The Power of Whispers:
A Theory of Rumor, Communication and Revolution*

HENG CHEN
University of Hong Kong
YANG K. LU
Hong Kong University of Science and Technology
WING SUEN
University of Hong Kong

September 14, 2012

Abstract. We study the role rumors play in revolutions using a global game model. Agents with diverse private information rationally evaluate the informativeness of rumors about the regime strength. Without communication among agents, wild rumors are discounted and agents are generally less responsive to rumors than to trustworthy news. When agents can exchange views on the informativeness of rumors, a rumor against the regime would coordinate a larger mass of attackers than that without communication. The effect of communication can be so large that rumors can have a greater impact on mobilization than does fully trustworthy information.

Keywords. coordination game, public signals, swing population, mixture distribution, censorship

JEL Classification. D74, D83

*We thank Roland Bénabou, Matthias Doepke, Christian Hellwig, Satoru Takahashi, Yikai Wang, Fabrizio Zilibotti, and seminar participants at NBER Summer Institute, UBC-HKU Microeconomics Workshop, University of Hong Kong, Toulouse School of Economics and European University Institute for their suggestions. Guangyu Pei deserves praise for providing excellent research assistance to this project. Our research is partially funded by the Research Grants Council of Hong Kong (Project No. HKU 742112B).

1. Introduction

Collective actions, such as riots, protests and political campaigns, are often immersed in rumors. Perhaps the most dramatic theater to witness rumors in action is a political revolution. Amid the recent Tunisian revolution, Ben Ali, the ex-Tunisian leader, was said to have fled his country. This was confirmed after conflicting rumors about his whereabouts, and finally led to the end of street protests. A while later in Egypt, it was widely reported that Mubarak's family had left for London, which was believed by many as a clear sign of fragility of the regime. Similar rumors about Qaddafi and his family appeared in Libya when the battle between the opposition and the regime intensified. Rumors are not unique to the series of revolutions in the Arab Spring. During the 1989 democracy movement in China, rumors repeatedly surfaced about the death of the leaders, Deng Xiaoping and Li Peng, as well as the divide among communist leaders, all of which indicated the vulnerability of the regime.[1]

Are rumors just rumors? In many cases, yes. Rumors that spread during turmoils often quickly disappear without leaving a trace. That seems to be natural as rational individuals may heavily discount unreliable information they receive in those situations. However many historical incidents suggest that rumors often turn out to be particularly effective in mobilization. Examples abound. The Velvet Revolution in Czechoslovakia was famously described as a "revolution with roots in a rumor" (Bilefsky 2009). At the dawn of the revolution in 1989, a prominent (false) rumor that a 19-year old college student was brutally killed by the police triggered many otherwise hesitant citizens to take to the streets. The revolution gained huge momentum right after that and the regime collapsed a few days later. In the Arab Spring, the news about Mubarak's family proved to be false, yet the opposition credited it for "mark[ing] a new phase" in their campaign.[2] Chinese history also offers many anecdotes in which rumors mobilized mass participation, including the Boxer Uprising, the Republican Revolution, and the May Fourth Movement.[3] Similarly, riots are often sparked by rumors as well: the 1921 Tulsa race riot, the 1967 Newark riot and the 2004 Rome riot provide dramatic examples.

A common interpretation of the role of rumors in mass movements is that individuals are just blindly herded by rumors. However we adopt the position that indi-

[1]There were widespread rumors of many variants that Deng died of illness during the protest and that Li was gun shot to death. It was also widely rumored that some senior leaders in the Communist Party wrote an open letter to oppose any actions against students. Some of these rumors were repeated in the print media. See, for example, the news report in the daily newspaper *Ming Pao* on June 6, 1989.

[2]*World Tribune* reported on January 28, 2011 that "... confirmed by a Western diplomat, ... Mubarak's wife, Suzanne, son, Gamal, and granddaughter arrived in London on a private jet as Egypt's defense minister secretly flew to the United States."

[3]Zhang (2009) documents these events in details, supplemented by several contemporary cases in China.

viduals are fully aware that rumors circulating in times of turmoil may or may not be well founded, and that they update their beliefs in a Bayesian manner. Since rumors are widely circulated and commonly observed, they may serve as a coordination device just like a public signal in a coordination game. We explain why some rumors are effective in mobilizing participation in collective actions while others are simply ignored.

In this paper, we focus on two key aspects of rumors: that they may be true or not true, and that people talk about them. Individuals in times of uncertainty and crisis often seek others' opinions and discuss with peers about their judgment and evaluation of rumors. Information from fellow participants can influence their beliefs and even actions. We show that such "idle talk" about rumors between individuals could potentially overcome their skepticism and amplify the impact of rumors on mobilization. It is even possible that rumors can have a greater impact on mobilization than does fully trustworthy information.

Specifically, we model political revolution as a coordination game among a large number of citizens, who decide whether to revolt against the regime or not. Citizens are uncertain about the regime's strength and possess dispersed information on it. In this model, global strategic complementarities arise, i.e., citizen's incentive to revolt increases with the aggregate action of all other citizens. If the number of participants is sufficiently high, the regime collapses; otherwise it survives. Before citizens take actions, they hear a rumor about the strength of the regime. This rumor is a publicly observed message, which could be either an informative signal on the regime's strength or an uninformative noise unrelated to fundamentals. Citizens assess the informativeness of the rumor based on their own private information on the regime. As a consequence of diverse private information among citizens, their assessment may also differ. Citizens are also allowed to communicate with one another and exchange their assessment on the rumor. In other words, they tell each other whether they believe the rumor or not.

In this model, citizens understand that rumors could be uninformative and therefore remain skeptical of them. They make an inference on how likely a rumor is informative based on their own private information, using Bayes' rule. The likelihood they assign that the rumor is informative is endogenous: if the rumor is far different from what the citizen personally knows about the regime, she tends to discount it more heavily. One obvious implication of this mechanism is that very extreme rumors—news that almost no citizens would believe—have little effects on equilibrium outcomes. Not surprisingly, due to their skepticism, citizens are less responsive to the news they hear: rumors against the regime mobilize less attackers, compared to the case where such news is known to be trustworthy.

When citizens are allowed to communicate with each other, a fraction of the population (those with intermediate private information) will decide to attack the regime depending on whether their peers tell them that they believe the rumor or not. Citizens who receive confirmatory messages believe that the regime strength is more likely to be close to what the rumor indicates. Therefore, when they form their posterior belief, they assign higher weight to states of nature which are close to the rumor. In other words, they will be more responsive to the rumor. Furthermore, when the rumor is actually close to the true regime strength, more citizens will believe it is sufficiently informative and send confirmatory messages to their peers. As a consequence, a higher fraction of citizens will receive confirmatory messages and attack the regime. Therefore, a rumor against the regime which is near the true state mobilizes more attackers given communication than without communication. By the same mechanism, when a negative rumor is far from the truth, citizens will be even more skeptical given communication because most of their peers do not believe it. In short, citizens on the whole are better informed about whether the rumor they hear is close to the underlying state than they are without communication.

Interestingly and surprisingly, we find that this communication effect could make rumors even more powerful than trustworthy news in mobilizing individuals. Suppose that citizens hear a rumor against the regime, and they are allowed to communicate with each other about its informativeness, it is possible that the regime would survive if all of them fully believe that the rumor is trustworthy or informative, but would collapse if they believe the rumor may be uninformative.

In this paper, we also investigate the role of sentiment—the public's perception of what untrustworthy news would sound like—and show that it is a double-edged sword. On the one hand, the regime could manipulate perceptions and discredit negative information against itself, so as to increase the likelihood of its survival. On the other hand, systematic propaganda attempts by the government to spread news about its alleged strength only make rumors about the regime's weakness more credible.

Given rumors can threaten the survival of the regime, it is natural to think the regime may increase its chance of survival by blocking any public information against it and only allowing positive rumors to circulate. We investigate this conjecture using our model and study the effects of censorship. We find that censorship does not necessarily help the regime to survive, as citizens would interpret the absence of rumors as a sign of bad news for the regime.

Our work enriches the global games literature in a couple of directions. We offer a specification of public signals that allow us to capture people's skepticism toward the public information they observe. A common implicit assumption in the literature is that citizens believe the public signal is informative. When they form the posterior belief on the fundamental, citizens assign a constant weight on the public signal based on its relative precision to private signals. The implication is that citizens would not adjust the weight, even though the public signal is remarkably different from what their own private information suggests. However, such updating rule seems to be counter-intuitive, in view of the fact that individuals tend to disbelieve information which is too different from their priors. Our model provides a formal justification using mixture distributions to explain why the standard linear updating rule may not be appropriate, and why skepticism toward rumors is qualitatively different from having a public signal with low precision.

In much of the global games literature, citizens are assumed to only respond to signals they observe, and further interaction between citizens are often left out for the sake of simplicity and tractability.[4] In reality, individuals do exchange information with each other before they make decisions and take actions. This is especially true in collective actions such as protests, demonstrations and revolutions. Our work extends the standard setup to study communication among citizens. We model direct interaction between citizens by allowing them to communicate privately, rather than just observing a public signal of what others are doing. We find that these interactions can have powerful effects that overcome individual skepticism to make rumors more powerful than trustworthy news at the aggregate level.

This paper should not be interpreted as contradicting the literature that stresses structural factors as root causes for a revolution (Skocpol 1979). Structural factors, such as the state of the economy and international pressure, are those that make a society "ripe" for revolution. However, it has been noted that structural factors are not sufficient for a successful revolution. In line with Bueno de Mesquita (2010), we argue that some random factors also play a role in determining the fate of a revolution. In our model, the realization of rumors serves as a source of randomness. In contrast to the literature with multiple equilibria, our model features unique equilibrium. We attribute the fundamental cause of success or failure of a revolution to the regime's strength. However, for the same strength of a regime, it would collapse when the realization of the rumor takes certain values, while it would survive when it takes some other values.

[4] Angeletos and Werning (2006) explicitly acknowledge the importance of direct interaction between agents in coordination models. In one extension of their model, agents are allowed to observe a public signal about the aggregate attack, which conveniently approximates the situation where agents could learn about the actions of others. We capture the direct interaction between citizens by allowing them to discover other's personal judgment through one-to-one private communication.

2. Related Literature

Our paper builds on the global games literature (Carlsson and van Damme 1993 and Morris and Shin 1998), which has been applied to analyze issues in political economy. Among recent examples are Boix and Svolik (2010), Chassang and i Miquel (2010) and Edmond (2011). In our model, the dispersion in private information is crucial to generate different assessment on the informativeness of rumors. Information heterogeneity provides a ground for the study of communication between citizens, which is at the core of this paper. We choose to interpret our model in the context of revolution, but it can also help to understand similar coordination games, such as bank runs and currency attacks.

Our work is related to a small economics literature on rumors. Banerjee (1993) develops a model where rumors on investment opportunity are passed on from one agent to another, but recipients do not necessarily believe those rumors. The probability that someone hears a rumor is positively related to the number of people who have heard it. Bommel (2003) studies the effects of rumors on stock prices. Informed investors with limited trading capacity profit from their private information, and they also spread rumors (i.e., give informative yet imprecise information to their followers) so as to profit from stock-price manipulation. Like their models, we also assume that a rumor heard by citizens cannot be verified, and that citizens use Bayesian updating in deciding whether to take action or not. Unlike their models, in which a rumor is passed on to other agents sequentially, we provide a model in a static setting, in which a rumor is heard by citizens simultaneously. We focus on communication among citizens about the rumor rather than the transmission of the rumor itself.

This paper also contributes to a growing literature on revolutions in economics. Edmond (2011) considers a coordination game where citizens' private information about the regime's strength is contaminated by the regime's propaganda. Citizens understand the signal-jamming technology used by the regime and form their beliefs accordingly. Our model differs in that private information is uncontaminated, but the public signal may be false and unrelated to fundamentals. Both Bueno de Mesquita (2010) and Angeletos and Werning (2006) study coordination games with two stages, where public signals arise endogenously from the collective action in the first stage. In our model, the "attack stage" is preceded by a "communication stage," where a private message endogenously arises and enlarges citizens' information set. New development of this literature puts emphasis on uncertain payoffs from revolt (Bernhardt and Shadmehr 2010 and Bueno de Mesquita 2010). Given that our focus is the effect of rumors and communication, we assume that only the strength of the regime is uncertain.

In other fields of social sciences, there is no lack of discussions on rumors (e.g., Allport and Postman 1947) and revolutions (e.g., Goldstone 1994). However there are few studies on the relationship between these two. The idea seems to have been "up in the air" that rumors work as a tool to motivate citizens to participate in social movements such as riots, demonstrations and revolts, but the precise mechanisms remain unspecified, through which the actions taken by citizens are related to the rumors they hear and talk about. Our model is a step toward formalizing one such mechanism to explain explicitly how rumors affect citizens' beliefs, actions, and therefore equilibrium outcomes in revolutions.

3. A Model of Rumors and Talk about Rumors

3.1. Players and payoffs

Consider a society populated by a unit mass of ex ante identical citizens, indexed by $i \in [0,1]$, who play against another player, the regime. Citizen i chooses one of two actions: revolt ($a_i = 1$) or not revolt ($a_i = 0$). The aggregate mass of population that revolt is denoted A. Nature selects the strength of the regime, θ, which is sometimes also referred to as the state. The regime survives if and only if $\theta > A$; otherwise it is overthrown. A citizen's payoff depends both on whether the regime is overthrown and on whether the citizen chooses to revolt. A positive cost, $c \in (0,1)$, has to be paid if she revolts. If the regime is overthrown, citizens who revolt receive a benefit, $b = 1$, and those who do not participate receive no benefit.[5] A citizen's net utility is therefore:

$$u(a_i, A, \theta) = \begin{cases} 1 - c, & \text{if } a_i = 1 \text{ and } A \geq \theta; \\ -c, & \text{if } a_i = 1 \text{ and } A < \theta; \\ 0, & \text{if } a_i = 0. \end{cases}$$

3.2. Information structure

Citizens are ex ante identical and have improper prior on θ. They become ex post heterogeneous after each of them observes a noisy private signal,

$$x_i = \theta + \varepsilon_i,$$

[5] We abstract from free-riding issues, which has been carefully addressed by Edmond (2011) and Bernhardt and Shadmehr (2010). The benefits from regime change can be modeled as a public good that all citizens would enjoy. Edmond (2011) offers a general payoff structure to accommodate this concern. He shows that a condition can be derived such that citizens still have incentives to act against the regime, despite the free-riding incentives. To avoid being repetitive and keep the results sharp, we adopt a simpler payoff structure in this paper, which is a special case of his.

where the idiosyncratic noise $\varepsilon_i \sim \mathcal{N}(0, \sigma_x^2)$ is normally distributed and independent of θ, and is independently and identically distributed across i. This assumption captures the situation that citizens have diverse assessment of the regime's strength, before they hear any rumor and communicate. This seemingly standard assumption in the global games literature turns out to be crucial for our model, for people will not exchange information in a society where everyone shares the same beliefs.

Our next assumption is that all citizens hear a rumor, z, concerning the strength of the regime. The key issue that we focus on in this paper is how citizens evaluate and react to the rumor, when the rumor could be totally uninformative. Towards this end, the rumor is modeled as a public signal, which could come from two alternative sources: either a source which offers an informative signal on the strength of the regime, or a source which only produces uninformative noise. Formally we model the random variable z as coming from a mixture distribution:

$$z \sim \begin{cases} I = \mathcal{N}(\theta, \sigma_z^2), & \text{with probability } \alpha; \\ U = \mathcal{N}(s, \sigma_U^2), & \text{with probability } 1 - \alpha, \end{cases}$$

where I indicates the informative source and U indicates the uninformative source. The parameters s and σ_U^2 are the mean and variance of the uninformative distribution, respectively. The rumor comes from an informative source with prior probability α. When this is the case, z is normally distributed with mean θ and variance σ_z^2. We assume that α, s, σ_z, and σ_U are commonly known to all citizens.

The parameter s can be interpreted as the "sentiment" of the public, which captures their perception of what uninformative messages would sound like.[6] For example, if the public is used to receiving propaganda materials telling that the regime is strong, then they may expect a high value of s. Alternatively, sentiments may be driven by external events: news that would have seemed completely implausible (e.g., Mubarak leaving the country) could suddenly become plausible when similar news was reported and confirmed in a neighboring country. In this paper, we do not model where the public's sentiment s comes from, but we provide comparative statics analysis of how shifts in s affect equilibrium outcomes (see Section 6.2).

We stress that our specification of rumor as a mixture distribution is different from an informative public signal with low precision. According to the linear Bayesian updating formula, all agents would react to an informative public signal in the same way regardless of their private information. In our specification, however, citizens

[6]The parsimonious assumption that s is common to all the citizens is made to simplify the exposition. Allowing them to possess diverse sentiments will be equivalent to adding noise to the uninformative distribution. Specifically we can assume that $U = \mathcal{N}(s_i, \sigma_u^2)$, where s_i is individual-specific sentiment and is distributed across the population according to $\mathcal{N}(s, \sigma_s^2)$. This will not change our results.

make an inference on how likely the rumor is informative. Since agents with different private signals have different views regarding what informative news would be like, they react to the same rumor differently. This mechanism plays a central role in our paper and cannot be replaced by modeling rumor simply as an informative public signal with a higher variance.

While not dismissing the relevance of rumormongers, we choose to model rumors as exogenous public signals in order to focus on their role in coordinating collective action. The origin and the content of rumors are assumed to be exogenous because of their sheer diversity and unpredictability. The possibility that rumors are cooked up strategically to influence individuals' belief is an important reason that people tend to be skeptical. But even in this case, the true source of rumors is often shrouded in obscurity, making it difficult to infer whether they are manufactured to defend the regime or to destabilize it.[7] Moreover, studies on rumors also show that they could be created unintentionally. For example, misunderstanding between individuals is a usual source of rumors.[8]

By modeling rumors as a public signal, we also abstract from the process of how rumors travel from one to another.[9] It is implicitly assumed that rumors could reach to every citizen in the game.[10] This assumption seems to be not unrealistic for many revolutions in history: rumors against authorities did gain a substantial, even huge, amount of publicity under very repressive regimes.[11]

We maintain the following parameter restrictions throughout this paper:

$$\sigma_x < \sigma_z^2 \sqrt{2\pi}; \tag{1}$$

$$\sigma_U^2 > \sigma_x^2 + \sigma_z^2 \equiv \sigma_I^2. \tag{2}$$

The first restriction is standard. When $\alpha = 1$, the model reduces to the standard

[7]See Knapp (1944), Nkpa (1977), Ley (1997), Grunden, Walker, and Yamazaki (2005), Elias and Scotson (1994), and Gambetta (1994) for related analysis. The incentives of rumormongers could be unpredictable in the sense that they might be motivated by many different reasons (Zhang 2009 and Turner, Pratkanis, Probasco, and Leve 1992).

[8]See detailed discussion in Allport and Postman (1947), Peterson and Gist (1951), and Buchner (1965).

[9]Sociological studies find two types of rumor propagation mechanisms: "snow balling pattern" (Peterson and Gist 1951) and "simplification pattern" (Allport and Postman 1947 and Buchner 1965). Modeling the process of diffusion of rumors in a non-reduced form deserves a separate endeavor. For example, see Acemoglu, Ozdaglar, and ParandehGheibi (2010) for a model where rumors spread in a network based economy.

[10]In principal, we could also assume that a certain fraction of citizens do not hear any rumor. This would not affect the main results in our model.

[11]The rumor that a college student was killed by the police, which ignited the Velvet Revolution in Czechoslovakia, was broadcast by Radio Free Europe. The internet offers additional channels for spreading rumors. In 2009, the Iranian post-election protest intensified after a rumor surfaced in the internet that police helicopters were pouring acid and boiling water on protesters (Esfandiari 2010).

Morris-Shin model with public signal. Condition (1) is both sufficient and necessary for uniqueness of equilibrium in that model; see the discussion in Angeletos and Werning (2006).

The second restriction captures the idea that uninformative noise exhibits greater variability than an informative signal. This assumption is justified by the fact that there are multiple possibilities that can generate an uninformative signal. When a signal is generated from an informative source, it is based on the true strength of the regime and its realization is anchored by the truth. When a signal is generated by an uninformative source, there is great uncertainty concerning where it originates: rumors may be made up by friends or enemies of the regime, or by people with unknown motives which are unrelated to regime survival. The possibility that rumors are often the result of mistakes adds to this uncertainty. In other words, since uninformative noise is drawn from a distribution which is not anchored by facts or fundamentals, it tends to be more unpredictable.[12] We will show that this assumption is responsible for the result that citizens tend to discount wild rumors.

3.3. Communication

After citizens observed their private signals and the rumor, they are randomly paired up to communicate with one another about the informativeness of the rumor. Specifically each citizen in a pair expresses her view on the likelihood that the rumor is drawn from an informative source, and hears her peer's view on the same matter. We assume that such communication is not frictionless: citizens can only convey to their peers whether they believe the rumor is informative or not in a binary fashion. Let y_i represent the message sent to citizen i by her peer k. The communication technology is characterized as follows.

$$y_i = \begin{cases} 1, & \text{if } \Pr[z \sim I | z, x_k] \geq \delta; \\ 0, & \text{if } \Pr[z \sim I | z, x_k] < \delta. \end{cases} \tag{3}$$

The threshold δ in this communication technology is common to all citizens and can be interpreted as a measure of their caution. Citizen k who sends the message $y_i = 1$ to citizen i can be interpreted as saying, "I believe the rumor is informative;" while the message $y_i = 0$ can be interpreted as "I don't believe it." A high value of δ means that citizens are unlikely to say they believe in the rumor unless they are sufficiently confident of their assessment. To rule out the possibility that agents will never say they believe in the rumor, we require an upper bound for the value of δ. Specifically

[12] Note that heterogeneity in individual sentiment also adds to enlarge the variance of the uninformative distribution.

we maintain the following assumption throughout the paper:

$$\delta < \frac{\alpha\sigma_I^{-1}}{\alpha\sigma_I^{-1} + (1-\alpha)\sigma_U^{-1}} \equiv \bar{\delta}. \tag{4}$$

The communication rule (3) is non-strategic. Given that it is a game with a continuum of agents, and that each agent only talks to one other agent, there is no incentive for citizens to strategically manipulate their peers' belief by lying. In reality, truthful revelation is probably a better description of the casual communication between acquaintances than is strategic revelation. Our communication rule can be formally justified if the dominant motive behind such communication is to avoid the expected cost (e.g., embarrassment) from saying the wrong thing.

Lemma 1. If an agent wants to minimize the expected loss from type 1 and type 2 errors in sending her binary message, then the optimal communication rule is given by (3).

Two distinct features of our communication rule (3) deserves discussion. The conversation between citizens is conducted in a coarsened fashion: the binary message space captures the coarsening of information in the communication process. This assumption is intended to capture interesting aspects of reality and to simplify our analysis. People usually do not fully express, explain and justify their own private assessment by conveying the exact probability. Information coarsening is a cost-effective way of exchanging casual information, and is commonly used in real life.[13]

What drives our results is the assumption that people talk about the *informativeness* of the rumor. It is motivated by the fact that rumors, similar to the weather or celebrities, often constitute a focal point of conversation because it is commonly observed by all. Even though citizens' assessment of the informativeness of the rumor depends on their private information, our communication rule is very different from simply exchanging private signals between citizens. Our communication rule anchors the content of communication to the realization of the rumor, thus producing interesting interactions between the public rumor and people's private signals. Section 6.1 will elaborate on this point.

[13]In Section 6.1, we show that the main insights of our model continue to hold if we allow people to exchange messages about the exact values of their probability assessment. Section 6.1 also shows that our results are robust when we relax other simplifying assumptions on communication protocol in the baseline case.

3.4. Decision rules

This model can be regarded as a two-stage game. In the communication stage, citizen i sends a message, $y_k \in \{0,1\}$, to her peer k based on the information set $\{z, x_i\}$, and receives a private message $y_i \in \{0,1\}$ from her peer k. In the attack stage, citizen i chooses to revolt or not, given the post-communication information set $\{z, x_i, y_i\}$.

In the communication stage, citizens are heterogeneous in only one dimension: their observed private signal x_i. Different people have different beliefs because they have different expectations of what informative news is like. They follow Bayes' rule to update their beliefs:

$$\Pr[z \sim I | z, x_i] = \frac{\alpha \sigma_I^{-1} \phi\left(\sigma_I^{-1}(z - x_i)\right)}{\alpha \sigma_I^{-1} \phi\left(\sigma_I^{-1}(z - x_i)\right) + (1-\alpha)\sigma_U^{-1} \phi\left(\sigma_U^{-1}(z - s)\right)} \equiv w(z, x_i), \quad (5)$$

where ϕ is the standard normal density function.

The function $w(\cdot, x_i)$ is single-peaked in z. Thus rumors which take extremely high or low values are not believed by most citizens. This feature is a consequence of our assumption that the distribution of U has fatter tails than the distribution of I. We also note that $w(z, \cdot)$ is symmetric and single-peaked in x_i, reaching a maximum at $x_i = z$. This means that citizens whose private information is consistent with the rumor are more likely to believe that the rumor is informative.[14]

Since $w(z, \cdot)$ is single-peaked in x_i and reaches a maximum at $x_i = z$, there exists $\underline{x}(z)$ and $\overline{x}(z)$ such that

$$w(z, \underline{x}(z)) = w(z, \overline{x}(z)) = \delta,$$

provided $w(z, z) > \delta$. Assumption (4) on the upper bound of δ ensures that this condition is satisfied for any z. A citizen with private information x_i between $\underline{x}(z)$ and $\overline{x}(z)$ would have belief $w(z, x_i)$ greater than the threshold δ. The communication decision rule in the communication stage can therefore be written as:

$$y(z, x_i) = \begin{cases} 1, & \text{if } x_i \in [\underline{x}(z), \overline{x}(z)]\,; \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

Finally, note that the symmetry of $w(z, \cdot)$ about the point $x_i = z$ implies that the interval $[\underline{x}(z), \overline{x}(z)]$ is centered at z. That is, we can write

$$\underline{x}(z) = z - \kappa(z) \quad \text{and} \quad \overline{x}(z) = z + \kappa(z),$$

[14] See also Gentzkow and Shapiro (2006) and Suen (2010) for models which share the same feature.

for some $\kappa(z) > 0$.

After communication, the information set of citizen i is enlarged by the private message y_i sent by her peer. Let $p(\theta|z, x_i, y_i)$ represent her posterior density function of θ. In the attack stage, she chooses a_i to maximize expected utility,

$$a(z, x_i, y_i) = \underset{a_i \in \{0,1\}}{\operatorname{argmax}} \left\{ \int_{-\infty}^{\infty} u(a_i, A(\theta, z), \theta) p(\theta|z, x_i, y_i) \, \mathrm{d}\theta \right\} \tag{7}$$

The aggregate mass of attackers depends on the distribution of private signals x_i and communication messages y_i. Given state θ, x_i is normally distributed with mean θ and variance σ_x^2. Given the decision rule $y(z, x_i)$ specified in equation (6), the probability that citizen i receives the message $y_i = 1$ is equal to the probability that x_k lies within $[\underline{x}(z), \overline{x}(z)]$. Let $J(\theta, z)$ represent this probability conditional on θ and z. Then the aggregate mass of attackers is given by:

$$A(\theta, z) = \int_{-\infty}^{\infty} [J(\theta, z) a(z, x_i, 1) + (1 - J(\theta, z)) \, a(z, x_i, 0)] \, \frac{1}{\sigma_x} \phi \left(\frac{x_i - \theta}{\sigma_x} \right) \, \mathrm{d}x_i, \tag{8}$$

with

$$J(\theta, z) = \Phi \left(\frac{\overline{x}(z) - \theta}{\sigma_x} \right) - \Phi \left(\frac{\underline{x}(z) - \theta}{\sigma_x} \right), \tag{9}$$

where Φ is the standard normal distribution function. See Figure 1 for illustration.

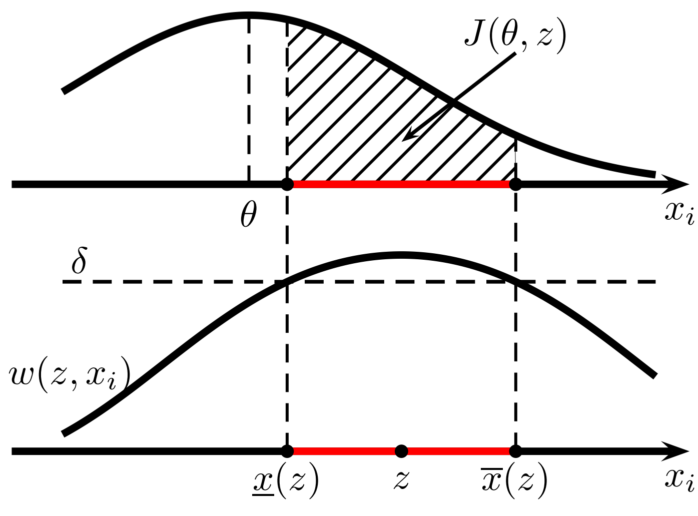


Figure 1. *Message sending rule and $J(\theta, z)$.*

3.5. Equilibrium

To summarize, the timing of events is the following. First, nature selects the strength of the regime. Then citizens receive private signals and hear a rumor. Afterwards, citizens are randomly matched in pairs and exchange their views on the informativeness of the rumor. Based on the updated information set, they choose to revolt or not. The regime survives or not, hinging on the mass of citizens who revolt.

Definition 1. *An equilibrium is a set of posterior beliefs $p(\theta|z, x_i, y_i)$, a message sending decision $y(z, x_i)$, a revolt decision $a(z, x_i, y_i)$, and a mass of attackers $A(\theta, z)$, such that (6), (7) and (8) hold and the beliefs are derived from Bayes' rule.*

We focus on monotone equilibrium in which, for each realization of the rumor z, there is a threshold strength of the regime $\theta^*(z)$ such that the regime collapses if and only if $\theta \le \theta^*(z)$. Citizens adopt the following monotonic decision rule:

$$a(z, x_i, y_i) = \begin{cases} 1, & \text{if } x_i \le x_I^*(z) \text{ and } y_i = 1, \\ & \text{or } x_i \le x_U^*(z) \text{ and } y_i = 0, \\ 0, & \text{otherwise;} \end{cases}$$

where x_I^* and x_U^* are a pair of cut-off rules. The equilibrium ordering of x_U^* and x_I^* depends on the realization of z, and is elaborated in Section 5.

In a monotone equilibrium, the cut-off types must be indifferent between attacking and not attacking. Let $P(\cdot|z, x_i, y_i)$ be the cumulative distribution corresponding to the posterior density $p(\cdot|z, x_i, y_i)$. Then the indifference conditions can be written as:

$$P(\theta^*|z, x_I^*, 1) = c, \tag{10}$$

$$P(\theta^*|z, x_U^*, 0) = c. \tag{11}$$

Let $\hat{A}(\theta, x_I^*, x_U^*, z)$ be the mass of attackers when the state is θ and when citizens adopt the cut-off rules x_I^* and x_U^*. Then,

$$\hat{A}(\theta, x_I^*, x_U^*, z) = J(\theta, z)\Phi\left(\frac{x_I^* - \theta}{\sigma_x}\right) + (1 - J(\theta, z))\Phi\left(\frac{x_U^* - \theta}{\sigma_x}\right),$$

where the function J is given by (9). The equilibrium regime survival threshold must satisfy

$$\hat{A}(\theta^*, x_I^*, x_U^*, z) = \theta^*. \tag{12}$$

A monotone equilibrium can be characterized by the triple (θ^*, x_I^*, x_U^*) that solves equations (10), (11), and (12).

4. Rumors without Communication

Our model departs from the standard global game model in two respects: (1) the public signal may be uninformative; and (2) citizens can exchange messages concerning the informativeness of the public signal. To highlight the effects of these two separate features, we discuss in this section a model with feature (1) only but without communication. Such a model can be obtained as a special case of our model by setting $\delta = 0$, so that the everyone always sends the same message $y(z, x_i) = 1$ and communication becomes irrelevant. We refer to this special case (i.e., $\delta = 0$) as the "mute model." The "communication model" with both features (1) and (2) (i.e., $\overline{\delta} > \delta > 0$) is left for Section 5.

The "mute model" nests two important benchmarks: $\alpha = 0$ and $\alpha = 1$. When $\alpha = 0$, citizens believe that the rumor is completely uninformative. In this case, the value of z is irrelevant and the model reduces to a standard Morris-Shin model without public signal. We refer to this as the "pure noise model" and use $(\theta^*_{ms}, x^*_{ms})$ to denote the equilibrium regime survival threshold and cut-off agent type in this model.

The other important benchmark is the case of $\alpha = 1$. In this case, the rumor is known to be informative, and the model reduces to the standard Morris-Shin model with public signal. We refer to this case as the "public signal model" and use $(\theta^*_{ps}, x^*_{ps})$ to denote the equilibrium regime survival threshold and cut-off agent type. Obviously in this case these equilibrium values depend on the realization of z.

The following result is standard; the proof is relegated to the Technical Appendix.

Proposition 1. *In the "pure noise model" (i.e., $\alpha = 0$), the equilibrium is $\theta^*_{ms} = 1 - c$ and $x^*_{ms} = (1 - c) - \sigma_x \Phi^{-1}(c)$. In the "public signal model" (i.e., $\alpha = 1$), there exists a unique equilibrium $\theta^*_{ps}(z)$ and $x^*_{ps}(z)$ such that:*

1. *$\theta^*_{ps}(z)$ and $x^*_{ps}(z)$ are decreasing in z.*
2. *(a) $\lim_{z \to \infty} \theta^*_{ps}(z) = 0$ and $\lim_{z \to \infty} x^*_{ps}(z) = -\infty$;*
   (b) $\lim_{z \to -\infty} \theta^*_{ps}(z) = 1$ and $\lim_{z \to -\infty} x^*_{ps}(z) = \infty$.
3. *There exists a unique $\tilde{z}$ such that $\theta^*_{ps}(\tilde{z}) = \theta^*_{ms}$ and $x^*_{ps}(\tilde{z}) = x^*_{ms}$.*

In the "public signal model," citizen i's posterior mean of θ upon observing public signal z is

$$X_i = \beta x_i + (1 - \beta) z, \tag{13}$$

where $\beta = \sigma_z^2 / (\sigma_z^2 + \sigma_x^2)$ (and the posterior variance is $\beta \sigma_x^2$). A lower value of z indicates that the regime is more fragile. Other things equal, this would result in more agents revolting against the regime (x^*_{ps} increases), making the regime more likely to collapse (θ^*_{ps} increases). In the limit, as z becomes very small, almost all citizens would revolt and the regime with type $\theta < 1$ would collapse almost surely (θ^*_{ps} goes to 1).

Part (1) and part (3) of Proposition 1 imply that $\theta^*_{ps}(z) > \theta^*_{ms}$ and $x^*_{ps}(z) > x^*_{ms}$ for all $z < \tilde{z}$. In what follows we say that a rumor is "against the regime" if $z < \tilde{z}$. If $z > \tilde{z}$, we say that the rumor is "for the regime."

In the "mute model," $\alpha \in (0, 1)$ and $\delta = 0$, so that citizens are skeptical of rumors but communication is ineffective. By Bayes' rule, the posterior belief about θ upon hearing a rumor z is a mixture of the posterior distribution in the "public signal model" and that in the "pure noise model," with weights given by the posterior belief that the rumor is informative or not. In other words,

$$P(\theta|z, x_i) = w(z, x_i)\Phi\left(\frac{\theta - X_i}{\sqrt{\beta}\sigma_x}\right) + (1 - w(z, x_i))\Phi\left(\frac{\theta - x_i}{\sigma_x}\right), \tag{14}$$

where the weight function $w(z, x_i)$ is given by (5) and where the posterior mean X_i is given by (13) of the "public signal model." The crucial feature of this mixture distribution for θ is that the relative weights are not fixed (even though the mixture distribution for z have fixed weights α and $1 - \alpha$). Instead the weights depend on both the content of the rumor and on each citizen's private information.

Despite the dependence of the weights of the mixture distribution on private information, the posterior belief about θ is still stochastically increasing in x_i. To see this, note that the density function of x_i given state θ and rumor z is $\sigma_x^{-1}\phi(\sigma_x^{-1}(x_i - \theta))$. For any $\theta_1 > \theta_0$, the likelihood ratio,

$$\frac{\phi(\sigma_x^{-1}(x_i - \theta_1))}{\phi(\sigma_x^{-1}(x_i - \theta_0))},$$

is increasing in x_i. This monotone likelihood ratio property implies that the posterior cumulative distribution, $P(\theta|z, x_i)$, is decreasing in x_i (Milgrom 1981). Thus, for any regime survival threshold θ, the expected payoff from revolt is decreasing in the value of the private information x_i. This justifies our focus on monotone equilibrium.

Let (θ^*_m, x^*_m) represent the equilibrium regime survival threshold and cut-off agent type in the "mute model." The cut-off type must be indifferent between attacking and not attacking:

$$P(\theta^*_m|z, x^*_m) = c; \tag{15}$$

and the mass of attackers in state θ^*_m must be equal to the regime survival threshold:

$$\hat{A}_m(\theta^*_m, x^*_m) = \Phi\left(\frac{x^*_m - \theta^*_m}{\sigma_x}\right) = \theta^*_m. \tag{16}$$

The following result characterizes the equilibrium $\theta_m^*(z)$ and compares it with the counterpart in the benchmark models.

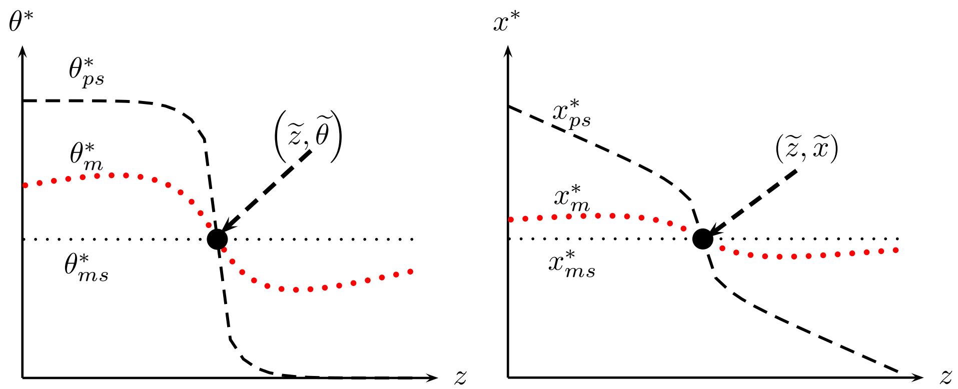
Proposition 2. *In the "mute model" (i.e., $\alpha \in (0,1)$ and $\delta = 0$), the equilibrium regime survival threshold $\theta_m^*(z)$ satisfies:*

1. $\theta_m^*(\tilde{z}) = \theta_{ps}^*(\tilde{z}) = \theta_{ms}^*$.
2. *(a) If the rumor is against the regime (i.e., $z < \tilde{z}$), then $\theta_m^*(z) \in (\theta_{ms}^*, \theta_{ps}^*(z))$;*
   (b) if the rumor is for the regime (i.e., $z > \tilde{z}$), then $\theta_m^*(z) \in (\theta_{ps}^*(z), \theta_{ms}^*)$.
3. $\lim_{z\to-\infty} \theta_m^*(z) = \lim_{z\to\infty} \theta_m^*(z) = \theta_{ms}^*$.
4. $\theta_m^*(z)$ *is increasing then decreasing then increasing.*

Figure 2(a) illustrates the properties of the equilibrium regime survival threshold described in this proposition.[15] Figure 2(b) shows the qualitative properties of the equilibrium cut-off agent type. Observe that the properties of $x_m^*(z)$ are similar to those of $\theta_m^*(z)$. In particular, $x_m^*(z)$ is increasing then decreasing then increasing, with $\lim_{z\to\pm\infty} x_m^*(z) = x_{ms}^*$. This connection follows from the attack equation (16), which can be written as:

$$x_m^*(z) = \theta_m^*(z) + \sigma_x \Phi^{-1}(\theta_m^*(z)).$$

Since $x_m^*(z)$ is increasing in $\theta_m^*(z)$, they share the same qualitative properties as functions of z.



(a) regime survival threshold $\theta^*(z)$ (b) cut-off agent type $x^*(z)$

Figure 2. *Comparing the "mute model" with benchmark "pure noise model" and "public signal model"*

Interestingly, part (1) of Proposition 2 says that the equilibrium pair $(\theta_{ms}^*, x_{ms}^*)$ of the "pure noise model" also solves the "mute model" when the rumor is neutral (neither "for the regime" nor "against the regime"), i.e., when $z = \tilde{z}$. Recall that in the

[15]Unless otherwise specified, we use the following set of parameters as benchmark to compute numerical examples: $c = 0.5$, $s = 0.5$, $\alpha = 0.5$, $\delta = 0.5$, $\sigma_U^2 = 1$, $\sigma_z^2 = 0.5$, and $\sigma_x^2 = 0.4$.

"pure noise model" an agent with private information x^*_{ms} believes that the regime will collapse with probability c given regime survival threshold θ^*_{ms}. Given the same θ^*_{ms}, this agent also believes that the regime will collapse with the same probability c in the "public signal model" if the public signal is $\tilde{z}$. Thus, for $z = \tilde{z}$, it does not matter for this agent whether the rumor is informative or not. If the regime survival threshold is θ^*_{ms}, she believes that the regime will collapse with probability c. Thus the indifference condition (15) is satisfied. Given that the cut-off agent type is x^*_{ms} at $z = \tilde{z}$, the threshold θ^*_{ms} satisfies the attack equation (16) as well. Thus $(\theta^*_{ms}, x^*_{ms})$ is an equilibrium of the "mute model" at $z = \tilde{z}$.

In this model, citizens are skeptical of the rumor, i.e., they take into account the possibility that the rumor could just provide a noise. The degree of skepticism is characterized by the weight given to the rumor being informative. The effect of skepticism manifests itself in the fact that the equilibrium regime survival threshold is less sensitive to changes in z around $\tilde{z}$ than it is in the "public signal model," and more responsive than in the "pure noise model." Specifically,

$$0 = \frac{\mathrm{d}\theta^*_{ms}}{\mathrm{d}z} > \frac{\mathrm{d}\theta^*_{m}(\tilde{z})}{\mathrm{d}z} > \frac{\mathrm{d}\theta^*_{ps}(\tilde{z})}{\mathrm{d}z}, \tag{17}$$

which is implied by part (2) of Proposition 2.

To understand why (17) holds, let $\hat{x}(\theta, z)$ be defined as the cut-off agent type who is indifferent between attacking and not attacking if the regime survival threshold is θ and the rumor realization is z. In the "mute model," as well as in the benchmark models, the response of equilibrium threshold to a change in the realization of the rumor can be calculated by,

$$\frac{\mathrm{d}\theta^*(\tilde{z})}{\mathrm{d}z} = \frac{\frac{1}{\sigma_x}\phi(\cdot)\frac{\partial \hat{x}}{\partial z}}{1 - \frac{1}{\sigma_x}\phi(\cdot)\left(-1 + \frac{\partial \hat{x}}{\partial \theta}\right)}, \tag{18}$$

where $\phi(\cdot)$ is evaluated at the equilibrium value of $(x^* - \theta^*)/\sigma_x$ of the relevant model. The effect of rumor on the equilibrium regime survival threshold can be decomposed into two components. At $z = \tilde{z}$, a small decrease in the value of z means that the regime is relatively weaker. In order to keep the expected payoff from attacking fixed, the cut-off agent type $\hat{x}$ must have a higher value of private information to compensate for the lower value of z. This direct effect tends to raise the mass of attackers because $\partial \hat{x}/\partial z < 0$. It is reflected in the numerator of (18). Furthermore, as more people attack the regime and the regime survival threshold rises, the payoff from attacking also rises, which raises $\hat{x}$ further, that is, $\partial \hat{x}/\partial \theta > 0$. This multiplier effect is reflected in the denominator of (18).

When $z = \tilde{z}$, the equilibrium values of θ^* and x^* are the same across the three models. Therefore, in equation (18), the term $\phi(\cdot)$ does not depend on the model being considered. The critical factors in determining the magnitude of $\mathrm{d}\theta^*/\mathrm{d}z$ are $\partial\hat{x}/\partial z$ and $\partial\hat{x}/\partial\theta$, which correspond to the direct and multiplier effects. At $z = \tilde{z}$, we have:

$$\frac{\partial \hat{x}_{ms}}{\partial \theta} = 1 < \frac{\partial \hat{x}_m}{\partial \theta} = \frac{w + (1-w)\sqrt{\beta}}{w\beta + (1-w)\sqrt{\beta}} < \frac{\partial \hat{x}_{ps}}{\partial \theta} = \frac{1}{\beta}. \tag{19}$$

We label this ordering the "skepticism effect," because it is driven by the fact that the weight w attached to the rumor being informative is between 0 and 1. Further, at $z = \tilde{z}$, $\partial\hat{x}/\partial z + \partial\hat{x}/\partial\theta = 1$ for all the three models. Therefore, (19) implies:

$$0 = \frac{\partial \hat{x}_{ms}}{\partial z} > \frac{\partial \hat{x}_m}{\partial z} > \frac{\partial \hat{x}_{ps}}{\partial z}.$$

The comparison of $\mathrm{d}\theta^*_m(\tilde{z})/\mathrm{d}z$ with the counterparts in the benchmark models given in (17) then follows immediately.

The comparison in (17) shows that, at the point $z = \tilde{z}$ the sensitivity of equilibrium outcome to the rumor in the "mute model" is between that in the "pure noise model" and in the "public signal model." More generally, part (2) of Proposition 2 states that the value of $\theta^*_m(z)$ is always between θ^*_{ms} and $\theta^*_{ps}(z)$ for any z. This, however, does not mean that the "mute model" is simply a "public signal model" with a noisier public signal.

Part (4) of Proposition 2 states that $\theta^*_m(z)$ is non-monotone, while Proposition 1 states that $\theta^*_{ps}(z)$ is monotonically decreasing (regardless of the precision of the public signal). Note that from equation (14),

$$\frac{\partial P(\theta^*_m | z, x_i)}{\partial z} = -w \frac{\beta}{\sqrt{\beta}\sigma_x} \phi_I + (\Phi_I - \Phi_U) \frac{\partial w}{\partial z},$$

where the subscripts I and U means that the functions are evaluated at the points $(\theta^*_m - X_i)/(\sqrt{\beta}\sigma_x)$ and $(\theta^*_m - x_i)/\sigma_x$, respectively. If the rumor is informative, an increase in z is indication that the regime is strong, which lowers the probability that the regime will collapse. Hence the first term is negative. However, since $w(z, x_i)$ is single-peaked in z, $\partial w/\partial z < 0$ for z sufficiently large, reflecting greater skepticism toward the rumor when it moves farther away from the agent's private signal. When the rumor is for the regime, the probability that the regime will collapse is lower if the rumor is informative than if it is uninformative; that is, $\Phi_I < \Phi_U$. Therefore the second term is positive for z sufficiently large. Moreover, the magnitude of the first term is small for z sufficiently large. When the second term dominates the first term, an increase in the value of z actually increases agents' assessment of the likelihood that the regime would collapse. As a result, the marginal agent who is indifferent between attacking and not attacking must have a higher private information about the strength of the regime. In other words, $\partial\hat{x}_m/\partial z > 0$ for z large, which implies $\mathrm{d}\theta^*_m(z)/\mathrm{d}z > 0$ by equation (18).

On the other hand, when $z = \tilde{z}$, we have $\Phi_I = \Phi_U$ because it does not matter to the cut-off type agent whether the rumor is informative or not. For values of z close to $\tilde{z}$, the magnitude of $\Phi_I - \Phi_U$ is small, and the first effect dominates. In this case, $\partial\hat{x}_m/\partial z < 0$, which implies $\mathrm{d}\theta^*_m(z)/\mathrm{d}z < 0$. This explains the non-monotonicity of $\theta^*_m(z)$.

The non-monotonicity result also explains why the limit behavior of the "mute model" is qualitatively different from that of the "public signal model." When the rumor is extreme, i.e., $z \to -\infty$ or $+\infty$, the probability that it comes from an informative source goes to 0. As a result, people disregard the rumor and the equilibrium is identical to that in the "pure noise" case: $\theta^*_m(z)$ goes to θ^*_{ms} and $x^*_m(z)$ goes to x^*_{ms}. In contrast, in the "public signal model", the equilibrium threshold monotonically deceases from 1 to 0 as z goes from $-\infty$ to ∞.

5. Rumors with Communication

In this section, we focus on the "communication model," with $\delta \in (0, \overline{\delta})$ and $\alpha \in (0, 1)$. In this model, citizens are still skeptical of the rumor they hear, but are allowed to communicate with one another about its informativeness. We show that communication could overcome the effect of skepticism and spark a sharp reaction among citizens. More surprisingly, a rumor against the regime could be more effective in mobilization than a negative trustworthy news that all citizens fully believe to be informative.

5.1. Posterior Beliefs

By Bayes' rule, agent i who receives the message $y_i = 1$ from her peer revises her belief about the state to:

$$P(\theta|z, x_i, 1) = \frac{\int_{-\infty}^{\theta} J(t,z)p(t|z,x_i)\,\mathrm{d}t}{\int_{-\infty}^{\infty} J(t,z)p(t|z,x_i)\,\mathrm{d}t}.$$

where $p(\cdot|z, x_i)$ is the density associated with the belief $P(\cdot|z, x_i)$ in the "mute model" given by equation (14), and where $J(t, z)$ is the probability that a randomly selected agent would send a message to her peer that confirms the rumor z in state t, as given by equation (9). The denominator is equal to $\Pr[y_i = 1|z, x_i]$, that is, the probability that agent i expects to receive $y_i = 1$ from her peer. Note that $\Pr[y_i = 1|z, x_i] = \mathrm{E}[J(t, z)|z, x_i]$. The term $J(t, z)$ is increasing then decreasing in t, with a peak at $t = z$. Therefore, $J(t, z)/\Pr[y_i = 1|z, x_i]$ is higher than 1 when the state t is close to z, and is lower than 1 when t is far away from z. In other words, upon receiving the message $y_i = 1$, the posterior density becomes more concentrated around the rumor z than its counterpart in the "mute model." Similarly, agent i who receives the message $y_i = 0$ revises her belief to:

$$P(\theta|z, x_i, 0) = \frac{\int_{-\infty}^{\theta}(1 - J(t,z))p(t|z, x_i)\,\mathrm{d}t}{\int_{-\infty}^{\infty}(1 - J(t,z))p(t|z, x_i)\,\mathrm{d}t}.$$

Let $\hat{x}_I(\theta, z)$ be the value of x_i that solves $P(\theta|z, x_i, 1) = c$. That is, an agent with private information $\hat{x}_I$ is indifferent between attacking or not attacking if the regime survival threshold is θ, the rumor realization is z, and the peer's message is that she "believes the rumor." Similarly, let $\hat{x}_U(\theta, z)$ be the value of x_i that solves $P(\theta|z, x_i, 0) = c$. Denote $\theta^*(z)$ to be the equilibrium regime survival threshold when the rumor is z. The equilibrium cut-off types are $x_I^*(z) = \hat{x}_I(\theta^*(z), z)$ and $x_U^*(z) = \hat{x}_U(\theta^*(z), z)$.

5.2. Extreme and Neutral Rumors

Proposition 3. *In the "communication model" (i.e., $\alpha \in (0,1)$ and $\delta \in (0, \bar{\delta})$), the equilibrium triple (θ^*, x_I^*, x_U^*) has the following properties:*

1. $\lim_{z\to\pm\infty} \theta^*(z) = \theta^*_{ms}$, $\lim_{z\to\pm\infty} x_U^*(z) = x^*_{ms}$, *and* $\lim_{z\to\pm\infty} x_I^*(z) = \mp\infty$.
2. *There exists a z' such that $\theta^*(z') = \theta^*_m(z')$ and $x_I^*(z') = x_U^*(z') = x^*_m(z')$.*
3. *If $c = 0.5$, then $z' = \tilde{z}$ and $\theta^*(z') = \theta^*_m(z') = \theta^*_{ps}(z') = \theta^*_{ms}$. Furthermore, $x_I^*(z) > x_U^*(z)$ for all $z < z'$ and $x_I^*(z) < x_U^*(z)$ for all $z > z'$.*

The limit behavior of the "communication model" is summarized in part (1). When the rumor takes extreme values, almost everyone says "I don't believe it" (because the limit of $J(\theta, z)$ is 0 for any finite θ). The skepticism of an agent (captured by the fact that the limit of $w(z, x_i)$ is 0) is reinforced by the skepticism of her peer. This explains why θ^* and x_U^* are the same as those in the "pure noise model."

The limit behavior of x_I^*, however, is qualitatively different. If the rumor indicates that the regime is very weak, an agent i is extremely unlikely to receive a confirmatory message from her peer. But in the unlikely event that she does, such an extreme event will overcome her initial skepticism toward the rumor to the extent that, for any finite value of x_i, she becomes almost sure that the regime will collapse. Therefore, in equilibrium, the cutoff type $x_I^*(z)$ must increase without bound as z becomes very low, meaning that almost anyone who hears a confirmatory message about an extreme rumor against the regime will revolt.

Part (1) of Proposition 3 immediately implies, via the intermediate value theorem, that there exists z' such that $x_I^*(z') = x_U^*(z') = x'$. Part (2) of the proposition says that such an x' turns out to be equal to $x_m^*(z')$ of the "mute model." To see why this is true, observe that an agent's decision to revolt does not depend on what her peer says when $x_I^* = x_U^*$. At z', therefore, the "communication model" is observationally equivalent to the "mute model": the mass of attackers is always the same in both models, regardless of the realization of θ.

Interestingly, we find that z' coincides with $\tilde{z}$ in the case where $c = 0.5$. This means that rumors for which cut-off type citizens do not care about what their peers say ($y = 1$ or $y = 0$) are also rumors for which they do not care about from which source the rumor is drawn ($z \sim I$ or $z \sim U$). This special case offers analytical convenience and tractability, since equilibrium thresholds are the same across all the four models when $z = z' = \tilde{z}$. We study the role played by communication in the "communication model" by letting z deviate from z' and comparing the equilibrium results with the other three benchmarks.[16] Therefore, in the remaining analysis, we focus on this case.

5.3. Swing Population

In a monotone equilibrium, there are three possible cases for the ordering of the cut-off types: $x_I^* > x_U^*$, $x_I^* < x_U^*$, or $x_I^* = x_U^*$. Suppose $x_I^* > x_U^*$. Then, citizens with private information $x_i < x_U^*$ attack the regime regardless of the message they receive. We label this group of citizens *revolutionaries*. Similarly, citizens with private information $x_i > x_I^*$ would not revolt regardless. This group of citizens is labeled *bystanders*. Citizens with $x_i \in [x_U^*, x_I^*]$ choose to revolt if they receive $y_i = 1$ from their peers, and not to revolt otherwise. Essentially, this group of citizens' revolt decisions are influenced by their peers' assessment of the rumor. We call this group the *swing population*. See Figure 3. The equilibrium ordering of $x_I^*(z)$ and $x_U^*(z)$ depends on z. Part (3) of Proposition 3 establishes that $x_I^*(z) > x_U^*(z)$ for all $z < z'$. The opposite is true if $z > z'$.

When $z < z'$, a citizen from the swing population attacks the regime when she receives a message that confirms the rumor and does not attack otherwise. It is straightforward to show that $P(\theta^*(z)|z, x_i, 1) > c > P(\theta^*(z)|z, x_i, 0)$ in this case. Intuitively, if the rumor says the regime is weak, anyone from the swing population finds that the expected payoff of attacking is higher than c when receiving a confirmatory message,

[16] It is difficult to directly compare $\theta^*(z)$, $\theta_m^*(z)$ and $\theta_{ps}^*(z)$ because none of these equilibrium values admit a closed-form solution. Our approach is to evaluate these functions at the point $z = z' = \tilde{z}$ and then compare their derivatives at that point. Since the equilibrium thresholds in all the different models are the same at $z = z' = \tilde{z}$, comparing the derivatives at that point allows us to make conclusions about the levels of these functions for values of z near z'. Qualitatively, our conclusions still hold when $z' \neq \tilde{z}$. The value of participation cost c does not affect the properties of equilibrium.

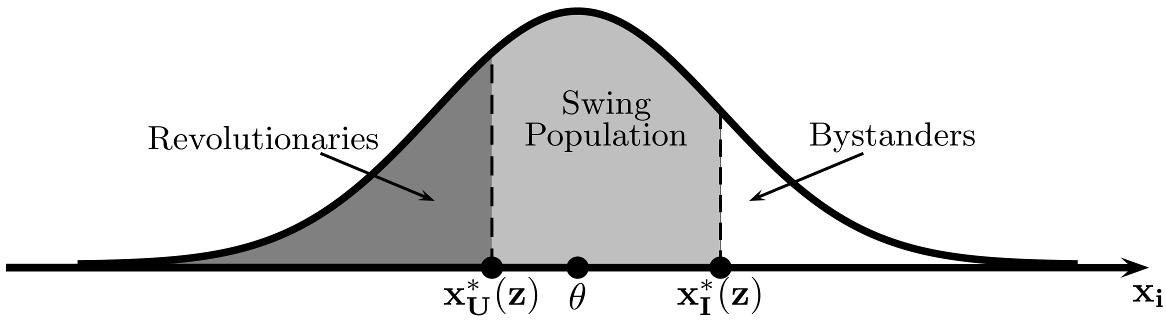


Figure 3. *An example of population distribution: revolutionaries, swing population and bystanders, when* $x_U^* < x_I^*$.

but is lower than c when the rumor is not confirmed by her peer.

5.4. The Communication Effect

We have explained that in the "mute model," two effects arise when z deviates from z'. The direct effect comes from the fact that a lower z has to be compensated by a higher realization of the private signal to keep the cut-off type indifferent. The magnitude of the direct effect depends on the magnitude of $\partial \hat{x}_m / \partial z$. The multiplier effect comes from the fact that any increase in the regime survival threshold would increase the payoff from revolting and hence further raise the cut-off type. The magnitude of the multiplier effect depends on $\partial \hat{x}_m / \partial \theta$. To compare the "communication model" with the "mute model," the first step is to study the effects of introducing communication on the magnitude of the direct and multiplier effects.

Lemma 2. At $z = z' = \tilde{z}$ and $\theta = \theta'$, the cut-off types who are indifferent between attacking and not attacking satisfy:

$$\frac{\partial \hat{x}_I}{\partial \theta} > \frac{\partial \hat{x}_m}{\partial \theta} > \frac{\partial \hat{x}_U}{\partial \theta} > 0, \tag{20}$$

$$\frac{\partial \hat{x}_I}{\partial z} < \frac{\partial \hat{x}_m}{\partial z} < \frac{\partial \hat{x}_U}{\partial z} < 0. \tag{21}$$

Lemma 2 states that the sensitivity of the cut-off type to changes in the regime survival threshold and to the rumor depends on the content of the communication, with citizens who receive $y_i = 1$ being more sensitive than citizens who receive $y_i = 0$, and with citizens in the "mute model" being in between. We label this ordering the "communication effect."

To understand the intuition behind Lemma 2, take the ordering of $\partial \hat{x} / \partial \theta$ (multiplier effect) first. Note that when $z' = \tilde{z}$, we also have $z' = \theta'$. With a confirmatory message that her peer believes the rumor z' is informative, a citizen would assign a higher probability weight (density) to those possible realizations of the state that are close to θ'. In particular, the posterior densities of the "communication model" and the "mute model" satisfy:

$$\frac{p(\theta'|z',x',1)}{p(\theta'|z',x')} = \frac{J(\theta',z')}{\Pr[y_i = 1|z',x']} > 1,$$

since $J(\theta,z')$ reaches a maximum at $\theta = \theta' = z'$ and the denominator is just the expected value of $J(\theta,z')$. The value of $p(\theta'|z',x',1)$ gives the marginal increase in payoff from attacking for the cut-off type who receives a confirmatory message when the regime survival threshold is raised slightly above θ'. Since $p(\theta'|z',x',1) > p(\theta'|z',x')$, this means that raising the regime survival threshold increases the incentive to attack by a greater amount for citizens who receive a confirmatory message than for citizens who receive no message in the "mute model." As a result, $\hat{x}_I$ also has to increase by a greater amount than $\hat{x}_m$ does to keep both cut-off types indifferent.

Next, consider the ordering of $\partial\hat{x}/\partial z$ (direct effect). As we have shown, a confirmatory message ($y_i = 1$) leads to a more concentrated posterior density for states around $\theta = z'$. This posterior distribution is more responsive to changes in the rumor than its counterpart in the "mute model." See Figure 4 for an illustration. The same increase in the value of z shifts the density function $p(\cdot|z,x_i,1)$ to the right by a greater amount than it does to the density function $p(\cdot|z,x_i)$. In other words, a citizen who receives $y_i = 1$ relies more on the rumor because it is confirmed by her peer. That explains why $\hat{x}_I$ has to decrease by a larger amount than $\hat{x}_m$ in order to balance the indifference condition, in response to the same amount of change in z.

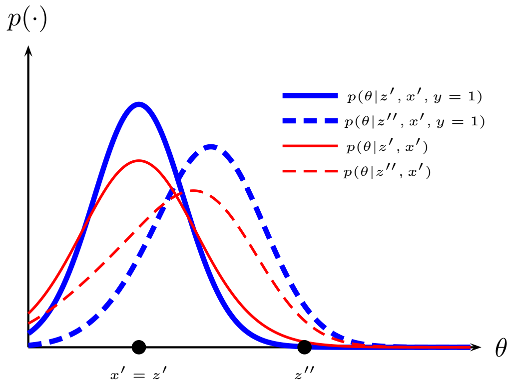


Figure 4. *Posterior density in the "communication model" and the "mute model." When $z = z'$, both densities are symmetric around z'. When $z'' > z'$, the density function in the "communication model" (for the case $y = 1$) shifts by a greater amount than does density function in the "mute model."*

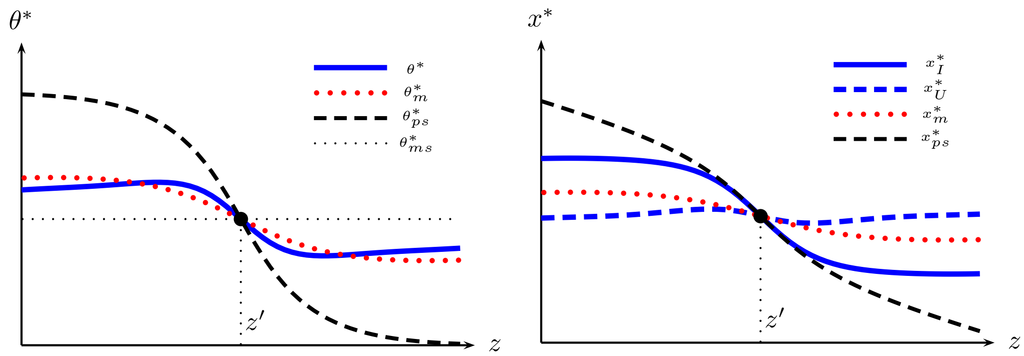
5.5. The Power of Whispers

The following proposition establishes another key result: when the rumor is close to neutral, the equilibrium regime survival threshold is more sensitive to the rumor under the "communication model" than under the "mute model." In other words, when the rumor z is against the regime and is sufficiently close to z', a regime with strength $\theta \in (\theta_m^*(z), \theta^*(z))$ would collapse if communication among citizens is allowed, but would survive if it is not.

Proposition 4. *For z near z' (which equals $\tilde{z}$), a small change in the rumor causes a bigger change in equilibrium regime survival threshold under the "communication model" than under the "mute model," i.e.,*

$$\frac{\mathrm{d}\theta^*(z')}{\mathrm{d}z} < \frac{\mathrm{d}\theta_m^*(z')}{\mathrm{d}z} < 0, \tag{22}$$

Figures 5(a) and 5(b) present the equilibrium thresholds θ^* and the equilibrium cut-off rules x^* for the four models. Around z', the equilibrium θ^* in the "communication model" is steeper than that in the "mute model," as is described in Proposition 4.



(a) regime survival thresholds   (b) cut-off agent types

Figure 5. *Equilibrium regime survival thresholds and equilibrium cut-off agent types in the "communication model" in comparison to other models*

The size of $\mathrm{d}\theta^*/\mathrm{d}z$ is determined by the magnitude of the direct and multiplier effects. Using the fact that $\hat{x}_I(\theta', z') = \hat{x}_U(\theta', z')$, we can differentiate the attack equation (12) to obtain:

$$\frac{\mathrm{d}\theta^*(z')}{\mathrm{d}z} = \frac{\frac{1}{\sigma_x}\phi(\cdot)\left(J\frac{\partial \hat{x}_I}{\partial z} + (1-J)\frac{\partial \hat{x}_U}{\partial z}\right)}{1 - \frac{1}{\sigma_x}\phi(\cdot)\left(-1 + J\frac{\partial \hat{x}_I}{\partial \theta} + (1-J)\frac{\partial \hat{x}_U}{\partial \theta}\right)}, \tag{23}$$

where $\phi(\cdot)$ is evaluated at the point $(x' - \theta')/\sigma_x$. Comparing (23) with its counterpart (18) for the "mute model," we obtain Proposition 4 by showing in the proof that, at the point $\theta = \theta'$ and $z = z'$,

$$J\frac{\partial \hat{x}_I}{\partial \theta} + (1-J)\frac{\partial \hat{x}_U}{\partial \theta} > \frac{\partial \hat{x}_m}{\partial \theta} > 1; \tag{24}$$

$$J\frac{\partial \hat{x}_I}{\partial z} + (1-J)\frac{\partial \hat{x}_U}{\partial z} < \frac{\partial \hat{x}_m}{\partial z} < 0. \tag{25}$$

The direct effect of an increase in z on the mass of attackers is the weighted average of $\partial \hat{x}_I/\partial z$ and $\partial \hat{x}_U/\partial z$, with the weights being J and $1-J$, respectively. By (25), this effect is larger in magnitude with communication than without. The multiplier effect—the effect of an increase in regime survival threshold on the mass of attackers—is determined by the weighted average of $\partial \hat{x}_I/\partial \theta$ and $\partial \hat{x}_U/\partial \theta$. By (24), this effect is also larger in magnitude with communication than without. These two effects combine to give a greater sensitivity of the equilibrium survival threshold to the rumor.

When the true state of nature is $\theta = \theta'$, the mass of attackers is equal to the equilibrium regime survival threshold. Therefore, Proposition 4 also implies that the mass of attackers in state $\theta = \theta'$ is more responsive to the rumor in the "communication model" than in the "mute model." In the "communication model," a slight decrease in z from z' leads to an increase in the size of swing population by $\sigma_x^{-1}\phi(\cdot)(\mathrm{d}x_I^*/\mathrm{d}z - \mathrm{d}x_U^*/\mathrm{d}z)$. A fraction J of the swing population would receive a confirmatory message from their peers and attack the regime. The change in the mass of revolutionaries is $\sigma_x^{-1}\phi(\cdot)\mathrm{d}x_U^*/\mathrm{d}z$. In the "mute model," the increase in the mass of attackers is $\sigma_x^{-1}\phi(\cdot)\mathrm{d}x_m^*/\mathrm{d}z$. At $z = z'$, we must have

$$\frac{1}{\sigma_x}\phi(\cdot)\left[J(\theta', z')\left(\frac{\mathrm{d}x_I^*}{\mathrm{d}z} - \frac{\mathrm{d}x_U^*}{\mathrm{d}z}\right) + \frac{\mathrm{d}x_U^*}{\mathrm{d}z}\right] < \frac{1}{\sigma_x}\phi(\cdot)\frac{\mathrm{d}x_m^*}{\mathrm{d}z} < 0.$$

To illustrate, we plot the mass of attackers $\hat{A}(\theta)$ against the state θ, holding the cut-off rules constant. The equilibrium regime survival threshold is given by the intersection of $\hat{A}(\theta)$ and the 45-degree line. For values of z slightly below z', Figure 6(a) shows that a larger fraction of attackers are mobilized than that in the "mute model" when the regime strength θ is near z. Note that communication does not necessarily increase the mass of attackers at all states and it actually lowers the mass of attackers when θ is far from z.

Communication allows citizens on the whole to be better informed about the true state whenever the true state is near what is suggested by the rumor. Because the function $J(\theta, z)$ peaks at $\theta = z$, more citizens will hear a confirmatory message when the true regime strength θ is close to the rumor z. If the rumor suggests that the regime is weak, this mechanism causes more people to attack when the regime is indeed weak

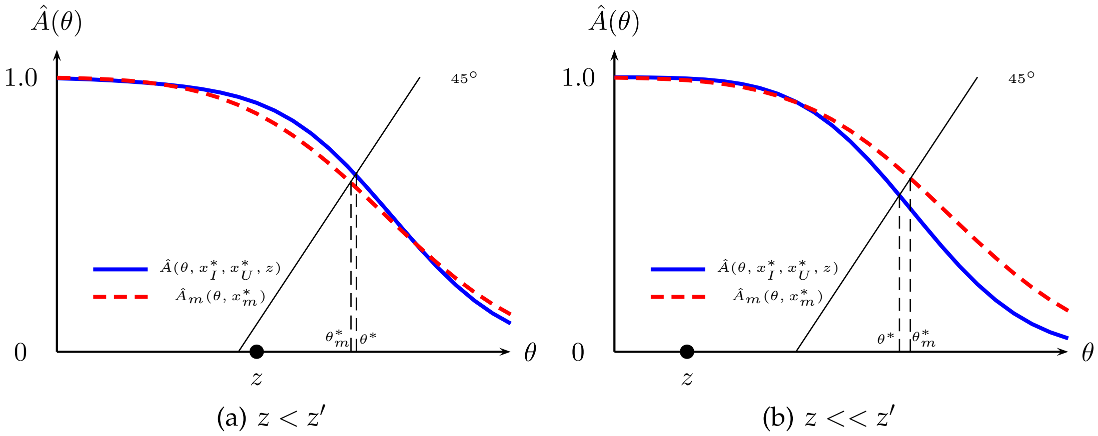


Figure 6. *Mass of attackers in different states in the "mute model" and in the "communication model"*

and close to what the rumor indicates. This explains why $\hat{A}(\theta')$ is higher in the "communication model" than that in the "mute model" at values of θ close to z, or why $\theta^*(z) > \theta_m^*(z)$ for z slightly lower than z'.

For the same reason, the ordering of $\theta^*(z)$ and $\theta_m^*(z)$ may be reversed for z much below z'. This is illustrated in Figure 6(b). When the true regime strength θ is much higher than what the rumor z suggests, $J(\theta, z)$ will be small and many citizens will hear that their peers do not believe the rumor. In this case, communication coordinates citizens not to attack. Figure 6(b) shows that the mass of attackers falls relative to that in the "mute model" when regime is indeed strong and far away from what the rumor suggests.

5.6. Rumors vs. Trustworthy News

In Section 4, we have shown that, due to citizens' skepticism toward rumors, the magnitude of the response of θ_m^* to a change in z is smaller in the "mute model" than that in the "public signal model" at $z = z'$. In the previous subsection, we establish that the effect of skepticism can be, to a certain extent, undone by communication, so that θ^* is more sensitive than θ_m^* is to changes in z around z'.

Interestingly, the effect of communication can be so large that θ^* in the "communication model" is more sensitive to the rumor than is its counterpart θ_{ps}^* in the "public signal model." In other words, when a rumor against the regime is heard by citizens who are allowed to communicate about its informativeness, the regime could survive when all agents believe that the rumor is trustworthy, but could collapse when citizens know that the rumor may be uninformative. Specifically, the slope of equilibrium θ^* can be steeper than θ_{ps}^* at $z = z'$, so that the regime survival threshold θ^* can be higher

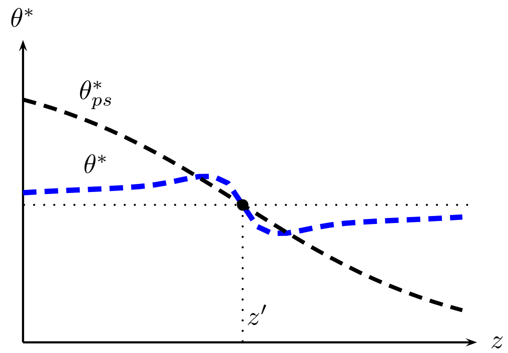


Figure 7. Comparing the "communication model" to the "public signal model" when private information has reasonably high precision

than θ^*_{ps} for z slightly below z'. Figure 7 shows such a possibility.[17]

To understand why this possibility could arise, recall that the sensitivity of θ^* to a change in z around z' is governed by the magnitude of the direct effect and the multiplier effect. Further, since these two effects sum to a constant, we may just focus on the multiplier effect, i.e., $\partial \hat{x}/\partial\theta$. It is shown in equation (19) that, due to the skepticism effect, i.e., $w(z', \hat{x}_m) < 1$, we have

$$\frac{\partial \hat{x}_{ps}}{\partial \theta} > \frac{\partial \hat{x}_m}{\partial \theta}.$$

It is also shown in equation (20) that, due to the the effect of communication, i.e., $J(\theta', z')/\Pr[y_i = 1|z', x'] > 1$, we have

$$\frac{\partial \hat{x}_I}{\partial \theta} > \frac{\partial \hat{x}_m}{\partial \theta}.$$

When the communication effect is large enough, the effect of skepticism can be overcome such that the following holds,

$$\frac{\partial \hat{x}_I}{\partial \theta} > \frac{\partial \hat{x}_{ps}}{\partial \theta}.$$

Moreover, the fraction of population who receive $y = 1$ can be so large that the average response of the population is larger than its counterpart in the "public signal model". That is,

$$J(\theta', z')\frac{\partial \hat{x}_I}{\partial \theta} + (1 - J(\theta', z'))\frac{\partial \hat{x}_U}{\partial \theta} > \frac{\partial \hat{x}_{ps}}{\partial \theta}. \tag{26}$$

Using the same logic as in Proposition 3, this inequality leads to the situation shown in Figure 7, that is,

$$\frac{d\theta^*(z')}{dz} < \frac{d\theta^*_{ps}(z')}{dz} < 0.$$

[17] In plotting this figure, we use the value of $\sigma_x^2 = 0.2$, which is lower than that used in Figure 5(a).

It is interesting to investigate under what conditions inequality (26) could hold. Toward this end, we study how the multiplier effect changes in each model as the variance of private information increases.

With less precise private signals, citizens in the "public signal model" are more responsive to the public information. Specifically the weight given to private information in the standard linear Bayesian updating formula is β. The value of $\partial \hat{x}_{ps}/\partial\theta$ is equal to $1/\beta$, which increases in σ_x^2; see the dashed line Figure 8.

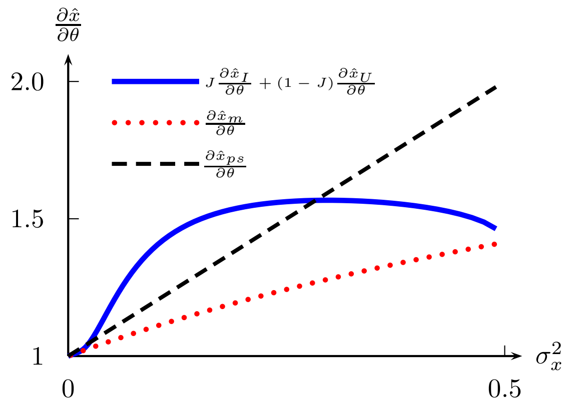


Figure 8. *The magnitude of the multiplier effect in different models as functions of the variance of private information.*

Similarly, $\partial \hat{x}_m/\partial\theta$ increases in σ_x^2 because a less precise private signal implies a smaller weight given to the private signal; see the dotted line in Figure 8. The widening gap between $\partial \hat{x}_m/\partial\theta$ and $\partial \hat{x}_{ps}/\partial\theta$ reflects the effect of skepticism, since an increase in σ_x^2 reduces the weight $w(z', x')$ that citizens attach to the rumor being informative. In other words, when the precision of private information decreases, skepticism becomes stronger, which further dampens citizens' response to public information.

In the "communication model," an increase in σ_x^2 raises both $\partial \hat{x}_I/\partial\theta$ and $\partial \hat{x}_U/\partial\theta$, partially due to the decrease in β. More importantly, the difference $\partial \hat{x}_I/\partial\theta - \partial \hat{x}_{ps}/\partial\theta$ increases with σ_x^2. This reflects the effect of communication, since an increase in σ_x^2 raises $J(\theta', z')/\Pr[y_i = 0|z', x']$.

For σ_x^2 close to 0, $\partial \hat{x}_{ps}/\partial\theta > \partial \hat{x}_I/\partial\theta$. In other words, the communication effect is dominated by skepticism. When σ_x^2 is sufficiently small, the fraction of population receiving confirmatory message approaches 1. Intuitively, if everybody is expected to send $y = 1$, then this message does not reveal much new information and takes little effect when one receives it. Therefore, the weighted average in concern in the "communication model," must be lower than that in the "public signal model."

When σ_x^2 is big, the communication effect could dominate skepticism, i.e., $\partial \hat{x}_I/\partial\theta > \partial \hat{x}_{ps}/\partial\theta$. Intuitively, when citizens are very skeptical of the rumor, it is unlikely to receive $y = 1$. Therefore, receiving such a confirmatory message can take a stronger effect.[18]

However, as σ_x^2 increases, less weight is attached to the larger value $\partial \hat{x}_I / \partial \theta$ because J falls. These two opposing effects account for the hump shape of $J \partial \hat{x}_I / \partial \theta + (1 - J) \partial \hat{x}_U / \partial \theta$ shown by the solid curve in Figure 8. For very small or very large values of σ_x^2, this curve is close to $\partial \hat{x}_m / \partial \theta$, and is always above it by equation (24). For intermediate values of σ_x^2, the effect of communication could be so large that inequality (26) holds.

The hump-shaped response is novel. It means that an increase in precision of private information can lead to a stronger reaction to public information, as in the downward-sloping part of the solid curve in Figure 8. In other words, citizens as a whole act as if they were more responsive to the rumor when their private information is more precise, which is opposite to what the standard linear Bayes' updating formula predicts. In this "communication model," public and private information can be either substitutes and complements, depending on the level of private information precision.[19]

6. Discussion

6.1. Mechanisms and Robustness

In the baseline model, we have assumed that communication takes the form of exchanging coarse (binary) signals about the informativeness of the rumor. In this subsection, we demonstrate that coarsening is not crucial for our results, although it simplifies the analysis and captures realistic features of information exchange among individuals. We show that it is the exchange of assessment of the rumor's informativeness which drives our results, and that this mechanism is qualitatively different from the case where the content of communication is unrelated to the rumor. To stress this point, we explore two alternative communication protocols, which allow citizens to exchange the exact value of $w(z, x_i)$, or the exact value of x_i.

Specifically, suppose citizens can tell each other the value of w (instead of just whether w is greater than or less than δ). Citizens i updates her belief in a Bayesian

[18] In this case receiving a contradictory message, $y = 0$, takes little effect. That is, $\partial \hat{x}_U / \partial \theta$ is lower than but very close to $\partial \hat{x}_m / \partial \theta$.

[19] The complementarity between public and private information has been recently discovered in some learning models (Duffie, Malamud, and Manso 2010 and Veldkamp 2011). Our mechanism differs from previous work in that the complementarity is a result of changes in the size of different groups who receive different messages when communication is allowed.

fashion when she receives the message $w_k = w(z, x_k)$ from her peer k:

$$P(\theta|z, x_i, w_k) = \frac{\int_{-\infty}^{\theta} j(t, z, w_k) p(t|z, x_i) \, \mathrm{d}t}{\int_{-\infty}^{\infty} j(t, z, w_k) p(t|z, x_i) \, \mathrm{d}t},$$

where

$$j(t, z, w_k) = \frac{1}{\sigma_x} \phi \left(\frac{x_l - t}{\sigma_x} \right) + \frac{1}{\sigma_x} \phi \left(\frac{x_r - t}{\sigma_x} \right),$$

with x_l and x_r being the two solutions to the equation $w(z, x_k) = w_k$.

The function $j(t, z, w_k)$ reaches a local maximum at $t = z$ if w_k is high (x_l and x_r are near z), or a local minimum at $t = z$ if w_k is low (x_l and x_r are far from z). Moreover, at $z = z'$, there exists a $\hat{w}$ such that the ratio, $j(\theta', z', w_k) / \int_{-\infty}^{\infty} j(t, z', w_k) p(t|z', x') \, \mathrm{d}t$, is greater than 1 if $w_k > \hat{w}$. Therefore, citizens who receive a message $w_k > \hat{w}$ in this model are similar to citizens who receive a message $y = 1$ in the baseline model, with $J(\theta', z') / \Pr[y = 1|z', x'] > 1$ that gives rise to the communication effect. Similarly, the ratio $j(\theta', z', w_k) / \int_{-\infty}^{\infty} j(t, z', w_k) p(t|z', x') \, \mathrm{d}t$ is less than 1 if $w_k < \hat{w}$. Therefore, citizens who receive a message $w_k < \hat{w}$ in this model are similar to citizens who receive a message $y = 0$ in the baseline model, with $(1 - J(\theta', z')) / \Pr[y = 0|z', x'] < 1$.

Intuitively, when citizen k conveys a high assessment that the rumor is likely to be informative, i.e., $w_k > \hat{w}$, her peer i assigns a higher probability weight (density) to states close to z. Similarly, when citizen k does not believe the rumor is sufficiently informative, i.e., $w_k < \hat{w}$, citizen i assigns lower weight to states close to z and higher weight to states far away from z. Given that this probability re-weighting is the key mechanism that drives our results in the baseline case, a model of exchanging the exact probability assessments of the rumor's informativeness is not qualitatively different from a model of exchanging coarse information about these assessments.

Now we turn to the case where citizens directly exchange their private information concerning the strength of the regime. To be more general, we allow the communication process to be noisy: each citizen receives her peer's private signal with an additive noise. That is, the message y_i received by the citizen i from her peer k is given by:

$$y_i = x_k + \xi_k,$$

where the noise $\xi_k \sim \mathcal{N}(0, \sigma_\xi^2)$ is normally distributed and independent across k. At the end of communication stage, each citizen possesses an information set which consists of the rumor z and two private signals x_i and y_i of different qualities. This information set is equivalent to (z, xy_i), where xy_i is a private signal with higher precision than x_i. That is,

$$xy_i \equiv \frac{\sigma_x^2 + \sigma_\xi^2}{2\sigma_x^2 + \sigma_\xi^2} x_i + \frac{\sigma_x^2}{2\sigma_x^2 + \sigma_\xi^2} y_i,$$
$$\sigma_{xy}^2 \equiv \frac{\sigma_x^2 + \sigma_\xi^2}{2\sigma_x^2 + \sigma_\xi^2} \sigma_x^2 < \sigma_x^2.$$

In other words, this setting is observationally equivalent to the "mute model," with private signals of a higher quality. An increase in the precision of private information causes citizens to be less responsive to the rumor in the "mute model." The resulting change in equilibrium thresholds is shown in Figure 9, which presents the equilibrium survival threshold in the "communication model," the "mute model," and the model with the modified assumption concerning the exchange of private signals (labeled θ_{ex}^*). The same set of parameters is used throughout these examples; the only difference is the communication protocol.[20]

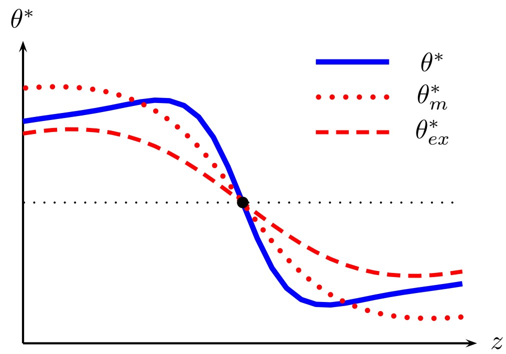


Figure 9. *Equilibrium survival threshold for regime strength is less response to rumors when citizens communicate with one another about their private signals.*

The contrast between these two alternative models highlights that the mode of communication matters: when we allow citizens to exchange what they privately know, they rely more on private information; when we allow them to exchange views on the public signal that they commonly observe, they rely more on public siganl.[21]

Our baseline specification on communication is the simplest possible one that allows citizens to exchange views on the informativeness of the rumor. We can relax

[20]The same conclusion still holds if we consider a model in which citizens exchange coarsened information about their private signals by sending the message $y = 1$ if $x > \bar{x}$, or $y = 0$ if $x \leq \bar{x}$, for some cutoff value $\bar{x}$.

[21]We can also have a hybrid model by allowing citizen i to exchange private information with a random peer k and then randomly meet another citizen k' to exchange views on informativeness of the rumor. Interestingly, given the result in Section 5.6 (see Figure 8), more precise private information arising from the exchange of private information does not necessarily undermine citizens' overall response to the rumor.

various simplifying assumptions without changing the main results of our analysis. For example, in the benchmark model, we assume that every citizen talks about the rumor and communicates truthfully without any error. If citizens meet one another with probability $q < 1$ or mis-communicate with a small probability $\mu < 0.5$ (i.e., mistakenly send a confirmatory message when a contradictory message is intended, or vice versa), then the effect of communication increases in q and decreases in μ.

We also investigate a specification where finer messages are exchanged between citizens. We allow citizens to justify and explain why they do not believe the rumor is informative if they send a contradictory message. Specifically, citizen k can send her peer a message $y = 0L$, interpreted to mean "the rumor is not informative because it indicates that the regime is too much stronger than I believe," when $x_k < \underline{x}(z)$; she sends $y = 0R$ when $x_k > \overline{x}(z)$; and she sends $y = 1$ when $x_k \in [\underline{x}(z), \overline{x}(z)]$. In the baseline model, the key mechanism, which leads to the result that rumors can be even more powerful than trustworthy news, is that the effect of communication could dominate skepticism when $y = 1$ is received, i.e., $\partial x_I / \partial \theta > \partial x_{ps} / \partial \theta$, and the fraction of population who receives $y = 1$, i.e., $J(\theta', z')$, can be sufficiently large. The same driving force could also dominate in this specification and lead to a similar result.

6.2. Sentiment

We interpret the mean of the uninformative distribution as sentiment, i.e., the public's perception of what uninformative messages would sound like. Given that the uninformative distribution U has a density that peaks at $z = s$, the probability that the rumor is informative decreases in the distance between z and s. If a rumor is close to what the public perceives as untrustworthy information, it will be given less credibility by citizens. Therefore, sentiment critically affects citizens' evaluation of the rumor's informativeness. For example, if citizens have lived with systematic government propaganda for a long time, any claims about its alleged strength would be considered less credible.

Suppose $s = z'$. We call this value "neutral sentiment," since the equilibrium regime survival threshold is symmetric in s about the point $s = z'$. If s is exogenous decreased from the neutral value, then for any $z > z'$, $\theta^*(z)$ goes down, indicating that the regime is more likely to survive. The key intuition is that when s is low, citizens believe that a rumor in favor of the regime ($z > z'$) is likely to have come from an informative source. As a result, they believe the regime is strong and are therefore less prepared to attack. When the rumor is against the regime ($z < z'$), citizens tend to attribute the rumor to an uninformative source. Because they do not believe that the regime is weak, they are again less likely to attack. Similar analysis shows the opposite happens when s increases exogenously from the neutral level. See the left panel of

Figure 10 for an illustration of these two cases.

The right panel of Figure 10 shows the ex ante survival probability for each θ, which is computed by integrating over z using the mixture distribution. When s is exogenously shifted up, it is less likely for the regime to survive.

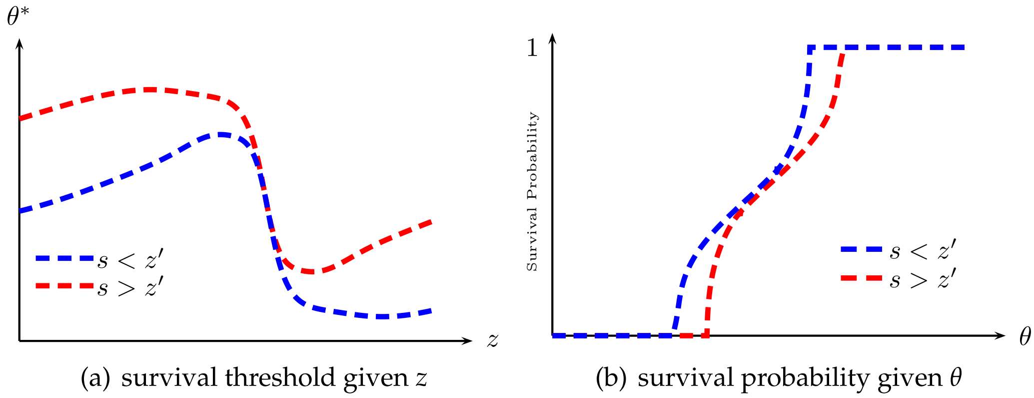


Figure 10. *Lower value of s tends to increase the chance of regime survival*

Thus autocratic regimes that understand the importance of public perception could increase their chances of survival by manipulating it. During the post-election Iranian protest in 2009, the Iranian government deliberately created rumors against itself and then disproved it on national television, which turned out to be effective in discrediting the opposition (Esfandiari 2010).

In reality, sudden shifts in sentiment often take place during revolutions. For instance, the rumor that Mubarak's family had fled Egypt would have sounded ridiculous without the previous successful Tunisian revolution, during which Ben Ali actually escaped. The fact that it was widely reported and believed during the Egyptian unrest was a sign of sentiment shift. In this case, the confirmed rumor in Tunisia shifted the sentiment upward in Egypt and made the regime more vulnerable.

6.3. Censorship: The Power of Silence

Our model shows that rumors against a regime could coordinate more citizens to attack, provided they are close to the truth. Thus autocratic governments may want to block negative rumors against it and to stop citizens from talking about these rumors. In reality, news censorship and the control of information flow are commonly adopted by many governments concerned about their survival. However, does it always help the regime to survive if censorship is adopted to screen out all negative information? In this subsection we investigate this issue with our model. When citizens hear no rumor about the regime, silence itself becomes a public signal about the regime strength, and it is not obvious that news censorship necessarily increases the chances of regime survival.

Assume that the regime blocks any rumor z if $z < K$. In other words, only rumors that suggest the regime is stronger than K can be heard and discussed by citizens. We assume that citizens are aware of this censorship rule. Therefore it is common knowledge that $z < K$ when citizens do not hear a rumor. In this case, citizens cannot communicate about the informativeness of z, since they do not have any knowledge of it.

When z is observable, i.e., $z \geq K$, the equilibrium is exactly the same as in our "communication model." When citizens hear no rumor, they understand that the event $z < K$ has taken place but the authority has censored the news. Taking θ_c as the threshold for regime survival, they calculate the expected payoff of revolt in a Bayesian fashion,

$$\Pr[\theta \leq \theta_c | z < K, x_i] = \frac{\int_{-\infty}^{\theta_c} \left[\alpha\Phi\left(\frac{K-t}{\sigma_z}\right) + (1-\alpha)\Phi\left(\frac{K-s}{\sigma_U}\right)\right] \frac{1}{\sigma_x}\phi\left(\frac{t-x_i}{\sigma_x}\right) \mathrm{d}t}{\int_{-\infty}^{\infty} \left[\alpha\Phi\left(\frac{K-t}{\sigma_z}\right) + (1-\alpha)\Phi\left(\frac{K-s}{\sigma_U}\right)\right] \frac{1}{\sigma_x}\phi\left(\frac{t-x_i}{\sigma_x}\right) \mathrm{d}t}.$$

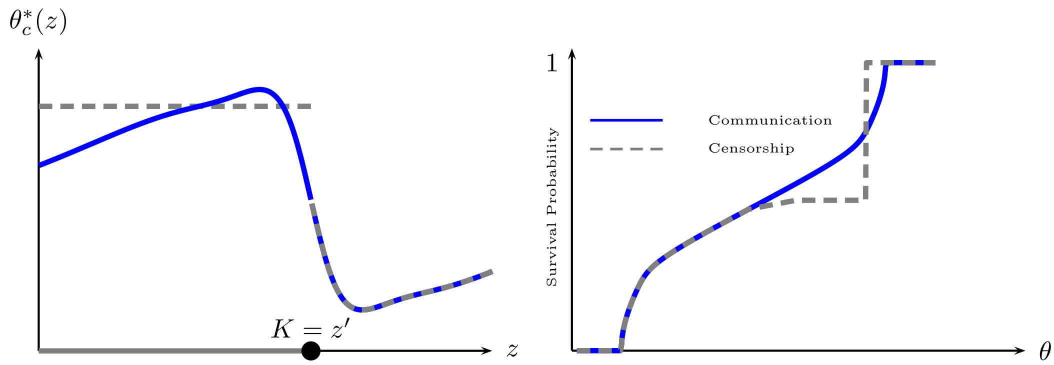
The equilibrium values of θ_c^* depends on the censorship rule K. When K is very small, it corresponds to the case where there is no effective censorship. When K is very high, the regime blocks all the public information, no matter whether it is for or against the regime. Therefore, in either case, θ_c^* converges to the value θ_{ms}^* in the "pure noise model."

We find that θ_c^* is increasing then decreasing in K. To understand why θ_c^* can be increasing, suppose that the censorship rule is raised from an initially low value of K to a slightly higher value K'. This means that rumors with realizations in $[K, K']$ are now screened out. In a model with trustworthy news only, $z \in [K, K']$ is better news for the regime than $z < K$. So the expected strength of the regime should be higher conditional on $z < K'$ than conditional on $z < K$. In our model, however, rumors with $z \in [K, K']$ can be *worse* news for the regime than those with $z < K$, because rumors strongly against the regime are deemed incredible. This explains why citizens' expectation about the strength of the regime conditional on hearing no news can be *lower* when the censorship rule is tightened. As the expected payoff from attacking upon hearing no news rises, the equilibrium threshold for regime survival θ_c^* becomes higher.

Now we return to the issue raised at the beginning of this subsection and examine the effect of censorship on regime survival in the case where $K = z'$, i.e., the regime blocks any rumors against it. See Figure 11(a) for illustration. The effect of censorship is represented by the difference between the solid line for the "communication model" and the dashed line for the model with censorship in the left panel.

It is worth pointing out that the "censorship model" is a variant of the "mute model" when $z < K$, since citizens cannot communicate when there is no rumor to talk about. In a sense, citizens pool the effects of negative rumors, both more and less dangerous ones, and use an average to make decision, given they cannot observe the specific realization of z. It turns out that censorship does help the regime to survive when the realizations of rumors are the most dangerous ones (i.e., when $\theta^*(z)$ is around the peak value). But when z is very far or close to z', censorship hurts the regime.

Figure 11(b) displays the ex ante survival probability in different states with and without censorship. When θ is sufficiently low, censorship makes no difference: if θ is lower than both θ_c^* and θ^*, the regime collapses in both cases. If regime strength is sufficiently high but lower than θ_c^*, censorship actually lowers the probability of survival. Not surprisingly, censorship becomes beneficial to the regime when $\theta > \theta_c^*$: the regime must survive with censorship, while there are chances that it collapses without it.



(a) the effects of censorship (b) survival probability in different states

Figure 11. *The "censorship model"*

7. Conclusion

Social interaction is an important source of information for individuals, especially when the coordination motive is important. Our paper highlights the significance of this channel and contributes to this interesting but under-exploited topic.

It is not news that revolutions in history are often intertwined with rumors. It is also not surprising that many rumors are intended to spur individuals' participation. However, what strikes us is why some rumors, which often turned out to be false later, could be so effective for mobilization, while others were simply ignored and left no trace at all. We offer an analysis of this phenomenon in a global game framework. In this model, individuals' skepticism toward rumors arises as a rational response, instead of a behavioral assumption. Moreover, we explicate a novel mechanism, where the effect of citizens' skepticism can be undone or reinforced by communication among themselves. We stress that communication about the informativeness of the rumor is the key.

To the best of our knowledge, our model is the first attempt to explicitly investigate the role of rumors in a regime change game. In this paper, we choose to focus on two key aspects of rumors: that they may or may not be true, and that people talk about them. We provide a framework using mixture distributions to study unreliable information. Such a framework may be useful in other applications. Our theory is interpreted in the context of political revolution, but it can also be extended to model rumors in bank runs, financial crises or currency attacks. We have not, however, addressed questions about how rumors originate or how they spread. Although we explore a number of communication protocols in this paper, our analysis is confined to decentralized communication with random pairwise matching. The role played by social networks, mass media, and modern communication technologies in promulgating or abating rumors remains to be studied.

Appendix

Proof of Lemma 1. Let the type 1 error cost (saying one believes the rumor when it turns out to be uninformative) and the type 2 error cost (saying one does not believe the rumor when it turns out to be informative) be c_1 and c_2, respectively. We allow probabilistic decision rules of the form $Y(z, x_i) \in [0, 1]$, where Y is the probability of sending the message $y_k = 1$ to one's peer k. Agent i chooses a decision rule $Y(z, x_i)$ to minimize her expected loss:

$$\alpha \int_{-\infty}^{\infty} (1 - Y(z, x_i))\, c_2 \frac{1}{\sigma_I} \phi \left(\frac{z - x_i}{\sigma_I} \right) \mathrm{d}z + (1 - \alpha) \int_{-\infty}^{\infty} Y(z, x_i) c_1 \frac{1}{\sigma_U} \phi \left(\frac{z - s}{\sigma_U} \right) \mathrm{d}z.$$

The objective function is linear in Y pointwise. Hence the optimal solution is a bang-bang solution:

$$Y(z, x_i) = \begin{cases} 1, & \text{if } \alpha c_2 \frac{1}{\sigma_I} \phi \left(\frac{z - x_i}{\sigma_I} \right) \geq (1 - \alpha) c_1 \frac{1}{\sigma_U} \phi \left(\frac{z - s}{\sigma_U} \right); \\ 0, & \text{otherwise.} \end{cases}$$

This solution can be re-arranged to give $Y(z, x_i) = 1$ if and only if:

$$\frac{\alpha \sigma_I^{-1} \phi \left(\sigma_I^{-1} (z - x_i) \right)}{\alpha \sigma_I^{-1} \phi \left(\sigma_I^{-1} (z - x_i) \right) + (1 - \alpha) \sigma_U^{-1} \phi \left(\sigma_U^{-1} (z - s) \right)} \geq \frac{c_1}{c_1 + c_2},$$

and $Y(z, x_i) = 0$ otherwise. The left-hand side of the above is simply $\Pr[z \sim I | z, x_i]$. If we let $\delta = c_1 / (c_1 + c_2)$, then the optimal decision rule is the same as the communication rule (3). ∎

Proof of Proposition 2. *Part (1).* Since $(\theta^*_{ms}, x^*_{ms})$ satisfies the indifference condition of the "pure noise model," we have

$$\Phi \left(\frac{\theta^*_{ms} - x^*_{ms}}{\sigma_x} \right) = c.$$

By part (3) of Proposition 1, $(\theta^*_{ms}, x^*_{ms})$ also satisfies the indifference condition of the "public signal model" when $z = \tilde{z}$. Therefore,

$$\Phi \left(\frac{\theta^*_{ms} - (\beta x^*_{ms} + (1 - \beta) \tilde{z})}{\sqrt{\beta} \sigma_x} \right) = c.$$

The posterior belief $P(\theta^*_{ms} | \tilde{z}, x^*_{ms})$ in the "mute model" is just the weighted average of the left-hand-side of the two equations above. Hence $P(\theta^*_{ms} | \tilde{z}, x^*_{ms}) = c$. Furthermore, $(\theta^*_{ms}, x^*_{ms})$ satisfies the attack equation of the "mute model." Therefore, it is an equilibrium for $z = \tilde{z}$.

Part (2). We first show that if $z < \tilde{z}$, then $\theta^*_m(z) < \theta^*_{ps}(z)$. From the attack equation (16), we can write $\Phi\left(\sigma_x^{-1}(\theta^*_m - x^*_m)\right) = 1 - \theta^*_m$. Therefore, the indifference condition (15) can be written as:

$$c = w(z, x^*_m)\Phi\left(\frac{\theta^*_m - (\beta x^*_m + (1-\beta)z)}{\sqrt{\beta}\sigma_x}\right) + (1 - w(z, x^*_m))(1 - \theta^*_m).$$

From the indifference condition of the "public signal model" and from the fact that $1 - \theta^*_{ms} = c$, we also have

$$c = w(z, x^*_m)\Phi\left(\frac{\theta^*_{ps} - (\beta x^*_{ps} + (1-\beta)z)}{\sqrt{\beta}\sigma_x}\right) + (1 - w(z, x^*_m))(1 - \theta^*_{ms}).$$

These two equations, together with the fact that $\theta^*_{ps} > \theta^*_{ms}$ when $z < \tilde{z}$, imply

$$\begin{aligned} & w(z, x^*_m)\Phi\left(\frac{q(\theta^*_m) - (1-\beta)z}{\sqrt{\beta}\sigma_x}\right) + (1 - w(z, x^*_m))(1 - \theta^*_m) \\ > \; & w(z, x^*_m)\Phi\left(\frac{q(\theta^*_{ps}) - (1-\beta)z}{\sqrt{\beta}\sigma_x}\right) + (1 - w(z, x^*_m))(1 - \theta^*_{ps}), \end{aligned}$$

where

$$q(\theta^*) \equiv \theta^* - \beta x^* = \theta^* - \beta(\theta^* + \sigma_x \Phi^{-1}(\theta^*)).$$

To show $\theta^*_m < \theta^*_{ps}$ from the above inequality, it suffices to show that $\mathrm{d}q(\theta^*)/\mathrm{d}\theta^* \leq 0$. We have

$$\frac{\mathrm{d}q(\theta^*)}{\mathrm{d}\theta^*} = 1 - \beta - \frac{\beta\sigma_x}{\phi\left(\Phi^{-1}(\theta^*)\right)} \leq 1 - \beta - \beta\sigma_x\sqrt{2\pi},$$

which is negative by assumption (1). Hence $\theta^*_m < \theta^*_{ps}$ when $z < \tilde{z}$.

Next, we show that $\theta^*_m(z) > \theta^*_{ms}$ if $z < \tilde{z}$. Suppose this is not true. Then $\Phi(\sigma_x^{-1}(\theta^*_m - x^*_m)) = 1 - \theta^*_m \geq 1 - \theta^*_{ms} = c$, which implies

$$\Phi\left(\frac{\theta^*_m - (\beta x^*_m + (1-\beta)z)}{\sqrt{\beta}\sigma_x}\right) \leq c.$$

Moreover, the fact that $(\theta^*_{ms}, x^*_{ms})$ satisfies the indifference condition of the "public signal model" at $z = \tilde{z}$ implies

$$\Phi\left(\frac{\theta^*_{ms} - (\beta x^*_{ms} + (1-\beta)\tilde{z})}{\sqrt{\beta}\sigma_x}\right) = c.$$

These two conditions can be combined to give

$$q(\theta_m^*) - q(\theta_{ms}^*) \leq (1-\beta)(z-\tilde{z}) < 0.$$

Since $\mathrm{d}q(\theta^*)/\mathrm{d}\theta^* < 0$, this inequality implies $\theta_m^* > \theta_{ms}^*$, a contradiction. Thus, when $z < \tilde{z}$, we must have $\theta_m^*(z) \in (\theta_{ms}^*, \theta_{ps}^*(z))$. The proof of part (2b) is symmetric.

Part (3). Combine equations (15) and (16) to get

$$\theta_m^* = (1-c) + \left[\Phi\left(\frac{\theta_m^* - [\beta x_m^* + (1-\beta)z]}{\sqrt{\beta}\sigma_x}\right) - c\right] \cdot \frac{w(z, x_m^*)}{1 - w(z, x_m^*)}.$$

As z goes to infinity, $x_m^*(z)$ must remain finite, otherwise it would violate equation (16). For any finite x_m^*, $\lim_{z\to\infty} w(z, x_m^*) = 0$. Therefore,

$$\lim_{z\to\infty} \theta_m^*(z) = 1 - c = \theta_{ms}^*.$$

A similar argument establishes that $\lim_{z\to-\infty} \theta_m^*(z) = \theta_{ms}^*$.

Part (4). We show that (a) $\theta_m^*(z)$ is increasing then decreasing for $z \in (-\infty, \tilde{z}]$; and (b) $\theta_m^*(z)$ is decreasing then increasing for $z \in [\tilde{z}, \infty)$.

Fix a $z_0 \in (-\infty, \tilde{z}]$. Define

$$f(z) \equiv P(\theta_m^*(z_0)|z, x_m^*(z_0)).$$

We show that f is single-peaked in z for $z \in (-\infty, \tilde{z}]$. It suffices to verify that $\mathrm{d}f(z)/\mathrm{d}z = 0$ implies $\mathrm{d}^2 f(z)/\mathrm{d}z^2 < 0$. To simplify the notation, we use the subscript I or U to denote the posterior distribution (or density) when z is known to be informative or uninformative, respectively. We have

$$\begin{aligned}\frac{\mathrm{d}f(z)}{\mathrm{d}z} &= (\Phi_I - \Phi_U)\frac{\partial w}{\partial z} - w\frac{\partial \Phi_I}{\partial z} \\ &= \left[(1-w)\left(\frac{z-s}{\sigma_U^2} + \frac{x_m^*(z_0) - z}{\sigma_I^2}\right)\left(1 - \frac{\Phi_U}{\Phi_I}\right) - \frac{1-\beta}{\sqrt{\beta}\sigma_x}\frac{\phi_I}{\Phi_I}\right] w\Phi_I.\end{aligned}$$

When $\mathrm{d}f(z)/\mathrm{d}z = 0$, the second derivative is given by:

$$\begin{aligned}\frac{1}{w\Phi_I}\frac{\mathrm{d}^2 f(z)}{\mathrm{d}z^2} = &-w(1-w)\left(\frac{z-s}{\sigma_U^2}+\frac{x_m^*(z_0)-z}{\sigma_I^2}\right)^2\left(1-\frac{\Phi_U}{\Phi_I}\right)\\ &+(1-w)\left(\frac{1}{\sigma_U^2}-\frac{1}{\sigma_I^2}\right)\left(1-\frac{\Phi_U}{\Phi_I}\right)\\ &+(1-w)\left(\frac{z-s}{\sigma_U^2}+\frac{x_m^*(z_0)-z}{\sigma_I^2}\right)\left(\frac{\mathrm{d}\left(1-\Phi_U/\Phi_I\right)}{\mathrm{d}z}\right)\\ &-\left(\frac{1-\beta}{\sigma_x\sqrt{\beta}}\right)\frac{\mathrm{d}\left(\phi_I/\Phi_I\right)}{\mathrm{d}z}.\end{aligned}$$

(i) The first term is negative because $\Phi_I > \Phi_U$ for $z < \tilde{z}$. To see this, suppose the contrary is true. Then $\Phi_I \leq \Phi_U$ implies

$$\theta_m^*(z_0) + \sqrt{\beta}x_m^*(z_0) \leq (1+\sqrt{\beta})z.$$

But for $z_0 \leq \tilde{z}$ the left-hand side is greater than $\theta_{ms}^* + \sqrt{\beta}x_{ms}^*$, which is equal to $(1+\sqrt{\beta})\tilde{z}$, a contradiction. (ii) The second term is negative because $\sigma_U^2 > \sigma_I^2$ from assumption (2). (iii) The third term is negative because $(z-s)/\sigma_U^2 + (x_m^*(z_0)-z)/\sigma_I^2 > 0$ whenever $\mathrm{d}f(z)/\mathrm{d}z = 0$ and because $1-\Phi_U/\Phi_I$ is decreasing in z. (iv) The fourth term is negative because the function ϕ_I/Φ_I is increasing in z.

The single-peakedness of $f(z)$ for $z \in (-\infty, \tilde{z}]$ implies that in this range there can be at most one $z_1 \neq z_0$ such that $\theta_m^*(z_1) = \theta_m^*(z_0)$. Suppose otherwise. Let $z_1 \neq z_2 \neq z_0$ be such that $\theta_m^*(z_1) = \theta_m^*(z_2) = \theta_m^*(z_0)$. By the attack equation (16), this implies $x_m^*(z_1) = x_m^*(z_2) = x_m^*(z_0)$. Since $(\theta_m^*(z_0), x_m^*(z_0))$ satisfies the equilibrium conditions for $z \in \{z_1, z_2, z_0\}$, the equation $f(z) = c$ has at least three solutions, which contradicts the single-peakedness of f.

Parts (1), (2a), and (3) of the proposition, coupled with the fact that there is at most one $z_1 \neq z_0$ such that $\theta_m^*(z_1) = \theta_m^*(z_0)$, together establish that $\theta_m^*(z)$ is increasing then decreasing for $z \in (-\infty, \tilde{z}]$.

For the case $z \in [\tilde{z}, \infty)$, write:

$$\frac{\mathrm{d}f(z)}{\mathrm{d}z} = \left[(1-w)\left(\frac{z-s}{\sigma_U^2}+\frac{x_m^*(z_0)-z}{\sigma_I^2}\right)\left(\frac{1-\Phi_U}{1-\Phi_I}-1\right)-\frac{1-\beta}{\sqrt{\beta}\sigma_x}\frac{\phi_I}{1-\Phi_I}\right]w(1-\Phi_I).$$

We can show that the bracketed term is increasing when it is equal to zero, because $\Phi_I < \Phi_U$ when $z > \tilde{z}$ and because $\phi_I/(1-\Phi_I)$ is decreasing in z. Hence $f(z)$ must be decreasing then increasing in z in this range, which, together with parts (1), (2b), and

(3) of the proposition, imply that $\theta_m^*(z)$ is decreasing then increasing for $z \in [\tilde{z}, \infty)$. ■

Proof of Proposition 3. *Part (1).* We only prove the limiting properties of $x_U^*(z)$ and $\theta^*(z)$ here. The limiting properties of $x_I^*(z)$ require comparing the rates of convergence of different functions, and they are formally established in Lemma 3 in the Technical Appendix.

Suppose the limit values of both $x_U^*(z)$ and $\theta^*(z)$ are finite. The indifference condition (11) in "communication model" requires

$$\lim_{z\to\infty} \int_{-\infty}^{\theta^*} \frac{1 - J(t,z)}{\Pr[y_i = 0 | z, x_U^*]} p(t|z, x_U^*)\, \mathrm{d}t = c.$$

By Lemma 3 (Claim 1) in the Technical Appendix, both $\underline{x}(z)$ and $\overline{x}(z)$ go to infinity as z goes to infinity. Therefore, for any $t \leq \theta^*$, the probability that x_j does not belong to $[\underline{x}(z), \overline{x}(z)]$ goes to one. We thus have

$$\lim_{z\to\infty} \frac{1 - J(t,z)}{\Pr[y_i = 0 | z, x_U^*]} = 1.$$

The indifference condition for type x_U^* becomes:

$$\lim_{z\to\infty} \int_{-\infty}^{\theta^*} p(t|z, x_U^*)\, \mathrm{d}t = \Phi\left(\frac{\theta^* - x_U^*}{\sigma_x}\right) = c,$$

where the first equality follows because $w(z, x_U^*)$ goes to 0 as z goes to infinity.

When z goes to infinity, $J(\theta^*, z)$ goes to zero. Therefore the attack equation (12) for the "communication model" becomes,

$$\lim_{z\to\infty} J(\theta^*, z)\Phi\left(\frac{x_I^* - \theta^*}{\sigma_x}\right) + (1 - J(\theta^*, z))\Phi\left(\frac{x_U^* - \theta^*}{\sigma_x}\right) = \Phi\left(\frac{x_U^* - \theta^*}{\sigma_x}\right) = \theta^*.$$

Given (θ^*, x_U^*) solves the same equation system as that in the "pure noise model," we conclude that $\lim_{z\to\infty} x_U^*(z) = x_{ms}^*$ and $\lim_{z\to\infty} \theta^*(z) = \theta_{ms}^*$. The proof of the case for the limit as z goes to minus infinity is analogous.

Part (2). From part (1), $x_I^*(z) > x_U^*(z)$ for z sufficiently small and $x_I^*(z) < x_U^*(z)$ for z sufficiently large. Both $x_I^*(z)$ and $x_U^*(z)$ are continuous. Therefore there exists a z' such that $x_I^*(z') = x_U^*(z')$.

Let $\theta^*(z') = \theta'$ and $x_I^*(z') = x_U^*(z') = x'$. We proceed to establish that (θ', x') solves the "mute model" as well. To see this, we first note that $x_I^* = x_U^*$ implies that the mass of the swing population is zero. Hence, the attack equation (12) of the "communication model" reduces to the attack equation (16) of the "mute model." Next, note that for any value of z', x' and θ', we have

$$P(\theta'|z',x') = \Pr[y_i = 1|z',x']P(\theta'|z',x',1) + \Pr[y_i = 0|z',x']P(\theta'|z',x',0).$$

Therefore, if z', x', and θ' satisfy the indifference conditions $P(\theta'|z',x',1) = c$ and $P(\theta'|z',x',0) = c$ in the "communication model," then they must satisfy the indifference condition $P(\theta'|z',x') = c$ in the "mute model" as well.

Part (3). Let $\theta^*_m(\tilde{z}) = \tilde{\theta}$ and $x^*_m(\tilde{z}) = \tilde{x}$. We need to show that $(\tilde{\theta},\tilde{x},\tilde{x})$ solves the "communication model" at $z = \tilde{z}$. Toward this end, it suffices to show that $P(\tilde{\theta}|\tilde{z},\tilde{x},1) = c$. Given that $(\tilde{\theta},\tilde{x})$ solves the "mute model" at $z = \tilde{z}$, we have $P(\tilde{\theta}|\tilde{z},\tilde{x}) = c$. These two conditions would imply that $P(\tilde{\theta}|\tilde{z},\tilde{x},0) = c$. Hence the indifference condition (11) for the "communication model" is satisfied. Given that $x^*_I(\tilde{z}) = x^*_U(\tilde{z}) = \tilde{x}$, the attack equation (12) holds as well.

Both $J(t,\tilde{z})$ and $p(t|\tilde{z},\tilde{z})$ are symmetric about the point $t = \tilde{z}$, which gives

$$\int_{-\infty}^{\tilde{z}} J(t,\tilde{z})p(t|\tilde{z},\tilde{z})\,\mathrm{d}t = 0.5\int_{-\infty}^{\infty} J(t,\tilde{z})p(t|\tilde{z},\tilde{z})\,\mathrm{d}t$$

Hence, $P(\tilde{z}|\tilde{z},\tilde{z},1) = 0.5$. For $c = 0.5$, $\tilde{\theta} = \tilde{x} = \tilde{z}$. Therefore, we have $P(\tilde{\theta}|\tilde{z},\tilde{x},1) = c$. This establishes that $\theta^*(\tilde{z}) = \theta^*_m(\tilde{z})$ and $x^*_I(\tilde{z}) = x^*_U(\tilde{z}) = x^*_m(\tilde{z})$.

To show that $x^*_I(z) > x^*_U(z)$ for any $z < \tilde{z}$, suppose the contrary is true. Then, since $x^*_I(z) > x^*_U(z)$ in the limit as z goes to minus infinity, there must be an $z_0 < \tilde{z}$ such that $x^*_I(z_0) = x^*_U(z_0)$. The argument in part (2) shows that the "communication model" and the "mute model" are equivalent whenever $x^*_I = x^*_U$. We therefore have $x^*_I(z_0) = x^*_U(z_0) = x^*_m(z_0) \equiv x_0$ and $\theta^*(z_0) = \theta^*_m(z_0) \equiv \theta_0$.

Since (θ_0, x_0) solves "mute model," by Proposition 2, we must have $\theta_0 > \tilde{z} > z_0$ and $x_0 > \tilde{z} > z_0$. Let $X_0 = \beta x_0 + (1-\beta)z_0 > z_0$.

Since (θ_0, x_0, x_0) solves "communication model" as well, we must have

$$P(\theta_0|z_0,x_0,1) = c = P(\theta_0|z_0,x_0).$$

Lemma 4 in the Technical Appendix shows that the function $P(\cdot|z_0,x_0,1) - P(\cdot|z_0,x_0)$ crosses zero once and from below. Hence $P(\theta_0|z_0,x_0,1) - P(\theta_0|z_0,x_0) = 0$ implies

$$P(z_0|z_0,x_0,1) - P(z_0|z_0,x_0) < 0,$$

because $z_0 < \theta_0$.

We next show that the above inequality is a contradiction. Write the left-hand side of the inequality as:

$$
\begin{aligned}
&\frac{\int_{-\infty}^{z_0} J(t,z_0)p(t|z_0,x_0)\,\mathrm{d}t}{\int_{-\infty}^{\infty} J(t,z_0)p(t|z_0,x_0)\,\mathrm{d}t} - P(z_0|z_0,x_0) \\
= \quad & \frac{(1-P(z_0|z_0,x_0))\int_{-\infty}^{z_0} J(t,z_0)p(t|z_0,x_0)\,\mathrm{d}t - P(z_0|z_0,x_0)\int_{z_0}^{\infty} J(t,z_0)p(t|z_0,x_0)\,\mathrm{d}t}{\int_{-\infty}^{\infty} J(t,z_0)p(t|z_0,x_0)\,\mathrm{d}t} \\
= \quad & \frac{\int_0^{\infty} J(z_0-\tau,z_0)h(\tau)\,\mathrm{d}\tau}{\int_{-\infty}^{\infty} J(t,z_0)p(t|z_0,x_0)\,\mathrm{d}t},
\end{aligned}
$$

where we have made use of the symmetry of $J(t,z_0)$ about the point $t=z_0$, and where

$$h(\tau) \equiv p(z_0-\tau|z_0,x_0)(1-P(z_0|z_0,x_0)) - p(z_0+\tau|z_0,x_0)P(z_0|z_0,x_0).$$

Lemma 5 in the Technical Appendix shows that there exists a τ' such that $h(\tau)>0$ for $\tau<\tau'$ and $h(\tau)<0$ for $\tau>\tau'$. Therefore, the fact that $J(z_0-\tau,z_0)$ is decreasing in τ for $\tau>0$ implies that

$$\int_0^{\infty} J(z_0-\tau,z_0)h(\tau)\,\mathrm{d}\tau > J(z_0-\tau',z_0)\int_0^{\infty} h(\tau)\,\mathrm{d}\tau = 0.$$

Thus, $P(z_0|z_0,x_0,1) - P(z_0|z_0,x_0) > 0$, which establishes the contradiction. ■

Proof of Lemma 2. We proceed in a number of steps.

Claim 1. $\Pr[y_i=1|z,x_i]$ is increasing in x_i for $x_i<z$ and decreasing in x_i for $x_i>z$.

Take derivative of $\Pr[y_i=1|z,x_i]$ with respect to x_i to get:

$$
\begin{aligned}
\frac{\partial \Pr[y_i=1|z,x_i]}{\partial x_i} = & -w\frac{\beta}{\sqrt{1+\beta}\sigma_x}\left[\phi\left(\frac{\overline{x}(z)-X_i}{\sqrt{1+\beta}\sigma_x}\right)-\phi\left(\frac{\underline{x}(z)-X_i}{\sqrt{1+\beta}\sigma_x}\right)\right] \\
& -(1-w)\frac{1}{\sqrt{2}\sigma_x}\left[\phi\left(\frac{\overline{x}(z)-x_i}{\sqrt{2}\sigma_x}\right)-\phi\left(\frac{\underline{x}(z)-x_i}{\sqrt{2}\sigma_x}\right)\right]+\frac{\partial w}{\partial x_i}F(\beta,z,x_i)
\end{aligned}
$$

where

$$F(\beta,z,x_i) = \Phi\left(\frac{\overline{x}(z)-X_i}{\sqrt{1+\beta}\sigma_x}\right)-\Phi\left(\frac{\underline{x}(z)-X_i}{\sqrt{1+\beta}\sigma_x}\right)-\Phi\left(\frac{\overline{x}(z)-x_i}{\sqrt{2}\sigma_x}\right)+\Phi\left(\frac{\underline{x}(z)-x_i}{\sqrt{2}\sigma_x}\right),$$

and $X_i=\beta x_i+(1-\beta)z$. It is straightforward to show that $F(\beta,z,x_i)$ is decreasing in β, with $F(1,z,x_i)=0$. Thus, $F(\beta,z,x_i)>0$ for $\beta<1$. Since $\overline{x}(z)+\underline{x}(z)=2z$, the first two terms are positive if $x_i<z$. Since $\partial w/\partial x_i>0$ for $x_i<z$, the third term is positive as well. Therefore, the derivative is positive. If $x_i>z$, then the opposite is true.

Claim 2. When $c = 0.5$,

$$\frac{\partial \hat{x}_I(\theta', z')}{\partial \theta} = \frac{-p(\theta'|z', x')}{\int_{-\infty}^{\theta'} \frac{J(t,z')}{J(\theta',z')} \frac{\partial p(t|z',x')}{\partial x} \, \mathrm{d}t} \quad \text{and} \quad \frac{\partial \hat{x}_U(\theta', z')}{\partial \theta} = \frac{-p(\theta'|z', x')}{\int_{-\infty}^{\theta'} \frac{1-J(t,z')}{1-J(\theta',z')} \frac{\partial p(t|z',x')}{\partial x} \, \mathrm{d}t}.$$

When $c = 0.5$, the value of x' that satisfies $P(\theta'|z', x', 1) = c$ is $x' = z'$. By Claim 1, $\partial \Pr[y_i = 1|z', x']/\partial x_i = 0$ at this point. Since

$$P(\theta'|z', x', 1) = \frac{\int_{-\infty}^{\theta'} J(t, z') p(t|z', x') \, \mathrm{d}t}{\Pr[y_i = 1|z', x']},$$

the claim follows by the implicit function theorem. The expression for $\partial x_U/\partial \theta$ is derived in a similar fashion.

Claim 3. When $c = 0.5$,

$$\int_{-\infty}^{\theta'} \frac{J(t, z')}{J(\theta', z')} \frac{\partial p(t|z', x')}{\partial x} \, \mathrm{d}t > \int_{-\infty}^{\theta'} \frac{\partial p(t|z', x')}{\partial x} \, \mathrm{d}t > \int_{-\infty}^{\theta'} \frac{1 - J(t, z')}{1 - J(\theta', z')} \frac{\partial p(t|z', x')}{\partial x} \, \mathrm{d}t.$$

When $c = 0.5$, $\theta' = z' = x'$. Therefore, $J(t, z') < J(\theta', z')$ for all $t < \theta'$. Moreover,

$$\frac{\partial p(t|z', x')}{\partial x} = -w \frac{1}{\sigma_x^2} \phi_I' - (1 - w) \frac{1}{\sigma_x^2} \phi_U' + \frac{\partial w}{\partial x} \left(\frac{1}{\sqrt{\beta}\sigma} \phi_I - \frac{1}{\sigma} \phi_U \right),$$

where the subscript I indicates that the corresponding function is evaluated at $(t - X')/(\sqrt{\beta}\sigma_x)$. Since $t < \theta' = X'$, ϕ_I' is positive. Likewise, the subscript U indicates that the corresponding function is evaluated at $(t - x')/\sigma_x$. Since $t < \theta' = x'$, ϕ_U' is also positive. Finally, since $\partial w/\partial x = 0$ at $x' = z'$, we have $\partial p(t|z', x')/\partial x < 0$. Thus, the first inequality of the claim follows. The second inequality can be established in a similar way.

Since x' satisfies the indifference condition $P(\theta'|z', x') = c$ of the "mute model," by the implicit function theorem we obtain:

$$\frac{\partial \hat{x}_m(\theta', z')}{\partial \theta} = \frac{-p(\theta'|z', x')}{\int_{-\infty}^{\theta'} \frac{\partial p(\theta|z',x')}{\partial x} \, \mathrm{d}\theta}.$$

The ranking of the partial derivatives in the lemma then follows by Claims 2 and 3. Finally, since $P(\theta|z, x, 0)$ increases in θ and decreases in x, we have $\partial \hat{x}_U/\partial \theta > 0$. This proves inequality (20) the lemma. Inequality (21) then follows immediately from Lemma 6 in the Technical Appendix, which shows that $\partial \hat{x}/\partial z = 1 - \partial \hat{x}/\partial \theta$. ■

Proof of Proposition 4. By Claim 2 in the proof of Lemma 2,

$$
\begin{aligned}
& J\frac{\partial \hat{x}_I}{\partial \theta} + (1-J)\frac{\partial \hat{x}_U}{\partial \theta} \\
= & -p(\theta'|z',x')\left[J(\theta',z')\frac{1}{\int_{-\infty}^{\theta'}\frac{J(t,z')}{J(\theta',z')}\frac{\partial p(t|z',x')}{\partial x}\,\mathrm{d}t} + (1-J(\theta',z'))\frac{1}{\int_{-\infty}^{\theta'}\frac{1-J(t,z')}{1-J(\theta',z')}\frac{\partial p(t|z',x')}{\partial x}\,\mathrm{d}t}\right] \\
> & -p(\theta'|z',x')\frac{1}{J(\theta',z')\int_{-\infty}^{\theta'}\frac{J(t,z')}{J(\theta',z')}\frac{\partial p(t|z',x')}{\partial x}\,\mathrm{d}t + (1-J(\theta',z'))\int_{-\infty}^{\theta'}\frac{1-J(t,z')}{1-J(\theta',z')}\frac{\partial p(t|z',x')}{\partial x}\,\mathrm{d}t} \\
= & \frac{-p(\theta|z',x')}{\int_{-\infty}^{\theta'}\frac{\partial p(t|z',x')}{\partial x}\,\mathrm{d}t} \\
= & \frac{\partial \hat{x}_m}{\partial \theta},
\end{aligned}
$$

where the inequality follows from Jensen's inequality and the fact that the function $1/t$ is concave for $t < 0$. This establishes the first inequality of (24). Moreover, since

$$
P_x = -p + (1-\beta)w\phi\left(\frac{\theta' - X'}{\sqrt{\beta}\sigma_x}\right) < 0,
$$

we have $\partial \hat{x}_m/\partial \theta > 1$. This establishes the second inequality of (24).

Inequality (25) follows from (24) and from Lemma 6 in the Technical Appendix, which shows that

$$
\frac{\partial \hat{x}_I}{\partial z} + \frac{\partial \hat{x}_I}{\partial \theta} = \frac{\partial \hat{x}_m}{\partial z} + \frac{\partial \hat{x}_m}{\partial \theta} = \frac{\partial \hat{x}_U}{\partial z} + \frac{\partial \hat{x}_U}{\partial \theta} = 1.
$$

Finally, inequalities (24) and (25), together with the comparison of the decomposition equation (18) of the "mute model" and the corresponding decomposition equation (23) of the "communication model," establish the proposition. ∎

References

Acemoglu, D., A. Ozdaglar, and A. ParandehGheibi (2010). Spread of (mis)information in social networks. *Games and Economic Behavior 70*(2), 194–227.

Allport, G.W. and L.J. Postman (1947). *The Psychology of Rumor*. New York,: H. Holt and Company.

Angeletos, G.M. and I. Werning (2006). Crises and crices: information aggregation, multiplicity, and volatility. *American Economic Review 96*(5), 1720–1736.

Banerjee, A.V. (1993). The economics of rumours. *Review of Economic Studies 60*(2), 309–327.

Bernhardt, D. and M. Shadmehr (2010). Collective action with uncertain payoffs: Coordination, public signals and punishment dilemmas. *SSRN eLibrary*.

Bilefsky, D. (2009). Celebrating revolution with roots in a rumor. *The New York Times*, November 18, p. 16.

Boix, C. and M. Svolik (2010). The foundations of limited authoritarian government: Institutions and power-sharing in dictatorships. *SSRN eLibrary*.

Bommel, J.V. (2003). Rumors. *Journal of Finance 58*(4), 1499–1519.

Buchner, H.T. (1965). A theory of rumor transmission. *Public Opinion Quarterly 29*(1), 54–70.

Bueno de Mesquita, E. (2010). Regime change and revolutionary entrepreneurs. *American Political Science Review 104*(03), 446–466.

Carlsson, H. and E. van Damme (1993). Global games and equilibrium selection. *Econometrica 61*(5), 989–1018.

Chassang, S. and G. P. i Miquel (2010). Conflict and deterrence under strategic risk. *Quarterly Journal of Economics 125*(4), 1821–1858.

Duffie, D., S. Malamud, and G. Manso (2010). The relative contributions of private information sharing and public information releases to information aggregation. *Journal of Economic Theory 145*(4), 1574–1601.

Edmond, C. (2011). Information manipulation, coordination, and regime change. Working Paper 17395, National Bureau of Economic Research.

Elias, N. and J.L. Scotson (1994). *The Established and The Outsiders: A Sociology Enquiry into Community Problems*, 2d ed. SAGE Publications.

Esfandiari, G. (2010). The Twitter devolution. *Foreign Policy*, June 7.

Gambetta, D. (1994). Godfather's gossip. *European Journal of Sociology 35*(02), 199–223.

Gentzkow, M. and J.M. Shapiro (2006). Media bias and reputation. *Journal of Political Economy 114*(2), 280–316.

Goldstone, J.A. (1994). *Revolutions: Theoretical, Comparative, and Historical Studies*, 2d ed. Fort Worth: Harcourt Brace.

Grunden, W.E., M. Walker, and M. Yamazaki (2005). Wartime nuclear weapons research in Germany and Japan. *Osiris 20*, 107–130.

Knapp, R. H. (1944). A psychology of rumor. *The Public Opinion Quarterly 8*(1), 22–37.

Ley, R. (1997). Köhler and espionage on the island of Tenerife: A rejoinder to Teuber. *American Journal of Psychology 110*(2), 277–284.

Milgrom, P.R. (1981). Good news and bad news: Representation theorems and applications. *Bell Journal of Economics 12*(2), 380–391.

Morris, S. and H. Shin (2003). Global games: Theory and applications. In M. Dewatripont, L.P. Hansen, and S.J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, Vol. 1. Cambridge: Cambridge University Press.

Morris, S. and H.S. Shin (1998). Unique equilibrium in a model of self-fulfilling currency attacks. *American Economic Review 88*(3), 587–597.

Nkpa, N. K. U. (1977). Rumors of mass poisoning in Biafra. *Public Opinion Quarterly 41*(3), 332–346.

Peterson, W.A. and N.P. Gist (1951). Rumor and public opinion. *American Journal of Sociology 57*(2), 159–167.

Skocpol, T. (1979). *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. New York: Cambridge University Press.

Suen, W. (2010). Mutual admiration clubs. *Economic Inquiry 48*(1), 123–132.

Turner, M.E., A.R. Pratkanis, P. Probasco, and C. Leve (1992). Threat, cohesion, and group effectiveness: Testing a social identity maintenance perspective on group think. *Journal of Personality and Social Psychology 63*(5), 781–796.

Veldkamp, L. (2011). *Information Choice in Macroeconomics and Finance.* Princeton: Princeton University Press.

Zhang, N. (2009). Rumors and mobilization in China's contentions politics. Ph.D. dissertation, Chinese University of Hong Kong.

Technical Appendix

(Not intended for publication)

Lemma 3. In the "communication model," $\lim_{z\to-\infty} x_I^*(z) = +\infty$ and $\lim_{z\to\infty} x_I^*(z) = -\infty$.

Proof. We only prove the first part of this lemma; the proof of the second part is similar.

Claim 1. (a) $\lim_{z\to-\infty} \underline{x}(z)/z = 1 + \sigma_I/\sigma_U$; and (b) $\lim_{z\to-\infty} \overline{x}(z)/z = 1 - \sigma_I/\sigma_U$.

Recall that $\overline{x}(z)$ is the larger solution to $w(z,x) = \delta$. Solving this equation gives

$$\overline{x}(z) = z + \sqrt{\frac{\sigma_I^2}{\sigma_U^2}(z-s)^2 - 2\sigma_I^2 \log \frac{\sigma_I^2 \delta(1-\alpha)}{\sigma_U^2(1-\delta)\alpha}} \equiv z + \kappa(z).$$

Since $\lim_{z\to\infty} \kappa(z)/z = \sigma_I/\sigma_U$, this establishes part (a). Part (b) also follows because $\underline{x}(z) = z - \kappa(z)$.

Claim 2. For any finite $\hat{\theta}$ and any $x_I^* \neq \infty$, $\lim_{z\to-\infty} P(\hat{\theta}|z, x_I^*, y_i = 1, z \sim I) = 1$.

Let $\lim_{z\to-\infty} x_I^*(z)/z = \gamma \geq 0$.

Consider the complementary probability,

$$\Pr[\theta > \hat{\theta}|z, x_I^*, y_i = 1, z \sim I] = \frac{\int_{\hat{\theta}}^{\infty} J(t,z) \frac{1}{\sqrt{\beta}\sigma_x} \phi\left(\frac{t-z-\beta(x_I^*-z)}{\sqrt{\beta}\sigma_x}\right) \mathrm{d}t}{\Phi\left(\frac{\overline{x}(z)-z-\beta(x_I^*-z)}{\sqrt{1+\beta}\sigma_x}\right) - \Phi\left(\frac{\underline{x}(z)-z-\beta(x_I^*-z)}{\sqrt{1+\beta}\sigma_x}\right)}.$$

Since $J(t,z)$ is decreasing in t for $t > z$, we have

$$\lim_{z\to-\infty} \Pr[\theta > \hat{\theta}|z, x_I^*, y_i = 1, z \sim I] \leq \lim_{z\to-\infty} \frac{J(\hat{\theta},z)\left(1 - \Phi\left(\frac{\hat{\theta}-z-\beta(x_I^*-z)}{\sqrt{\beta}\sigma_x}\right)\right)}{\Phi\left(\frac{\overline{x}(z)-z-\beta(x_I^*-z)}{\sqrt{1+\beta}\sigma_x}\right) - \Phi\left(\frac{\underline{x}(z)-z-\beta(x_I^*-z)}{\sqrt{1+\beta}\sigma_x}\right)}.$$

Note that $\lim_{z\to-\infty}(\overline{x}(z) - z - \beta(x-z))/z = \beta(1-\gamma) - \sigma_I/\sigma_U$. There are two cases to consider. If $\beta(1-\gamma) - \sigma_I/\sigma_U \leq 0$, then the denominator of the term above does not vanish as z goes to minus infinity, while the numerator goes to zero. So the limit of the ratio is 0. If $\beta(1-\gamma) - \sigma_I/\sigma_U > 0$ then both denominator and numerator vanishes. However, $J(\hat{\theta}, z)$ goes to 0 at the rate at which $(\overline{x}(z) - \hat{\theta})/\sigma_x$ goes to minus infinity, which is equal to $(1 - \sigma_I/\sigma_U)/\sigma_x$. The denominator goes to 0 at the rate at which

$(\overline{x}(z) - z - \beta(x - z))/(\sqrt{1+\beta}\sigma_x)$ goes to minus infinity, which is

$$\frac{\beta(1-\gamma) - \sigma_I/\sigma_U}{\sqrt{1+\beta}\sigma_x} < \frac{1-\sigma_I/\sigma_U}{\sigma_x}.$$

Hence, in both cases,

$$\lim_{z\to-\infty} \frac{J(\hat{\theta}, z)}{\Phi\left(\frac{\overline{x}(z)-z-\beta(x_I^*-z)}{\sqrt{1+\beta}\sigma_x}\right) - \Phi\left(\frac{\underline{x}(z)-z-\beta(x_I^*-z)}{\sqrt{1+\beta}\sigma_x}\right)} = 0.$$

This implies that $\lim_{z\to-\infty} P(\hat{\theta}|z, x_I^*, y_i = 1, z \sim I) = 1$.

Claim 3. For any finite $\hat{\theta}$ and any $x_I^* \neq \infty$, $\lim_{z\to-\infty} P(\hat{\theta}|x_I^*, y_i = 1, z \sim U) = 1$.

Let $\lim_{z\to-\infty} x_I^*(z)/z = \gamma \geq 0$.

Consider limit of the complementary probability,

$$\begin{aligned}\lim_{z\to-\infty} \Pr[\theta > \hat{\theta}|x_I^*, y_i = 1, z \sim U] &= \lim_{z\to\infty} \frac{\int_{\hat{\theta}}^{\infty} J(t,z)\frac{1}{\sigma_x}\phi\left(\frac{t-x_I^*}{\sigma_x}\right)\,\mathrm{d}t}{\Phi\left(\frac{\overline{x}(z)-x_I^*}{\sqrt{2}\sigma_x}\right) - \Phi\left(\frac{\underline{x}(z)-x_I^*}{\sqrt{2}\sigma_x}\right)} \\ &\leq \lim_{z\to-\infty} \frac{J(\hat{\theta},z)\left(1-\Phi\left(\frac{\hat{\theta}-x_I^*}{\sigma_x}\right)\right)}{\Phi\left(\frac{\overline{x}(z)-x_I^*}{\sqrt{2}\sigma_x}\right) - \Phi\left(\frac{\underline{x}(z)-x_I^*}{\sqrt{2}\sigma_x}\right)}.\end{aligned}$$

The term $J(\hat{\theta}, z)$ goes to 0 at the rate $(1-\sigma_I/\sigma_U)/\sigma_x$. If $1-\sigma_I/\sigma_U > \gamma$, the denominator goes to 0 at the rate

$$\frac{1-\gamma-\sigma_I/\sigma_U}{\sqrt{2}\sigma_x} < \frac{1-\sigma_I/\sigma_U}{\sigma_x}.$$

Therefore the ratio goes to 0 as z goes to minus infinity. If $1-\sigma_I/\sigma_U \leq \gamma$, the denominator does not vanish. So the ratio again goes to zero.

To prove the lemma, note that $P(\hat{\theta}|z, x_I^*, y_i = 1)$ is just a weighted average of $P(\hat{\theta}|z, x_I^*, y_i = 1, z \sim I)$ and $P(\hat{\theta}|z, x_I^*, y_i = 1, z \sim U)$. By Claims 2 and 3, we must have

$$\lim_{z\to-\infty} P(\hat{\theta}|z, x_I^*, y_i = 1) = 1 > c,$$

for any finite $\hat{\theta}$ and any $x_I^* \neq \infty$. We know from part (1) of Proposition 3 that the limit of $\theta^*(z)$ is finite. Therefore the indifference condition (10) cannot hold unless $\lim_{z\to-\infty} x_I^*(z) = \infty$. ■

Lemma 4. The function,

$$g(\theta) \equiv P(\theta|z_0, x_0, 1) - P(\theta|z_0, x_0),$$

crosses zero once and from below.

Proof. The function $g(\theta)$ has the same sign as

$$\int_{-\infty}^{\theta} (J(t, z_0) - \Pr[y_i = 1|z_0, x_0]))p(t|z_0, x_0)\,\mathrm{d}t.$$

Note that $J(t, z_0)$ is increasing then decreasing, with $\lim_{t\to\pm\infty} J(t, z_0) = 0$. Moreover, $\Pr[y_i = 1|z_0, x_0] = \mathrm{E}[J(t, z_0)]$ (where the expectation is taken using the probability density $p(\cdot|z_0, x_0)$). Therefore, there exist t' and t'' such that the integrand is negative for $t < t'$, positive for $t \in (t', t'')$, and negative for $t > t''$. This implies that $g(\theta)$ is decreasing for $\theta < t'$, increasing for $\theta \in (t', t'')$, and decreasing for $\theta > t''$. Finally, note that $g(0) = 0$ and $g(\infty) = 0$. Therefore, there is an unique θ' such that $g(\theta) < 0$ for $\theta < \theta'$ and $g(\theta) > 0$ for $\theta > \theta'$. ∎

Lemma 5. For $z_0 < x_0$, the function,

$$h(\tau) \equiv p(z_0 - \tau|z_0, x_0)(1 - P(z_0|z_0, x_0)) - p(z_0 + \tau|z_0, x_0)P(z_0|z_0, x_0),$$

crosses zero once and from above.

Proof. Let $X_0 = \beta x_0 + (1 - \beta)z_0$. Write $h(\tau)$ in the following form:

$$\frac{w}{\sqrt{\beta}\sigma_x}\left[\phi\left(\frac{z_0 - \tau - X_0}{\sqrt{\beta}\sigma_x}\right)(1 - P(z_0|z_0, x_0)) - \phi\left(\frac{z_0 + \tau - X_0}{\sqrt{\beta}\sigma_x}\right)P(z_0|z_0, x_0)\right]$$
$$+ \frac{1 - w}{\sigma_x}\left[\phi\left(\frac{z_0 - \tau - x_0}{\sigma_x}\right)(1 - P(z_0|z_0, x_0)) - \phi\left(\frac{z_0 + \tau - x_0}{\sigma_x}\right)P(z_0|z_0, x_0)\right].$$

The sign of the first term is the same as the sign of

$$l_1(\tau) - \frac{P(z_0|z_0, x_0)}{1 - P(z_0|z_0, x_0)},$$

where

$$l_1(\tau) = \frac{\phi\left(\frac{z_0 - \tau - X_0}{\sqrt{\beta}\sigma_x}\right)}{\phi\left(\frac{z_0 + \tau - X_0}{\sqrt{\beta}\sigma_x}\right)} = \exp\left\{\frac{2(x_0 - z_0)\tau}{\sigma_x^2}\right\}.$$

Since $z_0 < X_0 < x_0$,

$$P(z_0|z_0, x_0) = w\Phi\left(\frac{z_0 - X_0}{\sqrt{\beta}\sigma_x}\right) + (1 - w))\,\Phi\left(\frac{z_0 - x_0}{\sigma_x}\right) < 0.5.$$

Moreover, $l_1(\tau)$ is strictly increasing with $l_1(0) = 1$ and $\lim_{\tau\to\infty} l_1(\tau) = \infty$. Therefore, the first term crosses zero from above.

Similarly, the sign of the second term is the same as the sign of

$$l_2(\tau) - \frac{P(z_0|z_0, x_0)}{1 - P(z_0|z_0, x_0)},$$

where

$$l_2(\tau) = \frac{\phi\left(\frac{z_0-\tau-x_0}{\sigma_x}\right)}{\phi\left(\frac{z_0+\tau-x_0}{\sigma_x}\right)} = \exp\left\{\frac{2(x_0 - z_0)\tau}{\sigma_x^2}\right\} = l_1(\tau).$$

Therefore, the second term also crosses zero once and from above. Moreover it crosses zero at exactly the same point as the first term does. This implies that the function $h(\tau)$ itself crosses zero once and from above. ■

Lemma 6. At $z = z' = \tilde{z}$ and $\theta = \theta'$,

$$\frac{\partial \hat{x}_I}{\partial z} + \frac{\partial \hat{x}_I}{\partial \theta} = \frac{\partial \hat{x}_m}{\partial z} + \frac{\partial \hat{x}_m}{\partial \theta} = \frac{\partial \hat{x}_U}{\partial z} + \frac{\partial \hat{x}_U}{\partial \theta} = 1$$

Proof. Write the relevant indifference conditions in the following form:

$$\begin{aligned}
\tau_m(\theta, z, x) &\equiv P(\theta|z, x) - c = 0, \\
\tau_I(\theta, z, x) &\equiv \int_{-\infty}^{\theta} J(t, z)p(t|z, x)\,\mathrm{d}t - c\Pr[y_i = 1|z, x] = 0, \\
\tau_U(\theta, z, x) &\equiv \int_{-\infty}^{\theta} (1 - J(t, z))p(t|z, x)\,\mathrm{d}t - c(1 - \Pr[y_i = 1|z, x]) = 0.
\end{aligned}$$

Claim 1. At $z = z' = \tilde{z}$ and $\theta = \theta'$, $\partial\hat{x}_m/\partial z + \partial\hat{x}_m/\partial\theta = 1$.

In the "mute model," we have

$$\begin{aligned}
\frac{\partial \tau_m(\theta', z', x')}{\partial z} &= -w\frac{1-\beta}{\sqrt{\beta}\sigma_x}\phi\left(\frac{\theta' - X'}{\sqrt{\beta}\sigma_x}\right) + \frac{\partial w}{\partial z}\left(\Phi\left(\frac{\theta' - X'}{\sqrt{\beta}\sigma_x}\right) - \Phi\left(\frac{\theta' - x'}{\sigma_x}\right)\right) \\
&= -w\frac{1-\beta}{\sqrt{\beta}\sigma_x}\phi\left(\frac{\theta' - X'}{\sqrt{\beta}\sigma_x}\right),
\end{aligned}$$

where the last equality follows because $X' = \beta x' + (1-\beta)z' = x'$. Similarly,

$$\frac{\partial \tau_m(\theta', z', x')}{\partial x} = -w\frac{\beta}{\sqrt{\beta}\sigma_x}\phi\left(\frac{\theta' - X'}{\sqrt{\beta}\sigma_x}\right) - (1-w)\frac{1}{\sigma_x}\phi\left(\frac{\theta' - x'}{\sigma_x}\right)$$

It is straightforward to see that at the point (θ', z', x'),

$$\frac{\partial \tau_m}{\partial z} + \frac{\partial \tau_m}{\partial x} = -\frac{\partial \tau_m}{\partial \theta}.$$

The claim then follows by the implicit function theorem.

Claim 2. At $z = z' = \tilde{z}$ and $\theta = \theta'$,

$$\begin{aligned}\frac{\partial \tau_I}{\partial x} &= J(\theta', z')\frac{\partial P(\theta'|z', x')}{\partial x} - D,\\ \frac{\partial \tau_U}{\partial x} &= (1 - J(\theta', z'))\frac{\partial P(\theta'|z', x')}{\partial x} + D;\end{aligned}$$

with $D < 0$.

First, note that $\partial \Pr[y_i = 1|z', x']/\partial x = 0$ by Claim 1 in the proof of Lemma 2. Rewrite τ_I using integration-by-parts and then take derivative with respect to x to get

$$D = \int_{-\infty}^{\theta'} \frac{\partial J(t, z')}{\partial \theta}\frac{\partial P(t|z', x')}{\partial x}\,\mathrm{d}t.$$

This term is negative because $\partial J/\partial \theta > 0$ for $t < \theta'$ and $\partial P/\partial x < 0$. The derivation of $\partial \tau_U/\partial x$ follows the same lines.

Claim 3. At $z = z' = \tilde{z}$ and $\theta = \theta'$,

$$\begin{aligned}\frac{\partial \tau_I}{\partial z} &= J(\theta', z')\frac{\partial P(\theta'|z', x')}{\partial z} + Q,\\ \frac{\partial \tau_U}{\partial z} &= (1 - J(\theta', z'))\frac{\partial P(\theta'|z', x')}{\partial z} - Q.\end{aligned}$$

where $Q = D$.

Let $T_1 = \Pr[y_i = 1, \theta \leq \theta'|z', x', z \sim I]$, $T_2 = \Pr[y_i = 1|z', x', z \sim I]$, $T_3 = \Pr[y_i = 1, \theta \leq \theta'|z', x', z \sim U]$, and $T_4 = \Pr[y_i = 1|z', x', z \sim U]$. We can write

$$\tau_I(\theta, z, x) = w(T_1 - cT_2) + (1 - w)(T_3 - cT_4).$$

Therefore,

$$\frac{\partial \tau_I(\theta', z', x')}{\partial z} = w\frac{\partial}{\partial z}(T_1 - cT_2) + (1 - w)\frac{\partial}{\partial z}(T_3 - cT_4),$$

with a term involving $\partial w/\partial z$ that vanishes because $T_1 - cT_2 = T_3 - cT_4 = 0$ when

$c = 0.5$. Consider first the derivative of the term $T_1 - cT_2$:

$$\frac{\partial}{\partial z}(T_1 - cT_2) = \int_{-\infty}^{\theta'} J(t,z') \frac{1}{\sqrt{\beta}\sigma_x} \frac{\partial \phi\left(\frac{t-X'}{\sqrt{\beta}\sigma_x}\right)}{\partial z} \,\mathrm{d}t + \int_{-\infty}^{\theta'} \frac{\partial J(t,z')}{\partial z} \frac{1}{\sqrt{\beta}\sigma_x} \phi\left(\frac{t-X'}{\sqrt{\beta}\sigma_x}\right) \mathrm{d}t$$
$$- c\left[\frac{\frac{\mathrm{d}\kappa}{\mathrm{d}z} + \beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\overline{x}(z') - X'}{\sqrt{1+\beta}\sigma_x}\right) - \frac{-\frac{\mathrm{d}\kappa}{\mathrm{d}z} + \beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\underline{x}(z') - X'}{\sqrt{1+\beta}\sigma_x}\right)\right].$$

Use integration-by-parts on the first term to get

$$\frac{\partial}{\partial z}(T_1 - cT_2) = \frac{1}{w} J(\theta', z') \frac{\partial P(\theta'|z', x')}{\partial z}$$
$$+ \int_{-\infty}^{\theta'} \left((1-\beta)\frac{\partial J(t,z')}{\partial \theta} + \frac{\partial J(t,z')}{\partial z}\right) \frac{1}{\sqrt{\beta}\sigma_x}\phi\left(\frac{t-X'}{\sqrt{\beta}\sigma_x}\right) \mathrm{d}t$$
$$- c\left[\frac{\frac{\mathrm{d}\kappa}{\mathrm{d}z} + \beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\overline{x}(z) - X'}{\sqrt{1+\beta}\sigma_x}\right) - \frac{-\frac{\mathrm{d}\kappa}{\mathrm{d}z} + \beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\underline{x}(z) - X'}{\sqrt{1+\beta}\sigma_x}\right)\right].$$

From this, we obtain:

$$\frac{\partial}{\partial z}(T_1 - cT_2) = \frac{1}{w} J(\theta', z') \frac{\partial P(\theta'|z', x')}{\partial z}$$
$$+ \frac{\frac{\mathrm{d}\kappa}{\mathrm{d}z} + \beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\overline{x}(z') - X'}{\sqrt{1+\beta}\sigma_x}\right)\Phi\left(\frac{\theta' - \frac{X'+\beta\overline{x}(z')}{1+\beta}}{\sqrt{\frac{\beta}{1+\beta}}\sigma_x}\right)$$
$$- \frac{-\frac{\mathrm{d}\kappa}{\mathrm{d}z} + \beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\underline{x}(z') - X'}{\sqrt{1+\beta}\sigma_x}\right)\Phi\left(\frac{\theta' - \frac{X'+\beta\underline{x}(z')}{1+\beta}}{\sqrt{\frac{\beta}{1+\beta}}\sigma_x}\right)$$
$$- c\left[\frac{\frac{\mathrm{d}\kappa}{\mathrm{d}z} + \beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\overline{x}(z') - X'}{\sqrt{1+\beta}\sigma_x}\right) - \frac{-\frac{\mathrm{d}\kappa}{\mathrm{d}z} + \beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\underline{x}(z') - X'}{\sqrt{1+\beta}\sigma_x}\right)\right]$$
$$= \frac{1}{w} J(\theta', z') \frac{\partial P(\theta'|z', x')}{\partial z}$$
$$+ \frac{\beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\overline{x}(z') - X'}{\sqrt{1+\beta}\sigma_x}\right)\left[\Phi\left(\frac{\theta' - \frac{X'+\beta\overline{x}(z')}{1+\beta}}{\sqrt{\frac{\beta}{1+\beta}}\sigma_x}\right) - \Phi\left(\frac{\theta' - \frac{X'+\beta\underline{x}(z')}{1+\beta}}{\sqrt{\frac{\beta}{1+\beta}}\sigma_x}\right)\right],$$

where the second equality uses the fact that $z' = \theta' = X'$ and $c = 0.5$. Similarly,

$$\begin{aligned}
\frac{\partial}{\partial z}(T_3 - cT_4) &= \frac{1+\frac{\mathrm{d}\kappa}{\mathrm{d}z}}{\sqrt{2}\sigma_x}\phi\left(\frac{\overline{x}(z') - x'}{\sqrt{2}\sigma_x}\right)\Phi\left(\frac{\theta' - \frac{x'+\overline{x}(z')}{2}}{\sqrt{\frac{1}{2}}\sigma_x}\right) \\
&\quad - \frac{1-\frac{\mathrm{d}\kappa}{\mathrm{d}z}}{\sqrt{2}\sigma_x}\phi\left(\frac{\underline{x}(z') - x'}{\sqrt{2}\sigma_x}\right)\Phi\left(\frac{\theta' - \frac{x'+\underline{x}(z')}{2}}{\sqrt{\frac{1}{2}}\sigma_x}\right) \\
&\quad - c\left[\frac{1+\frac{\mathrm{d}\kappa}{\mathrm{d}z}}{\sqrt{2}\sigma_x}\phi\left(\frac{\overline{x}(z') - x'}{\sqrt{2}\sigma_x}\right) - \frac{1-\frac{\mathrm{d}\kappa}{\mathrm{d}z}}{\sqrt{2}\sigma_x}\phi\left(\frac{\underline{x}(z') - x'}{\sqrt{2}\sigma_x}\right)\right] \\
&= \frac{1}{\sqrt{2}\sigma_x}\phi\left(\frac{\overline{x}(z') - x'}{\sqrt{2}\sigma_x}\right)\left[\Phi\left(\frac{\theta' - \frac{x'+\overline{x}(z')}{2}}{\sqrt{\frac{1}{2}}\sigma_x}\right) - \Phi\left(\frac{\theta' - \frac{x'+\underline{x}(z')}{2}}{\sqrt{\frac{1}{2}}\sigma_x}\right)\right].
\end{aligned}$$

Combining the two terms, and noting that

$$\begin{aligned}
Q &= \frac{w\beta}{\sqrt{1+\beta}\sigma_x}\phi\left(\frac{\overline{x}(z') - X'}{\sqrt{1+\beta}\sigma_x}\right)\left[\Phi\left(\frac{\theta' - \frac{X'+\beta\overline{x}(z')}{1+\beta}}{\sqrt{\frac{\beta}{1+\beta}}\sigma_x}\right) - \Phi\left(\frac{\theta' - \frac{X'+\beta\underline{x}(z')}{1+\beta}}{\sqrt{\frac{\beta}{1+\beta}}\sigma_x}\right)\right] \\
&\quad + \frac{(1-w)}{\sqrt{2}\sigma_x}\phi\left(\frac{\overline{x}(z') - x'}{\sqrt{2}\sigma_x}\right)\left[\Phi\left(\frac{\theta' - \frac{x'+\overline{x}(z')}{2}}{\sqrt{\frac{1}{2}}\sigma_x}\right) - \Phi\left(\frac{\theta' - \frac{x'+\underline{x}(z')}{2}}{\sqrt{\frac{1}{2}}\sigma_x}\right)\right] \\
&= D,
\end{aligned}$$

we obtain $\partial\tau_I/\partial z = J\partial P/\partial z + D$. Moreover, since $\tau_U = \tau_m - \tau_I$, this implies $\partial\tau_U/\partial z = (1-J)\partial P/\partial z - D$.

Claims 2 and 3 imply that

$$\frac{\partial\tau_I}{\partial x} + \frac{\partial\tau_I}{\partial z} = J\left(\frac{\partial P(\theta'|z',x')}{\partial x} + \frac{\partial P(\theta'|z',x')}{\partial z}\right).$$

From Claim 1, the term in parenthesis is equal to $-p(\theta'|z',x')$. Therefore, $\partial\tau_I/\partial x + \partial\tau_I/\partial z = -\partial\tau_I/\partial\theta$. By the implicit function theorem, the lemma follows. ∎

Proof of Proposition 1. In the "pure noise model," the indifference condition for the cut-off type x^*_{ms} satisfies

$$\Phi\left(\frac{\theta^*_{ms} - x^*_{ms}}{\sigma_x}\right) = c,$$

while the attack equation for the equilibrium regime survival threshold θ^*_{ms} satisfies

$$\Phi\left(\frac{x^*_{ms} - \theta^*_{ms}}{\sigma_x}\right) = \theta^*_{ms}.$$

Solving these two equations gives the equilibrium values of θ^*_{ms} and x^*_{ms} as stated in the proposition.

In the "public signal model," let $X^*_{ps} = \beta x^*_{ps} + (1-\beta)z$ denote the posterior mean of θ for the cut-off type x^*_{ps}. The posterior variance is $\beta\sigma_x^2$. The indifference condition and the attack equation can be written, respectively, as:

$$\Phi\left(\frac{\theta^*_{ps} - X^*_{ps}}{\sqrt{\beta}\sigma_x}\right) = c,$$
$$\Phi\left(\frac{x^*_{ps} - \theta^*_{ps}}{\sigma_x}\right) = \theta^*_{ps}.$$

Eliminate x^*_{ps} from these two equations to get

$$\theta^*_{ps} - \frac{\beta\sigma_x}{(1-\beta)}\Phi^{-1}(\theta^*_{ps}) = \frac{\sqrt{\beta}\sigma_x}{(1-\beta)}\Phi^{-1}(c) + z.$$

Assumption (1) implies that the left-hand side of the equality above is decreasing in θ^*_{ps}. Since the right-hand side is increasing in z, we have $\mathrm{d}\theta^*_{ps}/\mathrm{d}z < 0$. From the indifference condition and the definition of X^*_{ps}, we have

$$\frac{\mathrm{d}\theta^*_{ps}}{\mathrm{d}z} - \beta\frac{\mathrm{d}x^*_{ps}}{\mathrm{d}z} - (1-\beta) = 0.$$

Hence $\mathrm{d}x^*_{ps}/\mathrm{d}z < 0$. This establishes part (1).

Taking limit of the attack equation gives $\lim_{z\to\infty}\theta^*_{ps}(z) = 0$, which then implies $\lim_{z\to\infty} x^*_{ps}(z) = -\infty$. This establishes part (2a). Part (2b) is obtained analogously.

Since $\lim_{z\to\infty}\theta^*_{ps}(z) < \theta^*_{ms} < \lim_{z\to-\infty}\theta^*_{ps}(z)$, and since $\theta^*_{ps}(z)$ is strictly decreasing, there exists a unique $\tilde{z}$ such that $\theta^*_{ps}(\tilde{z}) = \theta^*_{ms}$. From the attack equations in the "pure noise model" and in the "public signal model," $\theta^*_{ps}(\tilde{z}) = \theta^*_{ms}$ implies $x^*_{ps}(\tilde{z}) = x^*_{ms}$. This establishes part (3).

Finally, equilibrium uniqueness is established by Morris and Shin (2003) and Angeletos and Werning (2006). ■