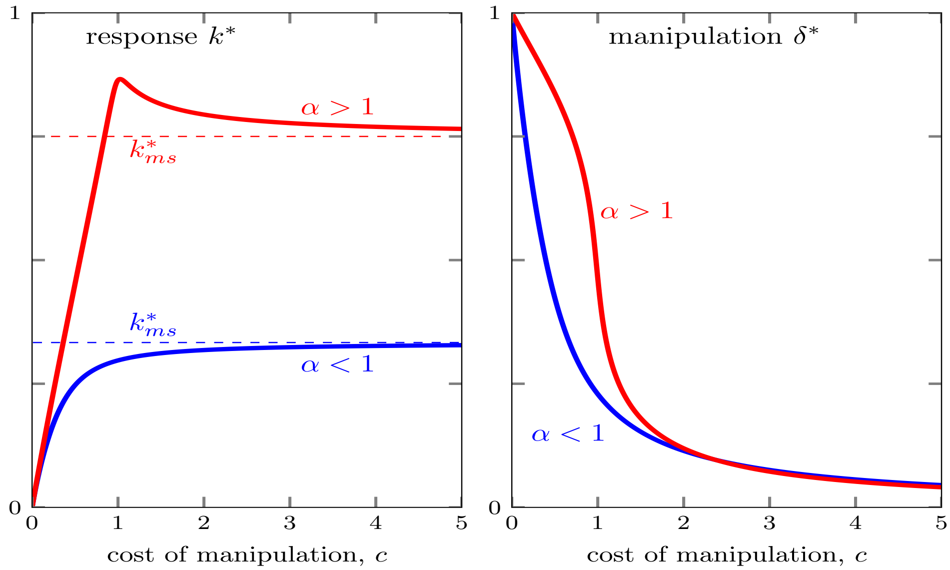

Figure 3: Changes in the sender's cost of manipulation c.

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^*_{ms} 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^*_{ms} from above as $c \to \infty$.

For any given level of the receivers' $k \in (0, 1)$, 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 2 we say that the precision effect dominates in equilibrium when α and c are such that

$$\delta^*(\alpha, c) > \hat{\delta}(\alpha) \tag{41}$$

In terms of primitives, if $\alpha \leq 1$ then from Lemma 2 we know 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. 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. Figure 3 illustrates.

When is k^* less than k^*_{ms}? Lemma 4 tells us that the receivers' responsiveness k^* is increasing in the relative precision α. But since the benchmark responsiveness $k^*_{ms} = \alpha/(\alpha+1)$ is also increasing in α, this does not yet tell us whether in equilibrium k^* is less or more than

k^*_{ms}. In the following lemma, we provide a simply necessary and sufficient condition for $k^* < k^*_{ms}$. In particular:

LEMMA 6. Receivers are less responsive to their signals with manipulation

$$k^*(\alpha, c) < k^*_{ms}(\alpha) \qquad \text{if and only if} \qquad c < c^*_{ms}(\alpha) \tag{42}$$

where

$$c^*_{ms}(\alpha) = \begin{cases} \dfrac{\alpha}{\alpha - 1}\left(\dfrac{\alpha}{\alpha+1}\right)^2 & \text{if } \alpha > 1 \\ \\ +\infty & \text{if } \alpha \leq 1 \end{cases} \tag{43}$$

In other words, if $\alpha \leq 1$ then we know that $k^* < k^*_{ms}$ regardless of c but if $\alpha > 1$ then whether or not the receivers' k^* is less than k^*_{ms} does depend on c. The key to this result is Lemma 2 which tells us that in order for $k(\delta) < k^*_{ms}$ it must be the case that the sender has manipulation δ that is sufficiently large, in particular, must have manipulation

$$\delta > 1 - 1/\alpha$$

(which is trivial if $\alpha \leq 1$). 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, $k^* < k^*_{ms}$. For $\alpha > 1$, the critical cost $c^*_{ms}(\alpha)$ given in (43) is the unique c that delivers $\delta^*(\alpha, c) = 1 - 1/\alpha$.

With this simple formula for $c^*_{ms}(\alpha)$ in hand, we know that for pairs α and c such that $c < c^*_{ms}(\alpha)$ the receivers are less responsive to their signals than they would be absent manipulation. For pairs α and c such that $c > c^*_{ms}(\alpha)$, the receivers are more responsive to their signals than they would be absent manipulation. Graphically, the function $c^*_{ms}(\alpha)$ is at first steeply decreasing in α, crosses $c^*_{ms}(\alpha) = 1$ 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^*_{ms} 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 $k^*_{ms} - k^*$. In particular, as $\alpha \to \infty$ we have $k^* \to \min[c, 1]$ and $k^*_{ms} \to 1$ so that:

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

$$\lim_{\alpha\to\infty}\left(k^*_{ms}-k^*\right)=\begin{cases}1-c & \text{if } c<1\\[2ex] 0 & \text{if } c>1\end{cases} \tag{44}$$

while the equilibrium amount of manipulation is

$$\lim_{\alpha\to\infty}\delta^*=\begin{cases}1 & \text{if } c<1\\[2ex] 0 & \text{if } c>1\end{cases} \tag{45}$$

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 ($k^* = c < 1$).[4]

Nearly cheap talk. Now consider the further special case that $c \to 0$ so that the sender's manipulation is nearly costless. Then with arbitrarily precise signals, we have $k^* \to c$ with $c \to 0$ so that in this *nearly cheap talk* limit the receivers do not respond to their signals at all, $k^* \to 0$. By contrast, in the absence of manipulation they would respond with $k^*_{ms} \to 1$. This combination of $c \to 0$ with $\alpha \to \infty$ gives the largest possible reduction in the receivers' responsiveness to their signals, and this will turn out to be the situation where the sender gains the most from their technology for manipulation.

The key point is that a combination of high α with low c allows the sender to achieve a large reduction in the receivers' responsiveness. This is because, although the *intrinsic precision* of the signals is high, the low cost of manipulation c makes k^* and δ^* strategic complements for the sender so that high α also implies high δ^*, thereby reducing the *equilibrium precision* of the receivers' signals. In the limit as $\alpha \to \infty$ with $c \to 0$, the receivers end up making decisions *based on their prior alone* even though their signals are intrinsically precise.

5 Welfare analysis

Our goal now is to establish the conditions under which the sender is made better off from their ability to manipulation and the conditions under which the receivers' are made worse off by the sender's manipulation.

[4] We discuss the knife-edge case $c = 1$ in the Appendix.

5.1 Sender's payoff with and without manipulation

Preliminaries. Let $V(\delta, k)$ denote the sender's ex ante expected utility if they choose manipulation δ when the receivers have response k. Taking expectations of (33) with respect to the prior that θ is normally distributed with mean z and precision α_z we get, in slight abuse of notation,

$$V(\delta, k) = \frac{1}{\alpha_z}\left(B(\delta, k) - C(\delta)\right) + \frac{1}{\alpha_x}k^2 \tag{46}$$

where, as before, $B(\delta, k) = (k\delta + 1 - k)^2$ and $C(\delta) = c\delta^2$ capture the sender's gross benefits and cost of manipulation.

Observe that whereas the equilibrium responsiveness of the receivers k^* and the manipulation of the sender δ^* depend only on the composite parameter $\alpha := (1-\lambda)\alpha_x/\alpha_z$ and the cost of manipulation c, the payoffs for the sender depend on the actual levels of the intrinsic signal precision α_x and the prior precision α_z separately.

Payoff without manipulation, v_{ms}^*. Now consider the special case where the sender cannot manipulate information, $\delta = 0$. If the receivers have response coefficient k then the sender has expected utility

$$v_{ms}(k) := V(0, k) = \frac{1}{\alpha_z}\left(B(0, k) - C(0)\right) + \frac{1}{\alpha_x}k^2 \tag{47}$$

which simplifies to

$$v_{ms}(k) = \frac{1}{\alpha_z}(1-k)^2 + \frac{1}{\alpha_x}k^2 \tag{48}$$

This is a quadratic in k and decreases from $v_{ms}(0) = 1/\alpha_z$ till it reaches its global minimum at $k = \alpha_x/(\alpha_x + \alpha_z)$ before increasing again to $v_{ms}(1) = 1/\alpha_x$. The benchmark payoff for the sender is the function $v_{ms}(k)$ evaluated at the point $k = k_{ms}^*$. Using $k_{ms}^* := \alpha/(\alpha+1)$ and $\alpha = (1-\lambda)\alpha_x/\alpha_z$, this works out to be

$$v_{ms}^* = \frac{(1-\lambda)^2\alpha_x + \alpha_z}{((1-\lambda)\alpha_x + \alpha_z)^2} \tag{49}$$

The sender's benchmark payoff v_{ms}^* is decreasing in the intrinsic signal precision α_x, increasing in the receivers' preference for coordination λ, and decreasing in prior precision α_z so long as the prior is not too diffuse. We now need to compare the sender's benchmark payoff v_{ms}^* to their payoff when they *can* manipulate information.

Payoff with manipulation, v^*. Turning now to the general case where the sender can manipulate information, let $v(k)$ and $\delta(k)$ denote the sender's maximum value function and best response if the receivers have response coefficient k, namely

$$v(k) := \max_{\delta\in[0,1]} V(\delta, k), \qquad \delta(k) := \operatorname*{argmax}_{\delta\in[0,1]} V(\delta, k) \tag{50}$$

Using the fact that $\delta(k) = (k - k^2)/(c - k^2)$, as in (31) above, the value function evaluates to

$$v(k) = \frac{1}{\alpha_z}\,(1-k)^2\left(\frac{c}{c-k^2}\right) + \frac{1}{\alpha_x}k^2, \qquad c - k^2 \geq 0 \tag{51}$$

Let $v^* = v(k^*)$ denote the sender's equilibrium payoff with manipulation.

Main properties of v^* and v^*_{ms}. With these definitions in hand, we can show that:

LEMMA 7.

(i) The sender's payoff with and without manipulation, v^* and v^*_{ms}, are both strictly decreasing in the intrinsic signal precision α_x.

(ii) The sender's payoff has limits

$$\lim_{\alpha_x \to 0^+} v^* = \lim_{\alpha_x \to 0^+} v^*_{ms} = \frac{1}{\alpha_z} \tag{52}$$

$$\lim_{\alpha_x \to \infty} v^* = \max\left[\,0\,,\ \frac{1-c}{\alpha_z}\,\right] \qquad \text{and} \qquad \lim_{\alpha_x \to \infty} v^*_{ms} = 0 \tag{53}$$

Both v^* and v^*_{ms} depend on the intrinsic signal precision α_x directly and also indirectly through the receivers' response coefficient. In both scenarios, the direct effect of increasing α_x is to reduce the sender's payoff — since a higher α_x makes the receivers' signals x_i more clustered around the true θ and for a given k this makes the receivers' actions a_i also more clustered, which reduces the sender's payoff. Likewise in both scenarios, the indirect effect of increasing α_x is to increase the receivers' k which also reduces the sender's payoff. Hence unambiguously an increase in α_x reduces the sender's payoff in both scenarios. But the *magnitude* of the reduction in the sender's payoff differs across these scenarios and depends sensitively on the sender's costs c. In particular, if $c < 1$ then v^* asymptotes to the constant $(1-c)/\alpha_z > 0$ as $\alpha_x \to \infty$ whereas v^*_{ms} asymptotes to zero. If instead $c > 1$, then both v^* and v^*_{ms} asymptote to zero.

Comparison of v^* and v^*_{ms}. How does the *level* of v^* compare to the benchmark payoff absent manipulation, v^*_{ms}? Since $v(k)$ is the maximum value function, for any particular response k we know that

$$v(k) := \max_{\delta \in [0,1]} V(\delta, k) \geq V(0, k) =: v_{ms}(k) \tag{54}$$

But this *does not* guarantee that the sender gains from manipulation. The receivers' response k^* generally differs from the level of responsiveness they would have absent manipulation, k^*_{ms}. And because of this, v^* can likewise be either less than or more than v^*_{ms}. In short, the technology for manipulating information need not make the sender better off.

For future reference we decompose the sender's gain from manipulation as follows:

$$v^* - v^*_{ms} = (v(k^*) - v_{ms}(k^*)) + (v_{ms}(k^*) - v_{ms}(k^*_{ms})) \tag{55}$$

This leads to the following sufficient conditions for the sender to gain from their manipulation:

PROPOSITION 2.

(i) The sender gains from manipulation $v^* > v^*_{ms}$, if either $c < c^*_{ms}(\alpha)$ or $\lambda < 1/2$.

(ii) In the limit as $\alpha_x \to \infty$, the sender's gain from manipulation is

$$\lim_{\alpha_x \to \infty} (v^* - v^*_{ms}) = \begin{cases} \dfrac{1-c}{\alpha_z} & \text{if } c < 1 \\[2ex] 0 & \text{if } c > 1 \end{cases} \tag{56}$$

The first part of the proposition uses the fact that $v(k) \geq v_{ms}(k)$ for all k and hence the first term in (55) is unambiguously positive. So for the sender to gain it is sufficient that that the second term in (55) is also positive. Since $v_{ms}(k)$ is strictly decreasing on $(0, k^*_{ms})$ the second term is positive if $k^* < k^*_{ms}$. In short, if $k^* < k^*_{ms}$ then $v^* > v^*_{ms}$. In turn, Lemma 6 tell us that $k^* < k^*_{ms}$ if and only if $c < c^*_{ms}(\alpha)$ where $c^*_{ms}(\alpha)$ is the critical cost given in equation (43) above. This is only a sufficient condition, however. Even if $k^* > k^*_{ms}$ it can still be the case that $v^* > v^*_{ms}$. For example, in the special case that $\lambda = 0$ the benchmark response coefficient k^*_{ms} turns out to be the minimizer of $v_{ms}(k)$ so that $v^*_{ms} \leq v_{ms}(k)$ for all k. Consequently, if $\lambda = 0$ we still have $v^* > v^*_{ms}$ even if $k^* > k^*_{ms}$. More generally, so long as the receivers' coordination motive is relatively weak, $\lambda < 1/2$, then the benchmark payoff v^*_{ms} is sufficiently low that v^* is greater than v^*_{ms} even if $k^* > k^*_{ms}$. The second part of the proposition simply uses Lemma 7 to conclude that the gain from manipulation $v^* - v^*_{ms}$ asymptotes to $(1-c)/\alpha_z$ if $c < 1$ but asymptotes to zero if $c > 1$.

Figure 4 illustrates, showing v^* and v^*_{ms} as functions of α_x for various levels of c. Both v^* and v^*_{ms} are monotonically decreasing in α_x starting from the common initial point $1/\alpha_z$ when $\alpha_x = 0$. For low levels of α_x we know $k^* < k^*_{ms}$ and hence for low levels of α_x we have $v^* > v^*_{ms}$ even as both v^* and v^*_{ms} decrease. Now consider the limiting case as $\alpha_x \to \infty$. If the sender cannot manipulate, then we get $k^*_{ms} \to 1$ so that the sender's benchmark payoff is driven to $v^*_{ms} \to 0$, the worst possible outcome for the sender. But if the sender can manipulate, then we get $k^* \to \min[c, 1]$ so that if $c < 1$ the receiver's responsiveness asymptotes to $k^* = c < k^*_{ms}$ which opens up a permanent payoff gap between the two scenarios. In particular, if the cost of manipulation is relatively low, $c < 1$, then the sender's payoff asymptotes to $v^* = (1-c)/\alpha_z > v^*_{ms}$ and the sender gains from manipulation.

Nearly cheap talk limit. To appreciate the potential magnitude of the gain from manipulation, observe that for any configuration of parameters we have

$$v^* \leq v(0) = \frac{1}{\alpha_z} \tag{57}$$

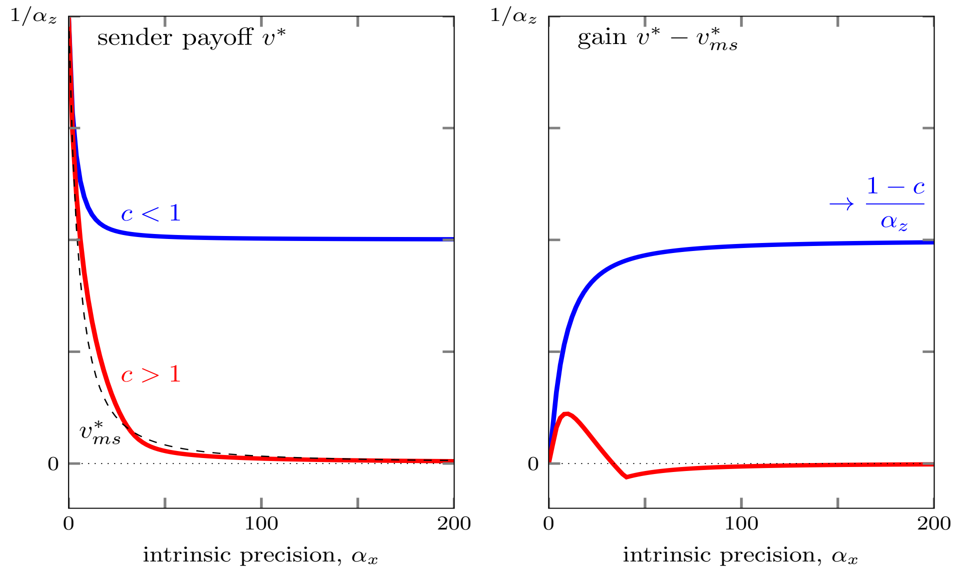

Figure 4: Sender gains most when α_x high and c low.

Sender's equilibrium payoff v^* (left panel) and gain from manipulation $v^* - v^*_{ms}$ (right panel) as functions of the intrinsic precision α_x for various levels of the cost of manipulation c. Both v^* and v^*_{ms} are monotonically decreasing in α_x and v^*_{ms} asymptotes to zero as $\alpha_x \to \infty$. If $c > 1$, then v^* also asymptotes to zero and the sender's manipulation backfires, $v^* < v^*_{ms}$, for high α_x. But if $c < 1$ then v^* asymptotes to $(1-c)/\alpha_z > 0$ as $\alpha_x \to \infty$ and the sender gains, $v^* > v^*_{ms}$, for high α_x. In this sense, the sender gains when α_x is high and c is low.

In this sense, the best possible payoff the sender can obtain is $1/\alpha_z$. This payoff can be obtained if $\alpha_x = 0$, in which case $k^* = 0$ and receivers make decisions based only on their prior, which has precision α_z. In the absence of manipulation, this $\alpha_x = 0$ scenario is the only way for the sender to get a payoff of $1/\alpha_z$. But in the presence of manipulation, this payoff can also be obtained even as $\alpha_x \to \infty$. Indeed we have just seen that if $c < 1$ then in the limit as $\alpha_x \to \infty$ the sender's payoff approaches $v^* = (1-c)/\alpha_z$. Thus if in addition $c \to 0$ we get $v^* \to 1/\alpha_z$. So in this *nearly cheap talk* limit, the sender is obtaining its best possible payoff $1/\alpha_z$, the same as it would have if $\alpha_x = 0$, even though $\alpha_x \to \infty$. Put differently, if the cost of manipulation is low enough, then the sender is able to effectively insulate itself from the consequences of receivers having intrinsically precise signals.

Manipulation can backfire. The combination of high α_x and low c tends to make the sender better off when they can manipulate information. But they are not always better off. As shown in Figure 4, when α_x is high and c is high then we have both v^* and v^*_{ms} approaching zero but with $v^* < v^*_{ms}$ for sufficiently high α_x. For these parameters, the sender's ability to manipulate information *backfires* in the sense that the sender would be better off if they could credibly commit to not use their manipulation technology. More precisely:

Proposition 3. For each $c > 1$ and $\lambda > 1/2$ there exists a critical point

$$\alpha_x^* > \left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{\alpha_z}{1-\lambda}\right) \tag{58}$$

such that the sender's manipulation backfires, $v^* < v_{ms}^*$, for all $\alpha_x > \alpha_x^*$.

To understand this condition, first observe from Proposition 2 that $c > c_{ms}^*(\alpha)$ and $\lambda > 1/2$ are *necessary* conditions for the sender's manipulation to backfire, $v^* < v_{ms}^*$. Moreover, using the formula for $c_{ms}^*(\alpha)$ in (43) it can be shown that whenever the composite parameter α is greater than the constant $(1+\sqrt{5})/2 \approx 1.62$ the critical cost is $c_{ms}^*(\alpha) < 1$ so that $c > 1$ suffices for $c > c_{ms}^*(\alpha)$. So the proposition says that if these necessary conditions are met, a sufficient condition for the sender's manipulation to backfire is for α_x to be large enough, $\alpha_x > \alpha_x^*$. Asymptotically, the sender's loss from this backfiring is negligible in that both $v^* = v_{ms}^* = 0$. But for more moderate levels of α_x the loss can be substantial.

A simple taxonomy. To summarize, the sender's payoffs are at their highest $v^* = v_{ms}^* = 1/\alpha_z$ when the intrinsic signal precision is low, $\alpha_x = 0$. When the intrinsic signal precision is high, $\alpha_x \to \infty$ we get $v^* = v_{ms}^* = 0$ if $c > 1$ but $v^* = (1-c)/\alpha_z > v_{ms}^*$ if $c < 1$. In this sense, the sender experiences large gains when α_x is high and c is low.

5.2 Receivers' loss with and without manipulation

Preliminaries. We measure the receivers' welfare using their ex ante expected loss. Let $L(k, \delta)$ denote their expected loss if they have response k when the sender chooses manipulation δ. This works out to be

$$L(k, \delta) = \frac{1-\lambda}{\alpha_z} B(\delta, k) + \frac{1}{\alpha_x} k^2 \tag{59}$$

where $B(\delta, k) = (k\delta + 1 - k)^2$ is the sender's gross benefit from manipulation (which tends to raise the receivers' loss).

Loss without manipulation, l_{ms}^*. Now consider the special case where the sender cannot manipulate information, $\delta = 0$. The receivers then have expected loss

$$l_{ms}(k) := L(k, 0) = \frac{1-\lambda}{\alpha_z}(1-k)^2 + \frac{1}{\alpha_x} k^2 \tag{60}$$

Evaluating at $k = k_{ms}^*$ and using $k_{ms}^* := \alpha/(\alpha+1)$ and $\alpha := (1-\lambda)\alpha_x/\alpha_z$, this is

$$l_{ms}^* = \frac{(1-\lambda)}{(1-\lambda)\alpha_x + \alpha_z} \tag{61}$$

Loss with manipulation, l^*. Turning now to the general case where the sender can manipulate information, let $l(\delta)$ and $k(\delta)$ denote the receivers' minimum loss function and best response if the sender has manipulation coefficient δ, namely

$$l(\delta) := \min_{k\in[0,1]} L(k,\delta), \qquad k(\delta) := \operatorname*{argmin}_{k\in[0,1]} L(k,\delta) \tag{62}$$

Using the solution for $k(\delta)$ from (28) above and $\alpha := (1-\lambda)\alpha_x/\alpha_z$ this evaluates to

$$l(\delta) = \frac{(1-\lambda)}{(1-\delta)^2(1-\lambda)\alpha_x + \alpha_z} \tag{63}$$

Let $l^* = l(\delta^*)$ denote the receivers' equilibrium loss with manipulation.

Main properties of l^* and l^*_{ms}. With these definitions in hand, we can show that:

LEMMA 8.

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

(ii) The receivers's loss with manipulation l^* is strictly decreasing in the intrinsic signal precision α_x if $c > 1$, but 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 has limits

$$\lim_{\alpha_x\to 0^+} l^* = \lim_{\alpha_x\to 0^+} l^*_{ms} = \frac{1-\lambda}{\alpha_z} \tag{64}$$

$$\lim_{\alpha_x\to\infty} l^* = \begin{cases} \dfrac{1-\lambda}{\alpha_z} & \text{if } c<1 \\ 0 & \text{if } c>1 \end{cases} \qquad \text{and} \qquad \lim_{\alpha_x\to\infty} l^*_{ms} = 0 \tag{65}$$

In both scenarios, the direct effect of increasing α_x is to reduce the receivers' loss — since a higher α_x increases the precision of their estimates of θ allowing them to take actions that are closer to θ and also to take actions that are closer to one another, both of which reduce their loss. When the sender can manipulate, there is an additional indirect effect from δ^* with higher values of δ^* increasing the receivers' loss. As shown in Lemma 4, 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 low, $c < 1$, then k^* and δ^* are strategic complements so that increasing α_x increases both k^* and δ^* together. 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 overall l^* is increasing in α_x even though l^*_{ms} is unambiguously decreasing. By contrast, if δ^* and k^* are strategic substitutes then δ^* is decreasing in α_x so that the indirect and direct effects of increasing α_x work together to reduce the receivers' loss.

Comparison of l^* and l^*_{ms}. Since in equilibrium the sender's manipulation $\delta^* \in [0,1]$ and the function $l(\delta)$ is strictly increasing in δ we immediately have:

Proposition 4. The receivers are worse off with manipulation, $l^* \geq l^*_{ms}$, strictly if $\delta^* > 0$.

So the only question for the receivers is just *how much* worse off will they be?

Proposition 5.

(i) For each $c < 1$ there is a critical point α_x^{**} such that the receivers' excess loss from manipulation $l^* - l^*_{ms}$ is strictly increasing for $\alpha_x > \alpha_x^{**}$.

(ii) In the limit as $\alpha_x \to \infty$, the receivers' excess loss from manipulation is

$$\lim_{\alpha_x \to \infty} (l^* - l^*_{ms}) = \begin{cases} \dfrac{1-\lambda}{\alpha_z} & \text{if } c < 1 \\ 0 & \text{if } c > 1 \end{cases} \tag{66}$$

Since l^*_{ms} is strictly decreasing in α_x, the excess loss from manipulation $l^* - l^*_{ms}$ will certainly be increasing if l^* is increasing in α_x. From Lemma 8 we know that l^* is strictly increasing if $c < 1$ and $\alpha_x > \alpha_x^{**}$. With this parameter configuration, the indirect effect of α_x via the sender's manipulation δ^* is large enough that the receivers are not just worse off with manipulation but are becoming increasingly worse off as α_x increases. If $c < 1$, the receivers' loss from manipulation asymptotes to $(1-\lambda)/\alpha_z > 0$ as $\alpha_x \to \infty$, i.e., the same loss the receivers would have if $\alpha_x = 0$. By contrast, if $c > 1$ then even though $l^* > l^*_{ms}$ for all α_x the gap between them becomes negligible as $\alpha_x \to \infty$.

Figure 5 illustrates, showing l^* and l^*_{ms} as functions of α_x for various levels of c. Both start from the common initial point $(1-\lambda)/\alpha_z$ when $\alpha_x = 0$ and are initially decreasing in α_x. If $c < 1$ then for high enough $\alpha_x > \alpha_x^{**}$ the strategic complementarity effect is large enough to dominate and l^* begins to increase in α_x. In this scenario, as $\alpha_x \to \infty$ we have $\delta^* \to 1$ and the receivers' loss is driven all the way back up to $l^* = (1-\lambda)/\alpha_z$ so that the receivers obtain no benefit from the high level of intrinsic signal precision α_x. Put differently, when α_x is high and c is low then the sender chooses high δ^* and the equilibrium precision of the receivers' signals is low so they make decisions based primarily on their prior and receive payoffs commensurate with the informativeness of their prior. In this sense, the receivers experience large losses from manipulation when α_x is high and c is low, i.e., the same circumstances as when the sender experiences large gains.

Jumps in the receivers' loss. Notice that in Proposition 5, the receivers' asymptotic loss is discontinuous in c in the vicinity of $c = 1$. A small decrease in costs from $c = 1 + \varepsilon$ to $c = 1 - \varepsilon$ increases the asymptotic loss sharply from 0 to $(1-\lambda)/\alpha_z > 0$ (i.e., all the way back to the loss they would have with intrinsically uninformative signals, $\alpha_x = 0$). We show

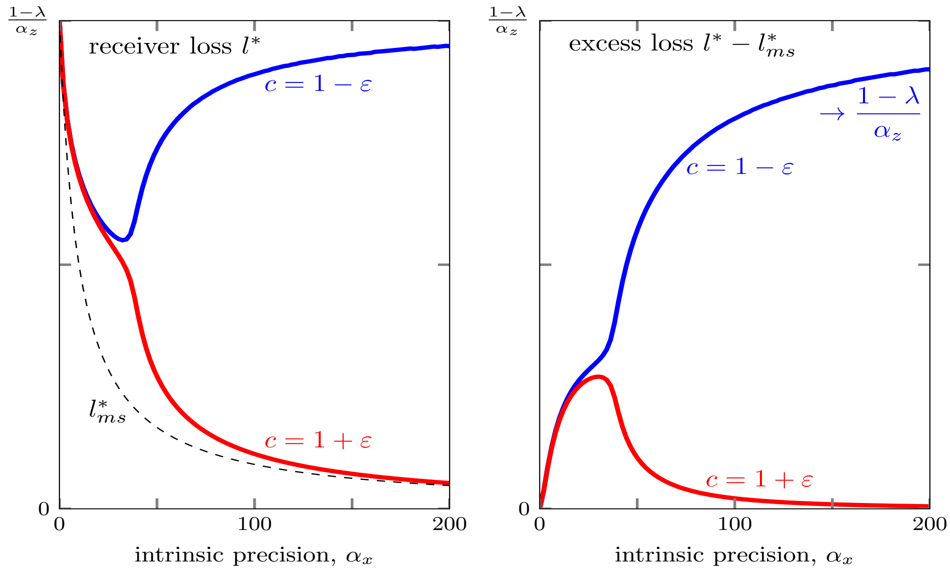

Figure 5: Receivers lose most when α_x high and c low.

Receivers' equilibrium loss l^* (left panel) and excess loss $l^* - l^*_{ms}$ (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 $\alpha_x \to \infty$. For $\alpha_x > 4\alpha_z/(1-\lambda)$ the receivers' loss is discontinuous in c at $c = 1$. A small decrease from $c = 1+\varepsilon$ to $c = 1-\varepsilon$ causes a discrete jump up in the receivers' loss.

in the Appendix that this discrete jump reflects a corresponding jump in the sender's δ^* in the vicinity of $c = 1$.

This jump in the receivers' loss in the vicinity of $c = 1$ is present for a large range of α_x and is not an artefact of the limit as $\alpha_x \to \infty$. In fact:

Remark 2. For each $\alpha > 4$, the receivers' loss l^* is discontinuous in c at $c = 1$. An ε-reduction from $c = 1+\varepsilon$ to $c = 1-\varepsilon$ causes the receivers' loss to jump up from

$$\underline{l}^* := \lim_{c \to 1^+} l^* = \frac{1-\lambda}{\alpha_z}\left(1 + \alpha\left(\frac{1+\sqrt{1-(4/\alpha)}}{2}\right)^2\right)^{-1}$$

to

$$\overline{l}^* := \lim_{c \to 1^-} l^* = \frac{1-\lambda}{\alpha_z}\left(1 + \alpha\left(\frac{1-\sqrt{1-(4/\alpha)}}{2}\right)^2\right)^{-1}$$

As explained in the Appendix, to get the jump we need the composite parameter $\alpha > 4$. Equivalently, we need the intrinsic signal precision $\alpha_x > 4\alpha_z/(1-\lambda)$. Observe that holding α_z and λ fixed and taking $\alpha_x \to \infty$ the limits become $\underline{l}^* \to 0$ and $\overline{l}^* \to (1-\lambda)/\alpha_z$ so that "at $\alpha_x = \infty$" the jump is from 0 to $(1-\lambda)/\alpha_z$, as in Proposition 5 above.

Manipulation is not zero-sum. Receivers are always worse off with manipulation but, depending on parameters, the sender can be worse off too. Hence there are circumstances in which *both* the receivers and the sender are worse off and hence the effects are not zero-sum. Even when the sender gains, the size of the sender's gains do not generally correspond to the size of the receivers' losses.

6 Implications of the model

We now discuss some key implications of our model. For this discussion, we suppose that the sender is a politician who seeks to influence the beliefs of a group of imperfectly informed receivers. We begin with an application of the model that interprets new social media technologies as changing both the intrinsic precision of information and the politician's costs of manipulation. We then discuss how the information manipulation technology in our model accords with the kinds of political messaging that are common in practice.

6.1 The challenge of social media technologies

Our goal now is to interpret the impact of new social media technologies using our model.

Pessimism and optimism about 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 new social media platforms like Google, Facebook, Twitter and YouTube has led to renewed pessimism about the implications of new media technologies (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 US presidential election and the 2016 UK referendum on leaving the European Union, has led to newly intense scrutiny of social media technologies (e.g., Faris et al., 2017).

Important features of social media technologies. As emphasized by Allcott and Gentzkow (2017), these new 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 meant 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 serves to blur the distinctions between traditional media outlets and their new competitors and to blur the distinctions between reporting and opinion.

In the context of our model, we interpret these two features as follows. First, we interpret these new social media technologies as facilitating the entry of new media outlets that results in an increase in the *number* of signals that each receiver obtains. Second, we interpret these technologies as reducing the costs to the sender of introducing a given amount of slant. The first interpretation is fairly direct. We suppose that each individual receiver $i \in [0,1]$ obtains n independent signals of the form $x_{ij} = \theta + s + \varepsilon_{ij}$ for $j = 1, ..., n$ where the ε_{ij} are IID normal with mean zero and precision α_x. Since the average $x_i := \frac{1}{n}\sum_{j=1}^{n} x_{ij}$ is a sufficient statistic for the unknown $\theta + s$, this is equivalent to the receiver having a single signal x_i with mean $\theta + s$ and precision $n\alpha_x$, increasing in n. The second interpretation is less direct. We suppose that the sender's cost of introducing slant s is $c(n)s^2$ where $c(n)$ is *decreasing* in n so that as n increases the cost of introducing a given amount of slant falls. The idea here is that these new technologies also lead to consumers acquiring their information in a feed that blurs the distinctions between different media outlets with an ecosystem of likes, faves and retweets serving to facilitate the diffusion of all kinds of reporting, both reliable and unreliable, and this serves to reduce the sender's cost of propagating their "alternative facts".

To summarize, we interpret the rise of new social media technologies as an increase in n that simultaneously (i) increases the intrinsic precision $n\alpha_x$ of the receivers' information and (ii) decreases the cost $c(n)$ of a given amount of manipulation. Then from Proposition 2 we know that for n high enough the sender will gain from their manipulation and that the size of their gain will be relatively large when n is large. For example, suppose $c(n) \to \underline{c} \in (0,1)$ as $n \to \infty$. Then the sender's gain from manipulation will approach

$$v^* - v^*_{ms} = \frac{1 - \underline{c}}{\alpha_z}$$

and this gain will be larger the smaller is $\underline{c}$. In this sense, the rise of new social media technologies makes information manipulation *increasingly tempting* to the politician. Moreover Proposition 5 tells us that this also causes the receivers' loss from manipulation to rise with n and to approach their worst possible welfare loss as $n \to \infty$, namely

$$l^* - l^*_{ms} = \frac{1 - \lambda}{\alpha_z}$$

To the extent that we view the interests of society at large as being roughly aligned with the receivers, these results suggest we should indeed be pessimistic about the impact of social media — the potential welfare gains from the entry of new sources of information are wiped out by the politician's "alternative facts" that can now be easily propagated. Or, in other words, these results suggest that with the rise of social media, societal efforts to keep information manipulation in check will be increasingly beneficial. In this sense, editorial judgement, fact-checking, and a reputation for reliable reporting are likely to be more important than ever.

Of course, one doesn't need this model to arrive at the conclusion that these new 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, our receivers are perfectly rational, ruthless Bayesian updating machines, utterly free from confirmation bias or other behavioral traits that might make them easily manipulated. In this sense, we view our results as understating the potential adverse implications of these new social media technologies.

Gains from groupthink. The strength of the receivers' coordination motive is important for the sender's ability to gain from manipulation. In particular, from Proposition 3 we know that when the coordination motive is sufficiently strong, $\lambda > 1/2$, the sender's manipulation can *backfire* if the cost of manipulation is relatively high, $c > 1$, and the signal precision α_x is high enough. In this situation, an ability to commit to *not* manipulate information would be valuable for the sender as well as for society. To the extent that the receivers' strong coordination motive helps align the interests of the sender with the rest of society (in preventing strategic information manipulation), this feature of our model suggests that we should be more optimistic about the rise of social media in societies that are intrinsically well coordinated — or perhaps simply *cooperative* — and more pessimistic about the rise of social media in societies that are less well coordinated.

In addition to helping align the interests of the sender and receivers, Lemma 8 shows that a strong coordination motive is intrinsically valuable to the receivers. The higher is λ, the lower the receivers' expected loss — i.e., the receivers actually prefer it when the coordination motive is strong. In this sense, the receivers experience *gains from groupthink*. These notional gains occur even if there is no information manipulation but are larger if the receivers can be manipulated. In the limit as $\lambda \to 1$, so that the receivers only care about coordinating with each other and not about the accuracy of their beliefs about θ, the receivers' are completely insulated from the welfare-reducing effects of the sender's manipulation because they ignore their signals and act only according to their common prior. It can be good to be dogmatic if people are trying to mislead you.

6.2 Interpreting the information manipulation technology

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

Changing the story. Public relations consultants typically advise that their clients "stay on message". This gives rise to pivots of the form "... but Megyn, I don't think that's the real question. Frankly, the real question is ...". Sometimes such pivots are because the client wants to actively push a specific message and this media opportunity is a valuable chance to do so. But often such pivots are attempts to "change the story" in an effort to mitigate damage from being unable to address an inconvenient fact. The current political environment is thick with examples. To take just one, in response to the Mueller investigation's indictment of former Trump campaign manager Paul Manafort, Trump administration surrogates pushed the widely-debunked hoax that the "real Russia scandal" is that, as secretary of state, "Hillary Clinton gave the Russians 20% of US uranium" (see e.g., Illing, 2017; Sullivan, 2017).

Expressions like "muddying the waters" and "throwing up chaff" are journalistic cliches precisely because these kinds of efforts to change the story are so routine.

Misdirection and bias. In such efforts to change the story, the *form* of the misdirection is a kind of slant or spin that is favourable to the client (or unfavourable to their opponents), at least when taken at face-value. But the *effect* of the misdirection is to create confusion. Strategic communications is rife with efforts to change the story where the form of such intervention is an alternative story with an opposing slant but where the effect of such interventions is to increase uncertainty. In our model, the truth is θ and the sender's intervention is some slant s such that the sender says "... no the truth is not what you think, it's actually $\theta + s$ ..." which individual receivers observe as $x_i = \theta + s + \varepsilon_i$.

But the receivers in our model do not take such slant at face-value. In equilibrium, the receivers in our model have *unbiased* beliefs in the sense that the posterior expectation $\mu_i := \mathbb{E}[\,\theta\,|\,x_i, z]$ is a statistically unbiased estimate of the true θ. The receivers know the sender's incentives and are able to use their knowledge of the sender's problem to neutralize any effort to shift their expectations about θ. Instead, the equilibrium effect of the sender's intervention is to increase the posterior variance of the receiver's beliefs (i.e., to increase the uncertainty in the receivers' estimates of θ relative to the benchmark of no manipulation). The equilibrium message is $y = \theta + s = (1-\delta)\theta + \delta z$ so the equilibrium slant is $s = \delta(z-\theta) = \delta\varepsilon_z$. The receivers' equilibrium signals are then $x_i = \theta + \delta\varepsilon_z + \varepsilon_i$ and the dependence on the common factor $\delta\varepsilon_z$ makes the signals x_i noisier sources of information about θ than they would be absent manipulation. The receivers' posterior variance increases from $1/(\alpha_x + \alpha_z)$ absent manipulation to $1/((1-\delta)^2\alpha_x + \alpha_z)$ with manipulation. As the amount of manipulation rises from $\delta = 0$ to $\delta = 1$, the receivers' posterior variance increases from the benchmark $1/(\alpha_x + \alpha_z)$ all the way up to the prior variance $1/\alpha_z$, i.e., when $\delta = 1$ it is as if $\alpha_x = 0$ even if the signals are actually intrinsically quite precise. The sender finds it optimal to pay the cost cs^2 of the slant s even though such slant has no equilibrium effect on mean beliefs because the slant makes the signals x_i endogenously noisier. The sender's gain from this endogenous noise is naturally larger when the intrinsic precision α_x is high and the cost of manipulation

c is low, because that is when the sender's manipulation leads the receivers to be much less responsive to their signals than the underlying informativeness of those signals warrants.

In short, although the form of the slant is a superficially mean-shifting message $y = \theta + s$ the equilibrium effect of the slant is no change in mean but instead an increase in variance.

Exploiting prior beliefs. The politician's scope to manipulate information is determined by the gap between the true θ and the receivers' prior expectation z. The politician's strategy

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

chooses a message y that lies between θ and z with small amounts of manipulation causing y to be close to θ and large amounts of manipulation causing y to be close to z. To the extent that z is itself close to θ, there is not much potential for the politician to manipulate and the message will inevitably also be close to θ. Ex ante, if α_z is high (so that the true θ is likely to be close to z), then the politician's expected gains from manipulation will be relatively low and the receivers' expected loss from manipulation will also be relatively low — i.e., the consequences of discounting the signals x_i on account of manipulation are modest if the prior is itself highly informative. But if α_z is low (so that the true θ may be far from z), then the politician's expected gains from manipulation will be relatively high and the receivers' expected loss from manipulation will also be relatively high — i.e., the consequences of discounting the signals x_i on account of manipulation are substantive if the prior is not informative. In this sense, the politician is exploiting the fact that the receivers' priors will not be, in general, perfectly aligned with the truth.[5]

To illustrate this, consider a US presidential election where the Republican candidate is communicating to Republican voters about the rival Democratic candidate, perhaps seeking to prevent marginally attached Republican voters from drifting to the Democratic candidate. Suppose the true θ is "Hillary is a centrist Democrat" but the Republican voters' prior z is "Hillary is a radical communist". Then the message y is that "Hillary is a left-wing Democrat" and in equilibrium the Republican voters have beliefs $\mu_i := \mathbb{E}[\,\theta \,|\, x_i, z]$ that are correct on average, i.e., on average the Republican voters *do* believe that "Hillary is a centrist Democrat", but they are more uncertain about their beliefs than they would be absent the messaging. A striking example of this is in the limit with very precise signals but very low costs of manipulation, in which case the Republican voters end up making decisions based entirely on their prior z. The messaging y creates a clear gap between the *actions* of the Republican voters (which reflect their prior) and the *beliefs* of the Republican voters (which are correct on average). In short, the Republican voters act according to their prior z that

[5] To be clear, the prior mean z constitutes an unbiased signal of the true θ in that $\mathbb{E}[\theta \,|\, z] = \theta$, but for any given realization $z \neq \theta$ almost surely. The distance $z - \theta$ then determines the sender's scope to manipulate and lower α_z implies that ex ante that distance is more likely to be large.

"Hillary is a radical communist" even though they tend to believe that "Hillary is a centrist Democrat". This misalignment between actions and beliefs is caused by the Republican candidate's messaging, which takes advantage of the misalignment between the Republican voters' prior and the truth. Of course this leaves out many interesting complications, especially the role of competition between rival communications and the role of heterogeneous priors. The point is simply that, in this model, the receivers' priors matter crucially for the sender's incentives to manipulate and for the welfare consequences of that manipulation.

7 Conclusion

We develop a sender/receiver game with many imperfectly informed and imperfectly coordinated receivers. The receivers try to take actions that align both with their beliefs about the true state of the world and also with the actions of other receivers. The sender engages in costly efforts to prevent the receivers from taking actions that align with the true state of the world. In this sense, the sender is at least partially *malevolent*. 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 receivers beliefs are unbiased.

But despite this lack of bias, the sender can gain substantially from the endogenous noise that is a byproduct of their manipulation. In particular, we find that the sender gains most when the receivers have intrinsically very precise information but the sender's costs of manipulation are low. In this situation, the receivers possess information that is a very reliable guide to the truth *but their actions do not reflect it*. We also find that the manipulation can *backfire* on the sender. This happens when the receivers' information is intrinsically precise, the costs of manipulation are high, and the receivers are sufficiently well coordinated. The receivers are always made worse off by the manipulation, even when it backfires on the sender (this is not a zero-sum game). We find that a small fall in the sender's costs of manipulation can lead to a discrete *jump* in the amount of manipulation and a correspondingly dramatic welfare loss for the receivers.

In our main application, we interpret the sender as a politician who benefits from the tumultuous noise created as they propagate their "alternative facts". And we argue that the rise of new social media technologies have indeed had the effect of increasing the intrinsic, potential, precision of available information while at the same time reducing the costs of propagating misinformation — i.e., the combination of parameter changes that our model suggests will be most beneficial to the sender and most harmful to everyone else. Of course, one doesn't need this model to arrive at the conclusion that these new social media technologies may have adverse implications. But we find it striking that severe adverse implications can arise even in a setting where the receivers are rational and not easy to deceive. In this sense, we view our results as something of an understatement.

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, we should note that our stylized model has broader applications. For example, consider a game with one market-maker and a continuum of traders. Suppose the traders care about both the true value of an asset and the price of that asset as determined by average trading activities in the market. It may then be in the interest of the market-maker to persuade the traders not to coordinate on the true value of the asset, so as to maximize trading volume. Another example is a game between a firm's owner and many creditors. Suppose the firm is teetering on the edge of bankruptcy. The individual creditors would like to pick the best bankruptcy plan taking into account both their assessment of the current profitability of the firm and the need to coordinate with the other creditors. The owner of the firm who wants to buy time or prevent the bankruptcy will then have incentive to persuade the creditors to not have similar assessments of the current profitability of the firm in order to make it harder for them to agree on a plan.

Appendix

A Proofs

A.1 Best responses

Proof of Lemma 1.

Recall that the sender's best response (20) requires $c \geq k^2$, otherwise the sender is at a corner with $\delta(k) = 0$. Hence we focus on $k \in [-\sqrt{c}, +\sqrt{c}]$ and we distinguish two cases, depending on the magnitude of c.

(i) If $c \geq 1$ such that $1 \leq \sqrt{c} \leq c$ then from the sender's best response (20) 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' best response (27) we have $k(\delta) < 0$ if $\delta > 1$ and $k(\delta) < 1$ if $\delta < 0$. Hence if $c \geq 1$ the only possible crossing points of $k(\delta)$ and $\delta(k)$ are in the unit square with $k^* \in [0, 1]$ and $\delta^* \in [0, 1]$.

(ii) If $c \in (0, 1)$ such that $0 < c < \sqrt{c} < 1$ then from the sender's best response (20) 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 $k \in [0, c]$. Again from the receivers' best response (27) we have $k(\delta) < 0$ if $\delta > 1$ and $k(\delta) < 1$ if $\delta < 0$. Hence if $c \in (0, 1)$ the only possible crossing points of $k(\delta)$ and $\delta(k)$ are in a *subset* of the unit square with $k^* \in [0, c]$ and $\delta^* \in [0, 1]$.

Hence in any equilibrium, $k^* \in [0, 1]$ and $\delta^* \in [0, 1]$. □

Proof of Lemma 2.

Differentiating the receivers' best response $k(\delta)$ in (28) with respect to δ gives

$$k'(\delta) = \alpha \frac{(1-\delta)^2\alpha - 1}{((1-\delta)^2\alpha + 1)^2}, \qquad \delta \in [0, 1], \qquad \alpha > 0 \tag{A1}$$

Hence

$$k'(\delta) > 0 \qquad \Leftrightarrow \qquad \delta < 1 - 1/\sqrt{\alpha} \tag{A2}$$

Now let $\hat{\delta}(\alpha) := \max[\,0, 1 - 1/\sqrt{\alpha}\,]$. If $\alpha \leq 1$, then the critical point is on the boundary $\hat{\delta}(\alpha) = 0$ and there is no $\delta \in [0, 1]$ such that $\delta < \hat{\delta}(\alpha)$ hence if $\alpha \leq 1$ we conclude that $k(\delta)$ is decreasing for all $\delta \in [0, 1]$. If $\alpha > 1$ then the critical point is in the interior $\hat{\delta}(\alpha) \in (0, 1)$ and we conclude that $k(\delta)$ is increasing on the interval $[0, \hat{\delta}(\alpha)]$ and decreasing on the interval $[\hat{\delta}(\alpha), 1]$. Plugging in $\delta = 0$ and $\delta = 1$ gives the boundary values $k(0) = \alpha/(\alpha + 1)$ and $k(1) = 0$ respectively. □

Proof of Lemma 3.

Differentiating the sender's best response $\delta(k)$ in (31) with respect to k gives

$$\delta'(k) = \left(\frac{1}{c - k^2}\right)^2 \left(k^2 - 2ck + c\right), \qquad k \in \mathcal{K}(c), \qquad c > 0 \tag{A3}$$

Hence

$$\delta'(k) > 0 \qquad \Leftrightarrow \qquad k^2 - 2ck + c > 0 \tag{A4}$$

Now consider the quadratic equation

$$q(k) = k^2 - 2ck + c = 0 \tag{A5}$$

Observe that if $c < 1$ then $q(k) = 0$ has no real roots and $q(k) > 0$ for all $k \in [0, 1]$, if $c = 1$ then $q(k) = 0$ has a unique root $k = 1$ and $q(k) \geq 0$ for all $k \in [0, 1]$ with equality only at $k = 1$, and if $c > 1$ then $q(k) = 0$ has two distinct real roots

$$k_1, k_2 \;:=\; c \pm \sqrt{c(c-1)}$$

with $k_1 < 1 < k_2$ such that if $c > 1$ then $q(k) > 0$ for all $k \in [0, k_1)$ and $q(k) < 0$ for all $k \in (k_1, 1]$.

Now let $\hat{k}(c)$ denote the largest admissible k such that $q(k) > 0$, that is

$$\hat{k}(c) := \max\{k \in \mathcal{K}(c) : q(k) > 0\} \tag{A6}$$

If $c \leq 1$, then the critical point is on the boundary $\hat{k}(c) = c$ and there is no $k \in \mathcal{K}(c)$ such that $k > \hat{k}(c)$ hence if $c \leq 1$ we conclude that $\delta(k)$ is increasing for all $k \in [0, c]$. If $c > 1$, then the critical point is in the interior $\hat{k}(c) = c - \sqrt{c(c-1)}$ and we conclude that $\delta(k)$ is increasing on the interval $[0, \hat{k}(c)]$ and decreasing on the interval $[\hat{k}(c), 1]$. 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 $\delta(1) = 0$. $\square$

A.2 Equilibrium results

Caveat on equilibrium results. As explained in Appendix B below, in the knife-edge case that $c = 1$ exactly there is no (linear) equilibrium 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 treat the knife-edge case $c = 1$ separately in Appendix B below.

Proof of Proposition 1.

An equilibrium is a pair k^*, δ^* simultaneously satisfying the receivers' $k(\delta)$ and the sender's $\delta(k)$. Plugging the expression for $\delta(k)$ from (31) into the sender's $k(\delta)$ from (28) and simplifying, we can write the equilibrium problem as finding $k^* \in \mathcal{K}(c)$ that satisfies

$$L(k) = R(k) \tag{A7}$$

where

$$L(k) := \frac{1}{\alpha}k, \qquad R(k) := c\frac{(c-k)(1-k)}{(c-k^2)^2} \tag{A8}$$

and recall that $k \in \mathcal{K}(c)$ implies $c - k^2 \geq 0$. Observe that $L(0) = 0$ and $L(\min(c, 1)) = \min(c, 1)/\alpha > 0$ and that $L'(k) = 1/\alpha > 0$. Observe also that $R(0) = c^2/c^2 = 1$ and that $R(\min(c, 1)) = 0$. Moreover

$$R'(k) = c\left(\frac{1}{c-k^2}\right)^3 P(k), \qquad P(k) := \left(2k^3 - 3k^2 - 3ck^2 + 6ck - c^2 - c\right) \tag{A9}$$

Hence the sign of $R'(k)$ is the same as the sign of the polynomial $P(k)$. Now let $\overline{P}(c)$ denote the largest possible value of this polynomial

$$\overline{P}(c) := \max_{k \in \mathcal{K}(c)} P(k) \geq P(k) \tag{A10}$$

Computing the maximum then gives

$$\overline{P}(c) = (2c - c^2 - 1)\max(c, 1) \leq 0 \tag{A11}$$

with equality only in the knife-edge case $c = 1$. Hence $P(k) \leq 0$ for all $k \in \mathcal{K}(c)$ and so we conclude $R'(k) \leq 0$ for all $k \in \mathcal{K}(c)$. Hence 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 response coefficient k^* such that $H(k^*) = 0$ or $L(k^*) = R(k^*)$. We can then recover the unique manipulation coefficient $\delta^* = \delta(k^*)$ from (31). Hence there is a unique pair k^*, δ^* simultaneously satisfying the receivers' $k(\delta)$ and the sender's $\delta(k)$. $\square$

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

$$\frac{\partial k^*}{\partial \alpha} = \left(\frac{1}{1-k'(\delta^*)\delta'(k^*)}\right)\frac{\partial k(\delta^*;\alpha)}{\partial \alpha} \tag{A12}$$

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 (28) that

$$\frac{\partial k(\delta;\alpha)}{\partial \alpha} = \frac{1-\delta}{((1-\delta)^2\alpha+1)^2} \in [0,1], \qquad \delta\in[0,1], \qquad \alpha>0 \tag{A13}$$

We also have formulas for $k'(\delta)$ in (A1) and for $\delta'(k)$ in (A3) above. We will now show that at equilibrium the product $k'(\delta^*)\delta'(k^*)$ is nonpositive. To do this, first use the definition of the receivers' best response $k(\delta)$ to write, after some rearrangements,

$$k'(\delta) = 2k(\delta)^2 - \frac{k(\delta)}{1-\delta} \tag{A14}$$

Evaluating this at the equilibrium k^*,δ^* and rearranging gives

$$k'(\delta^*) = \left(\frac{k^*}{c-k^*}\right)\left(2ck^* - k^{*\,2} - c\right) \tag{A15}$$

Notice that $k'(\delta^*)$ has the sign of $2ck^* - k^{*2} - c$ while $\delta'(k^*)$ has the sign of $k^{*2} - 2ck^* + c$. Indeed,

$$k'(\delta^*)\delta'(k^*) = -\left(\frac{k^*}{c-k^*}\right)\left(\frac{k^{*2}-2ck^*+c}{c-k^{*2}}\right)^2 \leq 0 \tag{A16}$$

Hence from (A12), (A13) and (A16) we can conclude that $k^*(\alpha,c)$ is strictly increasing in α. For part (ii) we use the sender's best response to calculate

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

From Lemma 3 we know that $\delta'(k) > 0$ if and only if $k < \hat{k}(c)$ where $\hat{k}(c)$ is the critical point defined in (32). Hence we can conclude

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

Now for any fixed $c>0$, the critical point $\hat{\alpha}(c)$ is constructed using the result from part (i) that $k^*(\alpha,c)$ is strictly increasing in α to find the smallest value of α such that $k^*(\alpha,c) \geq \hat{k}(c)$. If there is no such value, then we set $\hat{\alpha}(c) = +\infty$. If there is a finite value of α that achieves equality, $k^*(\alpha,c) = \hat{k}(c)$, then we take $\hat{\alpha}(c)$ to be that value. In particular, if $c<1$ then from (32) we have $\hat{k}(c) = c < 1$ and so for all finite α we have $k^*(\alpha,c) < \hat{k}(c)$ so that for $c<1$ we take $\hat{\alpha}(c) = +\infty$. Alternatively, if $c>1$ then from (32) we have $\hat{k}(c) \in (\frac{1}{2},1)$ and so there is a finite $\hat{\alpha}(c)$ such that $k^*(\hat{\alpha}(c),c) = \hat{k}(c)$. □

Proof of Lemma 5.

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

$$\frac{\partial \delta^*}{\partial c} = \left(\frac{1}{1-k'(\delta^*)\delta'(k^*)}\right)\frac{\partial \delta(k^*;c)}{\partial c} \tag{A19}$$

We already know from (A16) that $k'(\delta^*)\delta'(k^*) \leq 0$. And from (31) observe that

$$\frac{\partial \delta(k;c)}{\partial c} = -\frac{k-k^2}{(c-k^2)^2} < 0 \tag{A20}$$

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

$$\frac{\partial k^*}{\partial c} = k'(\delta^*)\frac{\partial \delta^*}{\partial c} \tag{A21}$$

From Lemma 2 we know that $k'(\delta) < 0$ if and only if $\delta > \hat{\delta}(\alpha)$ where $\hat{\delta}(\alpha)$ is the critical point defined in (29). Hence we can conclude

$$\frac{\partial k^*}{\partial c} > 0 \qquad \Leftrightarrow \qquad \delta^*(\alpha, c) > \hat{\delta}(\alpha) \tag{A22}$$

Now for any fixed $\alpha > 0$, the critical point $\hat{c}(\alpha)$ is constructed using the result from part (i) that $\delta^*(\alpha, c)$ is strictly decreasing in c to find the smallest value of c such that $\delta^*(\alpha, c) \leq \hat{\delta}(\alpha)$. If there is no such value, then we set $\hat{c}(\alpha) = +\infty$. If there is a finite value of c that achieves equality, $\delta^*(\alpha, c) = \hat{\delta}(\alpha)$, then we take $\hat{c}(\alpha)$ to be that value. In particular, if $\alpha < 1$ then from (29) we have $\hat{\delta}(\alpha) = 0$ and so for all finite c we have $\delta^*(\alpha, c) > \hat{\delta}(\alpha)$ so that for $\alpha < 1$ we take $\hat{c}(\alpha) = +\infty$. Alternatively, if $\alpha > 1$ then from (29) we have $\hat{\delta}(\alpha) \in (0, 1)$ and so there is a finite $\hat{c}(\alpha)$ such that $\delta^*(\alpha, \hat{c}(\alpha)) = \hat{\delta}(\alpha)$. □

Proof of Lemma 6.

Recall that $k^*_{ms}(\alpha) := \alpha/(\alpha + 1)$. We focus on the case that $\alpha > 1$. If $\alpha \leq 1$ then any $c < +\infty$ implies $\delta^*(\alpha, c) > 0$ and hence $k^*(\alpha, c) < k^*_{ms}(\alpha)$. With $\alpha > 1$ we find combinations of (α, c) that give $k^*(\alpha, c) = k^*_{ms}(\alpha)$. From the receivers' best response in Lemma 2 we can determine the amount of manipulation required to equate $k(\delta; \alpha)$ and $k^*_{ms}(\alpha)$, this works out to be

$$\delta^*(\alpha, c) = \frac{\alpha - 1}{\alpha}, \qquad \alpha > 1 \tag{A23}$$

Then since $\delta^*(\alpha, c) = \delta(k^*(\alpha, c); c)$ and at this boundary $k^*(\alpha, c) = k^*_{ms}(\alpha)$ this level of manipulation satisfies

$$\delta(k^*_{ms}(\alpha); c) = \frac{\alpha - 1}{\alpha} \tag{A24}$$

Plugging $k^*_{ms}(\alpha)$ into the formula (31) for the sender's best response $\delta(k; c)$ gives the condition

$$\frac{\dfrac{\alpha}{\alpha+1} - \left(\dfrac{\alpha}{\alpha+1}\right)^2}{c - \left(\dfrac{\alpha}{\alpha+1}\right)^2} = \frac{\alpha - 1}{\alpha} \tag{A25}$$

Solving this for c in terms of α we get

$$c = \frac{\alpha}{\alpha - 1}\left(\frac{\alpha}{\alpha+1}\right)^2 =: c^*_{ms}(\alpha) \tag{A26}$$

(with $c^*_{ms}(\alpha) = +\infty$ for $\alpha \leq 1$). Now we show that $k^*(\alpha, c) < k^*_{ms}(\alpha)$ if and only if $c < c^*_{ms}(\alpha)$. Let

$$\delta^*_{ms}(\alpha) := \delta^*(\alpha, c^*_{ms}(\alpha)) \tag{A27}$$

(with $\delta^*_{ms}(\alpha) = 0$ for $\alpha \leq 1$). Observe that

$$\delta^*_{ms}(\alpha) = \frac{\alpha - 1}{\alpha} > \hat{\delta}(\alpha) \tag{A28}$$

where $\hat{\delta}(\alpha)$ is the critical point defined in Lemma 2. Then the receivers' best response is decreasing in δ for any $\delta \geq \delta^*_{ms}(\alpha)$. Further observe that $k(\delta^*_{ms}(\alpha); \alpha) = k^*_{ms}(\alpha)$ so that $k^*(\alpha, c) < k^*_{ms}(\alpha)$ if and only if $\delta^*(\alpha, c) > \delta^*_{ms}(\alpha)$. From Lemma 5 we know that $\delta^*(\alpha, c)$ is strictly decreasing in c hence any $c < c^*_{ms}(\alpha)$ is equivalent to $\delta^*(\alpha, c) > \delta^*_{ms}(\alpha)$. □

A.3 Welfare results

Proof of Lemma 7.

For part (i), recall that in the absence of manipulation, the sender's equilibrium payoff is

$$v^*_{ms} = \frac{(1-\lambda)^2\alpha_x + \alpha_z}{((1-\lambda)\alpha_x + \alpha_z)^2} \tag{A29}$$

Differentiating with respect to α_x and collecting terms we can write

$$\frac{dv^*_{ms}}{d\alpha_x} = -\frac{(1-\lambda)^3\alpha_x + (1-\lambda)(1+\lambda)\alpha_z}{((1-\lambda)^2\alpha_x + \alpha_z)((1-\lambda)\alpha_x + \alpha_z)}\, v^*_{ms} < 0 \tag{A30}$$

Hence v^*_{ms} is strictly decreasing in α_x.

With manipulation, the sender's equilibrium payoff is $v^* = v(k^*;\alpha_x)$ where $v(k;\alpha_x)$ is the sender's value function, namely

$$v(k;\alpha_x) = \frac{1}{\alpha_z}(1-k)^2\left(\frac{c}{c-k^2}\right) + \frac{1}{\alpha_x}k^2 \tag{A31}$$

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

$$\frac{dv^*}{d\alpha_x} = v'(k^*)\frac{\partial k^*}{\partial \alpha_x} + \frac{\partial v(k^*;\alpha_x)}{\partial \alpha_x} \tag{A32}$$

where, in slight abuse of notation, $v'(k^*)$ denotes the derivative of the sender's value function evaluated at the equilibrium k^*. From Lemma 4 we have

$$\frac{\partial k^*}{\partial \alpha_x} > 0 \tag{A33}$$

(since $\alpha = (1-\lambda)\alpha_x/\alpha_z$, so the derivative of k^* with respect to α_x is proportional to the derivative with respect to α). The other derivatives we need are

$$\frac{\partial v(k^*;\alpha_x)}{\partial \alpha_x} = -\left(\frac{k^*}{\alpha_x}\right)^2 < 0 \tag{A34}$$

and

$$v'(k^*) = 2\left(\frac{k^*}{\alpha_x} - \frac{R(k^*)}{\alpha_z}\right) = -2\frac{\lambda}{1-\lambda}\left(\frac{k^*}{\alpha_x}\right) < 0 \tag{A35}$$

where $R(k)$ is given in (A8) above and we use the equilibrium condition from (A8) that $\frac{1}{1-\lambda}\frac{\alpha_z}{\alpha_x}k^* = R(k^*)$. Hence from (A33), (A34) and (A34) we can conclude that v^* is also strictly decreasing in α_x.

For part (ii) we can directly compute the limits of v^*_{ms} from (A29), namely

$$\lim_{\alpha_x \to 0^+} v^*_{ms} = \frac{\alpha_z}{\alpha_z^2} = \frac{1}{\alpha_z} \tag{A36}$$

And dividing numerator and denominator of (A29) by $\alpha_x^2 > 0$ we have

$$\lim_{\alpha_x \to \infty} v^*_{ms} = \lim_{\alpha_x \to +\infty} \frac{(1-\lambda)^2\frac{1}{\alpha_x} + \frac{\alpha_z}{\alpha_x^2}}{(1-\lambda)^2 + 2(1-\lambda)\frac{\alpha_z}{\alpha_x} + \frac{\alpha_z^2}{\alpha_x^2}} = 0 \tag{A37}$$

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

$$\lim_{\alpha_x \to 0^+} v^* = \frac{1}{\alpha_z}(1-0)^2\left(\frac{c}{c-0^2}\right) + \lim_{\alpha_x \to 0^+}\frac{k^{*2}}{\alpha_x} = \frac{1}{\alpha_z} + 0 = \frac{1}{\alpha_z} \tag{A38}$$

where we have used L'Hôpital's rule and (A12) and (A13) to calculate that

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

where the limit follows because $\delta^*\in[0,1]$ for all α_x and $k^*\to 0$ and hence from (A16) $k'(\delta^*)\delta'(k^*)\to 0$ 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

$$\lim_{\alpha_x\to\infty}v^*=\begin{cases}\frac{1}{\alpha_z}(1-c)^2\frac{c}{c-c^2} & =\frac{1-c}{\alpha_z} \qquad \text{if } c<1\\ \frac{1}{\alpha_z}(1-1)^2\frac{c}{c-1} & =0 \qquad \text{if } c>1\end{cases}\tag{A39}$$

□

Proof of Proposition 2.

Recall that the sender's value function with manipulation is

$$v(k):=\max_{\delta\in[0,1]}V(\delta,k)=\frac{1}{\alpha_z}(1-k)^2\left(\frac{c}{c-k^2}\right)+\frac{1}{\alpha_x}k^2\tag{A40}$$

whereas the sender's value function without manipulation is

$$v_{ms}(k):=V(0,k)=\frac{1}{\alpha_z}(1-k)^2+\frac{1}{\alpha_x}k^2\leq v(k)\tag{A41}$$

We decompose the sender's gain from manipulation as

$$v^*-v^*_{ms}=(v(k^*)-v_{ms}(k^*))+(v_{ms}(k^*)-v_{ms}(k_{ms}))\tag{A42}$$

For part (i), first observe that $v(k^*)\geq v_{ms}(k^*)$ for any k^* hence for $v^*>v^*_{ms}$ it suffices that $v_{ms}(k^*)>v_{ms}(k_{ms})$. Then observe that $v'_{ms}(k)<0$ for all k in the interval $(0,\frac{\alpha_x}{\alpha_x+\alpha_z})$. Hence if $k^*\in(0,k^*_{ms})$ then $v_{ms}(k^*)>v_{ms}(k_{ms})$ and we conclude that $v^*>v^*_{ms}$. Then from Lemma 6 we know that a necessary and sufficient condition for $k^*<k^*_{ms}$ is that the cost of manipulation be $c<c^*_{ms}(\alpha)$ where $c^*_{ms}(\alpha)$ is the critical cost given in (43). This parameter configuration is sufficient for $v^*>v^*_{ms}$ but is not necessary. We now show that if $\lambda<\frac{1}{2}$ then $v^*>v^*_{ms}$ even if $k^*>k^*_{ms}$. To see this, observe from (A8) above that in equilibrium

$$\frac{1}{1-\lambda}\frac{\alpha_z}{\alpha_x}k^*=c\frac{(c-k^*)(k^*-1)}{(c-k^{*2})^2}\tag{A43}$$

With can use this to write the sender's equilibrium payoff as

$$v^*=\frac{1}{(1-\lambda)\alpha_x}\left\{k^*-\lambda k^{*2}+\frac{k^{*2}(1-k^*)^2}{c-k^*}\right\}\tag{A44}$$

Likewise the benchmark coefficient k^*_{ms} satisfies the analogous condition

$$\frac{1}{1-\lambda}\frac{\alpha_z}{\alpha_x}k^*_{ms}=1-k^*_{ms}\tag{A45}$$

and we can write the sender's equilibrium payoff absent manipulation as

$$v^*_{ms}=\frac{1}{(1-\lambda)\alpha_x}\left\{k^*_{ms}-\lambda k^{*2}_{ms}\right\}\tag{A46}$$

Hence the sender gains from manipulation, $v^*>v^*_{ms}$ if and only if

$$f(k^*)+g(k^*)>f(k^*_{ms})\tag{A47}$$

where

$$f(k) := k - \lambda k^2, \qquad g(k) := \frac{k^2(1-k)^2}{c-k} \geq 0 \tag{A48}$$

Since $g(k) \geq 0$, a sufficient condition for $v^* > v^*_{ms}$ is that $f(k^*) > f(k^*_{ms})$. Now observe that $f(k)$ is strictly concave and is strictly increasing in k on $[0, \frac{1}{2\lambda})$ and hence if $\lambda < \frac{1}{2}$ we have that $f(k)$ is strictly increasing for all $k \in [0,1]$ and hence even if $k^* > k^*_{ms}$ we can conclude that $\lambda < \frac{1}{2}$ is sufficient to guarantee $v^* > v^*_{ms}$.

For part (ii), simply observe from Lemma 7 that as $\alpha_x \to \infty$ we have $v^*_{ms} \to 0^+$ while $v^* \to (1-c)/\alpha_z$ if $c < 1$ but $v^* \to 0^+$ if $c > 1$. □

Proof of Proposition 3.

From the proof of Proposition 2, the sender's manipulation backfires, $v^* < v^*_{ms}$, if

$$g(k^*) < f(k^*_{ms}) - f(k^*) \tag{A49}$$

where again

$$f(k) := k - \lambda k^2, \qquad g(k) := \frac{k^2(1-k)^2}{c-k} \geq 0 \tag{A50}$$

and we know that necessary conditions for this are that $c > c^*_{ms}(\alpha)$ so that $k^* > k^*_{ms}$ and that $\lambda > \frac{1}{2}$. Supposing $k^* > k^*_{ms}$, we can rewrite the inequality in (A49) as

$$\frac{k^{*2}(1-k^*)^2}{k^* - k^*_{ms}} < (\lambda(k^*_{ms} + k^*) - 1))(c - k^*) \tag{A51}$$

and moreover from the fixed point conditions (A43) and (A45) that characterize k^* and k^*_{ms} we can write

$$\frac{1-k^*_{ms}}{1-k^*} = \frac{k^*_{ms}}{k^*}\left(c\frac{(c-k^*)}{(c-k^{*2})^2}\right) \tag{A52}$$

so that (A51) is equivalent to the inequality

$$\frac{k^{*2}(1-k^*)}{\frac{k^*_{ms}}{k^*}\left(c\frac{(c-k^*)}{(c-k^{*2})^2}\right) - 1} < (\lambda(k^*_{ms} + k^*) - 1))(c - k^*) \tag{A53}$$

We will now show that if in addition to $c > c^*_{ms}(\alpha)$ and $\lambda > \frac{1}{2}$ we also have $c > 1$, then, for α_x large enough, the LHS of (A53) will be strictly less than the RHS of (A53). In particular, if $c > 1$ and $\lambda > \frac{1}{2}$ then as $\alpha_x \to \infty$ the RHS of (A53) converges to a strictly positive constant

$$\lim_{\alpha_x \to \infty} (\lambda(k^*_{ms} + k^*) - 1))(c - k^*) = (\lambda 2 - 1)(c - 1) > 0 \tag{A54}$$

(since $k^* \to 1$ if $c > 1$). But the LHS of (A53) converges to zero

$$\lim_{\alpha_x \to \infty} \frac{k^{*2}(1-k^*)}{\frac{k^*_{ms}}{k^*}c\frac{c-k^*}{(c-k^{*2})^2} - 1} = \frac{0^+}{\frac{c}{(c-1)} - 1} = 0^+ \tag{A55}$$

so that if $c > 1$ and $\lambda > \frac{1}{2}$ then there exists α^*_x such that for $\alpha_x > \alpha^*_x$ the LHS of (A53) is strictly less than the RHS of (A53) so that the sender's manipulation backfires, $v^* < v^*_{ms}$.

Finally, observe that this argument requires $c > 1$ in addition to $c > c^*_{ms}(\alpha)$ and $\lambda > \frac{1}{2}$. This means we need α_x large enough to ensure that $c > 1$ also implies $1 > c^*_{ms}(\alpha)$ where α and α_x are linked by $\alpha = (1-\lambda)\alpha_x/\alpha_z$. In particular, from (43) we have $1 > c^*_{ms}(\alpha)$ if $\alpha > 1$ and if

$$\alpha^2 - \alpha - 1 > 0 \tag{A56}$$

The roots of the quadratic on the LHS are

$$\alpha_1, \alpha_2 = \frac{1}{2} \pm \frac{1}{2}\sqrt{5} \tag{A57}$$

with $\alpha_1 < 1 < \alpha_2$. So we need $\alpha > \alpha_2 = (1+\sqrt{5})/2$ to ensure that $1 > c^*_{ms}(\alpha)$ so that $c > 1$ is sufficient for $k^* > k^*_{ms}$. Since $\alpha = (1-\lambda)\alpha_x/\alpha_z$, the critical point α^*_x must be at least

$$\alpha^*_x > \left(\frac{1+\sqrt{5}}{2}\right)\left(\frac{\alpha_z}{1-\lambda}\right) \tag{A58}$$

□

Proof of Lemma 8.

For part (i), observe that in the absence of manipulation, the receivers' equilibrium loss is

$$l^*_{ms} = \frac{(1-\lambda)}{(1-\lambda)\alpha_x + \alpha_z} \tag{A59}$$

Hence in the absence of manipulation, the receivers' loss $l^*_{ms}(\alpha_x)$ is strictly decreasing in α_x

$$\frac{dl^*_{ms}}{d\alpha_x} = -\frac{(1-\lambda)^2}{((1-\lambda)\alpha_x + \alpha_z)^2} < 0 \tag{A60}$$

For part (ii), observe that with manipulation, the receivers' equilibrium loss is $l^* = l(\delta^*(\alpha_x); \alpha_x)$ where $l(\delta; \alpha_x)$ is the receivers' expected loss function, namely

$$l(\delta; \alpha_x) = \frac{(1-\lambda)}{(1-\delta)^2(1-\lambda)\alpha_x + \alpha_z} \tag{A61}$$

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

$$\frac{dl^*}{d\alpha_x} = l'(\delta^*)\frac{d\delta^*}{d\alpha_x} + \frac{\partial l(\delta^*; \alpha_x)}{\partial \alpha_x} \tag{A62}$$

We will now show that if $c > 1$ the receivers' loss with manipulation l^* is strictly decreasing in α_x but for each $c < 1$ there exists a critical point α^{**}_x such that l^* is strictly decreasing for all $\alpha_x < \alpha^{**}_x$ and strictly increasing for all $\alpha_x > \alpha^{**}_x$. Calculating the derivatives in (A62) and rearranging, we obtain

$$\frac{dl^*}{d\alpha_x} > 0 \qquad \Leftrightarrow \qquad (1-\delta^*) - 2\alpha_x\frac{d\delta^*}{d\alpha_x} < 0 \tag{A63}$$

Equivalently, if and only if

$$\frac{d\delta^*}{d\alpha_x} > \frac{1}{2\alpha_x}(1-\delta^*) > 0 \tag{A64}$$

Now recall that in equilibrium the sender's manipulation depends on α_x only via the receivers' response coefficient, $\delta^*(\alpha_x) = \delta(k^*(\alpha_x))$, so that

$$\frac{d\delta^*}{d\alpha_x} = \delta'(k^*)\frac{dk^*}{d\alpha_x} \tag{A65}$$

So we can write condition (A64) as

$$\delta'(k^*)\frac{dk^*}{d\alpha_x} > \frac{1}{2\alpha_x}(1-\delta^*) > 0 \tag{A66}$$

Applying the implicit function theorem to the equilibrium condition (A8) above we have

$$\frac{dk^*}{d\alpha_x} = \frac{\frac{\alpha_z}{(1-\lambda)\alpha_x}\frac{k^*}{\alpha_x}}{\frac{\alpha_z}{(1-\lambda)\alpha_x} - R'(k^*)} > 0 \tag{A67}$$

where $R(k)$ is defined in (A8). Plugging this into (A66) and simplifying we get the equivalent condition

$$\frac{\alpha_z}{(1-\lambda)\alpha_x}\left(\delta'(k^*)k^* - \frac{1}{2}(1-\delta^*)\right) > -\frac{1}{2}(1-\delta^*)R'(k^*) \tag{A68}$$

Now observe from (A3) above that

$$\delta'(k)k - \frac{1}{2}(1-\delta) = \frac{1}{2}\left(\frac{1}{c-k^2}\right)^2\left(k^3 - 3ck^2 + 3ck - c^2\right) \tag{A69}$$

and that using the formula for $R'(k)$ given in (A9) above we can calculate that

$$\frac{1}{2}(1-\delta)R'(k) = \frac{1}{2}\left(\frac{1}{c-k^2}\right)^2 R(k)\frac{1}{1-k}P(k) \tag{A70}$$

where $P(k)$ is also defined in (A9) above. Plugging these calculations back into (A68) gives

$$\frac{\alpha_z}{(1-\lambda)\alpha_x}\left(\frac{1}{2}\left(\frac{1}{c-k^{*2}}\right)^2\left(k^{*3} - 3ck^{*2} + 3ck^* - c^2\right)\right) > -\frac{1}{2}\left(\frac{1}{c-k^{*2}}\right)^2 R(k^*)\frac{1}{1-k^*}P(k^*) \tag{A71}$$

Cancelling common terms gives the condition

$$\frac{\alpha_z}{(1-\lambda)\alpha_x}\left(k^{*3} - 3ck^{*2} + 3ck^* - c^2\right) > -R(k^*)\frac{1}{1-k^*}P(k^*) \tag{A72}$$

Using the equilibrium condition $L(k^*) = R(k^*)$ from (A8) and $\alpha = (1-\lambda)\alpha_x/\alpha_z$ gives

$$\frac{\alpha_z}{(1-\lambda)\alpha_x}\left(k^{*3} - 3ck^{*2} + 3ck^* - c^2\right) > -\frac{\alpha_z}{(1-\lambda)\alpha_x}\frac{k^*}{1-k^*}P(k^*) \tag{A73}$$

Using the definition of $P(k)$ and cancelling more common terms gives the condition

$$k^{*4} - 2k^{*3} + 2ck^* - c^2 > 0 \tag{A74}$$

To summarize, we have shown that

$$\frac{dl^*}{d\alpha_x} > 0 \qquad \Leftrightarrow \qquad F(k^*) := k^{*4} - 2k^{*3} + 2ck^* - c^2 > 0 \tag{A75}$$

We will now show that if $c > 1$ then it cannot be the case that $F(k^*) > 0$ but that for $c < 1$ it will be the case that $F(k^*) > 0$ for sufficiently large α_x.

To see this, write $F(k) = J(k;c) - G(k)$ where $J(k;c) := 2ck - c^2$ and $G(k) := 2k^3 - k^4$. Observe that $G(0) = 0$, $G(1) = 1$, $G(k) < k$ for all k; $G'(k) = 2k^2(3-2k) \geq 0$ with $G'(0) = 0$ and $G'(1) = 2$; and $G''(k) = 12k(1-k) \geq 0$ so that $G'(k) \leq G'(1) = 2$ for all k. Further observe that $J(0;c) = -c^2 < 0$, $J(1;c) = 2c - c^2 \leq 1$ (with equality if $c = 1$) and $J'(k;c) = 2c > 0$ for all k so that $J(k;c) \leq J(1;c) = 2c - c^2 \leq 1$ for all k, c. These imply $F(0) = J(0;c) - G(0) = -c^2 < 0$ and $F(1) = J(1;c) - G(1) = 2c - c^2 - 1 \leq 0$ (with equality if $c = 1$); $F'(k) = J'(k;c) - G'(k) = 2c - G'(k)$ and $F''(k) = -G''(k) \leq 0$. Since $G'(k) \leq 2$ we have

$$F'(k) = J'(k;c) - G'(k) = 2c - G'(k) \geq 2c - 2 = 2(c-1) \tag{A76}$$

Now consider $c > 1$. Then $F'(k) \geq 2(c-1) > 0$ so $F(k)$ is strictly increasing from $F(0) = -c^2 < 0$ to $F(1) = 2c - c^2 - 1 < 0$ so that $F(k) < 0$ for all k and hence it is not possible for $F(k^*) > 0$ and hence for $c > 1$ the receivers' loss is unambiguously decreasing in α_x.

Now consider $c < 1$. Then since $G'(k)$ is monotone increasing from $G'(0) = 0$ to $G'(1) = 2$ there is a unique critical point $\tilde{k}$ such that

$$F'(\tilde{k}) = 0 \qquad \Leftrightarrow \qquad 2c = G'(\tilde{k}) \tag{A77}$$

Since $F''(k) \leq 0$, this critical point maximizes $F(k)$ hence

$$F(k) \leq \max_{k\in[0,1]} F(k) = F(\tilde{k}) \tag{A78}$$

and observe that if we take $k = c < 1$ (which is feasible since here $c < 1$) then we have

$$F(c) = J(c;c) - G(c) = 2c^2 - c^2 - G(c) = c^2 - 2c^3 + c^4 = c^2(1 - 2c + c^2) > 0 \tag{A79}$$

so that indeed

$$F(\tilde{k}) \geq F(c) > 0 \tag{A80}$$

Hence for $c < 1$ there exist k such that $F(k) > 0$. More precisely, the function $F(k)$ increases from $F(0) = -c^2 < 0$ to a lower cutoff $\underline{k} \in (0, \tilde{k})$ defined by $F(\underline{k}) = 0$. The function $F(k)$ keeps increasing until it reaches the critical point $\tilde{k}$ at which $F'(\tilde{k}) = 0$ and $F(\tilde{k}) > 0$. From there $F(k)$ decreases, crossing zero again at a higher cutoff $\overline{k} \in (\tilde{k}, 1)$ defined by $F(\overline{k}) = 0$ and keeps decreasing until $F(1) = 2c - c^2 - 1 < 0$ (since $c < 1$).

So for $c < 1$ 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. For $c < 1$ these critical points are defined by the roots of $F(k; c) = 0$. Observe that since $F(c) > 0$ yet $\underline{k}$ is the first k for which $F(k) = 0$ it must be the case that $\underline{k} < c$. Likewise since $F(\overline{k}) = 0$ it must also be the case that $\overline{k} > c$. In short, 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

$$k^*(\alpha_x^{**}, c) = \underline{k} \tag{A81}$$

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) we can directly compute the limits of l^*_{ms} from (A59), these are:

$$\lim_{\alpha_x \to 0^+} l^*_{ms} = \frac{1-\lambda}{\alpha_z} \tag{A82}$$

and

$$\lim_{\alpha_x \to \infty} l^*_{ms} = 0 \tag{A83}$$

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

$$\lim_{\alpha_x \to 0^+} l^* = \frac{(1-\lambda)}{(1-0)^2(1-\lambda)0 + \alpha_z} = \frac{1-\lambda}{\alpha_z} \tag{A84}$$

At the other extreme, as $\alpha_x \to \infty$ we have $\delta^* \to 1$ if $c < 1$ but $\delta^* \to 0$ if $c > 1$. If $c < 1$ then in the limit as $\alpha_x \to \infty$ we have

$$\lim_{\alpha_x \to \infty} l^* = \lim_{\alpha_x \to \infty} \frac{(1-\lambda)}{(1-\delta^*)^2(1-\lambda)\alpha_x + \alpha_z} = \frac{1-\lambda}{\alpha_z} \tag{A85}$$

where we have used L'Hôpital's rule and (A65) and (A67) to calculate that

$$\begin{aligned}
\lim_{\alpha_x \to \infty} (1-\delta^*)^2 \alpha_x = \lim_{\alpha_x \to \infty} 2(1-\delta^*) \frac{d\delta^*}{d\alpha_x} \alpha_x^2 &= \lim_{\alpha_x \to \infty} 2(1-\delta^*) \delta'(k^*) \frac{dk^*}{d\alpha_x} \alpha_x^2 \\
&= \lim_{\alpha_x \to \infty} 2(1-\delta^*) \delta'(k^*) \left(\frac{\frac{\alpha_z}{(1-\lambda)} k^*}{\frac{\alpha_z}{(1-\lambda)\alpha_x} - R'(k^*)} \frac{1}{\alpha_x^2} \right) \alpha_x^2 \\
&= 2(1-1) \frac{1}{c-c^2} \left(\frac{\frac{\alpha_z}{(1-\lambda)} c}{0 + \frac{1}{c-c^2}} \right) \\
&= 0
\end{aligned}$$

(since if $c < 1$ we have $\delta^* \to 1$ and $k^* \to c$ and from (A3) and (A9) 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 from (A61) we simply have

$$\lim_{\alpha_x \to \infty} l^* = \lim_{\alpha_x \to \infty} \frac{(1-\lambda)}{(1-\delta^*)^2(1-\lambda)\alpha_x + \alpha_z} = 0 \tag{A86}$$

Putting (A85) and (A86) together we finally have

$$\lim_{\alpha_x \to \infty} l^* = \begin{cases} \dfrac{1-\lambda}{\alpha_z} & \text{if } c < 1 \\ 0 & \text{if } c > 1 \end{cases} \tag{A87}$$

□

Proof of Proposition 4.

Recall that the receivers' expected loss function is

$$l(\delta) = \frac{(1-\lambda)}{(1-\delta)^2(1-\lambda)\alpha_x + \alpha_z} \tag{A88}$$

And that $l^* = l(\delta^*)$ and that $l^*_{ms} = l(0)$. Since

$$l'(\delta) = \frac{2(1-\delta)(1-\lambda)^2\alpha_x}{((1-\delta)^2(1-\lambda)\alpha_x + \alpha_z)^2} > 0 \tag{A89}$$

we have that $l^* \geq l^*_{ms}$, strictly if $\delta^* > 0$. □

Proof of Proposition 5.

For part (i), since l^*_{ms} is strictly decreasing in α_x the excess loss $l^* - l^*_{ms}$ is surely increasing in α_x if l^* is. But by Lemma 8 we know that if $c < 1$ then l^* is strictly increasing in α_x for all $\alpha_x > \alpha_x^{**}$ where α_x^{**} is the critical point implicitly defined in (A81) above. Hence $\alpha_x > \alpha_x^{**}$ is also sufficient for the excess loss $l^* - l^*_{ms}$ to be strictly increasing in α_x. For part (ii), simply observe from Lemma 8 that as $\alpha_x \to \infty$ we have $l^*_{ms} \to 0^+$ while $l^* \to (1-\lambda)/\alpha_z$ if $c < 1$ but $l^* \to 0^+$ if $c > 1$. □

B Discontinuity at $c = 1$ and jumps in manipulation

Preliminaries. There is no issue with $c = 1$ if the relative precision $\alpha \leq 4$. The issues with $c = 1$ arise only if $\alpha > 4$. To see this, first recall from Lemma 2 that if $\alpha > 1$ the receivers' best response $k(\delta; \alpha)$ is increasing in δ on the interval $[0, \hat{\delta}(\alpha)]$ and obtains its maximum at $\delta = \hat{\delta}(\alpha) = 1 - 1/\sqrt{\alpha} \in (0, 1)$. At the maximum, the receivers' best response takes on the value $k(\hat{\delta}(\alpha); \alpha) = \sqrt{\alpha}/2$. Hence for $\alpha > 4$ the maximum value exceeds 1. Moreover, by continuity of the best response in δ if $\alpha > 4$ there is an interval of δ such that $k(\delta; \alpha) > 1$. The boundaries of this interval $(\underline{\delta}(\alpha), \overline{\delta}(\alpha))$ are given by the roots of $k(\delta; \alpha) = 1$, which work out to be

$$\underline{\delta}(\alpha)\,,\; \overline{\delta}(\alpha) = \frac{1}{2}\left(1 \pm \sqrt{1 - (4/\alpha)}\right), \qquad \alpha \geq 4 \tag{B1}$$

Observe that this interval is symmetric and centred on $\frac{1}{2}$ with a width of

$$\overline{\delta}(\alpha) - \underline{\delta}(\alpha) = \sqrt{1 - (4/\alpha)} \geq 0, \qquad \alpha \geq 4 \tag{B2}$$

If $\alpha = 4$, we have $\underline{\delta}(4) = \overline{\delta}(4) = \frac{1}{2}$ but as α increases the width of the interval $(\underline{\delta}(\alpha), \overline{\delta}(\alpha))$ expands around $\frac{1}{2}$ with the boundaries $\underline{\delta}(\alpha) \to 0^+$ and $\overline{\delta}(\alpha) \to 1^-$ as $\alpha \to \infty$. Now recall from Lemma 1 that only $k \in [0, 1]$ and $\delta \in [0, 1]$ are candidates for an equilibrium. So if $\alpha > 4$ then none of the values of $\delta \in (\underline{\delta}(\alpha), \overline{\delta}(\alpha))$ are candidates for an equilibrium.

Cost of manipulation, $c > 0$. Now consider the sender's best response $\delta(k; c)$ parameterized by $c > 0$ and suppose $\alpha > 4$. As illustrated in Figure 6, if $c < 1$ the sender's best response $\delta(k; c)$ is strictly increasing in k and hence the equilibrium point k^*, δ^* must be on the "upper branch" where $\delta^* > \overline{\delta}(\alpha)$. But for the same value of α and instead $c > 1$ the equilibrium point k^*, δ^* must be on the "lower branch" where $\delta^* < \underline{\delta}(\alpha)$.

Knife-edge case, $c = 1$. Now consider the case $c = 1$ exactly. The sender's best response is then $\delta(k; 1) = (k - k^2)/(1 - k^2) = k/(1 + k)$ which takes on the value $\delta(1; 1) = \frac{1}{2}$ at $k = 1$. This value $\delta(1; 1) = \frac{1}{2}$ is in $(\underline{\delta}(\alpha), \overline{\delta}(\alpha))$ so that if $\alpha > 4$ there is no (linear) equilibrium. This case is negligible in the sense that for c arbitrarily close to 1 there is a unique (linear) equilibrium regardless of α, it is only the case $c = 1$ that is troublesome. In particular, if $\alpha > 4$ and we take $c \to 1^-$ the unique equilibrium point k^*, δ^* is where $\delta(k; 1^-)$ intersects $k(\delta; \alpha)$ on the "upper branch" where $\delta^* > \overline{\delta}(\alpha)$ but if we take $c \to 1^+$ the unique equilibrium point k^*, δ^* is where $\delta(k; 1^+)$ intersects $k(\delta; \alpha)$ on the "lower branch" where $\delta^* < \underline{\delta}(\alpha)$.

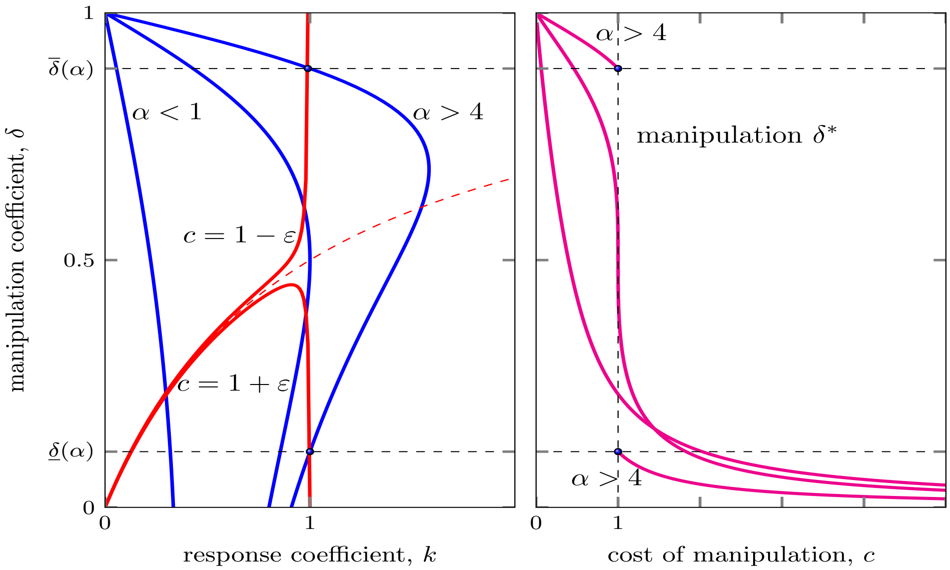

Figure 6: Discontinuity at $c = 1$ and jump in the amount of manipulation δ^*

The left panel shows the receivers' best response $k(\delta;\alpha)$ for $\alpha < 1$, $\alpha = 4$ and $\alpha > 4$ (blue) and the sender's best response $\delta(k;c)$ for $c = 1-\varepsilon$, $c = 1$, and $c = 1+\varepsilon$ (red). For $\alpha > 4$ and $c = 1$ there is no (linear) equilibrium. For $\alpha > 4$, in the limit as $c \to 1^-$ the equilibrium is at $k^* = 1, \delta^* = \overline{\delta}(\alpha)$ but in the limit as $c \to 1^+$ the equilibrium is at $k^* = 1, \delta^* = \underline{\delta}(\alpha)$. The right panel shows the equilibrium manipulation δ^* as a function of c for $\alpha < 1$, $\alpha = 4$ and $\alpha > 4$. 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 possible amount, from $\delta^* = 0$ if $c < 1$ to $\delta^* = 1$ if $c > 1$.

Jumps in the amount of manipulation. This lack of existence at $c = 1$ means that, as a function of c, the equilibrium amount of manipulation δ^* is discontinuous at $c = 1$. If $\alpha > 4$, an ε-reduction from $c = 1+\varepsilon$ to $c = 1-\varepsilon$ causes the sender's equilibrium manipulation δ^* to jump up from

$$\lim_{c \to 1^+} \delta^* = \underline{\delta}(\alpha) = \frac{1}{2}\big(1 - \sqrt{1-(4/\alpha)}\big) \tag{B3}$$

to

$$\lim_{c \to 1^-} \delta^* = \overline{\delta}(\alpha) = \frac{1}{2}\big(1 + \sqrt{1-(4/\alpha)}\big) \tag{B4}$$

The jump in manipulation as c is perturbed from $c = 1+\varepsilon$ to $c = 1-\varepsilon$ is present for any $\alpha > 4$ but the size of the jump is large when α is large. As $\alpha \to \infty$, the manipulation δ^* jumps all the way from $\delta^* = \underline{\delta}(\infty) = 0^+$ to $\delta^* = \overline{\delta}(\infty) = 1^-$ as c is perturbed from $c = 1+\varepsilon$ to $c = 1-\varepsilon$. In this sense, a tiny change in costs of information manipulation can give rise to a very large change in the amount of manipulation being conducted. As discussed in the main text, Section 5.2 above, this jump in the amount of manipulation is inherited by the receivers' loss l^* which likewise jumps as c is perturbed from $c = 1+\varepsilon$ to $c = 1-\varepsilon$.

Not all the equilibrium outcomes inherit this discontinuity, however. For example, the receivers' response k^* *is* continuous in c in that $k^*(\alpha, c^-) = k^*(\alpha, c^+)$. Instead k^* has a *kink*, i.e., is *not differentiable*, at $c = 1$. This is easily seen in the special case that $\alpha \to \infty$ in which case $\delta^* = 1$ for $c < 1$ and $\delta^* = 0$ for $c > 1$. In this case, $k^* = \min(c, 1)$ which is continuous in c but not differentiable at $c = 1$. The continuity of k^* is inherited by the sender's payoff v^* which is likewise continuous in c.

What is the economic meaning of $c = 1$? So given that the equilibrium manipulation can be extremely sensitive to c in the vicinity of $c = 1$, what does $c = 1$ mean? Recall that in the sender's objective (4) the gross benefit $\int_0^1 (a_i - \theta)^2 \, di$ has a coefficient normalized to 1. If instead we had written the objective with $b \int_0^1 (a_i - \theta)^2 \, di$ for some $b > 0$ then throughout the analysis the relevant parameter would be the cost/benefit ratio c/b and the point of special significance would be where the cost/benefit ratio is $c/b = 1$. In this parameterization, the sender's equilibrium manipulation is extremely sensitive to changes in either c or b in the vicinity of $c/b = 1$. With α high and costs and benefits *evenly poised*, a small increase in b or small decrease in c would lead to a very large increase in manipulation.

C Coefficients sum to one

In this Appendix we show that writing the receivers' linear strategy as $a(x_i, z) = kx_i + (1-k)z$ is without loss of generality. To see this, suppose that the receivers' linear strategy is

$$a_i = \beta_0 + \beta_1 x_i + \beta_2 z$$

for some coefficients $\beta_0, \beta_1, \beta_2$. We will show that in any linear equilibrium $\beta_0 = 0$ and $\beta_1 + \beta_2 = 1$. With this strategy, the aggregate A is

$$A = \beta_0 + \beta_1 y + \beta_2 z$$

The sender's problem is then to choose y to maximize

$$\int_0^1 (a_i - \theta)^2 \, di - c(y-\theta)^2 = (\beta_0 + \beta_1 x_i + \beta_2 z - \theta)^2 + \frac{1}{\alpha_x}\beta_1^2 - c(y-\theta)^2$$

The solution to this problem is

$$y = \gamma_0 + \gamma_1 \theta + \gamma_2 z$$

where

$$\gamma_0 = \frac{\beta_0 \beta_1}{c - \beta_1^2} \tag{C1}$$

$$\gamma_1 = \frac{c - \beta_1}{c - \beta_1^2} \tag{C2}$$

$$\gamma_2 = \frac{\beta_1 \beta_2}{c - \beta_1^2} \tag{C3}$$

But if the sender has the strategy $y = \gamma_0 + \gamma_1 \theta + \gamma_2 z$, the receivers' posterior expectation of θ is

$$\begin{aligned} \mathbb{E}[\theta \,|\, x_i, z] &= \frac{\gamma_1 \alpha_x}{\gamma_1^2 \alpha_x + \alpha_z} \left(\frac{1}{\gamma_1}(x_i - \gamma_2 z) - \frac{\gamma_0}{\gamma_1} \right) + \frac{\alpha_z}{\gamma_1^2 \alpha_x + \alpha_z} z \\ &= \frac{\gamma_1 \alpha_x}{\gamma_1^2 \alpha_x + \alpha_z} x_i + \frac{\alpha_z - \gamma_1 \alpha_x \gamma_2}{\gamma_1^2 \alpha_x + \alpha_z} z - \frac{\gamma_1 \alpha_x}{\gamma_1^2 \alpha_x + \alpha_z} \gamma_0 \end{aligned}$$

And the equilibrium strategy of an individual receiver then satisfies

$$\begin{aligned} a_i &= \lambda \mathbb{E}[A \,|\, x_i, z] + (1-\lambda)\mathbb{E}[\theta \,|\, x_i, z] \\ &= \lambda \beta_1 \mathbb{E}[y \,|\, x_i, z] + (1-\lambda)\mathbb{E}[\theta \,|\, x_i, z] + \lambda \beta_2 z + \lambda \beta_0 \\ &= (\lambda \beta_1 \gamma_1 + (1-\lambda))\mathbb{E}[\theta \,|\, x_i, z] + \lambda(\beta_1 \gamma_2 + \beta_2) z + \lambda(\beta_1 \gamma_0 + \beta_0) \end{aligned}$$

Matching coefficients with $a_i = \beta_0 + \beta_1 x_i + \beta_2 z$ we then have

$$\beta_0 = -(\lambda \beta_1 \gamma_1 + (1-\lambda)) \frac{\gamma_1 \alpha_x}{\gamma_1^2 \alpha_x + \alpha_z} \gamma_0 + \lambda(\beta_1 \gamma_0 + \beta_0) \tag{C4}$$

$$\beta_1 = \quad (\lambda \beta_1 \gamma_1 + (1-\lambda)) \frac{\gamma_1 \alpha_x}{\gamma_1^2 \alpha_x + \alpha_z} \tag{C5}$$

$$\beta_2 = \quad (\lambda \beta_1 \gamma_1 + (1-\lambda)) \frac{\alpha_z - \gamma_1 \alpha_x \gamma_2}{\gamma_1^2 \alpha_x + \alpha_z} + \lambda(\beta_1 \gamma_1 + \beta_2) \tag{C6}$$

Now observe that equations (C1) and (C4) together imply that the intercepts are $\beta_0 = \gamma_0 = 0$. Now observe from (C2)-(C3) and (C5)-(C6) that $\gamma_1 + \gamma_2 = 1$ implies $\beta_1 + \beta_2 = 1$ and vice-versa. So in one equilibrium the receivers' strategy takes the form $a_i = kx_i + (1-k)z$ where $k = \beta_1$ and the sender's strategy takes the form $y = (1-\delta)\theta + \delta z$ where $\delta = \gamma_2$. Hence from (C3) and (C5) we can write

$$\delta = \frac{k - k^2}{c - k^2}, \qquad k = \frac{(1-\delta)\alpha}{(1-\delta)^2 \alpha + 1}$$

where $\alpha := (1-\lambda)\alpha_x / \alpha_z$. These are the same as the best response formulas equations (20) and (27) in the main text and from Proposition 1 we know that there is a unique pair k^*, δ^* satisfying these conditions.

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.

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.

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.

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.

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.

Illing, Sean, "A Giant Fog Machine: How Right-Wing Media Obscures Mueller and other Inconvenient Stories," *Vox*, October 2017.

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

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

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

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

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

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

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

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

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

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

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.

Sullivan, Eileen, "What Is the Uranium One Deal and Why Does the Trump Administration Care So Much?," *New York Times*, November 2017.

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

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.